3-16-2009

Cost Forecasting Models for the Air Force Flying Hour Program

Tyler J. Hess

COST FORECASTING MODELS FOR THE AIR FORCE FLYING HOUR PROGRAM

THESIS

Tyler J. Hess, First Lieutenant, USAF

AFIT/GCA/ENV/09-M07

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government.

AFIT/GCA/ENV/09- M07

COST FORECASTING MODELS FOR THE AIR FORCE FLYING HOUR PROGRAM

THESIS

Presented to the Faculty

Department of Systems and Engineering 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

Tyler J. Hess, BS

First Lieutenant, USAF

March 2009

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT/GCA/ENV/09- M 07

COST FORECASTING MODELS FOR THE AIR FORCE FLYING HOUR PROGRAM

Tyler J. Hess, BS
First Lieutenant, USAF

Approved:

Lt Col Eric J. Unger (Chairman)	12 Mar 09 Date
Edward D. White (Member)	16 Mar 09 Date

AFIT/GCA/ENV/09- M 07

Abstract

The fiscally constrained environment in which the Air Force executes its mission places great emphasis on accurate cost estimates for planning and budgeting purposes. Inaccurate estimates result in budget risks and undermine the ability of Air Force leadership to allocate resources efficiently. This thesis evaluates the current method used by the Air Force and introduces new methods to forecast future Flying Hour Program costs. The findings suggest the current forecasting method's assumption of a proportional relationship between cost and flying hours is inappropriate and the relationship is actually inelastic. Prior research has used log-linear least squares regression techniques to forecast Flying Hour Program cost, but has been limited by the occurrence of negative net costs in the underlying data. This research uses time series and panel data regression techniques while controlling for flying hours, lagged costs, and age to create net costs models and an alternative model by separately estimating the two components of net costs which are charges and credits. Finally, this research found neither the proportional, net costs, nor charge minus credit models is a superior forecaster. As such, the models introduced in this research may be used as a cross check for the current method.

AFIT/GCA/ENV/09- M 07

This work is dedicated to my wife and my children. Without their support, patience, and understanding I would never have been able to accomplish such an undertaking.

Acknowledgements

First I would like to thank my wife and my children. Their love, support, and understanding provided me with the motivation I needed to produce the best research product I knew how.

I would also like to extend my sincere gratitude to my thesis committee. The constant encouragement and advice from my advisor, Lt Col Eric Unger, was pivotal in moving this research along. In addition, the input from Dr. Tony White provided me with a fresh outlook on my work. I feel extremely fortunate to have worked with each of them.

Also, thank you to Billy Kirby at the Cost per Flying Hour CAM office. The time he spent with me was crucial in understanding how the Flying Hour Program works. I am also grateful to Mark Gossett and William "Crash" Lively for all of their help with accessing and understanding Flying Hour Program data.

Finally, I want to thank my classmates, Captains David Brown and Stephen Gray, for their assistance with editing the paper and helping me prepare for its defense.

Tyler Hess

Table of Contents

Page

Page

Page

List of Figures

List of Tables

Page

COST FORECASTING MODELS FOR THE AIR FORCE FLYING HOUR PROGRAM

I: Introduction

Background

In his leadership statement, the Acting Secretary of the Air Force for Financial Management and Comptroller, The Honorable John G. Vonglis stated, "In a constrained fiscal environment, our ability to provide accurate, timely and relevant financial data, from cost estimates to budget projections...is paramount to enabling Air Force leadership at all organizational levels to make informed decisions" (2009). Given the recent economic crises experienced by the U.S. and the trend of decreasing budgets experienced by the Air Force as a percentage of Gross Domestic Product (GDP) depicted in Figure 1, constrained resources are exactly what Air Force decision makers are dealing with.

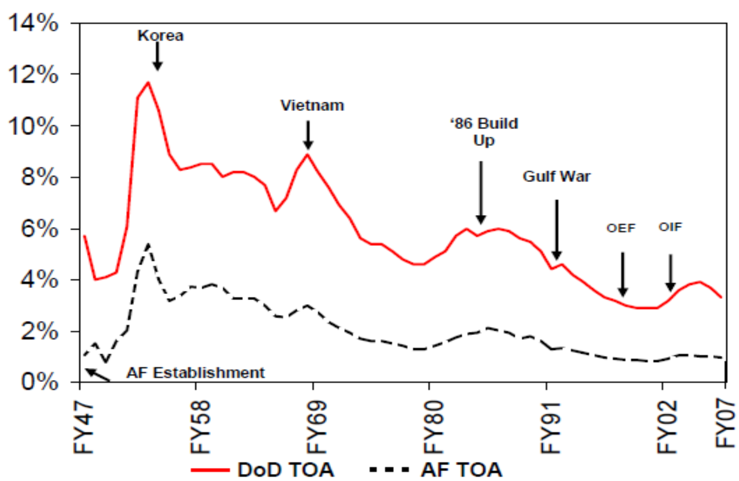


Figure 1: DoD and AF Total Obligation Authority (TOA) as a Percentage of U.S GDP Over Time (Faykes, 2007)

In addition to making decisions with limited resources, the cost of executing the Air Force's mission has become increasingly expensive. In Fiscal Year 2008 readiness made up roughly 27 percent of the Air Force's $110.7 billion base line budget (Faykes, 2007). The readiness portion of the budget represents the cost to operate and maintain the Air Force's weapon systems. Figure 2 breaks out Flying Hour (FH) Program, Depot Purchased Equipment Maintenance, and Contractor Logistic Support Cost from the overall readiness budgets in relation to the aircraft inventory over time. Figure 2 clearly demonstrates the fact that the Air Force has been spending more money on fewer systems.

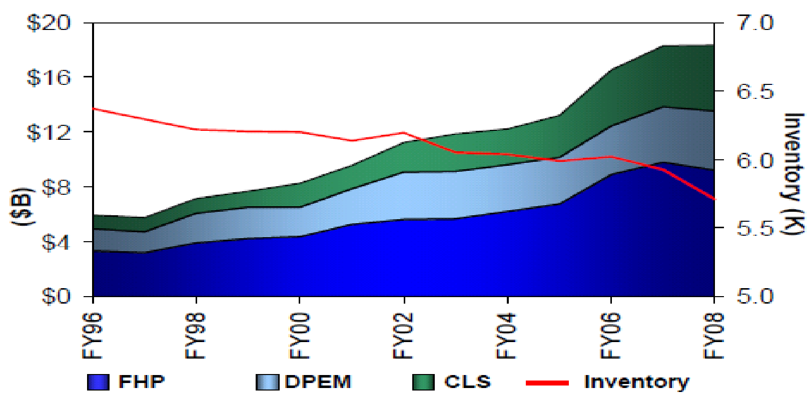


Figure 2: FHP, DPEM, and CLS Cost with Aircraft Inventory (Faykes, 2007)

In the Fiscal Year 2008 Air Force Posture Statement, then Secretary of the Air Force Michael W. Wynne stated that many factors ranging from increased fuel costs to an aging fleet have applied pressure on Air Force Budgets (Wynne and Moseley, 2008). A multitude of studies have shown that the aging of the Air Force's fleet is one of the reasons for Operation and Maintenance (O&M) cost growth (Hawkes and White, 2008; Unger, 2008; Hildebrandt and Sze, 1990). In addition, Hawkes and White showed that the variability in cost growth is larger for older airframes (2008). This means that as aircraft age, predicting their operating costs becomes more difficult. In order to maintain air dominance in the future and alleviate the problem of cost growth associated with aging aircraft, Air Force leaders have made the recapitalization and modernization of the Air Force's fleet one of their top priorities.

The culmination of constrained budgets, growing O&M costs, and recapitalization and modernization efforts places a great demand on the Air Force's financial managers to provide accurate cost and budget estimates. In 2000, the GAO reported that, unlike Research and Development Programs, little emphasis was placed on evaluating O&M costs. Given the fact that O&M costs make up such a large portion of the Air Force's overall budget, we find the previous statement alarming. Further, O&M costs represent the cost to fight today and are considered must pay items. In other words, the Air Force must fly and maintain its airframes and pay its personnel in order to accomplish the current mission. Unlike acquisition programs, these costs cannot be deferred which explains why resources that are originally intended for modernizing the Air Force's fleet often find themselves reallocated to pay for current operations.

With that said, we find it imperative for financial managers to reevaluate the way they estimate O&M costs to identify initiatives that will allow Air Force leaders to manage resources better. Underfunding programs causes the Air Force to either reprogram from other appropriations or ask Congress for additional funds. On the other hand, overfunding programs causes limited resources to be unavailable for use in other potentially value adding programs. Either way, inaccurate budget estimates create funding instability which has been found to cause cost growth in acquisition programs (Smirnoff and Hicks, 2008).

Purpose of This Study

In this study we reevaluate the way Air Force financial managers estimate and budget FH Program costs. In addition, we evaluate conflicting findings from past research in this same area and create econometric forecasting models to estimate future costs. Improved FH Program cost models can help to reduce instability and risk inherent in each year's budget and help to improve resource allocation decisions. While, the FH Program is only one portion of the overall O&M budget, it is a nontrivial segment. In Fiscal Year 2008, the FH Program was $7.4 billion which was 6.7 percent of that year's baseline budget. In order to assess the validity of the models currently used by the Air Force to estimate FH Program costs we answer the questions outlined in the following section.

Research Questions.

1. Should forecasting models use a top-level approach in which the relationship between costs and its predictors are averaged across all airframe types in the Air Force's fleet or estimated individually for specific airframe types?
2. What variables are significantly related to FH Program cost and can be used to help estimate future costs?
3. Due to the separate charges and credit components that comprise net cost, does predicting each component separately result in better forecasts than forecasting net cost alone?

4. Using our net cost models to control for other significant explanatory variables are flying hours and cost proportionally related such that cost increase by a constant factor in relation to flying hours?
5. Do the forecasting models we create in this study perform better than current, proportional models used by the Air Force?

Chapter Summary

The rest of this paper is structured as follows: Chapter II provides a review of past research done both on the FH Program and O&M cost in general. In Chapter III we build upon the prior literature and detail the methods we implement to answer our research questions as well as explain the structure and source of our data. Next, we present the results and analysis in Chapter IV. Finally we summarize the results and provide policy implications based on our findings in Chapter V.

II: Literature Review

In this chapter we provide the reader with a general overview of the Air Force's FH Program. We discuss the critical components of the method the Air Force uses to forecast FH Program costs. We also offer background and general discussion on previous research that has either attempted to improve upon or generate new models that can be used to forecast O&M or FH Program specific costs.

Flying Hour Program Overview

The Air Force FH Program encompasses the in house costs associated with flying and maintaining its airplanes. The Air Force estimates the budget for this program using a proportional cost model with two primary inputs: the number of flying hours to be flown and a cost per flying hour (CPFH) factor. The product of the two inputs results in the expected FH Program costs and estimated budget associated with each Air Force Mission Design Series (MDS) (e.g. F-15E or B-52H) as shown in Figure 3.

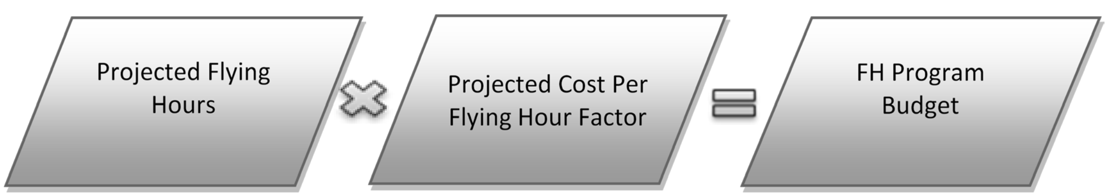


Figure 3: Cost per Flying Hour Budgeting Process

We define this as the proportional model because it assumes the cost to fly the Air Force's aircraft are proportionally related to the number of hours flown such that the relationship between flying hours and FH Program costs is a constant, linear relationship;

and when zero flying hours are flown, zero FH Program costs are incurred. In the following sections we provide more detail on how Air Force personnel generate each component of the proportional model.

Estimating Flying Hours.

In 1999, the General Accountability Office reported the Air Force's inability to execute the number of hours it requested each year. For example, from fiscal years 1995 through 1998, the Air Force flew fewer hours than it requested with a low of 89 percent of requested hours flown in fiscal year 1995 to a high of 94 percent in fiscal year 1996 (GAO, 1999:2). In efforts to achieve greater accuracy in flying hour estimates, each Major Command (MAJCOM) switched to standardized methodologies which reflected the mission of each respective MAJCOM. The new models calculate flying hours based on the number of pilots required to be combat mission-ready, basic mission-capable, or current with their training. The flying hour models also account for pilot experience, guidelines for mission types and weapons qualifications, special capability sorties, and collateral sorties (GAO, 1999:5-6).

Estimating CPFH Factors.

The second portion of the FH Program budgeting model is the CPFH factor. Three separate types of costs typically make up the CPFH factor: Depot Level Reparable (DLR) and consumable spares managed by the Material Support Division (MSD), consumable supplies (both General Support Division and those purchased via the Government Purchase Card), and aviation fuel (Rose, 1997:4). DLR spares are described as those items that are used in direct support of aircraft maintenance and can be repaired at an authorized maintenance facility. DLR spares include items such as engines and avionics equipment. Consumable items are expendable, non-reparable spares in direct support of the FH Program. Chargeable consumable items include items such as special solvents, nuts, bolts, and de-icing fluid (AFCAIG). We note that the terms DLR and consumables are generally used to identify parts managed by MSD and GSD respectively. Contrary to how the process has been described in previous literature, the calculation of the CPFH factors is complex and requires an involved bottom up methodology.

We describe the process as complex because there are thousands of types of DLR parts and the demand for each type is forecasted with a grass roots approach. The DLR CPFH factor calculation starts by collecting two years of demand data for every part used on the MDS being estimated. The total demand for each part is divided through by the number of hours flown over that same two year period. The result is a demand per flying hour factor for each part. Based on expected flying hours and the demand per flying hour factor, analysts from both the Central Asset Management (CAM) Office and the Spares Requirement Review Board project the total demand for each part and then adjust for known changes in maintenance procedures, warranties, part changes, or other factors. The final total demand for each part is then multiplied by the projected price of each part, which is provided by the Supply Management Activity Group (SMAG), and the result is the estimated flying hour budget for each specific MDS. The budget is then divided by the projected flying hours to arrive at a CPFH factor which can then be used to justify the budget and as an incremental analysis tool should the number of flying hours to be flown change in the future. These factors are then adjusted for inflation, sent through multiple levels for review, and finally approved by a General Officer/Senior

Executive Service level group known as the Air Force Cost Analysis Improvement Group (AFCAIG). Figure 4 illustrates how this process works for a single type of part.

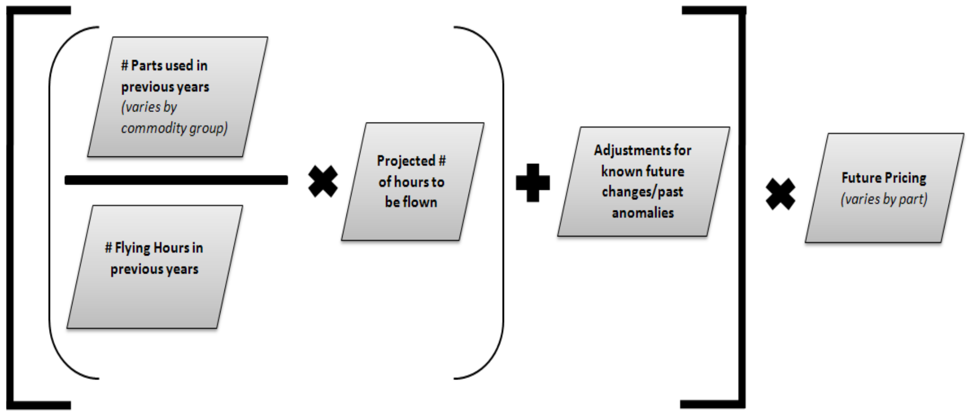


Figure 4: Flying Hour Program Budget Estimation Overview

The consumable CPH factor development process is less involved because the demand for the different types of consumable parts is much less volatile than the demand for DLR spares. Also, the DLR spares constitute a much larger portion of the FH Program costs. The consumable CPFH factor is calculated using three years of historical data as the ratio of normalized dollars spent on all consumables and total hours flown for each MDS. It is then changed from Constant Year dollars to Then Year dollars and sent through multiple levels of review for final approval by the AFCAIG (Kirby, 2008).

It is important to note that prior to FY08, the Air Force generated separate factors for every MAJCOM even if the same MDS was owned by more than one MAJCOM. In an effort to manage the FH Program more efficiently, the Air Force centralized the management and estimating processes in the CAM Program Office. The purpose behind this change was to create an office with sufficient expertise to manage the program while also reducing manpower and financial transaction costs. As a result, the Air Force now calculates each factor for each separate MDS and then applies these factors Air Force wide instead of at the individual MAJCOM level.

As these factor calculations demonstrate, the current process the Air Force uses to calculate flying hour costs is heavily dependent on three areas: prices of reparable and consumable spare parts, demand of reparable and consumable spare parts, and actual flying hours flown (GAO, 1999:3; Kirby, 2008). These three areas of variability make it difficult to accurately forecast flying hour cost. The GAO reported that price instability has been the biggest player in the inability to properly estimate flying hour costs. They note that the SMAG did not provide stable prices throughout the late 90's even though it is required to do so (1999).

The Air Force Repair Enhancement Program.

The figures projected in the FH Program budget represent a combination of expenditures and credits into an aggregated net sales figure. Therefore, we provide background on the Air Force Repair Enhancement Program (AFREP) because of the integral part it plays as the origin of credits in the FH Program and the fact that we later attempt to estimate charges separate from credits. According to Air Force Instruction 21-123, the objective of AFREP is to optimize "Air Force resources by increasing the wing-level repair capability of aerospace parts and equipment" (2002:3). The program encourages maintenance organizations to identify parts for repair. Repair processes for parts are approved locally by the base for base reparables or by the Single Manager Organization for all other items. Approval will be granted only if the total cost of repairing the part is less than the cost of purchasing a new part, the part is considered necessary to meet mission requirements, and the repair of the part does not introduce risk to mission performance. The repair and return of aircraft parts by maintenance organizations to the supply system generates a credit to capture the savings associated with repairing the part instead of purchasing a new one.

Previous Work on CPFH/FH Program Forecasting Models

A large amount of research has been previously conducted by the Air Force Institute of Technology (AFIT), Logistics Management Institute (LMI), and RAND. These studies have focused on various aspects of Operating and Support (O&S) Costs ranging from total Air Force O&S costs to MDS specific DLR or consumable parts costs. In general, the previous research has attempted to find cost drivers and improve upon the proportional model or create new forecasting models to aid the Air Force in its budgeting efforts. While much effort and some progress has been made by previous researchers, decision makers have yet to change the FH Program budget estimating process.

We speculate decision makers have been slow to implement new initiatives for three reasons. First, many of the studies focused on a single or very few MDS. Therefore, decision makers are left to wonder if those models can be generalized to other airframes. Any gained efficiencies in using the MDS specific models created by previous research may be outweighed by the time and manpower needed to calibrate and implement numerous and different models to each specific MDS. Second, many of the previous studies have failed to properly validate that the forecasts generated from their own models create forecasts superior to those generated by the proportional model. Finally, by using the bottom up method the AFCAIG's current approach assists logisticians by creating a valuable byproduct in the form of a part demand forecasts. This demand forecasts can be used to aid the supply system to meet expected demands and attempt to keep prices stable as required for the Air Force Working Capital Fund. While only a few of the studies may have created a policy impact on the FH Program, each of them has helped to further the research on how to best estimate the Air Force's FH Program budget. The following sections highlight some of the most significant findings of those studies.

Hildebrandt and Sze Create Cost Estimating Relationships for Operating and Support Costs and Its Various Components.

Using data from the Visibility and Management of Operating and Support Costs database in conjunction with aircraft characteristic information, Hildebrandt and Sze created aggregate cost estimating relationships to explain O&S costs. They found flying hours, flyaway costs, and Mission Design average fleet age to be statistically significant predictors for both total O&S costs per aircraft and O&S costs less fuel and personnel costs per aircraft. Hildebrandt and Sze used a log-log regression, to estimate the relationships. To avoid confusion, here and throughout the remainder of this paper we refer to a log-log regression as one in which the dependent variable and at least some of the independent variables have been adjusted using a natural log transformation. Hildebrandt and Sze chose this specification to allow for a nonlinear relationship between the independent and dependent variables as well as for other benefits which will be further discussed in Chapter III. The authors found that flyaway costs were a good proxy for the year of initial operation capability and aircraft type. Their findings support a less than proportional relationship between aggregate O&S costs, less personnel and fuel costs, and flying hours.

The Air Force has traditionally characterized the relationship between age and O&S costs as a bathtub effect or parabolic relationship. This characterization manifests itself from the belief that operation costs initially decline early in an aircraft's service life as learning benefits take effect. After the initial learning phase the aircraft moves into a steady state mid-life period in which O&S costs are fairly stable. Finally, it's assumed the aircraft will fail more often and incur modifications as the aircraft approaches its service life forcing the O&S costs to rise. Hildebrandt and Sze found no such bathtub effect, but rather a positive relationship between O&S costs and average age (1990). Figure 5 demonstrates the bathtub effect and breaks it apart into its three stages.

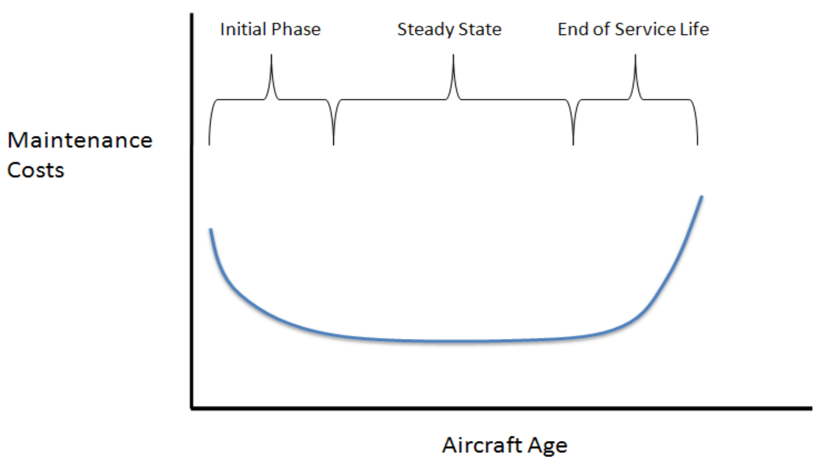


Figure 5: Bathtub Effect Demonstrating Parabolic Relationship between Age & Maintenance Costs

Hildebrandt and Sze conjectured that based on the time frame of the data used in their study, the costs had already moved past the initial and steady state periods and were in the upward portion of the aging effect. Later, Hawkes and White took this point into account by evaluating airframes that represented each of the stages shown in Figure 5 and showed that the cost per flying hour do follow the bathtub curve with respect to age (2008).

Wallace, Houser, and Lee Predict Removals Using Physics Based Constructs.

Wallace, Houser, and Lee found that during Operation Desert Storm the proportional CPFH model over predicted part demand by more than 200% and as a result, they attempted to find other factors that would be useful in calculating flying hour costs (2000:iii). Figure 6 graphically displays the proportional model's erroneous prediction of part demand during Operation Desert Storm.

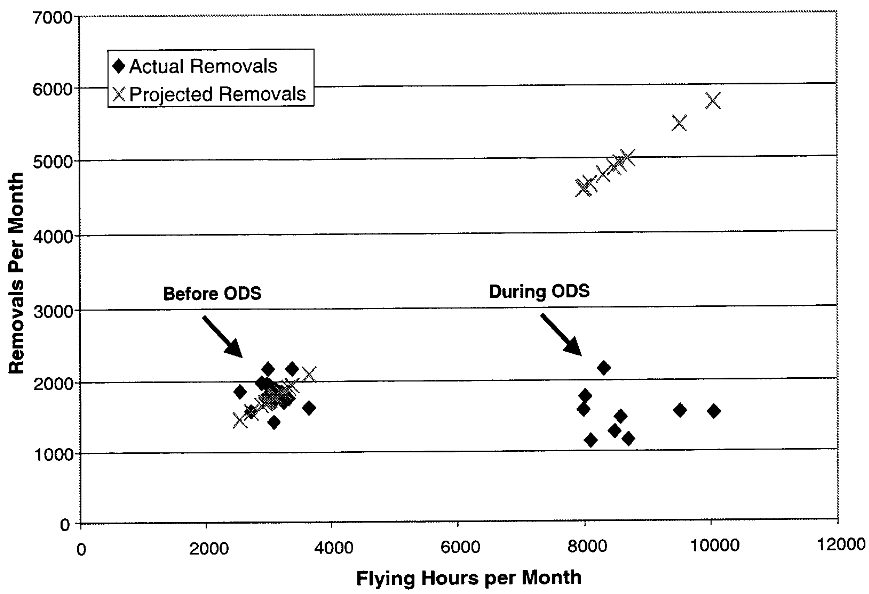


Figure 6: Proportional Model Projected Versus Actual C-5B Removals Prior to and During Operation Desert Storm (Wallace et al., 2007)

Wallace et al. used C-5B data from Operations Desert Storm to create a model that predicted part removals using maximum likelihood estimation. They validated their model with data from Kosovo for the KC-10, F-16C, and C-17. The model considered three separate failure modes that they believed cause removals: dormant, cycle induced, and operations based (2000:iii). The authors operationalized these theoretical constructs using ground days, cold and hot cycles (initial take-offs/final landing and intermediate take-offs/landings), and flying hours as independent variables in their model (2000:2-1).

Wallace et al. argue that the proportional CPFH model only captures one aspect of why part removals occur. They claim that ground time causes removals because environmental aspects such as dust and humidity degrade part integrity and that cycles cause removals by creating intense stress on the aircraft. They contend that as long as the flying behavior remains constant, then the failure point from each cause of removals remains constant. Therefore, it would be reasonable to use any one of the three failure modes to predict removals except for periods in which the flying behavior changes (2000:1-1). It is for this reason the authors maintained the proportional model performed poorly during contingency operations.

The authors stated that their physics-based model is more robust because it performed at least as well during peace time, but outperformed the proportional model during surges (2000:iii). The surges referred to any periods that include operations which do not coincide with routine flying hour operations such that the normal flying behavior is changed. However, we note assumptions made by the Air Force's proportional model is essentially a regression forced through the origin. As illustrated by Unger, when the number of flying hours increases beyond the average hours normally flown, the proportional model will begin to overestimate costs (2008). If an intercept, representing fixed costs, is included in the regression then the marginal effect of additional flying hours on cost will be attenuated.

Wallace et al. states that take-offs and landings accounted for fewer removals in the F-16 than the heavier tanker and transport planes. This point may be attributed to the fact that fighters are much lighter and undergo less stress during cold cycles (2000:4-19). Contrasting the findings from Hildebrandt and Sze, Wallace et al. reported the presence of long term increasing or decreasing trends in three of the four airframes they analyzed which they said supported the bathtub effect characterization. We note that the dependent variables for the two studies were different. Nonetheless, costs should behave similarly to removals because costs are essentially driven by removals. In addition, the dependent variable for one of the models studied by Hildebrandt and Sze contained costs associated with the types of removals Wallace et al. analyzed. The shortfall of this argument is that prices of different parts vary so we would need to know which types of parts are being removed. However, at an aggregate parts level the costs of various parts might average out. In fact, Wallace et al. use this theory because they do not actually predict which types of parts are removed and so their number of removals is an average across all parts.

Slay and Sherbrooke Focus On Predicting Removals As a Function Of Sortie Duration Instead Of Flying Hours.

Like Wallace et al., Slay and Sherbrooke also took note of how grossly the proportional flying hour model over predicted part removals during Operation Desert Storm. They stated that "although the sorties flown were much longer than their peacetime counterparts, demands per sortie remained about the same" (Slay and Sherbrooke, 2000:1-1). Based on this phenomenon Slay and Sherbrooke hypothesized that parts fail based on the number of sorties flown, not the number of hours flown. They argued that because the Air Force forecasts wartime demand based on peacetime data, and because predictions will drive inventory investment and capability assessment, it is important to know if failures result from flying hours, sorties, or a combination of the two. Through linear regression, using fighter aircraft data from 1993, Slay and Sherbrooke found sortie duration to be a significant predictor variable of part demand per sortie (2000:2-3). In addition, they also found the last sortie of the day, the mission type, and location to be significantly related to removals.

First, the last sortie of the day was associated with drastically more maintenance removals than other sorties. The authors argue that this was due to deferred maintenance. Second, they found that mission type also affected maintenance removals per sortie. For fighters, shorter missions are associated with combat training where pilots pull excessive Gravity Forces and the aircraft go through high levels of stress. On the other hand, cross country sorties are longer and are associated with much less stress on the aircraft. Therefore, the authors concluded mission type must be controlled for so that it does not overwhelm the effects of sortie length (2000:202). Finally, Slay and Sherbrooke stated to have found location effects when they evaluated the same type of MDS located at multiple bases. The authors reasoned that the location effect could have been due to the proximity of training ranges to the bases because bases that must fly further to reach the training ranges have higher average sortie durations (2000:2-3).

Sherbrooke and Slay generated a piecewise linear model that assumes demands are 40 percent flying hour/60 percent sortie dependent for sorties up to 1.4 hours and 6 percent flying hour/94 percent sortie dependent above 1.5 hours (2000:2-6). They selected their model by minimizing the mean squared error found from a validation data set. Figure 7 demonstrates the relationship between demands per sortie, flying hours, and sortie length found by Slay and Sherbrooke.

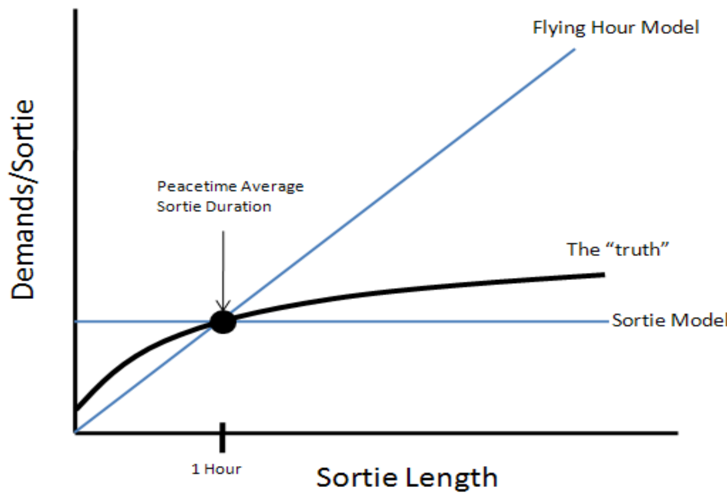


Figure 7: Slay and Sherbrooke's Demand Forecasting Model (Slay and Sherbrooke, 2000:2-5)

Laubacher, Hawkes, and Armstrong Each Attempt To Improve the Proportional Model By Better Predicting CPFH Rates.

Laubacher, Hawkes, and Armstrong generated a series of theses using various methods. Each of these studies was aimed at creating a model better capable of predicting CPFH factors and ultimately FH Program budgets (2004; 2005; 2006). These particular theses all focused on the same proportional model specification currently used by the Air Force to forecast costs such that the FH Program budget for a specific MDS is a result of the product of the number of flying hours expected to be flown and a CPFH factor. Methods used to calculate the CPFH factor included: simple forecasting techniques, multiple regression analysis, and panel data multiple regression analysis.

Laubacher (2004) analyzed three forecasting techniques: moving averages, single exponential smoothing, and Holt's Linear Method as ways to calculate the CPFH factors for each of the Air Force's rotary aircraft in each MAJCOM. By comparing the accuracy of his forecasted rates and the rates forecasted by the AFCAIG process using mean error, mean absolute error, mean percent error, and mean absolute percent error; Laubacher found that Holt's Linear Method provided the best estimates for 75 percent of the time series analyzed (the study analyzed data from 2001-2004) (Laubacher, 2004:iv). Laubacher argued Holt's Linear Method was superior because of its ability to capture trends.

Next, Hawkes (2005) built simple and multiple linear regression equations to forecasts CPFH DLR rates for all of the National Guard F-16 wings and 13 of 14 Active Duty F-16 wings. He found different explanatory variables drove the Active Duty and National Guard rates. Using data from fiscal years 1998 through 2004, Hawkes tested the explanatory power of the following nine variables: age of aircraft, average sortie duration, MAJCOM, base, utilization rate, percent engine type, percent block, percent deployed, and a lagged CPFH rate. Of the variables tested Hawkes found utilization rate, base, percent block, percent engine type, average age of aircraft, and the lagged CPFH rate variable to be significant predictors of CPFH rates. The percent block and percent engine type variables were used as moderator variables to capture possible differences in rates between various versions of the F-16. Also, the percent deployed variable was used because of the findings from Wallace et al. and Slay and Sherbrooke's work. Hawkes may not have found the percent deployed to be a significant variable because the F-16 is such a large fleet that any possible change in rate due to a change in flying behavior was overwhelmed by the number of aircraft that did not deploy in support of contingency operations and vary their flying behavior.

Finally, Armstrong (2006) used panel data multiple regression analysis with fixed effects to predict both DLR and consumable rates in an effort "to find a 'marginal CPFH' rate such that if a MAJCOM flies in excess of its programmed baseline (PB) direct hours, the additional funding to pay for contingencies etc. is commensurate with the additional (marginal) cost for the extra hours flown, not the full value of a flying hour for that weapon system" (2008:4). Aside from his labeling as a "marginal CPFH rate" he attempted to determine how well his models performed as forecasting tools by comparing the forecasted costs of the FH Program against the actual cost of the FH Program. He then contrasted his models performance to other models, including the Air Force's proportional model. While Armstrong claimed his models were superior to others, we find flaws in his argument because he came to that conclusion by comparing the accuracy of his F-15 model against Wallace's model, Hawkes' model, and a proportional model; all of which estimated rates for F-16s. Nonetheless, using monthly data from 2001 through 2005, Armstrong's research supports the idea that age, average sortie duration, and seasonality (monthly), affected the consumable and DLR CPFH rates (2008).

Hildebrandt Revisits His Previous Work, Focusing on Depot Level Reparable Costs.

Hildebrandt narrowed his previous research and created budget estimating relationships (BER) to predict DLR net sales. He employed pooled data from fiscal years 1998 through 2003 and applied longitudinal regression techniques to analyze all USAF MDS excluding those supported solely by Contractor Logistics Support. Hildebrandt's BER employed aircraft characteristics, operations tempo information, and time related variables to estimate DLR net sales. The specific variables which compose those three categories can be seen in Figure 8. All independent variables were found to be significant with the exception of the fiscal year indicators.

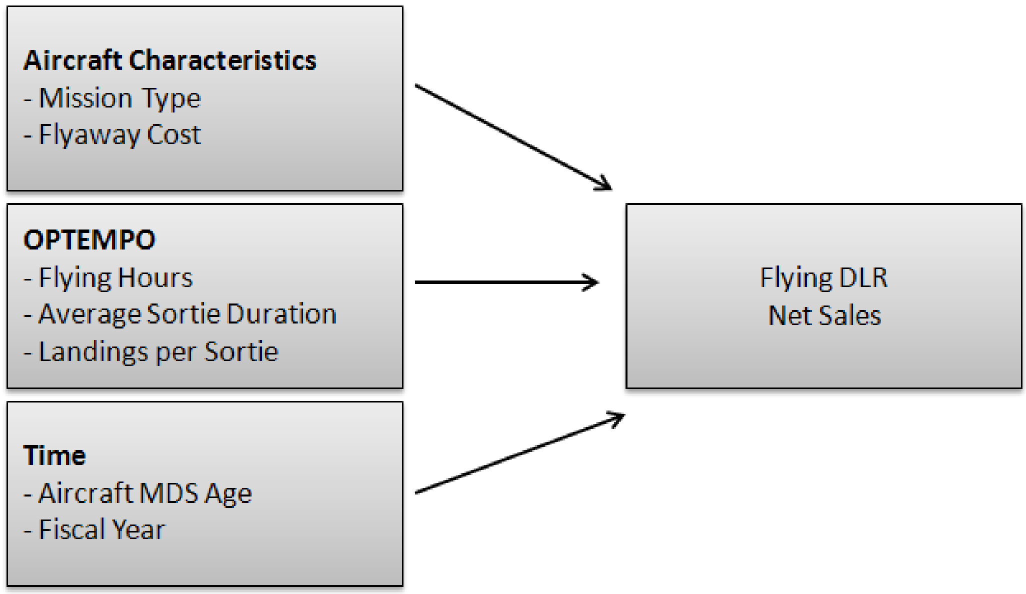


Figure 8: Hildebrandt's DLR Net Sales BER (2007:23)

Hildebrandt was able to capture variables found to generally affect FH Program costs or specifically affect DLR costs in previous research in one model. He used mission type and flyaway costs, as he and Sze had done in 1990, to capture the type of aircraft. Because the timeframe of his data provided variation in both average sortie duration and landings per sortie as a result of Operations Enduring Freedom and Iraqi Freedom he was able to capture the effects of flying in contingency operations discussed earlier with Sherbrooke, Slay, and Wallace et al. Finally, though he admits the aging effect is complex and in reality dependent upon other variables such as modifications and technology, he still finds a significant aging effect after controlling for other temporally dependent covariates (2007). Because Hildebrandt used a log-log model specification and the coefficient on the age variable was greater than one, Hildebrandt possibly captured the tail end of the bathtub effect discussed in previous research.

Another important finding from Hildebrandt's 2007 research was the idea that the AFCAIG process of forecasting FH Program costs as proportionally related to flying hours, described earlier in this chapter is appropriate (44). Of course Hildebrandt's BERs controlled for variables other than flying hours; nonetheless, his results supported a regression coefficient on flying hours that was not significantly different than 1 percent using a log-log model specification. In addition, Hildebrandt reasons that because the prices charged for DLR parts include overhead costs, which normally are considered fixed costs, the applicability of a fixed cost portion in the BER is questionable. However, the intercepts in Hildebrandt's regressions are all significant and support the inclusion of fixed costs. Finally, we consider noteworthy the fact that the coefficient on flying hours in his study is the average across all the MDS analyzed. Even though Hildebrandt controlled for the MDS using flyaway cost and mission type, the use of cross-sectional fixed effects only accounts for separate intercepts and does not account for differing slopes.

It is important to explore the appropriateness of accounting for various slopes across the various airframes because the slopes represent the estimated CPFH factors. While Hildebrandt's results suggest an averaged proportional relationship between flying hours and costs across all airframes, it is likely for individual MDS to have non-proportional relationships. This shortcoming is important because Hildebrandt's BER should not be used as a marginal tool focused on incremental analysis of a single MDS based on the fact that it uses averaged effects. In addition, because it is likely that each

MDS follows different budget estimating relationships it is also probably better to estimate the costs for each MDS separately.

Unger Updates Hildebrandt And Sze's Research By Evaluating O&S Cost Drivers.

In his study Unger sought to improve O&S resource allocation through better estimation methods (2008:1). Contrary to Hildebrandt's (2007) findings, Unger states that the major problems with the AFCAIG's proportional model are that the CPFH factor creates an average usage effect and the proportional model is incorrectly specified. First, because the proportional model creates an average effect, when decreasing the number of hours flown beyond the average hours used to calculate the CPFH factor the budget will be underestimated and vice versa when the hours flown are above the average number of hours used in the CPFH factor calculations (Unger, 2008:17). Figure 9 depicts the over and underestimation of flying hour costs when using the proportional model.

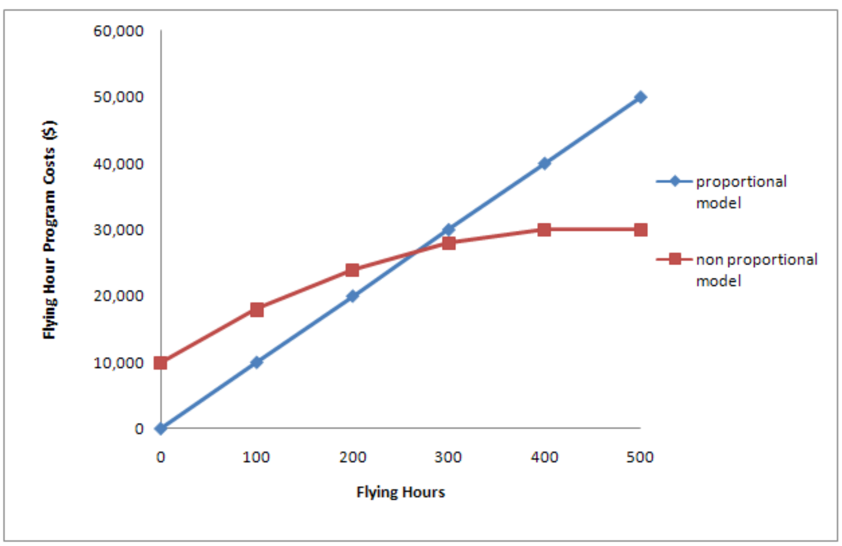


Figure 9: Proportional versus Non-proportional Models

The proportional model results in a poor marginal analysis and budget forecasting tool because the further the real number of hours moves away from the average number of hours the larger the error becomes. This statement can be visualized in Figure 9 by the separation between the proportional model and non-proportional model as the number of flying hours move away from the point where the two models intersect. Yet, Air Force Instruction 65-503, which governs the use of the CPFH Factors, advises the use of the factors as both a budgeting and an incremental analysis tool. Second, Unger argues the proportional model is incorrectly specified because it assumes a constant, linear relationship between flying hours and flying hour costs as well as having a lack of fixed costs. Figure 9 demonstrates the theoretical constant relationship and the theoretical curvilinear relationship between flying hours and FH Program costs for the proportional and non-proportional models respectively.

Among other models, Unger regressed O&S, DLR, and consumable costs on usage while controlling for other factors. He used logarithmic transformations which provided a constant elasticity, curvilinear, model and aided in meeting OLS assumptions by compressing the variance of the variables. The log-log specification allowed for nonlinear relationships between the dependent and independent variables. His results supported the inclusion of a nonzero intercept which theoretically would capture fixed costs associated with flying hour costs and ultimately dampen the effects of usage on costs (2008:4).

Different from Hildebrandt's 2007 results, Unger's findings showed an insignificant relationship between flying hours and DLR costs. Unger did find a significant and less than proportional relationship between flying hours and consumable costs (0.4 percent increase in consumable costs for every 1 percent increase in flying hours) (2008:74). Even though the usage effect was found to be insignificant with respect to DLR costs, the non-proportional usage effect associated with consumable costs serves to invalidate at least that piece of the proportional model.

Unger also tested for a linear and nonlinear aging effect using both age and its square as variables and found mixed results concerning age. He found that when the age variable was significant, it accounted for the majority of the effect on cost and therefore, the squared term was unwarranted. However, Unger did observe that because his models were aggregated at the Mission Design level, newer aircraft might have decreased the average age and biased any possible nonlinear aging effects (2008:42-43).

Van Dyk Continues Unger's Work, Focusing On DLR and Consumable Costs for The Air Force Bomber Fleet.

Van Dyk attempted to present improvements to the current method of forecasting flying hour costs. She focused on two model specifications directed at forecasting consumable and DLR costs for B-1B, B-2, and the B-52H each separately. Like Unger, she used a logarithmic transformation on the dependent variable and some of the independent variables in an attempt to capture possible nonlinear relationships. First, she tested the correctness of the proportional model by regressing the log of DLR and consumable costs on the log of flying hours and found that the intercept was significant at the 0.05 level in two of the three MDS evaluated for both types of costs. In addition, the flying hour variable was significant in two of the three MDS evaluated at the 0.01 level for both types of costs and statistically different than 1 percent. Therefore, contrary to

Hildebrandt (2007) her results support the prospect that the proportional model is generally inappropriate.

Van Dyk tested eleven variables other than flying hours in search of predictor variables to include: lagged costs, fiscal trends, sorties, average sortie duration, age utilization rate, mission capable rate, cannibalization rate, total ownership hours, crude oil prices, and temperature. Van Dyk's research supported mixed results as to which variables were valid predictors for both consumable and DLR costs across each of the three airframes she evaluated. However, her findings did contradict Unger's in that she found flying hours to be a significant predictor in seven of her eleven DLR cost models. In addition, Van Dyk found that lagged costs were helpful in predicting five of her eleven DLR cost models and six of her ten consumable cost models. Also, her findings supported Armstrong's (2006) results with the need to explain for a seasonal trend.

Finally, Van Dyk suggested that evaluating credits and expenditures separately may reduce some of the variance associated with both DLR and consumable costs. Often, expenditures and credits show co-movement and cancel each other out during a given time period, making it difficult to find the true variance in the variables as can be seen in Figure 10 (2008:81-85).

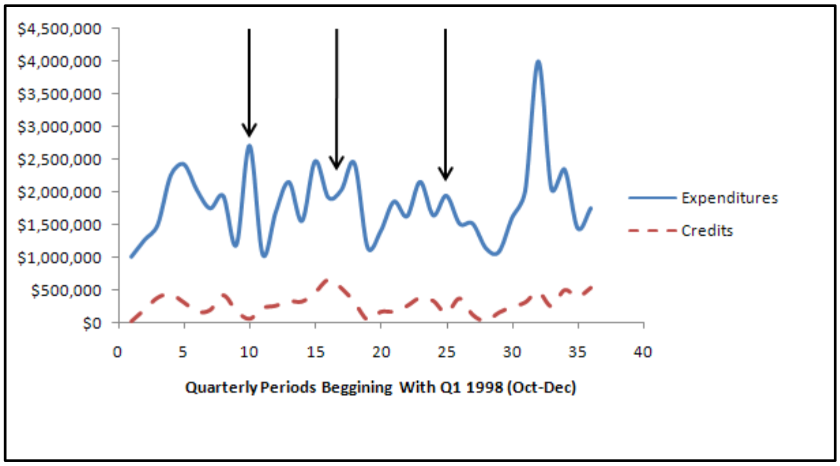


Figure 10: Quarterly Expenditures versus Credits for the B-2A (1998-2004)

The arrows in Figure 10 point to the 10th, 16th, and 25th periods and represent instances in which costs and credits move in opposite directions. When this occurs Van Dyk observed that the net sales consumable variable demonstrated very large variance from its normal behavior. This would essentially create observations in the data that can affect both the regression coefficient and its standard error and possibly affect the inferences drawn from the results.

Chapter Summary

In previous sections, we detailed how the Air Force currently estimates the FH Program Budget. We explained that the current method assumes a proportional relationship between flying hours and FH Program costs. In addition, we discussed circumstances and reasons why the current proportional model may not be the best method for forecasting the Air Force FH Program budget. We also discussed how previous research has sought to find different cost drivers and used different estimating techniques in search of a best forecasting method. Despite the copious amount of research already done we find the results of that research to be mixed and thus see the need for further analysis. Some of the contradictory examples can be seen in the list below:

- Hildebrandt (2007) found flying hours to be a significant predictor of DLR costs, but Unger (2008) found the opposite, and Van Dyke (2008) found flying hours to be significant in roughly half of her models.
- After controlling for other variables, Hildebrandt's (2007) results support a linear relationship between flying hours and DLR costs, yet Van Dyk (2008) found the relationship was better explained through a nonlinear specification.
- Laubacher, Hawkes, and Armstrong all considered the ability of multiple variables to explain changes in the CPFH factors, but their work hinges on a proportional relationship which we have already shown to be questionable.

We continue the work by first evaluating the relationship between predictor variables, which we discuss in the next chapter and DLR costs. In our models we predict costs like Hildebrandt (2007), Unger (2008), and Van Dyk (2008) instead of the CPFH factor like Laubacher (2004), Hawkes (2005), and Armstrong (2006). By predicting costs and including usage as an independent variable the model intrinsically estimates a CPFH factor as the usage regression coefficient.

In addition, based on the previous literature we control for other variables found to effect costs to include age, fiscal year/fiscal month, number of sorties, landings per sortie, MDS, base, and possible lag structures. The statistical techniques used by Hildebrandt (2007) and Unger (2008) generated average usage effects which are assumed to apply across airframes and locations such that a single usage coefficient was estimated and applied across all MDS types and all locations. We investigate this assumption by evaluating location and MDS type as moderators of the costs/usage relationship. In addition, we examine Van Dyk's (2008) suggestion by using a piecewise model in which expenditures and credits are estimated separately and then combined to calculate net costs.

We discussed how some of the model comparison techniques were unfair and also note now that the most recent research did not evaluate the forecasting abilities of their models. We objectively scrutinize accuracy of forecasts from our model with previous models and the current, proportional model to determine the best methodology.

Finally, many of the previous studies used time series data and included autoregressive terms in their models. Only Armstrong tested for stationarity of the variables, however. Hildebrandt created a first differenced model to control for nonstationarity in his appendix, but never determined if unit roots were in fact present. It is possible for time series to contain unit roots such that they are non stationary. If the variables are truly non stationary, and the nonstationarity is not accounted for in the estimating process, it is possible the results may lead to spurious correlation (Harris and Sollis, 2003:1). We test for unit roots in our work and if present adjust accordingly. We discuss stationarity in more detail in Chapter III.

Chapter III: Data Collection and Methodology

In this chapter we describe the data and how we propose to answer each of the research questions outlined in Chapter I. First we explain the variables and where the data for each variable was acquired. We discuss the shortcomings of the data. Lastly, we discuss the methods that we use in Chapter IV to analyze and interpret the results.

Data Sources and Variables

We gained data for this study using two databases: the Air Force Total Ownership Cost (AFTOC) database and the Air Force Reliability and Maintainability Information System (REMIS). AFTOC provides financial and inventory data, through the Supply Distribution Table (SDT), representing information from the Standard Base Supply System and the Wholesale and Retail Receiving/Shipping System. The inventory data provides information on quantity and part type for each transaction. The financial data provides prices for each part and designates charges versus credits for each transaction. Based on the information from various data feeds, the SDT uses allocates the financial and inventory data to organizations, Mission Design Series, types of cost, and time periods (AFTOC, 2008). We found all of the charge, credit, and net cost information used as dependent variables in this study using the AFTOC database. The REMIS database contains usage information and aircraft characteristic data used as the independent variables in this study.

Dependent Variables: Material Support Division (MSD) Fly DLR/Consumable Costs (Charges, Credits, and Net Costs)

Defining the dependent variables is critical to our research because the terms used to define FH Program costs and its components can vary greatly. In general, Depot Level Reparable costs found in AFTOC can be attributed to Element of Expense/Investment Code (EEIC) 644, also known as Material Support Division (MSD), and EEIC 645. Consumable costs can be found in EEIC 644 (MSD), but are also found in EEIC 609 (General Supply Division). The focus of this study is on EEIC 644, or MSD, portion of the CPFH factors, which we further disaggregate into our dependent variables: charges, credits, and net costs. Charges occur when parts are purchased, credits occur when parts are turned in, and net costs are charges minus credits for a particular MDS, location, and time period. Because not all items allocated to these EEICs are FH Program costs it is necessary to determine which costs are flying hour-driven costs.

We identify FH Program cost based on AFTOC business rules which are predicated upon supporting documentation associated with each transaction. The AFTOC database defines FH related program charges or credits from the Material Support Division (MSD) to the MAJCOMS under EEIC 644 as previously mentioned. Assignment as a MSD flying hour cost is contingent on the item being a Budget Code 8 plus a Type Organization Code (TOC) 3, 6, 7, 8, or 9. If the item is a Budget Code 8 item, but does not have a TOC of 3, 6, 7, 8, or 9 then it is allocated to EEIC 645. The EEIC 645 items might still be considered DLR items, but they are not "fly DLR" items and are not included as FH costs. Note that throughout this study we discuss DLR costs as they pertain to flying DLR costs only and we do not attempt to predict those items expensed under EEIC 645 considered non-fly DLRs. In addition, it is important to understand there are both consumable and flying DLR costs found within MSD (EEIC 644).

As previously discussed, in addition to the consumable items found under the MSD, there are also consumables found in the GSD. The GSD consumable items are designated by a Budget Code of 9 and are assigned to EEIC 609 or 605. We do not estimate GSD consumable items in this study. Figure 11 demonstrates how costs are allocated.

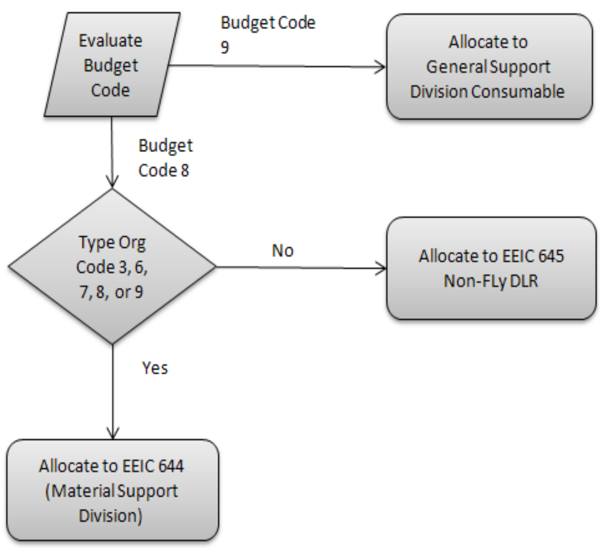


Figure 11: Cost Allocation Procedures

We used the Air Force Total Ownership Cost database in conjunction with Microsoft Access© to retrieve the cost information at various levels of aggregation by location and by time. We discuss both the advantages and disadvantages of data aggregation with more detail in later sections of this chapter. Table 1 shows a subset of selected fields of the raw data taken from the AFTOC database.

Table 1: Subset of Raw Cost Data Taken from AFTOC Database

MDS	Fiscal Year	FY_Month	Command	Base	Noun	Phrase	Net_Qty	Charge	Credit	Net_Cost	EEIC	TOC	ERRC
F-15C/D	1998	1	ACC	MOUNTAIN HOME AFB (ID)	RECEIVER-TRANSMITTE	CREDIT @ XCH	-1	$0.00	$1,720.92	($1,720.92)	644	7	XD2
B-52H	1999	6	ACC	BARKSDALE AFB (LA)	WHEEL,LANDING GEAR	CREDIT @ XCH	-1	$0.00	$583.01	($583.01)	644	7	XD2
B-2A	2006	4	ACC	WHITEMAN AFB (MO)	BLADE,FAN,AIRCRAFT GAS TURBINE	CREDIT @ XCH	-1	$0.00	$2,222.43	($2,222.43)	644	7	XD2
A-10	2003	9	ACC	MOODY AFB (GA)	GUN,AUTOMATIC,20 MILLIMETER	CREDIT @ XCH	-1	$0.00	$15,907.84	($15,907.84)	644	7	XD2
F-16C/D	2004	10	ACC	CANNON AFB (NM)	WHEEL ASSEMBLY,AIRC	CREDIT @ XCH	-1	$0.00	$2,041.77	($2,041.77)	644	7	XD2
KC-135R	1998	2	AMC	MACDILL AFB (FL)	TIRE,PNEUMATIC,AIRCRAFT	CREDIT @ XCH	-1	$0.00	$557.26	($557.26)	644	7	XD2
F-15E	1998	11	USAFE	RAF LAKENHEATH AIR BASE (UK	VALVE,LINEAR,DIRECT	CHARGE @ XCH	1	$5,109.59	$0.00	$5,109.59	644	7	XD2
F-16C/D	2000	12	PACAF	MISAWA AIR BASE (JAPAN)	TRANSFER UNIT,PROGRAMMABLE CAR	CHARGE @ XCH	1	$4,490.48	$0.00	$4,490.48	644	7	XD2
C-5A	2001	7	AMC TWCF	TRAVIS AFB (CA)	PRINTOUT UNIT	CHARGE @ XCH	1	$10,558.92	$0.00	$10,558.92	644	7	XD2
C-5C	2007	7	AMC TWCF	TRAVIS AFB (CA)	ELECTRONIC COMPONEN	CHARGE @ XCH	1	$4,031.53	$0.00	$4,031.53	644	7	XD2
B-52H	2005	8	ACC	BARKSDALE AFB (LA)	TRANSMITTER,COUNTERMEASURES	CHARGE @ XCH	1	$46,303.50	$0.00	$46,303.50	644	7	XD2
B-52H	2003	8	ACC	BARKSDALE AFB (LA)	CONTROL,COUNTERMEASURES TRANS	CHARGE @ XCH	1	$28,261.20	$0.00	$28,261.20	644	7	XD2
B-1B	1999	4	ACC	ELLSWORTH AFB (SD)	CONVERTER,SIGNAL DA	CHARGE @ XCH	1	$3,464.94	$0.00	$3,464.94	644	7	XD2
B-1B	2002	12	ACC	ELLSWORTH AFB (SD)	INDICATOR,MULTIPLE,	CHARGE @ XCH	1	$3,852.30	$0.00	$3,852.30	644	7	XD2
B-1B	2004	3	ACC	ELLSWORTH AFB (SD)	SUPPORT,STRUCTURAL	CHARGE @ STD	1	$669.30	$0.00	$669.30	644	7	XB3
F-15C/D	2003	10	ACC	EGLIN AFB (FL)	FLAMEHOLDER,AFTERBURNER,TURBIN	CREDIT @ STD	-1	$0.00	$10,938.57	($10,938.57)	645	G	XF3
F-15C/D	2006	10	ACC	EGLIN AFB (FL)	FLAMEHOLDER,AFTERBURNER,TURBIN	CREDIT @ STD	-1	$0.00	$10,938.57	($10,938.57)	645	G	XF3
F-15E	1998	9	ACC	SEYMOUR JOHNSON AFB (NC)	FLAMEHOLDER,AFTERBURNER,TURBIN	CREDIT @ STD	-1	$0.00	$10,938.57	($10,938.57)	645	G	XF3
F-15C/D	1998	4	ACC	LANGLEY AFB (VA)	WIRING HARNESS	CHARGE @ STD	1	$2,466.18	$0.00	$2,466.18	644	7	XF3
F-15C/D	2001	1	ACC	LANGLEY AFB (VA)	SWITCH,PROXIMITY	CREDIT @ LAC	-1	$0.00	$1,283.31	($1,283.31)	644	7	XB3

Until recent years, both the National Guard and Reserve units used slightly different cost allocation procedures than the Active Duty Air Force. Therefore, it would be inappropriate to use the cost definitions discussed for any portion of the Air Force other than the Active Duty portion. For these reasons we chose not to evaluate the National Guard and Reserve units.

Finally, we used inflation rates approved by the Office of the Secretary of Defense (OSD) to normalize the cost data and mitigate the effects of inflation. We obtained the rates from the SAF/FM inflation tutorial located on the Air Force Portal and converted all costs into CY08$.

Independent Variables

Our research explores the predictive power of usage and aircraft characteristic variables to predict the cost of the FH Program. We use these variables because of their ability in prior literature to predict FH Program costs discussed in the previous chapter. The following sections define each of the independent variables.

Usage Variables. The Air Force currently uses expected flying hours to create FH Program budgets. While the specifics of the estimated relationships in previous literature has varied, the presence of a correlation between flying hours and FH Program costs has been empirically validated. The literature also suggests that other usage variables play a role in FH Program costs, especially when the ratios between the usage variables changes as discussed by Wallace et al. (2007). In addition to flying hours, we also evaluate how landings and sorties contribute to the multi-dimensional influence of aircraft usage on costs. The Air Force does not currently estimate the number of future landings and sorties as it does for flying hours, but could potentially do so using historical information, training requirements, and contingency information. It would stand to reason that as usage increases, FH Program cost should also increase. However, there are more detailed effects that can be estimated when the usage variables are analyzed separately. We obtained monthly flying hour, sortie, and landing information for each tail number associated with the Mission Design groups evaluated in this study.

Age. Though age is not the primary variable of interest in this study it is important to control for age as demonstrated in previous research. In general, we expect that as aircraft increase in age they become more expensive to maintain. From the REMIS database, we obtained aircraft acceptance dates of each tail number in the Air

Force inventory for the MDS groups evaluated. We then used Microsoft Excel© to match the aircraft acceptance dates with the usage data via the aircraft tail numbers. The usage data is organized by date, so we were able to calculate each aircraft's age by subtracting the aircraft acceptance date from the usage date. Though we chose to calculate the age variable in days, the units in which the variable is analyzed will have no effect on the forecasting results because the regression coefficient adjusts appropriately according to units used. Finally, because we aggregate from the tail level to MDS groups and from months to quarters and years, the average age variable is an average of the aircraft for each MDS group for a given period of time.

Dummy Variables: Location and Aircraft Characteristics. We use dummy variables for each MDS group for every level of aggregation and for each location at the base and MAJCOM levels of aggregation. We created nine cross sectional dummy variables representing each MDS group and five representing each MAJCOM for the MAJCOM level of aggregation. In addition to the cross sectional dummy variables, we created seasonal dummy variables representing months and quarters for their respective levels of aggregation. Table 2 provides a list of all the dummy variables created. We discuss the purpose of using the dummy variables later in the Methodology section.

Table 2: List of Dummy Variables

MDS Specific	MAJCOM	Monthly	Quarterly
A-10	ACC	January	Qtr 1
B-1B	AETC	February	Qtr 2
B-2A	AMC	March	Qtr 3
B-52H	PACAF	April	Qtr 4
C-5	USAFE	May	
F-15C/D		June	
F-15E		July	
F-16C/D		August	
KC-135		September	
		October	
		November	
		December	

Data Aggregation

Based on the available data sources, it is possible to aggregate our data on three separate dimensions: aircraft type, assigned location, and units of time. First, it is possible to evaluate the different types of airframes ranging from MDS (e.g. F-16C or KC-135R), to MD (e.g. F-15 or KC-135). Unger argued it is inappropriate to analyze the effects of usage variables on costs at the MDS level because of intricacies associated with cost allocation in the Air Force. The crux of his argument was that for multiple MDS the cost may be recorded by a single Program Element Code and allocated to a single MD. In such an instance, costs are then allocated from the MD to the MDS level based on proportions of flying hours. The argument follows that if the MDS level is used the relationship between flying hours and cost would be overstated (2007). For this reason, we choose not to evaluate aircraft at the MDS level except where the accounting process properly allocates costs.

The CAM office and the A4/AMC office provided us with appropriate MDS groups (Kirby, 2008; Chamberlain, 2008). These groups allow us to evaluate each type of airframe at the lowest appropriate level, while still maintaining proper cost allocation. Table 3 demonstrates how these groups were made for the aircraft we evaluate in this study. Where MDS are very similar and shared accounting classifications they are grouped together. However, the MDS that are very different than others from within the same MD are kept in their own group. For example, the F-15E is very different than the F-15C and F-15D because of different missions, roles, and accounting classifications. On the other hand, all versions of the C-5 are so similar that we can appropriately group them together.

Table 3: MDS to MDS Groups

MDS	MDS Grouping
A-10A	A-10
OA-10A	A-10
B-1B	B-1B
B-2A	B-2A
B-52H	B-52H
C-5A	C-5
C-5B	C-5
C-5C	C-5
F-15C	F-15C/D
F-15D	F-15C/D
F-15E	F-15E
F-16C	F-16C/D
F-16D	F-16C/D
KC-135R	KC-135
KC-135T	KC-135

Second, we can evaluate the data at different levels of aggregation based on location. Previous work has discussed separate relationships between usage variables and costs for separate bases (Slay and Sherbrooke, 1997:1-4). In addition, in the past CPFH factors were calculated separately for each MAJCOM because the rates did not apply well across the entire Air Force. The data we obtained can be analyzed by MAJCOM, or aggregated to a non-locality level that estimates relationships within each MDS across the entire Air Force. We find that analyzing the AFTOC data at the base level of aggregation presents construct validity concerns that we discuss in later sections.

Finally, we can use monthly, quarterly, or annual time periods for our analysis. Armstrong's work was done on a monthly basis and found a fiscal trend based on the months (2005). On the other hand, we might argue that a month is not a long enough period of time to allow a true maintenance process to take form because of its stochastic nature. Shorter time periods may show more specificity, but they might also cause more noise in the data and suggests a longer time period where the random nature of part failures can average out.

Overall, we manipulated the data into nine levels of aggregation. Deciding which level of aggregation to use mandates a balance between the level of data that will be forecasted, properly specifying the model, and maintaining sufficient observations to properly carry out statistical tests. For example, we can initially rule out the Air Force by Year level because it has insufficient observations.

Combining the Cost and Usage Databases

Once the costs and usage databases were created for each level of aggregation, we then used Microsoft Excel© to create our databases. Table 4 is a subset from the final MAJCOM by quarter database where the dummy variables are not shown for simplicity.

Table 4: Subset of Final Majcom by Quarter Database

MD	Command	FY	FY_Quarter	Charges	Credits	Net Costs	FH	Landings	Sorties	Age (in days)
A-10	ACC	1998	1	$ 18,603,485	$ 6,195,516	$ 12,407,969	15556	7885	7885	5892
A-10	ACC	1998	2	$ 21,632,804	$ 6,686,494	$ 14,946,310	15845	8372	8352	5980
A-10	ACC	1998	3	$ 25,187,042	$ 5,638,942	$ 19,548,100	17120	8554	8541	6087
A-10	ACC	1998	4	$ 28,392,101	$ 5,604,210	$ 22,787,891	16147	7762	7760	6173
A-10	ACC	1999	1	$ 19,875,609	$ 4,179,532	$ 15,696,077	13367	7033	7032	6256
A-10	ACC	1999	2	$ 22,277,144	$ 4,531,408	$ 17,745,736	14280	7637	7637	6360
A-10	ACC	1999	3	$ 25,106,841	$ 4,220,671	$ 20,886,170	14346	7203	7164	6435
A-10	ACC	1999	4	$ 24,763,977	$ 5,184,037	$ 19,579,940	13642	6893	6792	6534
A-10	ACC	2000	1	$ 21,522,363	$ 5,273,269	$ 16,249,094	11911	6335	6216	6632
A-10	ACC	2000	2	$ 21,169,459	$ 4,301,334	$ 16,868,125	12528	6480	6450	6712
A-10	ACC	2000	3	$ 21,325,927	$ 5,433,414	$ 15,892,514	13285	7014	6995	6825
A-10	ACC	2000	4	$ 21,390,707	$ 5,278,018	$ 16,112,689	12144	6359	6278	6900

Location Based Construct Validity Concerns

When combining the costs information obtained from the AFTOC and REMIS databases we find that there are some construct validity issues with regards to the ability of determining cost and usage relationships based on location. The problem is rooted in the difference between the methods the two databases use to allocate costs and usage variables. The AFTOC database allocates cost based on the organization financing the maintenance action, while the REMIS database allocates usage based on a geographic locator code which indicates the owning organization of the specific tail number incurring the flying hour, sortie, or landing. In the Air Force, aircraft often fly to locations for training or other missions. Sometimes the aircraft need maintenance and receive repair from maintenance organizations other than their owning location. In these situations the financing organization may be different than the owning organization. We hypothesize that this may occur either when accounting records are incorrectly kept or when organizations request training with dissimilar aircraft and are responsible for paying for the operations. For example, Whiteman AFB does not own any A-10 aircraft. However, our data indicate that costs were incurred at Whiteman AFB for every year from FY1998 through FY2007. In addition, there is no corresponding usage information from REMIS for the A-10 at Whiteman AFB during that same period of time.

Table 5 demonstrates the percentage of costs that are misallocated as a percentage of total net costs over all MDS groups. Because the base is the lowest level of aggregation by location, we see that it has the highest percentage of misallocated costs. By moving from base to the MAJCOM level we alleviate much of the misallocation. Table 5 shows overall unallocated costs and Appendix A breaks out the information by MDS group and reveals much more dramatic differences in the percent of unallocated costs from the base to the MAJCOM level of aggregation.

Table 5: Costs Misallocation at MAJCOM and Base levels

Level of Aggregation	Total Net Costs	Mismatched Net Costs	% of Costs Unallocated
MAJCOM	$22,170,347,217	$6,396,655	0.03%
Base	$22,170,333,787	$1,394,715,803	6.29%

We hypothesize that because not all costs are captured, in general the relationship between the usage variables and cost might be underestimated. This information takes away from the construct validity of the expected location effect at the base level and supports the analysis of a location effect by MAJCOM or no location effect at all. Therefore, to avoid surrounding our results with construct validity concerns we do not analyze the base level of aggregation.

Methodology

As discussed in Chapter II, prior research has a used a variety of methodologies ranging from a moving average calculation to panel data analysis using fixed effects to estimate either costs of the FH Program or the CPFH factors themselves. Based on our panel data, the necessity to answer our research questions, and the ability to forecasts future FH Program costs we find a combination of ordinary least squares dummy variable regressions and panel regressions to be the most fitting methodology for this study.

Ordinary Least Squares (OLS) is a method used to estimate parameters of a linear regression model. The estimates are calculated by minimizing the sum of squared differences between the actual and predicted values of the model. These differences are often called the residuals. OLS is said to be the best linear unbiased estimator given the residuals are identically and independently normally distributed with zero conditional mean and constant variance or homoskedasticity (Woolridge, 2006). In addition, the regressor and regressands must be linearly related through the estimated parameters and the regressands must not demonstrate perfect collinearity or multicollinearity at the least. Finally, when dealing with time series OLS the series must be stationary. Violations of these assumptions will be discussed with the results in Chapter IV, but we highlight the stationary assumption, specifically discussing stationarity and tests for it later in this chapter because it is the only assumption that has not been properly tested for in the previous literature with few exceptions.

We utilize both time series and panel OLS models and choose which one to use based on the level of aggregation analyzed. For example, the Air Force by quarter level offers too few observations to estimate MDS group cost functions separately and so a panel model utilizing the MDS groups as the cross sections is convenient. On the other hand, if the results support estimating separate cost functions for each MDS then time series analysis for each MDS group would be appropriate for levels of aggregation such as the Air Force by month level. If MDS specific cost functions are more appropriate, the MAJCOM by quarter and MAJCOM by month levels offer enough observations to estimate separate cost functions for each MDS group, but using a panel model with MAJCOMs as cross sections makes this estimation more efficient and convenient. Equation 1 demonstrates our general use of a time series model:

$$\ln(\text{costs})_t = \beta_0 + \beta_1 \ln(\text{FH})_t + \beta_2 \text{age}_t + \delta_0 \text{season} + \varepsilon_t \tag{1}$$

where costs represents either net costs, charges, or credits at period *t*. Flying hours, landings, sorties, and age are continuous variables as described in previous sections. Season is a vector of dummy variables representing months or quarters. N-1 total dummy variables are included so that the base case is represented by β_0.

We use two general forms of panel models. Equation 2 demonstrates a least squares dummy variable regression (LSDV):

$$\begin{aligned}\ln(\text{costs})_{it} &= \beta_0 + \beta_1 \ln(\text{FH})_{it} + \beta_2 \text{age}_{it} + \beta_3 \text{landings}_{it} + \beta_4 \text{sorties}_{it} \\ &+ \delta_{i0} + \delta_{i1}\text{MDS}_i {*}\ln(\text{FH})_{it} + \delta_{i2}\text{MDS}_i {*}\text{age}_{it} + \delta_{i3}\text{MDS}_i {*}\text{landings}_{it} \\ &+ \delta_{i4}\text{MDS}_i {*}\text{sorties}_{it} + \delta_5 \text{season} + \varepsilon_{it}\end{aligned} \quad (2)$$

where cost may again take on one of the three dependent variables. MDS is a vector of dummy variables representing the different MDS groups. The δ_{ij} represent parameters to be estimated and are associated with MDS group specific effects on the dependent variable. The estimated β_i represent the base case parameters because N-1 MDS dummies are included in the regression. The error term varies both across MDS and across time. Interpretations of the LSDV are straightforward. If the KC-135 MDS group is the base group then a one percent increase in flying hours would increase costs by β_1. For a MDS group not used as the base case a one percent increase in flying hours would increase costs by $\beta_1 + \delta_i$ where i denotes a specific MDS group. This regression is convenient because separate intercepts and slopes can be estimated for each MDS group, but it does require the estimation of many parameters.

To avoid this, we can use a different panel model which we call the fixed effects model. The fixed effects model is seen as:

$$\ln(\text{costs})_{it} = \alpha + \beta_1 \ln(\text{FH})_{it} + \beta_2 \text{age}_{it} + \beta_3 \text{landings}_{it} + \beta_4 \text{sorties}_{it} + \varepsilon_{it} \quad (3)$$

Where now all the parameters estimated represent the average effect across all i cross sections. The average effect is accomplished through the use of a within transformation which essentially averages each observation on the ith individual over the Tth time periods for each variable and then subtract these averages from the actual observations.

Greene (2003) and Kennedy (2008) provide complete derivations for the within transformation. As apparent in the difference between equations 2 and 3, the fixed effects model reduces the number of parameters estimated. Deciding which panel model to use depends on the objective of the regression.

We use the LSDV regression to determine if the relationship between cost and predictor variables should be estimated as an average, with separate intercepts, or separately across MDS groups. We utilize F-test and t-test shown in equations 4 and 5 to test for the individual significance as well as the joint significance of variables with respect to their ability to predict costs:

$$F = \frac{\left(R_{ur}^2 - R_r^2\right)/q}{(1-R_{ur}^2)/(n-k-1)} \tag{4}$$

$$t_{\beta_i} = \beta_i / S.E.(\beta_i) \tag{5}$$

Based on the outcome of the previous research question we can then utilize equations 1 through 3 again to determine which variables are significant predictors of costs. Finally, after creating models for the different levels of aggregation, we can determine if flying hours are proportionally related to net costs by calculating confidence intervals around the estimated flying hour coefficients. We discuss more on how each model is used to specifically address the first three research questions in Chapter IV.

Forecasting Accuracy

We use the models created from answering the first three research questions to forecast the various types of costs. Given a specific level of aggregation we subtract the credits from the charges to calculate the charges minus credits forecast. The proportional model estimates are forecasted using CPFH factors from the 65-503 Cost Factors. In reality, the CPFH factors are continuously updated and these factors represent a snapshot in time. However, we argue budget estimates and submissions also represent a snapshot in time and therefore these factors are a fair representation of a simulated budgeting process. We also offer a third comparison by computing our own proportional CPFH factors using the prior two years of costs and flying hours to alleviate problems associated with the differences in historical costs used to calibrate the models. This version of the proportional model does not take into account adjustments made to the baseline values, however. Forecasting errors for the FH Program are often due to differences in predicted versus actual flying hours. We remove these effects by holding flying hours constant in our forecasts of both our proposed model and the proportional model. By doing this, any forecast errors are attributed to factors other than errors in the number of predicted flying hours.

We evaluate the forecasting accuracy of all competing models using root mean squared error (RMSE). We use this loss function to assess forecasting accuracy because forecasting is essentially an out of sample problem and we are forced to use prior data along with estimated future data to arrive at estimates. RMSE penalizes larger errors by squaring the error. RMSE is calculated as:

$$RMSE = \left(\frac{\sum_{i=1}^{n} \left(A_i - F_i \right)^2}{n} \right)^{1/2} \tag{6}$$

where A_i is the actual cost, F_i is the forecasted cost and n is the number of forecasts. As previously stated, because forecasting is an out of sample phenomenon we use an iterative calibration method. The models are first calibrated with data from Fiscal Years

1998 through 2004 and forecasts for 2005 are generated. The model is recalibrated by adding data from Fiscal Year 2005 to the calibration set and forecasting the costs for Fiscal Year 2006. The process is repeated to forecasts Fiscal Year 2007 costs. This process mirrors real life budget forecasting.

We find that having the lowest loss function score is necessary, but not sufficient in determining the superior forecasting model. It is possible the difference in loss function scores between the two models is not statistically different from zero. In 1995, Diebold and Mariano created a way to test the equality of forecasts. In the Diebold-Mariano (DM) Statistic the null hypothesis is defined as zero difference between forecasting model errors such that E[d_t] = 0. d_t is the loss differential defined as $g(e_{1t}) - g(e_{2t})$ where g is some loss function, and e_{1t} and e_{2t} are forecast errors for the given loss function in period t for two competing forecasting models. $\bar{d}$ is the average d_t and the DM Statistic is calculated as:

$$S_1 = \frac{\bar{d}}{[\hat{V}(\bar{d})]^{1/2}} \tag{7}$$

where $\hat{V}(\bar{d})$ is:

$$\hat{V}(\bar{d}) \approx n^{-1}[\gamma_0 + 2\sum_{k=1}^{h-1}\gamma_k] \tag{8}$$

and γ_k is the kth autocovariance of d_t. The autocovariance is:

$$\hat{\gamma}_k = n^{-1}\sum_{t=k+1}^{n}(d_t - \bar{d})(d_{t-k} - \bar{d}) \tag{9}$$

Diebold and Mariano show through Monte Carlo simulation that their statistic has an asymptotic standard normal distribution and through Monte Carlo simulations argue the performance of their statistic is robust to autocorrelation and non-normal distributions. However, Diebold and Mariano did explain that the test was oversized for small samples meaning their statistic generated Type I errors more often than would be expected for a given level of significance. As a result Harvey, Leybourne, and Newbold worked to improve the DM statistic by using an unbiased estimator of the variance of $\bar{d}$ and improved the finite sample performance (1997). Harvey et al.'s modification is found in equation 10 as:

$$S_1^* = \left[\frac{n+1-2h+n^{-1}h(h-1)}{n}\right]^{1/2} S_1 \qquad (10)$$

where S_1 is the original DM Statistic, n is the number of forecasts, and h represents how many periods ahead the forecast was used for. We use Harvey et al.'s modified DM Statistic, S_1^* to determine which forecast is statistically significantly better, if a superior forecasting model does exist. The null distribution of the modified DM Statistic is Student's t-distribution.

Natural Logarithmic Variable Transformation

We use a natural logarithmic transformation of our variables for three reasons. First, as we have discussed at length in the previous chapter, by taking logs we are able to capture possible non-linear relationships between our dependent and independent variables without using more complicated statistical procedures. Hu notes that in cost analysis we often deal with multiplicative error terms, "because experience tells us that the error of an individual observation (e.g., cost) is generally proportional to the magnitude of the observation (not a constant)" (2005). By transforming the variables from unit space into log space we are able to convert a multiplicative error term into an additive error term, supporting OLS assumptions. Second, when we transform both the dependent and independent variables the normal interpretations change and the estimated coefficients become cost elasticities. For example, in the model that regresses the natural log of cost on the natural log of flying hours the estimated regression coefficient, β_{FH}, would be interpreted as: a one percent increase in flying hours results in a β_{FH} percent increase in cost. The conversion of the regression coefficient to an elasticity allows us to more easily interpret the regression results because we no longer deal with units. Third, the use of a natural log transformation helps our OLS models to be more robust in meeting the model assumptions of homoskedasticity and normality of the error terms (Habing, 2004; Osborne 2002). While the log transformation has advantages, it does have one specific disadvantage.

The natural log is undefined for numbers of zero and less. For our models which do not estimate charges and credits separately, it is possible that the Air Force may incur negative net costs for any one of the MDS evaluated at any given time. Previous research has dealt with this topic by adding a constant to every observation so that the lowest net cost is one. Van Dyk stated the estimated coefficients should not be affected because the constant is added to every observation. While this manipulation does work when estimating regression models in unit space, we find that this is in fact not the case in log space. The addition of the constant biases the size of the estimated parameters and the back transformed forecast even after subtracting out the original constant. The size of this bias changes with the size of the constant added and the magnitude of the true elasticity. Because we are concerned with the size of the estimated parameters as well as the forecast, when data for a given MDS group contains negative net costs we cannot estimate the net cost model for that group in that specific level of aggregation.

Testing For Unit Roots

Harris and Sollis loosely define a stationary variable as having a constant mean and constant variance (2003:29). On the other hand, statistical properties of non-stationary variables vary with time. In their 1974 paper entitled "Spurious Regressions in Econometrics," Granger and Newbold took the point of view of a "statistical time series analyst, rather than the more classic econometric approach" and criticized much of the econometric research done at that time for drawing inappropriate inferences founded on unsound statistics. Specifically, they simulated regressions of independent, non-stationary time series to investigate the likelihood of finding spurious regressions. Granger and Newbold concluded that if we regress non-stationary variables, "it will be the rule rather than the exception to find spurious relationships" (1974:117). Non-stationary variables have a tendency to grow over time. Though two separate series might grow for unrelated reasons, a normal OLS regression will more often than not incorrectly find a significant relationship between them. The spurious regression problem is multiplied by the fact that the t-test and F- tests no longer have the normal distributions associated with stationary series. Therefore, we find it imperative to test for stationarity so that we might avoid making a Type I error.

We use the Augmented-Dickey Fuller test for individual time series and the Im, Pesaran, and Shin (IPS) unit root test for panels to determine if our series are stationary. The IPS test assumes the data generating processes for the cross sections are heterogeneous and that each cross section contains a unit root (Im, Pesaran, and Shin, 2003). The IPS performs the test in equation 11 for each i cross section:

$$\Delta y_{it} = \rho_i y_{i,t-1} + \sum_{L=1}^{\rho_i - 1} \theta_{iL} \Delta y_{i,t-L} + z_{it}' \gamma + u_{it} \tag{11}$$

where $\sum_{L=1}^{\rho_i} \theta_{iL} \Delta y_{i,t-L}$ represents lagged dependent variables for each i cross section, t period, and L lags; $z_{it}'\gamma$ represent a constant and possibly a deterministic trend; and ρ_i is the main coefficient of interest. The IPS averages the t-test on ρ_i from the individual ADF as $\bar{t} = \frac{1}{N}\sum_{i=1}^{N} t_{\rho_i}$. If $\bar{t}$ is found to be significantly different than zero, then we reject the null of a non-stationary series and can proceed with our estimation. If we are unable to reject the null, then researchers must resort to other techniques such as first differencing or estimation via an error correction model. The ADF test is essentially the IPS test with two differences. First, instead of running the regression in equation 11across multiple i, it is only done on a single cross section. Second, the ADF test does not follow a standard t-distribution and so the critical values have been estimated using Monte Carlo techniques.

Chapter Summary

In this chapter, we discussed how we collected data and manipulated it into a usable form. We then detailed construct validity issues where the data we collected disconnected from our theoretical expectations. Next, we showed how to use a dummy variable OLS regression to create a FH Program cost model. In addition, we discussed how to compare our models with the Air Forces current proportional model. Finally, we outlined a key assumption, stationarity, which the majority of past CPFH research has ignored and showed how we plan to test for it. In the next chapter we provide our results.

Chapter IV: Analysis and Results

In this chapter we provide analysis and results for each of the five research questions detailed in Chapter I and again in Chapter III. First, we extend the work done by Unger (2008) and Hildebrandt (2007). We use deductive reasoning to move from their generalized models and attempt to determine if it is more appropriate to estimate the FH Program costs for different airframes separately as opposed to a common model. Second, we generate net cost, charges, and credits models and find significant predictors for each of the three types of cost over various levels of aggregation. Third, we shift our analysis to question the appropriateness of the proportional Flying Hour model specification currently in use by the U. S. Air Force. Fourth, we evaluate alternative dependent variables in an attempt to increase forecast accuracy. Specifically, we consider a model in which we predict charges and credits separately and then combine them to arrive at net cost. Finally, we compare the forecast accuracy of our models and the proportional model.

Common versus Individual Airframe Flying Hour Program Cost Models

We first estimate a panel model like Unger (2008) using data from the Air Force by quarter level of aggregation with net costs as the dependent variable and flying hours, age, landings, and sorties as the independent variables. We use this level of aggregation as a starting point because the Air Force by annual level of aggregation only contains 90 observations which is an insufficient amount for this part of the analysis. Also, we only analyze eight of the nine MDS groups because the B-2A contains negative net costs, which does not work with our current specification.

Immediately we find problems with multicollinearity. Table 6 is a correlation matrix for all of the variables in this regression. Of note are the extremely high correlations between flying hours, sorties, and landings.

Table 6: Correlation Matrix of Variables at the Air Force by Quarter Level of Aggregation

	CREDITS	CHARGES	AGE	LANDINGS	SORTIES	FH
NET_COSTS	0.8198	0.9554	-0.3343	0.5975	0.6785	0.5105
CREDITS		0.9524	-0.2715	0.2508	0.3099	0.1589
CHARGES			-0.3180	0.4475	0.5212	0.3538
AGE				-0.1165	-0.2999	-0.0173
LANDINGS					**0.9591**	**0.9695**
SORTIES						**0.9016**

Calculating Variance Inflation Factors (VIF) for the independent variables, we find that each of the usage variables has a VIF greater than 50 and the age variable has a VIF of 20. Common VIF rules of thumb vary between acceptable levels of equal to or less than 5 or 10. It is possible that the variation between the usage variables might become greater at a lower level of aggregation, but we find that at lower levels of aggregation the VIF's only decrease slightly. Consequentially, we find it inappropriate to include more than one of the usage variables in light that doing so would invalidate our ability to determine the effects of individual predictors. We use flying hours because it is the variable of greatest use in answering our research questions.

The P-value for the test statistic shown in Table 7 shows that we reject the IPS test null hypothesis that the natural log of net costs is non-stationary at the 0.01 level.

Table 7: Panel Unit Root Test of Net costs for the Air Force by Quarter Level of Aggregation

Panel unit root test: Summary
Series: ln(Net Costs)
Sample: 1998Q1 2007Q4
Exogenous variables: Individual effects, individual linear trends
Automatic selection of lags based on SIC: 0 to 2

Method	Statistic	Prob.	Cross-sections	Obs
Im, Pesaran and Shin W-stat	-5.39684	0.0000	8	307

The results from the Air Force by quarter fixed effects panel model are shown in Table 8. In this model we include a lagged dependent variable as a regressor because without it the model exhibits first order autocorrelation as evidenced by a Durbin – Watson statistic of 0.81 which is significant at the 0.05 level.

Table 8: Air Force by Quarter Net Cost Fixed Effects Panel Model

Dependent Variable: ln(Net Costs)
Sample (adjusted): 1998Q2 2007Q4
Periods included: 39
Cross-sections included: 8
Total panel (balanced) observations: 312
Robust Standard Errors

Variable	Coefficient	Std. Error	t-Statistic	Prob.
constant	3.347538	0.825784	4.053768	< 0.001
ln(FH)	0.365704	0.061534	5.943074	< 0.001
age/365	0.021176	0.006070	3.488610	< 0.001
ln(net costs((-1))	0.588651	0.056065	10.49939	< 0.001
Effects Specification				
Cross-section fixed (dummy variables)				
R-squared	0.954134	Adjusted R-squared		0.952611

All of the predictor variables are significant at the 0.01 level of significance and are positively related to net costs. We can infer that hours have a ceteris paribus effect on net costs such that as the flying hours for the airframes analyzed increase by 1 percent, net costs increase by 0.37 percent. In addition, as age increases by one year net costs increase by 2.1 percent.

We now turn away from the fixed effects panel model which estimates a single, averaged coefficient for each MDS group to a LSDV model so that we may assess the significance of cross section specific factors by allowing for differences in intercepts. The results are displayed in Table 9. In this regression the KC-135 is the base group and the dummies represent differences in intercepts from the KC-135. Note that the coefficients for age, flying hours, and lagged net costs are all equivalent to those found in the fixed effects panel model.

Table 9: Air Force by Quarter Net cost Model w/ MDS Specific Intercepts

Dependent Variable: ln(Net Costs)
Sample (adjusted): 1998Q2 2007Q4
Periods included: 39
Cross-sections included: 8
Total panel (balanced) observations: 312
Robust Standard Errors

Variable	Coefficient	Std. Error	t-Statistic	Prob.
constant	2.304970	0.765978	3.009187	0.0028
A10	0.777641	0.158403	4.909250	< 0.001
B1B	1.647484	0.261378	6.303064	< 0.001
B52H	0.673461	0.108620	6.200173	< 0.001
C5	1.234427	0.207214	5.957271	< 0.001
F15CD	1.333260	0.232392	5.737119	< 0.001
F15E	1.495909	0.273645	5.466603	< 0.001
F16CD	1.178360	0.248631	4.739386	< 0.001
ln(FH)	0.365704	0.061534	5.943074	< 0.001
age/365	0.021176	0.006070	3.488610	< 0.001
ln(net costs((-1))	0.588651	0.056065	10.49939	< 0.001
R-squared	0.954134	Adjusted R-squared		.952611

Not only is the intercept for the KC-135 significant at the 0.01 level, but we also find that each of the other air frame's intercept is significantly different than that of the

KC-135. We can further analyze differences in cost estimating relationships between the different airframes by allowing for differences in age, flying hours, and the lagged dependent variable across airframes. However, with only 312 observations and the inclusion of interaction variables we move to a lower level of aggregation and estimate at the Air Force by month level.

At this lower level of aggregation and for this particular research question we remove the B-2A along with the B-1B because they both contain negative net costs which are not appropriate for this section of the analysis. We again start with the IPS test for unit roots and find that at this level of aggregation the null of a non-stationary series is rejected at the 0.01 level of significance. We follow the same procedures here as we did before with the Air Force by quarter level of aggregation by first estimating a common, panel model. Table 10 shows the results from the Air Force by month level of aggregation fixed effects panel model.

Table 10: Air Force by Month Net Cost Fixed Effects Panel Model

Dependent Variable: LN_NET_COSTS
Method: Panel Least Squares
Sample (adjusted): 1998M02 2007M12
Periods included: 119
Cross-sections included: 7
Total panel (balanced) observations: 833
Robust Standard Errors

Variable	Coefficient	Std. Error	t-Statistic	Prob.
constant	4.853955	0.588972	8.241406	< 0.001
ln(FH)	0.287524	0.038522	7.463947	< 0.001
age/365	0.027936	0.003751	7.448205	< 0.001
ln(net costs((-1))	0.520181	0.036847	14.11748	< 0.001
Effects Specification				
Cross-section fixed (dummy variables)				
R-squared	0.934513	Adjusted R-Squared		0.933796

The coefficients from this model are slightly different than those computed in the Air Force by quarterly level of aggregation, but calculating confidence intervals reveals that they overlap and none of the variables are significantly different over the separate levels of aggregation. The partially factored LSDV regression for the Air Force by month level of aggregation is shown in Table 11. The inferences gained from this regression mirror those from the same model at the Air Force by quarterly level of aggregation.

Table 11: Air Force by Month Net Cost Model with MDS Specific Intercepts

Dependent Variable: ln(Net Costs)
Method: Panel Least Squares
Sample (adjusted): 1998M02 2007M12
Periods included: 119
Cross-sections included: 7
Total panel (balanced) observations: 833
Robust Standard Errors

Variable	Coefficient	Std. Error	t-Statistic	Prob.
constant	3.752928	0.540911	6.938161	< 0.001
A10	0.919024	0.096377	9.535692	< 0.001
B52H	0.550033	0.071513	7.691369	< 0.001
C5	1.352341	0.120276	11.24368	< 0.001
F15CD	1.594138	0.144536	11.02933	< 0.001
F15E	1.749015	0.164101	10.65813	< 0.001
F16CD	1.542636	0.162894	9.470203	< 0.001
ln(FH)	0.287524	0.038522	7.463947	< 0.001
age/365	0.027936	0.003751	7.448205	< 0.001
ln(net costs((-1))	0.520181	0.036847	14.11748	< 0.001
R-squared	0.934513	Adjusted R-squared		0.033796

We now evaluate differences in cost estimating relationships across the MDS groups. We test for a difference between the A-10 groups and the other groups by testing the joint significance of the cross section specific interaction variables.

Given equation 12:

$$\begin{aligned}\ln\left(\text{net costs}\right)_{it} &= \beta_0 + \beta_1\ln\left(\text{FH}\right)_{it} + \beta_2\text{age}_{it} + \beta_3\ln\left(\text{net costs}\right)_{i,t-1} + \\ &\delta_0 + \delta_1\text{A10log}\left(\text{FH}\right)_{it} + \delta_2\text{A10age}_{it} + \delta_3\text{A10}\ln\left(\text{net costs}\right)_{i,t-1} + \varepsilon_{it}\end{aligned} \tag{12}$$

The null, H_0: $\delta_1 = 0$, $\delta_2 = 0$, $\delta_3 = 0$, is evaluated against the alternative that there is no difference in the effect of age, flying hour, or a lagged period of net costs on net costs. We do not include δ_0 because we already found evidence of significantly different intercepts between the cross sections and do not want to bias the joint test with the power of the intercept. First we estimate the unrestricted model from equation 12 and then estimate the restricted model where all of the A-10 interaction variables are absent. From the unrestricted and restricted models we calculate an F-statistic of 9.42 which is significant with (3, 825) degrees of freedom at the 0.01 level. Accordingly, we reject the null of similar cost estimating relationships between the A-10 and the other MDS groups. We repeat this same procedure for each of the six other MDS groups we analyze in this portion of the study and the results are summarized in Table 12.

Table 12: Summarized Results for Test of Different Cost Estimating Relationships across MDS Groups

t-test	**A10**	**B52H**	**C5**	**F15CD**	**F15E**	**F16CD**	**KC135**
Separate Constant	X	X*	X*		X*	X*	X*
ln(FH)	X*	X*	X	X*	X*	X*	X
age	X*	X*		X*	X*	X*	X*
lagged D.V.	X*	X*	X*	X*	X*	X*	X*
F-test							
Slope differences	X*	X*	X*	X*	X*	X*	X*

x = significant at the .05 level
x* = significant at the .01 level

The results show that for each group the interaction variables are individually significant at the 0.05 level except for the C-5 age interaction variable. In addition, the analysis provides further support for estimating separate cost estimating relationships for each of the groups with significant joint tests of the interaction variables for every MDS group at the 0.01 level.

While we have found statistically significant evidence to support estimating separate cost estimating relationships for each MDS group, we find it important to note the non trivial size of the differences as well. Because our models are estimated in log space it is difficult to comprehend the magnitude of the differences. For example, the LSDV regression including only intercept differences at the Air Force by month level of aggregation supports a ceteris paribus interpretation that net cost for the A-10 is 0.92 more than the KC-135 in log space. When we use these models to forecast the cost of the FH Program we are dealing with millions of dollars for any given month. One million dollars in unit space translates roughly to 13.82 in log space. Adding the 0.92 difference for the A-10 estimate in log space increases a million dollar forecast to over 2.5 million dollars in unit space. So we see that very small differences in log space can result in large differences when we back transform the forecast.

Which Variables are Significant Predictors of Flying Hour Program Cost

From the previous section we have already found that we cannot use the number of landings and sorties as we had hoped because of their high degree of correlation with flying hours. Furthermore, we have also seen various examples where flying hours, age, and a lagged dependent variable have proven to be significant predictors of net costs. Here we further explore the significance of a number of variables to predict net costs, charges, and credits for each of the MDS groups across five levels of aggregation.

We start by looking at the Air Force by month level of aggregation and estimate each MDS separately using equation 13:

$$\ln(\text{costs})_t = \beta_0 + \beta_1 \ln(\text{FH})_t + \beta_2 (\text{age})_t + \beta_3 \ln(\text{costs}_{t-1}) + \delta_i \text{season}_t + \varepsilon_{it} \quad (13)$$

Here we exclude B-1B and B-2A net cost models because they have negative net cost at this level of aggregation. The costs variable takes on the form of net cost, charges and credits to create three different models. The β_i are the normal coefficients for previously discussed variables and season is a vector of monthly dummy variables with October as the base. Using the net cost A-10 model as an example, *a priori* estimation we check for unit roots and find that we reject the null of a non-stationary series using the ADF test. The results of the ADF test for the A-10 are found in Table 13.

Table 13: A-10 ADF Test of Net Cost for Air Force by Month Level of Aggregation

Null Hypothesis: ln(Net Costs) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic based on SIC, MAXLAG=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-7.328989	<0.001
Test critical values:	1% level	-4.036983	
	5% level	-3.448021	
	10% level	-3.149135	

*MacKinnon (1996) one-sided p-values.

With evidence of net cost as a stationary series we proceed with the model estimation. The natural log of flying hours, age, and a lagged dependent variable are all significant at the 0.05 level. In addition, November and December are both significant at the 0.05 level demonstrating significant seasonal effects for the A-10. The results of the estimated model are shown below in Table 14.

Table 14: A-10 Net Cost Model for the AF by Month Level of Aggregation

Dependent Variable: ln(NET_COSTS)
Method: Least Squares
Sample: 1998M02 2007M12
Included observations: 119

Variable	Coefficient	Std. Error	t-Statistic	Prob.
constant	6.412204	1.492056	4.297562	< 0.001
ln(FH)	0.372604	0.111186	3.351177	0.0011
age/365	0.045358	0.007243	6.261917	< 0.001
ln(net costs((-1))	0.345550	0.079170	4.364640	< 0.001
Nov or Dec	-0.122463	0.049616	-2.639814	0.0112
R-squared	0.678734	Adjusted R-squared		0.664519

We use test for heteroskedasticity, serial correlation, and normality to assess the OLS assumptions and find that each of the assumptions are valid for this model. A brief discussion of the diagnostic tests and their results can be found in Appendix B. We continue the same methods to estimate separate net cost, charge, and credit models for each MDS in the Air Force by month level of aggregation. In each model we start with the full complement of regressors and eliminate non significant variables to achieve parsimonious models.

We also estimate models for the four remaining levels of aggregation, but use fixed effects panel models as described in the previous chapter. The β coefficients do not vary across cross sections and the reported constant is the average of the fixed effects. For the Air Force by quarter level of aggregation we use a common fixed effects panel model with the MDS groups representing the cross sections. For the MAJCOM by year level of aggregation we again estimate a common fixed effects panel model, but this time use MDS groups by MAJCOM as the cross sections. For the MAJCOM by quarter and by month levels of aggregation we estimate each MDS group separately, but use fixed effects panel models with the major commands as the cross sections.

Of the 56 models we created, the *a priori* tests for stationarity never failed to reject the null of non-stationary series. For the *a posteriori* diagnostic tests we rely on Newey-West Heteroskedasticity and Autocorrelation (HAC) consistent standard errors for the time series models and White robust standard errors for the panel models if either heteroskedasticity or serial correlation is present. We utilize robust standard errors for every panel model and for some of the time series models. Those time series models that required robust standard errors are reported in Appendix C along with the specific results for the remaining net cost, charge, and credit models.

Flying hours is a significant predictor in each of the 22 net cost models, 16 of the 17 charges models, and 14 of the 17 credits models. In general, the results show strong support for the use of flying hours as a predictive variable. In addition, the highest estimated flying hour coefficient is 0.724 and is from the Air Force by quarter net cost model. These results indicate that all three dependent variables are positively related, but inelastic with respect to flying hours, which we touch more on in the next section of this chapter.

The age variable is a significant predictor in 19 of the 22 net cost models, 15 of the 17 charges models, and 15 of the 17 credits models. Six of the seven times the age variable was insignificant can be attributed to the C-5 and the other event can be attributed to the F-15E in the MAJCOM by quarter net cost model. Holding the number of flying hours and lagged net cost constant the effect of a one year increase in age on net cost ranges from a 2.7 percent increase in net cost for the F-15 Air Force by month model to a 9.3 percent increase in net cost for the B-52H Air Force by month model. The ceteris paribus effect of age on net cost for both the Air Force by quarter and MAJCOM by year levels of aggregation support an average increase in net cost of 3.4 percent for a one year increase in age.

Next we find that using a lagged dependent variable to predict itself is significant in every model except the KC-135 MAJCOM by quarter charges model. In addition to being a significant predictor, including a lagged dependent variable also helps meet OLS assumptions in many cases. Net cost, charges, and credits are all positively related, but inelastic for every model with respect to their own lagged values. We also created credit models with lagged charges instead of lagged credits with the idea that higher costs might be associated with higher numbers of parts purchased. As the purchased parts break more opportunities for credits should arise as the broken parts are refurbished. In the end the lagged credit variable dominated the lagged charges variable in the credit models.

Finally, the results support significant seasonal effects at the monthly level of aggregation. Each of the significant months has a small negative effect on the dependent variable in log space, but as we have seen in earlier sections the difference can result in large dollar amounts in unit space. November is significant more often than any other month with significance in 57 percent of the monthly models. Aside from November, September and December are also significant in many of the monthly models. Past studies have attributed seasonal effects to the fiscal cycle, but we hypothesize that November and December are commonly significantly less than other months because of the national holidays observed by the Air Force. With the exception of combat areas many bases virtually shut down during Thanksgiving and Christmas. Therefore, it makes sense that those months would be associated with lower costs.

In summary, we have found the log of flying hours, age, lagged cost variables, and seasonal effects to be significant predictors of the various types of cost we analyze. The significance and directional relationship of the discussed independent variables with the dependent variables is generally robust to all levels of aggregation we assessed.

Evaluating the Appropriateness of the Proportional Model Specification

We now turn to testing the validity of the proportional model. We rely on the net cost models generated in the previous section to test the assumption of the proportional model which assumes that every hour flown increases FH Program costs by the same amount.

The null hypothesis is H_0: $\beta_{FH} = 1$ and failure to reject the null would support the specification of the proportional model because we would experience unit elasticity between flying hours and costs. In other words, a one percent increase in flying hours would result in a ceteris paribus one percent increase in net cost. To understand why this test works we first assume that the age variable is log transformed like the other variables. Transforming the age variable does not affect the other coefficients, but does change the size and interpretation of its own coefficient as well as the size of the intercept.

The proportional model's specification states that holding other variables constant, as we increase flying hours by one unit, net cost increase by some constant factor. Back transforming our previous models from log space to unit space we achieve the multiplicative model shown in equation 14:

$$\text{net cost}_t = (e^{\beta 0})(\text{FH}_t^{\beta 1})(\text{age}_t^{\beta 2})(\text{net costs}_{t-1}^{\beta 3}) \qquad (14)$$

Evaluating equation 14 we see that if we hold all else constant and ask ourselves how net cost change as we increase flying hours by one hour we find that the model basically reduces to net costs = $\omega(\text{FH}^{\beta 1})$, where the other terms are held constant and based on the distributive property their product can be reduced to ω. It is apparent that the estimate of β_{FH} will determine how costs change as we increase flying hours. If $\beta_{FH} = 1$, then a one unit increase from any number of flying hours will increase costs proportionally by ω. If $\beta_{FH} < 1$ then a one unit increase in flying hours will increase costs, but it does so at a diminishing rate relative to the amount of flying hours flown and to be flown. For example, if β_{FH} was estimated to be 0.5 and we increase the number of flying hours from 10 to 11 we find that cost would increase by 0.154ω. If we again increase the number of flying hours by one unit, but this time the increase is from 11 to 12 we find that costs increase by 0.147ω. We use the parsimonious models discussed in the previous section to obtain a ceteris paribus effect of flying hours on net cost and to reduce third variable bias. The confidence intervals around β_{FH} for each MDS are shown in Table 15.

Table 15: Summarized Test Results of Proportional Model FH Assumption

MD	FH	S.E.	99% C.I.
A-10	0.37	0.11	0.08 - 0.66
B-52H	0.30	0.08*	0.09 - 0.52
C-5	0.18	0.06*	0.01 - 0.35
F-15C/D	0.39	0.10	0.13 - 0.65
F-15E	0.50	0.16*	0.08 - 0.93
F-16C/D	0.42	0.10	0.15 - 0.69
KC-135	0.38	0.08	0.16 - 0.60

119 Degrees of Freedom
*Newey-West HAC robust standard errors estimated

Based on the confidence intervals it is obvious that we reject the null of a proportional relationship for each MDS. The F-15E is the only MDS that comes close to unit elasticity in the 99 percent confidence interval. We repeat this same procedure for the MAJCOM by quarter and the MAJCOM by month levels of aggregation, but use panel models with fixed effects to combine the MAJCOM cross sectional effects for each MDS group like we did in the previous section. These two are the only other levels of aggregation beside the AF by month level that allow us to estimate MDS specific flying hour coefficients with an arguably sufficient number of observations to carry out the tests. The results from these two levels of aggregation can be found in Appendix D and are robust with respect to level of aggregation. Of the 22 confidence intervals calculated over the various levels of aggregation, only the F-15C/D in the MAJCOM by quarter level of aggregation fails to provide significant evidence against the proportional model. Furthermore, it only fails to do so at the 0.1 level of significance. In short, our results provide substantial evidence against the proportional model assumption that net cost increase by a constant factor with respect to flying hours.

Forecasting Performance of the Net Cost and Charges minus Credits Models

Forecasts for both the net cost and charges minus credits models are dynamic one step ahead forecast for each level of aggregation below the annual level. By dynamic we mean that forecasts use actual lagged cost values for the first forecasted period, but then use forecasted values for lagged cost values in further forecasts until the entire year is estimated. The MAJCOM by year forecast is a static one step ahead forecast. The forecast period is from Fiscal Years 2005 through 2007. We evaluate the accuracy of the forecast using the RMSE of each MDS group at two levels. First we evaluate the accuracy of each forecast at the level of aggregation corresponding to the data used in the models formulation. This evaluation simulates each model's ability to forecast within the year of execution. Second we aggregate lower levels of aggregation to annual levels. This aggregated evaluation simulates the models' ability to forecast for an entire budget year. In addition to evaluating the forecasting accuracy of the individual MDS groups, we also aggregate the annual forecasts into overall forecasts which represent the projected costs of our sample Air Force fleet. Both annual and lower level analysis are important because the Air Force has a need to create annual budgets, but also conducts marginal analysis when executing within a fiscal year.

The results show that each model has benefits and disadvantages, but we are unable to conclusively state that one model is superior. The net cost model has a lower RMSE for individual MDS groups 62 percent of the time when comparing forecasts associated with less than annual forecasts. In addition, the net cost model has a lower RMSE for the individual MDS groups 70 percent of the time when comparing annual forecast. However, the charges minus credits model has a lower RMSE in three of the four forecast comparisons of overall annual estimates when combining MDS groups to simulate the fleet's budget.

Thus far we have only discussed forecasting accuracy based on the relative size of the forecast errors of the competing models. With this in mind, it is quite possible that the RMSE of one forecast is only trivially better than the other.

Figure 12 shows the annual forecasts from our competing models and the actual net costs based on the Air Force by month level of aggregation as the models' underlying data.

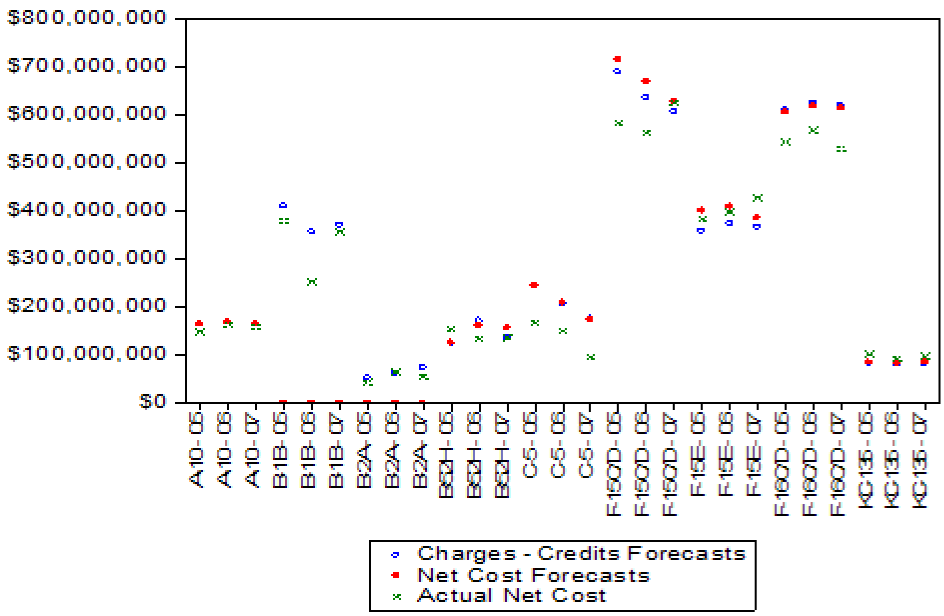


Figure 12: Comparison of Non-Proportional Model Annual Forecasts

Figure 12 helps to illustrate that while one model may have a lower RMSE, it appears as though the forecast from each models are generally very close. Therefore, we rely on the modified DM statistic to determine if the difference in forecast is due to chance rather than being significantly different than zero. Of the 35 instances in which the net cost model had a lower RMSE the difference was only significant 5 times. Four times for individual MDS specific forecast at lower than annual levels and once for individual MDS specific forecast at the annual forecast level. Of the 20 times the charges minus credits model had a lower RMSE the difference was significant 4 times. Twice for less than annual MDS specific forecast and twice for the overall annual forecast. Table 16 displays the RMSE and instances where the forecast were significantly better for a small subset of the forecasts. In Appendix E we provide results for RMSE for all the forecasts.

Table 16: Subset of Forecast Accuracy Results for Net Costs and Charges minus Credits Models

			RMSE Net Cost Model		RMSE Charges - Credits Model
AF by Month, Monthly Forecasts	A-10	✓	**1,994,520**		2,140,152
	B-1B				7,640,466
	B-2A				2,356,663
	B-52H		**3,022,145**		3,085,837
	C-5		6,497,772		**6,459,446**
	F-15C/D		9,670,983	✓	**8,310,215**
	F-15E		**5,660,208**		6,166,468
	F-16C/D		8,692,376		**7,937,276**
	KC-135	✓	**1,476,716**		1,609,425
AF by Month, Annual Forecasts	A-10		**9,750,849**		9,965,914
	B-1B				62,991,563
	B-2A				11,751,064
	B-52H		**25,320,516**		27,275,792
	C-5		71,903,651		**71,411,121**
	F-15C/D		97,669,353		**75,417,025**
	F-15E		**27,494,330**		40,958,511
	F-16C/D	✓	**66,727,579**		72,299,307
	KC-135		**13,428,595**		15,379,886
	Overall*		220,936,643	✓	**163,221,986**

Bold values represent lower RMSE between competing models and checked values represent where differences in forecasts are statistically significant
*B-1B and B-2A not included in overall forecast for each year

Based on these results the net cost model appears to do better with MDS specific forecast, however only a handful of its differences are significantly better than the charges minus credits model. From the standpoint of estimating the overall annual budget the charges minus credits model has a slight advantage with two significantly better forecasts in a total of only four calculated overall forecasts. In addition, in many instances we were unable to create a net cost model for the B-1B and the B-2A because of the presence of negative net costs and our model specification. The charges minus credits model conveniently eradicates this problem by forecasting charges and credits separately to arrive at a net cost figure, but does create extra effort for forecasters because it requires extra forecasts. For these reasons we conclude that there is insufficient evidence of preferring either the net cost or charges minus credits model in all circumstances and that the benefits and disadvantages of each model may help forecasters decide which to use.

Proportional versus Non-Proportional Model Forecasting Performance

Because we were unable to determine which of the non-proportional models has superior forecasting ability we compare both of them to a proportional model based on published factors and a proportional model based on factors calculated from our data set. The proportional model based on factors calculated from our data set has the lowest RMSE 35 out of the 61 forecasts. Only three of those forecasts are statistically significantly better than both of the non-proportional models' forecasts, however. In general, the non-proportional models performed better than the proportional model based on the published factors by outperforming for 45 of the 61forcasts. Out of the 45 times the non-proportional model was better the difference in forecast was only significant 9 times. The full complement of forecasting accuracy results and the comparison between each model for each forecast level can be found in Appendix F. Based on the general insignificance of the different forecasting models we again conclude that we are unable to confidently claim the superiority of one method over another. In addition, we are unable to determine trends in which one model is better than the other as we had previously suggested in the analysis of the net cost versus charges minus credits model.

We did find that in general the non-proportional models we created over predict and the proportional model calculated from our data set under predicts net cost. Figure 13 plots three of the competing models and the actual net cost for individual MDS specific forecast using the AF by month as the underlying data.

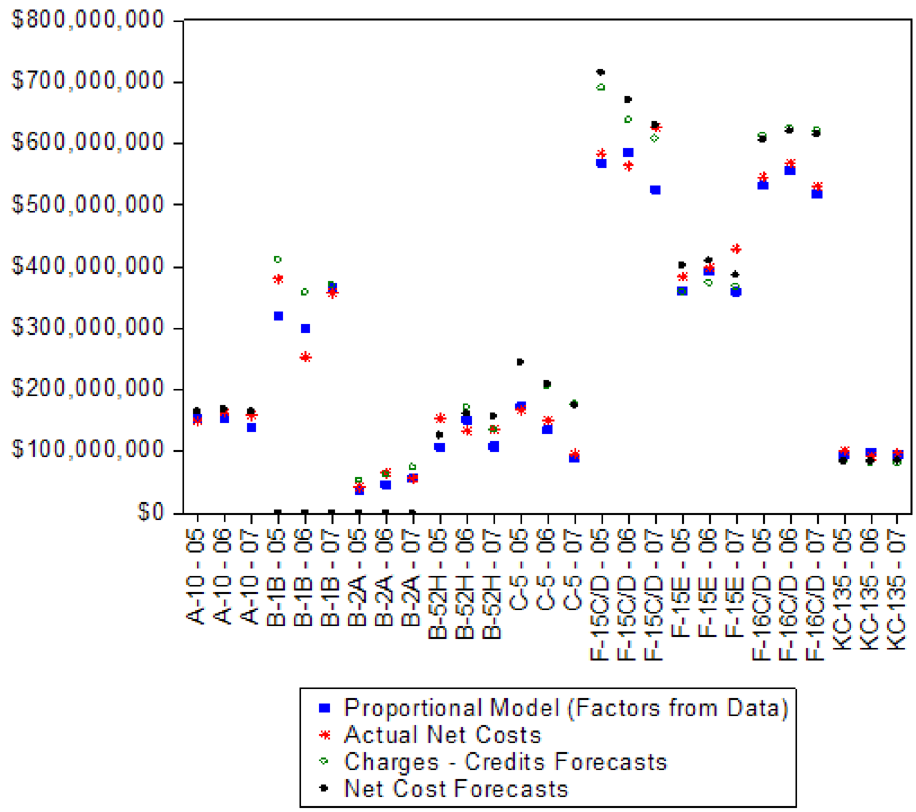


Figure 13: Comparison of Annual Forecast from Proportional and Non-Proportional Models

For this forecast level the non-proportional models over predict cost roughly 60 percent of the time and the proportional models under predict roughly 74 percent of the time. We would expect this to occur for the proportional models because the actual flying hours have generally decreased over the years between Fiscal Years 2003 and 2007 as evidenced in Figure 14.

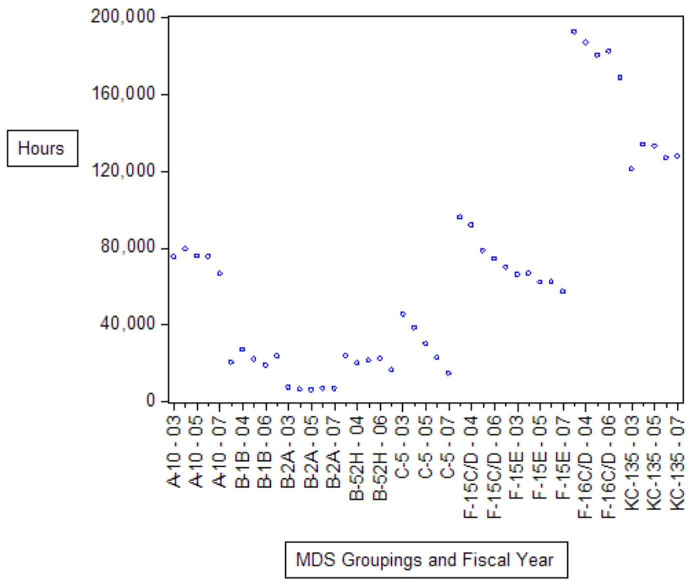


Figure 14: Flying Hours from FY03 to FY07

Based on this point and the inability to determine any single model as a superior forecasting model, we believe that the two models can potentially be used in conjunction with each other, using one as a floor and the other as a ceiling. Though, this prospect should only be considered in the event that flying hours continue to decrease. If flying hours either remain the same or increase then the prediction error of the proportional model will also change. For example, if the number of hours that are projected to be flown in Fiscal Year 2008 are greater than the average annual hours flown in Fiscal Years 2006 and 2007 the proportional models will most likely over predict FH Program costs.

Chapter Summary

In this chapter we have answered each of our research questions. First, we found evidence to support the estimation of cost relationships separately for each MDS group. As a result we then created various cost estimating relationships for each MDS group at various levels of data aggregation. We found that flying hours, age, and past cost data are generally significant predictors of cost. In addition, we found some evidence of a holiday seasonal effect on cost. We then used the estimated parameters for the flying hour variable and determined that the cost elasticity with respect to flying hours is inelastic, supporting a non-proportional relationship between flying hours and cost. Finally, we compared the forecasting accuracy of non-proportional net cost and charges minus credits models against each other and against two proportional models that follow the same methodology of the Air Force's current FH Program estimating model. In the next chapter we use our findings to discuss policy implications and discuss the strengths and limitations of this study.

Chapter V: Conclusions

In this chapter we highlight the strengths and limitations of our findings in an effort to guide further research in this area. In addition, we discuss how our findings can potentially result in policy implications.

Strengths, Limitations, and Policy Implications

Because the FH Program is a central and reoccurring piece of the Air Force's annual budget and because the Air Force needs seek out the most efficient means of managing its resources, we evaluated the Air Force's current method of forecasting costs associated with the FH Program. After conducting a literature review to identify what previous research had to offer, we built econometric models using various OLS regression techniques to ascertain which variables are significantly related to FH Program costs and how those costs should be predicted.

Some of the previous literature had created models similar to ours, but different in the point that they were so top-level that they missed the innate differences between the different types of aircraft. However, these very top level models have merit in the sense that they are very easy to use because one model captures every airframe. Savings in manpower and ultimately dollars might result from the reduced amount of effort required to estimate the common models. However, we show empirically that the relationship between cost and predictor variables is not equal across different types of airframes. Budget estimates may take more time in the estimation process, but ultimately time and money would be saved as a result of more efficient budget estimates when estimating different types of airframes separately.

Of the many variables evaluated by previous research flying hours and age appeared to occur and be significant most often. We also found flying hours and age to be positively related significant predictors of FH Program costs. In addition, we found a holiday seasonal effect where costs are generally lower in the months of November and December. As expected we found past values of costs to be positively related to current values of costs. Our findings are very robust in that we estimate many of our models over many different levels of aggregation and for every level of aggregation the results are generally the same. Our forecasting models are intuitive. It stands to reason that costs should increase as airframes get older and as they get used more. Based on limitations of our modeling techniques we were not able to control for the dynamic relationship found in some of the previous literature, however. For reasons discussed in Chapter II, it would be beneficial to forecast costs based on the ability to control for the interconnected relationship between flying hours, landings, and sorties. It may be possible to use factor analysis and produce a linear combination of the three variables into a single usage variable. In this way it might be possible to control for their dynamic relationship with costs and how it differs across airframes.

Previous literature had conflicted findings with respect to the relationship between flying hours and costs. We provided empirical evidence supporting a non-proportional relationship. Based on our findings the Air Force should reevaluate not only the way it budgets the FH Program, but also how it conducts marginal analysis for changes to execution year flying hour amounts. The non-proportional relationship is positive, but diminishing so that the effect of additional hours on cost dwindles as more hours are flown. Along those same lines, our models assume that the non-proportional relationship between hours and cost starts anew every fiscal year. In reality, the non-proportional relationship probably follows a lifetime curve and does not start over every fiscal year. A bright spot for further research might be to evaluate cumulative hours flown over a lifetime versus hours flown in a given period. Nonetheless, the Air Force may want to alter their forecasting method so that they can control for the variables we have found to be significant predictors of FH Program Costs and so that they can control for the non-proportional relationship between flying hours and costs.

Previous research that delved into lower levels of aggregation revealed limitations to the log-log model specification that we so highly touted. Based on occurrences in which the co-movement of charges and credits veered from its normal relationship to create outlier and negative observations we produced a viable work around by forecasting charges and credits separately. We found previous manipulations that attempted to account for negative net costs biased coefficients and forecasts. Based on tests for forecast accuracy we found the net cost model to generally perform no better than our charges minus credits model. A limitation of our charges minus credits model is that we use very simplistic models which estimates the charges and credits separately then add the point estimates together. As a result our charges minus credits model misses a portion of the variance between charges and credits. Further research may improve upon the charges minus credit model by using more elegant techniques such as simultaneous or systems methods that estimate the parameters for each of the dependent variables at the same time.

We finalized our research efforts by testing the forecasting accuracy of four competing models. We not only used loss functions like the previous literature, but we also added a new component to it with the modified DM statistics to test if the difference in the forecast errors were statistically different from zero. However, our results were mixed and we were unable to determine a single superior forecasting model. In Chapter IV we suggested that the proportional and non-proportional models be used in conjunction to obtain a ceiling and possible floor for the budget. With that information Air Force leaders would better understand the risk inherent in the annual budget. A major limitation of our forecast tests was that we could not obtain actual CPFH factors used to estimate the President's budget. We suggest that the Air Force initiate moratoriums on FH Program cost estimates. The moratoriums would only require minimal work and with it valuable information as to the true accuracy of the Air Force's model could be learned. In addition, if the factors also contained information about the underlying data used in its calculation researchers would easily be able to compare forecasting methods and be better capable of making suggestions as to which model the Air Force should use. In the future, if multiple methods are used to estimate costs and moratoriums performed the Air Force would gain a clear understanding as to which model is a better tool.

Finally, as part of the CAM initiative, in Fiscal Year 2009 the Air Force changed the way the Material and General Support Divisions of the Air Force Working Capital Fund are reimbursed for FH Program Supplies. In short, rather than paying for parts at the base level, the charges and credits are now managed centrally. Air Force leaders expect the transition to be seamless and urge maintenance organizations to continue moving parts through the repair cycle in a timely manner, limit purchases to items and quantities required to accomplish the mission, and continue to utilize the AFREP program. We question the idea of a seamless transition based on the lack of incentives provided for the base level maintenance organizations. Prior to CAM these organizations were incentivized to repair authorized parts in cost effective ways because they were often able to utilize AFREP credits to fund other items. In addition, the base level maintenance organizations were limited by their base level budget. With funds now managed at a centralized location it is extremely important for the Air Force to ensure the base level organizations are fiscally responsible. We hypothesize that FH Program Costs will potentially increase as a result of a lack of incentives for maintenance organizations to find ways to generate credits through the AFREP program. After the centralized funding process has been in place for a period of time we suspect a policy analysis of the impact that centralization has on FH Program costs would be interesting.

Appendix A: Cost Allocation Mismatches for Majcom and Base Levels of Aggregation

MDS Group	Level of Aggregation	Total Net Costs	Net Costs Not Matched	% Net Costs Mismatch	Delta Between MAJCOM and Base Level of Aggregation
A-10	MAJCOM	$1,234,938,492	$252,368	0.02%	
A-10	Base	$1,234,938,492	$193,258,475	15.65%	15.63%
B-1B	MAJCOM	$2,586,551,422	$602,333	0.02%	
B-1B	Base	$2,586,551,422	$505,346,681	19.54%	19.51%
B-2A	MAJCOM	$261,733,496	$30,894	0.01%	
B-2A	Base	$261,733,496	$72,335	0.03%	0.02%
B-52H	MAJCOM	$946,183,323	$168,393	0.02%	
B-52H	Base	$946,183,323	$125,690,980	13.28%	13.27%
C-5	MAJCOM	$2,061,928,351	$908,760	0.04%	
C-5	Base	$2,061,914,922	$185,857,210	9.01%	8.97%
F-15C/D	MAJCOM	$5,254,344,780	$51,894	0.00%	
F-15C/D	Base	$5,254,344,780	-$2,312,609	-0.04%	-0.05%
F-15E	MAJCOM	$3,404,238,774	$159,561	0.00%	
F-15E	Base	$3,404,238,774	$175,259,700	5.15%	5.14%
F-16C/D	MAJCOM	$4,632,602,156	$143,760	0.00%	
F-16C/D	Base	$4,632,602,156	$43,328,047	0.94%	0.93%
KC-135	MAJCOM	$757,826,422	$4,078,703	0.54%	
KC-135	Base	$757,826,422	$168,214,983	22.20%	21.66%

Appendix B: Sample of Time Series Regression Diagnostic Tests

A-10 Net Cost Model White Test for Heteroskedasticity

White's test for heteroskedasticity tests for an unknown general form of non constant variance. The null hypothesis is that the residual values are homoskedastic. Based on the F-statistic with (3,115) degrees of freedom the test fails to reject the null of constant variance.

Heteroskedasticity Test: White

F-statistic	1.301598	Prob. F(3,115)	0.2774
Obs*R-squared	3.907919	Prob. Chi-Square(3)	0.2716
Scaled explained SS	3.340706	Prob. Chi-Square(3)	0.3420

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Sample: 1998M02 2007M12
Included observations: 119

Variable	Coefficient	Std. Error	t-Statistic	Prob.
constant	0.095633	0.146655	0.652095	0.5156
(ln(FH))^2	0.000774	0.001242	0.623514	0.5342
(AGE/365)^2	-8.63E-06	3.51E-05	-0.246099	0.8060
(ln(NET_COSTS(-1)))^2	-0.000491	0.000494	-0.993796	0.3224

White's test for heteroskedasticity is repeated for each time series regression. For the panel models we use a test suggested by Kennedy that tests for differences between the variance of the cross sections. A full explanation of this test is detailed by Kennedy (2008:295).

A-10 Net Cost Model Breusch-Godfrey LM Test for Serial Correlation

The null hypothesis of the Breusch-Godfrey Serial Correlation LM Test is that the residual values are not serially correlated. Based on the low F-statistic with (6,109) degrees of freedom there is very little evidence against the null and so we fail to reject that the residuals are not serially correlated.

Breusch-Godfrey Serial Correlation LM Test:

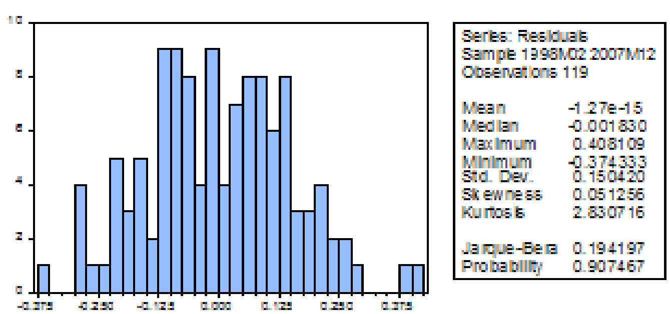
F-statistic	0.794950	Prob. F(6,109)	0.5758
Obs*R-squared	4.988975	Prob. Chi-Square(6)	0.5452

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Sample: 1998M02 2007M12
Included observations: 119

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.697350	3.056231	-0.555374	0.5798
ln(FH)	0.019932	0.110721	0.180019	0.8575
AGE/365	-0.006619	0.014153	-0.467677	0.6409
ln(NET_COSTS(-1))	0.102610	0.202948	0.505597	0.6142
RESID(-1)	-0.147179	0.226205	-0.650643	0.5166
RESID(-2)	-0.068558	0.120490	-0.568994	0.5705
RESID(-3)	0.039727	0.098049	0.405171	0.6861
RESID(-4)	0.180702	0.095306	1.896024	0.0606
RESID(-5)	0.047861	0.097602	0.490372	0.6249
RESID(-6)	0.044171	0.097831	0.451508	0.6525

A-10 Net Cost Model Jarque-Bera Test for Normality

Null hypothesis is that the residuals come from a normal distribution. The P-Value of 0.907 associated with the Jarque-Bera statistic provides virtually no evidence against the null and so we fail to reject that the residuals are normally distributed.



Appendix C: Summary of Regression Coefficients for All Models

Summarized Regression Coefficients for MDS Specific Net Cost Models

Dependent Variable: ln(net cost)

Level of Aggregation	Independent Variables	A-10	B-52H	C-5	F-15C/D	F-15E	F-16C/D	KC-135
Air Force by Month (Time Series)	constant	6.412**	4.938**	4.748**	4.358**	9.242**	8.024**	7.166**
	ln(FH)	0.373**	0.303**	0.176**	0.390**	0.505**	0.419**	0.383**
	age/365	0.045**	0.093**		0.046**	0.035**	0.041**	0.032**
	$\ln(\text{net costs})_{t-1}$	0.346**	0.307**	0.632**	0.506**	0.19*	0.287**	0.234**
	February		-0.122*				-0.167**	
	April						-0.129*	
	May		-0.124*				-0.132**	
	September		-0.220**					
	November	-0.128*	-0.129**	-0.124*	-0.153*		-0.197	
	December	-0.133*					-0.134*	
	n	119	119	119	119	119	119	119
	Adjusted R^2	0.665	0.838	0.538	0.698	0.358	0.575	0.621
MAJCOM by Quarter (Panel Model)	constant	6.544**		3.126**	7.557**	10.571**	8.620**	8.003**
	ln(FH)	0.394**		0.531**	0.491*	0.295	0.436**	0.366**
	age/365	0.045**			0.061**		0.047**	0.034**
	$\ln(\text{net costs})_{t-1}$	0.316**		0.556**	0.251**	0.146	0.228**	0.168**
	ln(FH)*PACAF					0.484**		
	(age/365)*ACC					0.074**		
	# cross sections	3		2	4	3	4	4
	n	117		77	156	116	156	133
	Adjusted R^2	0.964592		0.97689	0.898011	0.841344	0.89749	0.951157
MAJCOM by Month (Panel Model)	constant	8.37**	4.938**	8.263**	7.061**	9.470**	8.627**	6.632**
	ln(FH)	0.348**	0.303**	0.282**	0.443**	0.408**	0.363**	0.286**
	age/365	0.058**	0.093**		0.058**	0.027**	0.050**	0.053**
	$\ln(\text{net costs})_{t-1}$	0.174**	0.307**	0.333**	0.292**	0.197**	0.243**	0.230**
	February		-0.122*				-0.174**	-0.119*
	April						-0.130**	
	May		-0.124*				-0.098**	
	July							-0.114*
	September		-0.220**				-0.101*	-0.171*
	November	-0.118*	-0.129**		-0.157*		-0.185**	
	December	-0.141*					-0.191**	
	# cross sections	3	1	2	3	3	4	2
	n	357	119	235	357	352	476	238
	Adjusted R^2	0.924	0.838	0.927	0.700	0.719	0.818	0.923

*significant at the 0.05 level

**significant at the 0.01 level

Variables not significant at the 0.1 level are not included to achieve parsimonious models

Summarized Regression Coefficients for MDS Specific Charges Models

Dependent Variable: ln(charges)

Level of Aggregation	Independent Variables	A-10	B-1B	B-2A	B-52H	C-5	F-15C/D	F-15E	F-16C/D	KC-135
Air Force by Month (Time Series)	constant	6.094**	4.072**	9.471**	1.584	4.368**	3.035**	6.941**	10.311	6.416**
	ln(FH)	0.427**	0.249**	0.17†	0.300**	0.127**	0.481**	0.452**	0.446**	0.320**
	age/365	0.044**	0.018**	0.130**	0.050**		0.051**	0.044**	0.050**	0.031**
	$\ln(\text{charges})_{t-1}$	0.349**	0.649**	0.238**	0.326**	0.686**	0.546**	0.370**	0.147**	0.346**
	$\ln(\text{charges})_{t-2}$				0.316**					
	May		-0.137**		-0.149**					
	September				-0.186**					
	November	-0.114*	-0.151**			-0.128*	-0.142**		-0.1*	-0.090*
	December	-0.125**								
	n	119	119	119	118	119	119	119	119	118
	Adjusted R^2	0.697	0.756	0.653	0.899	0.600	0.805	0.647	0.543	0.804
MAJCOM by Quarter (Panel Model)	constant	6.440**				-0.003	3.787**	5.441**	9.818**	10.881**
	ln(FH)	0.314*				0.392*	0.478**	0.553**	0.368**	0.250*
	age/365	0.041**					0.051**	0.015*	0.048**	0.057**
	$\ln(\text{charges})_{t-1}$	0.383**				-0.335**	0.508**	0.425**	0.208*	
	# cross sections	3				2	4	3	4	5
	n	117				75	156	116	156	179
	Adjusted R^2	0.970				0.155	0.946	0.880	0.923	0.955

*significant at the 0.05 level

**significant at the 0.01 level

Variables not significant at the 0.1 level are not included to achieve parsimonious models

† P value of 0.17

Summarized Regression Coefficients for MDS Specific Credits Models

Dependent Variable: ln(credits)

Level of Aggregation	Independent Variables	A-10	B-1B	B-2A	B-52H	C-5	F-15C/D	F-15E	F-16C/D	KC-135
Air Force by Month (Time Series)	constant	4.682**	5.597**	6.723**	1.916*	8.361**	2.709	3.289*	8.704**	4.327**
	ln(FH)	0.593**	0.279**	0.125†	0.337**	0.132**	0.647**	0.379**	0.361*	0.264**
	age/365	0.043**	0.013*	0.307**	0.075**		0.071**	0.040**	0.047**	0.029**
	$\ln(\text{credits})_{t-1}$	0.283**	0.534**	0.309**	0.507**	0.412**	0.440**	0.595**	0.228*	0.481**
	September	-0.147*			-0.181*					
	November						-0.095*			-0.121**
	December									
	n	119	119	119	119	119	119	119	119	119
	Adjusted R^2	0.411	0.504	0.435438	0.801	0.292	0.748	0.746	0.386	0.714
MAJCOM by Quarter (Panel Model)	constant	5.606**				4.715**	2.745**	3.874**	12.775**	5.102**
	ln(FH)					0.184*	0.502**	0.442**		0.421**
	age/365	0.031**					0.057**	0.021*	0.047**	0.046**
	$\ln(\text{credits})_{t-1}$	0.510**				0.606**	0.527**	0.546**	0.172*	0.281**
	Qtr 1						0.112**			
	Qtr 2									-0.175**
	Qtr 3							-0.149*		
	# cross sections	3				2	4	3	4	5
	n	117				77	156	116	156	174
	Adjusted R^2	0.907				0.964	0.933	0.854	0.843	0.926

*significant at the 0.05 level
**significant at the 0.01 level
Variables not significant at the 0.1 level are not included to achieve parsimonious models
† P value of 0.18

Summarized Regression Coefficients for Common Panel Net Cost Models

Dependent Variable: ln(net cost)

Independent Variables	Air Force by Quarter (MDS Groups as Cross Sections)	MAJCOM by Year (MAJCOM by MDS Group as Cross Section)
constant	5.991**	4.188**
ln(FH)	0.373**	0.724**
age/365	0.034**	0.034**
$\ln(\text{net cost})_{t-1}$	0.386**	0.323**
ln(FH)*B1B	0.506*	
(age/365)*B-52H	0.048**	
(age/365)*C-5	-0.070**	
# Cross Sections	8	21
n	312	184
Adjusted R^2	0.961	0.976

*significant at the 0.05 level
**Significant at the 0.01 level

Summarized Regression Coefficients for Common Panel Charges Models

Dependent Variable: ln(charges)

Independent Variables	Air Force by Quarter (MDS Groups as Cross Sections)	MAJCOM by Year (MAJCOM by MDS Group as Cross Section)
constant	5.896**	4.986**
ln(FH)	0.331**	0.663**
age/365	0.029**	0.043**
$\ln(\text{charges})_{t-1}$	0.512**	0.323**
(age/365)*B-52H	0.040**	
(age/365)*C-5	-0.046**	
(age/365)*B-2A	0.141**	
$\ln(\text{charges})_{t-1}$*B-2A	-0.527**	
# Cross Sections	9	24
n	351	211
Adjusted R^2	0.980	0.986

*significant at the 0.05 level

**Significant at the 0.01 level

Summarized Regression Coefficients for Common Panel Credits Models

Dependent Variable: ln(credits)

Independent Variables	Air Force by Quarter (MDS Groups as Cross Sections)	MAJCOM by Year (MAJCOM by MDS Group as Cross Section)
constant	6.498**	5.171**
ln(FH)	0.341**	0.552**
age/365	0.039**	0.044**
$\ln(\text{credits})_{t-1}$	0.405**	0.335**
(age/365)*B-2A	0.102*	
ln(FH)*B-2A	-.733*	
(age/365)*B-52H	0.053**	
(age/365)*C-5	-.029*	
# Cross Sections	9	24
n	351	211
Adjusted R^2	0.960	0.973

*significant at the 0.05 level

**Significant at the 0.01 level

The Air Force by month time series models that failed to pass either a test for heteroskedasticity or serial correlation and used robust standard errors include:

1. B-52H net cost model
2. C-5 net cost model
3. F-15 net cost model
4. B-1B charges model
5. B-52H charges model
6. C-5 charges model
7. F-16C/D charges model
8. B-1B credits model
9. C-5 credits model
10. F-15C/D credits model
11. F-15E credits model

Appendix D: Tests for Proportional Model FH Assumption

MAJCOM by Year Level of Aggregation

	FH	S.E.*	99% C.I.	n
Average Effect	0.72	0.07	0.55 - 0.90	184

*robust standard errors estimated

MAJCOM by Quarter Level of Aggregation

MD	FH	S.E.	99% C.I.	n
A-10	0.39	0.13	0.06 - 0.73	117
C-5	0.53	0.09	0.29 - 0.77	77
F-15C/D	0.49	0.20	-0.02 - 1.01	119
F-15E**	0.29	0.16	-0.13 - 0.72	116
F-16C/D	0.44	0.09	0.19 - 0.68	156
KC-135	0.37	0.12	0.05 - 0.69	133

*robust standard errors estimated
**The F-15E coefficient only includes ACC and USAFE. A Wald Test shows that the coefficient for the F-15E PACAF FH coefficient is still significantly different than 1 at the 0.01 level with a F-statistic of 10.6 and (1, 109) degrees of freedom.

MAJCOM by Month Level of Aggregation

MD	FH	S.E.*	99% C.I.	n
A-10	0.35	0.08	0.14 - 0.55	357
B-52H	0.30	0.08	0.09 - 0.52	119
C-5	0.28	0.06	0.12 - 0.44	235
F-15C/D	0.44	0.08	0.22 - 0.66	357
F-15E	0.41	0.09	0.18 - 0.64	352
F-16C/D	0.36	0.07	0.17 - 0.55	476
KC-135	0.29	0.07	0.11 - 0.46	238

*robust standard errors estimated

Appendix E: Summary of Forecast Accuracy for Net Cost and Charges-Credits Models

AF by Month, Monthly Forecasts

Aircraft		RMSE Net Cost Model		RMSE Charges - Credits Model
A-10	✓	**1,994,520**		2,140,152
B-1B				7,640,466
B-2A				2,356,663
B-52H		**3,022,145**		3,085,837
C-5		6,497,772		**6,459,446**
F-15C/D		9,670,983	✓	**8,310,215**
F-15E		**5,660,208**		6,166,468
F-16C/D		8,692,376		**7,937,276**
KC-135	✓	**1,476,716**		1,609,425

MAJCOM by Quarter, Quarterly Forecasts

Aircraft		RMSE Net Cost Model		RMSE Charges - Credits Model
A-10	✓	**4,490,037**		4,804,069
B-1B				
B-2A				
B-52H				
C-5		**5,487,800**		7,252,261
F-15C/D		**18,610,011**		21,714,398
F-15E		44,037,585		**19,914,938**
F-16C/D		**16,597,783**		20,881,231
KC-135		5,177,654		**3,779,915**

AF by Month, Annual Forecasts

Aircraft		RMSE Net Cost Model		RMSE Charges - Credits Model
A-10		**9,750,849**		9,965,914
B-1B				62,991,563
B-2A				11,751,064
B-52H		**25,320,516**		27,275,792
C-5		71,903,651		**71,411,121**
F-15C/D		97,669,353		**75,417,025**
F-15E		**27,494,330**		40,958,511
F-16C/D	✓	**66,727,579**		72,299,307
KC-135		**13,428,595**	✓	**15,379,886**
Overall**		**220,936,643**	✓	**163,221,986**

MAJCOM by Quarter, Annual Forecasts

Aircraft		RMSE Net Cost Model		RMSE Charges - Credits Model
A-10		**14,151,324**		15,828,289
B-1B				
B-2A				
B-52H				
C-5		**18,116,527**		24,028,692
F-15C/D		**66,716,950**		73,193,911
F-15E		171,083,891		**68,351,936**
F-16C/D		**51,183,330**		72,284,989
KC-135		24,342,667		**15,192,484**
Overall***		185,693,085		**106,829,975**

AF by Quarter, Quarterly Forecasts

Aircraft		RMSE Net Cost Model		RMSE Charges - Credits Model
A-10		**2,741,825**		3,302,585
B-1B		11,317,884		**10,160,755**
B-2A				4,126,041
B-52H		**5,627,362**		6,080,072
C-5		9,130,564	✓	**6,124,526**
F-15C/D		18,679,760		**16,206,196**
F-15E		**12,540,912**		16,233,380
F-16C/D		**18,223,106**		18,882,096
KC-135	✓	**1,840,534**		2,895,285

MAJCOM by Year, Annual Forecasts

Aircraft		RMSE Net Cost Model		RMSE Charges - Credits Model
A-10		**8,029,615**		8,771,378
B-1B		46,811,324		**46,540,969**
B-2A		13,684,132		**11,519,517**
B-52H		33,182,601		**32,514,015**
C-5		**41,192,555**		48,471,828
F-15C/D		**53,979,031**		55,696,133
F-15E		**43,147,722**		46,281,165
F-16C/D		**34,262,764**		47,401,861
KC-135		**9,830,247**		12,724,489
Overall		**128,098,319**		144,423,301

AF by Quarter, Annual Forecasts

Aircraft		RMSE Net Cost Model		RMSE Charges - Credits Model
A-10		**9,650,151**		14,532,661
B-1B		**33,530,864**		33,648,676
B-2A				13,686,857
B-52H		**21,096,503**		22,864,360
C-5		37,621,277		**22,868,483**
F-15C/D		74,233,596		**66,019,903**
F-15E		**45,494,060**		67,920,905
F-16C/D		**67,309,655**		69,540,931
KC-135		**8,379,017**		11,810,456
Overall*		**231,343,666**	✓	**212,045,767**

Bold values represent lower RMSE between competing models and checked values represent where differences in forecasts are statistically significant

*B-2A not included in the total forecast for each year

**B-1B and B-2A not included in total forecast for each year

***B-1B, B-2A, and B-52H not included in total forcast for each year

Appendix F: Summary of Forecast Accuracy for Proportional and Non-proportional Models

		RMSE 65-503 Factor Model	RMSE Generated Factor Model	RMSE Net Cost Model	RMSE Charges - Credits Model
AF by Month, Monthly Forecasts	A-10	2,670,376	2,170,089	✓ **1,994,520**	2,140,152
	B-1B	7,869,805	7,709,582		**7,640,466**
	B-2A	2,368,878	2,398,528		**2,356,663**
	B-52H	3,704,132	3,626,240	**3,022,145**	3,085,837
	C-5	4,999,187	**4,101,521**	6,497,772	6,459,446
	F-15C/D	9,604,374	**8,188,307**	9,670,983	8,310,215
	F-15E	11,425,805	5,868,837	**5,660,208**	6,166,468
	F-16C/D	7,664,012	**5,209,637**	8,692,376	7,937,276
	KC-135	2,316,763	**1,145,443**	1,476,716	1,609,425
AF by Month, Annual Forecasts	A-10	21,143,263	13,935,916	**9,750,849**	9,965,914
	B-1B	61,666,115	**44,756,792**		62,991,563
	B-2A	**7,968,838**	11,841,145		11,751,064
	B-52H	34,771,809	33,041,345	**25,320,516**	27,275,792
	C-5	43,859,936	✓ **10,355,863**	71,903,651	71,411,121
	F-15C/D	79,302,378	**60,627,964**	97,669,353	75,417,025
	F-15E	122,303,183	42,693,145	**27,494,330**	40,958,511
	F-16C/D	65,054,183	**12,821,287**	66,727,579	72,299,307
	KC-135	23,932,460	**5,422,015**	13,428,595	15,379,886
	Overall	254,019,152	**166,115,484**		224,905,373
AF by Quarter, Quarterly Forecasts	A-10	6,407,345	4,894,089	**2,741,825**	3,302,585
	B-1B	20,283,475	16,728,716	11,317,884	**10,160,755**
	B-2A	✓ **3,998,424**	4,407,649		4,126,041
	B-52H	9,693,222	9,349,423	**5,627,362**	6,080,072
	C-5	11,645,532	**4,555,618**	9,130,564	6,124,526
	F-15C/D	22,627,272	18,396,071	18,679,760	**16,206,196**
	F-15E	31,621,751	13,244,285	**12,540,912**	16,233,380
	F-16C/D	19,172,400	✓ **10,509,718**	18,223,106	18,882,096
	KC-135	6,335,526	2,274,901	✓ **1,840,534**	2,895,285
AF by Quarter, Annual Forecasts	A-10	21,143,263	13,935,916	**9,650,151**	14,532,661
	B-1B	61,666,115	44,756,792	**33,530,864**	33,648,676
	B-2A	**7,968,838**	11,841,145		13,686,857
	B-52H	34,771,809	33,041,345	**21,096,503**	22,864,360
	C-5	43,859,936	**10,355,863**	37,621,277	22,868,483
	F-15C/D	79,302,378	**60,627,964**	74,233,596	66,019,903
	F-15E	122,303,183	**42,693,145**	45,494,060	67,920,905
	F-16C/D	65,054,183	**12,821,287**	67,309,655	69,540,931
	KC-135	23,932,460	**5,422,015**	8,379,017	11,810,456
	Overall	254,019,152	**166,115,484**		218,025,526

		RMSE 65-503 Factor Model	RMSE Generated Factor Model	RMSE Net Cost Model	RMSE Charges - Credits Model
MAJCOM by Quarter, Quarterly Forecasts	A-10	6,407,345	4,894,089	**4,490,037**	4,804,069
	B-1B				
	B-2A				
	B-52H				
	C-5	11,645,532	**4,555,618**	5,487,800	7,252,261
	F-15C/D	22,627,272	**18,396,071**	18,610,011	21,714,398
	F-15E	31,621,751	**13,244,285**	44,037,585	19,914,938
	F-16C/D	19,172,400	✓ **10,509,718**	16,597,783	20,881,231
	KC-135	6,300,018	**2,035,737**	5,177,654	3,779,915
MAJCOM by Quarter, Annual Forecasts	A-10	21,143,263	**13,935,916**	14,151,324	15,828,289
	B-1B				
	B-2A				
	B-52H				
	C-5	43,859,936	**10,355,863**	18,116,527	24,028,692
	F-15C/D	79,302,378	**60,627,964**	66,716,950	73,193,911
	F-15E	122,303,183	**42,693,145**	171,083,891	68,351,936
	F-16C/D	65,054,183	**12,821,287**	51,183,330	72,284,989
	KC-135	23,932,460	**5,422,015**	24,342,667	15,192,484
	Overall	255,948,945	126,325,270	133,706,379	**107,340,535**
MAJCOM by Year, Annual Forecasts	A-10	21,143,263	13,935,916	**8,029,615**	8,771,378
	B-1B	61,666,115	**44,756,792**	46,811,324	46,540,969
	B-2A	**7,968,838**	11,841,145	13,684,132	11,519,517
	B-52H	34,771,809	33,041,345	33,182,601	**32,514,015**
	C-5	43,859,936	**10,355,863**	41,192,555	48,471,828
	F-15C/D	79,302,378	60,627,964	**53,979,031**	55,696,133
	F-15E	122,303,183	**42,693,145**	43,147,722	46,281,165
	F-16C/D	65,054,183	**12,821,287**	34,262,764	47,401,861
	KC-135	23,932,460	**5,422,015**	15,291,041	14,156,251
	Overall	254,019,152	166,115,484	**130,395,979**	144,830,868

Bold values represent lower RMSE between competing models and checked values represent where differences in forecasts are statistically significant

References

Air Force Cost Analysis Improvement Group (AFCAIG), Cost per Flying Hour Process Guide, Nov 1999.

Armstrong, Patrick D. *Developing an Aggregate Marginal Cost Per Flying Hour Model for Air Force's F-15 Fighter Aircraft*. MS Thesis, AFIT/GCA/ENV/06M-01. School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2006 (ADA423137).

Department of the Air Force. *Air Force Repair Enhancement Program (AFREP)*. AFI 21-123. Washington: HQ USAF/ILM, 3 October 2002.

Diebold, F.X. and R.S. Mariano. "Comparing predictive accuracy," *Journal of Business and Economic Statistics,* 13: 253-263, (July 1995).

Faykes, Frank. FY08 President's Budget Rollout Brief. Powerpoint. http://www.saffm.hq.af.mil/shared/media/document/AFD-070212-012.pdf. 5 February 2007.

Government Accountability Office. *Air Force Operating and Support Cost Reductions Need Higher Priority.* Washington DC: Government Printing Office. GAO/NSIAD- 00-165. August 2000.

Government Accountability Office. *Observations on the Air Force Flying Hour Program*. Washington DC: Government Printing Office. GAO/NSIAD-99-165. July 1999.

Granger, C.W.J. and P. Newbold. "Spurious Regressions in Econometrics," *Journal of Econometrics*, 2: 111-120, (1974).

Greene, William H. *Econometric Analysis*. 5th ediction, Upper Saddle River, New Jersey: Pearson Education Inc. (2003).

Habing, Brian. "Transformation of Variables." Course supplement, Stats 516, Statistical Methods II. Department of Statistics, University of South Carolina. July 2004. http://www.stat.sc.edu/curricula/courses/516/516s7p8sup.pdf.

Harris, Richard and Robert Sollis. *Applied Time Series Modelling and Forecasting*. West Sussex, England: John Wiley & Sons Ltd., 2003.

Harvey, D.I., S.J. Leybourne, and P. Newbold. "Testing the equality of prediction mean squared errors," *International Journal of Forecasting*, 13:281-291 (June 1997).

Hawkes, Eric M. and Edward D. White. "Empirical Evidence Relating Aircraft Age and Operating and Support Cost Growth." *Journal of Cost Analysis and Parametrics*. 1: 31-44 (Fall 2008).

Hawkes, Eric M. *Predicting the Cost per Flying Hour for the F-16 using Programmatic and Operational Variables*. MS Thesis, AFIT/GOR/ENC/05-01. School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, June 2005 (ADA436138).

Hildebrandt, Gregory G. and Man-bing Sze. *An Estimation of USAF Aircraft Operating and Support Cost Relations*. Santa Monica, CA: RAND, May 1990 (N-3062-ACQ).

Hildebrandt, Gregory G. *Budget Estimating Relationships for Depot-Level Reparables in the Air Force Flying Hour Program.* Santa Monica, CA: RAND, 2007 (MG-355)

Hu, Shu-Ping. "The Impact of Using Log-Error CERS Outside the Data Range and Ping Factor," *National Society of Cost Estimating and Analysis Conference*, June, 2005.

Im, So Kyung, M. Hashem Pesaran, and Yongcheol Shin. "Testing for unit roots in heterogeneous panels," *Journal of Econometrics*, 115: 53-74, (2003).

Kennedy, Peter. *A Guide to Econometrics*. 6th edition, Malden, MA: Blackwell Publishing (2008).

Kirby, Billy L. Cost Per Flying Hour Analyst, Centralized Asset Management Program Office, Air Force Materiel Command, Wright-Patterson AFB OH. Personal Interview. 2 July 2008.

Laubacher, Matthew E. *Analysis and Forecasting of Air Force Operating and Support Cost for Rotary Aircraft*. MS Thesis, AFIT/GCA/ENV/04M-05. School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2004 (ADA423137).

Osborne, Jason W. "Notes on the use of data transformations." *Practical Assessment, Research & Evaluation*, 8(6), (December 2002).

Rose, Pat A. Jr. "Cost Per Flying Hour Factors: A Background and Perspective of How They Are Developed and What They Do," *Air Force Comptroller*, 31-1:4-9 (July 1997).

Slay, Michael F. and Craig C. Sherbrooke. *Predicting Wartime Demand for Aircraft Spares.* McLean, VA: Logistics Management Institute, April 1997 (AF501MR2).

Smirnoff, James P. and Michael J. Hicks. "The impact of acquisition factors and economic reforms on the cost of defense weapon systems." *Review of Financial Economics*, 17:3-13 (2008).

Unger, Eric J. *An Examination of the Relationship Between Usage and Operating and Support Costs for Air Force Aircraft.* Santa Monica, CA: RAND, 2008. (RGSD-229).

Van Dyk, Stefanie L. *Forecasting Flying Hour Costs of the B-1, B-2, and b-52 Bomber Aircraft.* MS Thesis, AFIT/GCA/ENV/08-M02. School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2008 (ADA483271).

Vonglis, John G. "Leadership Message." Excerpt from unpublished article. n. pag. http://www.saffm.hq.af.mil/.

Wallace, John M., Scott A. Hauser, and David A. Lee. *A Physics-Based Alternative to Cost-Per-Flying-Hour Models of Aircraft Consumption Costs*. McLean, VA: Logistics Management Institute, August 2000 (ADA387273).

White, H. (1980). "A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity." *Econometrica*, 48(4), 817−838.

Woolridge, Jeffrey M. *Introductory Econometrics A Modern Approach*. 3rd edition, Mason, Ohio: Southwestern-Thomson Learning, 2006.

Wynne, Michael W. and Michael T. Moseley. The FY 2008 Air Force Posture Statement. Excerpt from unpublished article: http://armed-services.senate.gov/statemnt/2007/March/Wynne-Moseley%2003-20-07.pdf. 20 March 2007.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 074-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
26-03-2009	Master's Thesis	Jun 2008 - Mar 2009

4. TITLE AND SUBTITLE	**5a. CONTRACT NUMBER**
Cost Forecasting Models For the Air Force Flying Hour Program	**5b. GRANT NUMBER**
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER**
Hess, Tyler J., First Lieutenant, USAF	**5e. TASK NUMBER**
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way WPAFB, OH 45433-7765	AFIT/GCA/ENV/09- M 07

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
Mr. Billy L. Kirby, Civilian, USAF AFMC A4FC/CAM Cost Per Flying Hour Analyst 4225 Logistics Avenue Bldg 262, Post19R WPAFB, OH 45433-7222 (937) 257-3718	A4FC CAM **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION/AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The fiscally constrained environment in which the Air Force executes its mission places great emphasis on accurate cost estimates for planning and budgeting purposes. Inaccurate estimates result in budget risks and undermine the ability of Air Force leadership to allocate resources efficiently. This thesis evaluates the current method used by the Air Force and introduces new methods to forecast future Flying Hour Program costs. The findings suggest the current forecasting method's assumption of a proportional relationship between cost and flying hours is inappropriate and the relationship is actually inelastic. Prior research has used log-linear least squares regression techniques to forecast Flying Hour Program cost, but has been limited by the occurrence of negative net costs in the underlying data. This research uses time series and panel data regression techniques while controlling for flying hours, lagged costs, and age to create net costs models and an alternative model by separately estimating the two components of net costs which are charges and credits. Finally, this research found neither the proportional, net costs, nor charge minus credit models is a superior forecaster. As such, the models introduced in this research may be used as a cross check for the current method.

15. SUBJECT TERMS
Cost Per Flying Hour, Flying Hour Program, Budget Estimates, Forecasting Models

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Lt Col Eric J. Unger (AFIT/ENV)
U	U	U	UU	114	**19b. TELEPHONE NUMBER** *(Include area code)* (937) 255-3636

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

Form Approved
OMB No. 074-0188