The Effects of Funding Gaps on Depot Maintenance Hours

THESIS

Carlo S. D'Amato, 1st Lieutenant, USAF

AFIT-ENV-MS-20-M-196

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENV-MS-20-M-196

THE EFFECTS OF FUNDING GAPS ON DEPOT MAINTENANCE HOURS

THESIS

Presented to the Faculty

Department of Cost Analysis

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

Carlo S. D'Amato, B.A.

1st Lieutenant, USAF

March 2020

AFIT-ENV-MS-20-M-196

THE EFFECTS OF FUNDING GAPS ON DEPOT MAINTENANCE HOURS

THESIS

Carlo S. D'Amato, B.A.
1st Lieutenant, USAF

Committee Membership:

Lt Col Scott Drylie, Ph.D.
Chair

Lt Col Clay M. Koschnick, Ph.D.
Reader

Dr. Jonathan D. Ritschel, Ph.D.
Reader

AFIT-ENV-MS-20-M-196

Abstract

The relationship between expenditures and readiness level is a topic of interest to military senior leaders, defense resource planners, and the American taxpayer alike. Senior leaders within the Air Force (AF) justify increased defense budgets by pointing to the potential adverse effects that decreased funding could have on military readiness. Resource planners within the AF are then tasked with the responsibility of ensuring that budgets are allocated most effectively to maximize the AF's ability to project airpower across a variety of contingency operations. This thesis investigates the relationship between budgets and readiness by examining the relationship between depot level funding and hours of aircraft downtime spent at the depot. Funding is analyzed in terms of the magnitude that the amount of funding receives deviates from the amount of funding requested by the planner. The analysis ultimately did not find any conclusive relationship between deviations from requested depot budget levels and the number of hours of downtime spent at the depot.

Table of Contents

List of Figures

List of Tables

THE EFFECTS OF FUNDING GAPS ON DEPOT MAINTENANCE HOURS

I. Introduction

1.1 Background

Stewardship of taxpayer dollars is a perennial concern within DoD, and is a common backdrop for empirical research, internal studies, and formal audits. A subject line of a 2018 memo from Secretary of Defense James Mattis reads: “Be Peerless Stewards of Taxpayers’ Dollars.” The memo is a call-to-arms for planners across the DoD to “gain the full value from every taxpayer dollar spent on defense” and “focus on lethality and affordability” [**?**] (Mattis, 2018). Mattis thereby identifies financial stewardship as a prime DoD objective and defines its purpose: the provision of the best warfighting capability with the budgets provided. This is perhaps a definition of stewardship that is most analogous to how businesses assess themselves. It frames taxes as costs and national defense as the product. For senior leaders, fighting capability and readiness are the “receipt” that the USAF provides to the American taxpayer (Donovan & Goldfein, 2019).

Itemizing this receipt is typically framed in terms of the USAF’s ability to put planes in the air (Fry, 2010; Losey, 2019). It comes as no surprise then that the ongoing decline in mission capable aircraft has made headlines (Losey, 2018; Losey, 2019; Mehta, 2018). That decline is seen in the Mission Capability (MC) metric. MC measures the percentage of aircraft at the unit level that are ready to conduct operations (Air Force Logistics Management Agency [AFLMA], 2009). The MC rate, in aggregate, declined from 2012 to 2018 (Losey, 2019). In Fiscal Year (FY) 2019, the rate increased trivially from 70.65% to 70.99%. Making matters worse is the fact that this negative trend comes despite a steady growth in the Air Force's top-level budget over that same period (Figure 1). The Air Force has been accordingly admonished in internal audits for overspending and under-delivering on key aircraft readiness metrics (GAO [Government Accountability Office], 2018).

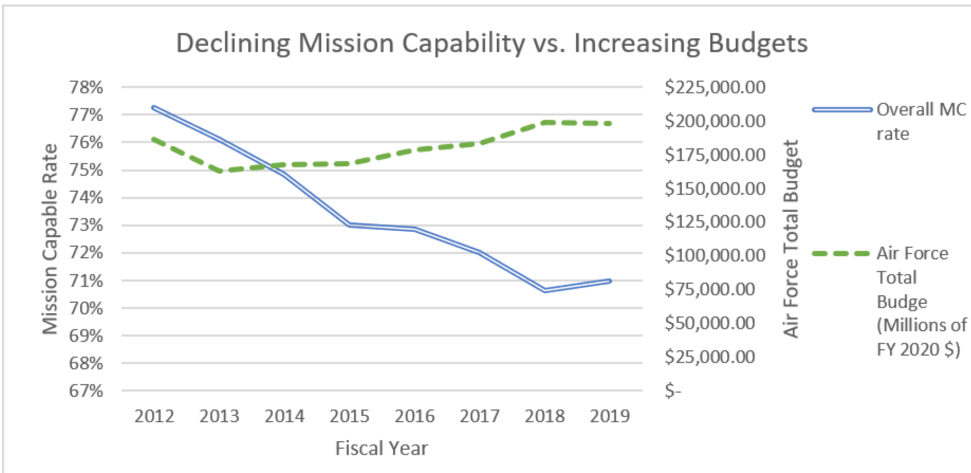

Figure 1. Mission Capability vs. Budget

The incongruity of growing budgets and declining readiness suggests that readiness may be more complex than what is revealed by the MC rate. The DoD officially frames readiness as "The ability of US military forces to fight and meet the demands of the national military strategy" (Department of Defense [DoD], 2019) This definition leaves it to the branches to determine how readiness should be measured. For its part, the Air Force has internally made the shift from the unit-focused MC metric to the total-inventory minded aircraft availability (AA) metric as the preferred measurement of readiness (Oliver, 2003; Meserve, 2007; Fry, 2010; Ritschel, Ritschel, and York, 2019).

The major difference between AA and MC is that AA explicitly includes aircraft that the unit has sent away to the depot for maintenance in its calculation of the quantity of operable aircraft (Fry, 2010). Under MC, only the aircraft in direct possession of the unit are used in the operability calculation. Including aircraft possessed by the depot ensures that aircraft undergoing more serious repair or maintenance are not left unaccounted for during readiness audits, as they are when using MC. The two equations, 1 and 2, are represented below. The mission capable hours, the total hours in the period that the airframe is operable and ready for use, are ultimately divided by a larger denominator in the AA metric:

$$MC = \frac{Mission\,Capable\,Hours}{Inventory\,Hours\,of\,Unit\,Possessed\,Aircraft} \tag{1}$$

$$AA = \frac{Mission\,Capable\,Hours}{Inventory\,Hours\,of\,Unit\,Possessed\,Aircraft + Inventory\,Hours\,of\,Depot\,Possessed\,Aircraft} \tag{2}$$

One implication of the AA metric is that the longer that aircraft are at the depot, the worse the readiness picture is. Indeed, top-level Air Force data from the last five years show that an average of 30% of all aircraft downtime is due to time spent at the depot for maintenance. The next logical step is to examine how defense planners, as peerless stewards of taxpayer dollars, can best allocate funds to minimize the amount of downtime that is spent at the depot, thereby reducing downtime as a whole. Doing so will shed light on the overall relationship between budgets and Air Force readiness.

1.2 Problem Statement

Defense economists fundamentally contend that it is the way budgets are allocated, not just the overall magnitude of the Air Force total budget authority, has an impactful role on readiness (Hartung, 1999; Biddle, 2004). In 1980, defense analyst Franklin C. Spinney released a report that depicted declining readiness as a function of underfunded operations and support (O&S) budgets (Spinney, 1985). Spinney observed that cyclical swells in the Air Forces total budget authority provoked strong short run investment in new capabilities-driven programs. In the long run, the OS tails of these same programs were not being protected during subsequent ebbs in funding. Spinney felt that the failure to protect O&S funding was linked to a decline in readiness. Programs were being funded with a stated level of O&S funding in mind and the level was not being maintained; were it maintained, readiness would not suffer.

Spinney's report, though controversial at the time, found widespread support within the Department of Defense (Hankins, 2018). In comments made to the Air Force Times in 2019, Heritage Foundation Fellow John Venable blamed the slide in availability on the prioritization of research and development over operations and maintenance during Air Force budget planning (Losey, 2019). Venable's comments echo Spinney's concerns from 40 years prior. Moreover, Venable's claims support Spinney's tacit implication that greater O&M budgets would lead to increased readiness.

The chart in Figure 2 suggests that some of the patterns observed in Spinney's initial report may still be observed today. Figure 2 shows how investment and OS budgets continue to be unstable from year to year. Investment budgets especially appear to fluctuate radically in a way that is not matched by steady growth in OS budgets.

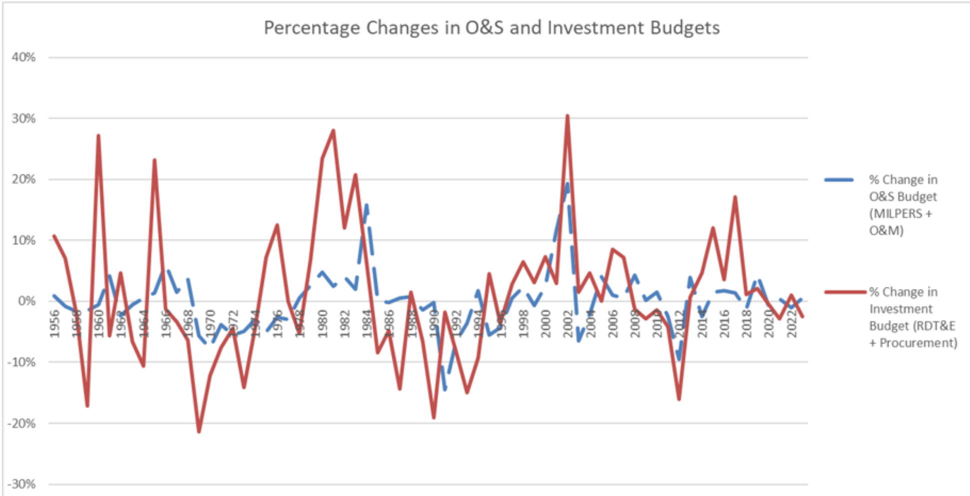

Figure 2. Trends in OS and Investment Budgets

Spinney's claims, however influential, remain only vaguely supported. Detailed investigation of the relationship between budgets and readiness requires linking aircraft availability to specific O&S expenditures, not just those at the top level. The budgetary readiness model will need to operationalize OS expenditure categories as independent variables within the equation. Data for such a task are, fortunately, available.

The Air Force's Funding Requirement Management platform (FRM), provides budget data that are indexed by Air Force Element of Expense/Investment Identification Code (AFEEIC). The AFEEIC can then be used to identify the category of good or service purchased. FRM also provides budget data both in terms of planned expenditure, as outlined in the unit's execution plan, and in terms of the actual funding level executed at the end of the fiscal period. The consolidation of planned requirements and executed obligation under the umbrella of a single system facilitates the rapid identification of budgetary shortfalls and surpluses across various Air Force platforms. Overall, the data provided by FRM allow for a granular analysis that is impossible with top-level OS data.

This research will seek to establish how budgetary gaps, the deviations between requested budgets and actual spending, in various budgetary categories influence the downtime hours caused by depot maintenance. A regression will be conducted to measure how budgets allocated for depot level repairs ultimately impact the total aircraft hours spent under depot possession. The analysis will necessarily identify and include the additional non-financial variables pertinent to aircraft downtime in order to control for their effects within the broader model.

More specifically, the FRM data will combine with the metrics already outlined within the body of defense readiness literature to provide a comprehensive view of the key determinants of reductions in unscheduled maintenance. The analysis will use organically supported Air Force attack, fighter, and bombing aircraft as its modelling cohort. The end result is a model that will provide insight into the functional relationship between OS funding and readiness.

The central question guiding the research is: *How do shortfalls or surpluses in depot-allocated budgets impact depot downtime hours?* Specifically, *which types of expenditures, by AFEEIC, have the greatest impact on depot maintenance hours?*

1.3 Outline of Thesis

The remaining chapters in the thesis will be organized as follows. Chapter II furnishes a review of the unique economic characteristics of the defense sector, as well as provide background on the usage of AA as a quantifiable metric of readiness and identify the key empirical drivers of AA. Chapter III contains the regression analysis of the data provided by FRM in conjunction with the relevant variables identified in Chapter II to create a regression model. Chapter IV presents the results of the model created in Chapter III. Chapter V provides concluding remarks on the relationship between OS expenditures and readiness are made.

II. Literature Review

> "The defense business, notwithstanding the rhetoric of the corporate managers involved in it, is not private enterprise in anything like the classic sense. But it is not public enterprise. It is *sui generis.* Therein lies much of the difficult rationalizing and reforming it, or even talking about it sensibly..."
>
> *-Dr. Robert Higgs*

2.1 The Defense Economic Framework

The market forces at play in the defense sector are markedly different from those in the private sector. Consumers in private markets reconcile supply and demand through their preferences: they spend money on what provides the greatest amount of utility in accordance with their needs (Mises, 1958; Kaldor, 1986). The utility of defense goods, however, can only truly be proven during war (Kaldor, 1986). It is therefore war, not the consumer, that mediates the supply and demand of defense goods (Kaldor, 1986).

Yet militaries must be maintained even in the absence of war in order to ensure preparedness should war break out (Eisenhower, 1961). Without the inputs of consumer preferences or the exigent needs of an armed conflict, the military is forced to forge ahead based on the best estimates of central planners within the defense apparatus (Kaldor, 1986; Hartley, 2012). The necessity of a peacetime force thereby makes the planner an inextricable piece of any analysis of the defense economy.

The planner's job is complicated by a variety of peculiarities of the defense market that are fundamentally rooted in its status as a public good. Planners theoretically develop budgets that will provide the military readiness required to meet national defense goals. Fiscal realities may necessitate changes to the "optimal" funding levels identified by the planners. If the planner has done the job correctly, deviations from what has been deemed necessary should result in some sort of fluctuation in readiness.

2.2 Defense as a Public Good

National defense, unlike any private sector output, is a public good. Public goods are defined by their non-rivalry and non-excludability. Essentially, they are goods that can be consumed regardless of the number of total consumers and whose consumption cannot be diminished. In the words of economist Robert Higgs, "public goods, if created for anyone, are created for all" (Higgs, 1990). Indeed, once provided, an individual cannot diminish the amount of national defense being supplied, nor prevent others from being protected by national defense (Hartley, 2012).

Although national defense is a desirable good, the inability to exclude non-paying consumers from the market means that private markets will struggle to provide the optimal amount of public goods (Hummel Lavoie, 1990). Moreover, no single individual can raise the funds necessary to protect the nation from threats (Higgs, 1990) The result is that the government must supply national defense to correct the private market's inevitable failure to provide the optimal amount of national defense.

The optimal provision of a public good is a difficult task. Private sector businesses can rely on market forces to guide the allocation of resources to their most productive purpose (Mises, 1958). Public goods, to include defense, must rely on the political process and planners to effectively allocate funding (Hartley, 2012). The intertangled web of regulations and bureaucracy complicates the effective allocation of resources (Coyne, Michaluk, and Reese, 2016). Hence, although government provides national defense to correct a failure in the private market, the governmental control of national defense contains its own inefficiencies and market failures.

2.3 The Economic Characteristics of National Defense

Despite frequent bipartisan calls to "run government as a business," profit and readiness are not analogous metrics (Dunne, 1995; Hartley Solomon, 2012; Hartley, 2012; Coyne et al. 2016; Mintzberg, 2017). While both profit and readiness can be considered the outputs of private and defense markets respectively, readiness is a far more ambiguous concept. Profit can be easily used to measure the productivity of an investment in the private sector. Moreover, the metric is readily defined as any revenues generated in excess of expenditures.

Readiness, meanwhile, is not as unambiguously calculated as profit. Readiness levels are foremost defined in the context of the threats a system will face (Hartung, 1999). A fighter squadron will excel against enemies with air-to-air capabilities, but be of only marginal use against a ground-based enemy shielded by jungle growth; two readiness levels exist for the same squadron. In economic terms, this means that the marginal productivity of a defense investment is defined in terms of its advantage over rival actors.

Goods whose value depends on the value of a rival's goods are known as tournament goods (Hove Lillekvelland, 2015). The concept of the tournament good was originally introduced by economists Edward Lazear and Sherwin Rosen to account for the high wages earned by senior executives in excess of their marginal productivity (Lazear Rosen, 1981). Lazear and Rosen posited that firms could elicit optimal levels of productivity from skilled workers by making workers compete for promotions. Their model found that rational workers will set their marginal effort equal to the marginal expected gains that their additional effort will yield (Lazear Rosen, 1981). In short, the value of skilled workers is accentuated when they are incentivized to stratify themselves above lower-performing colleagues. Moreover, the greater the rewards of the promotion, the greater the resources a skilled worker is willing to commit to secure it.

Lazear and Rosen's framework can be expanded to defense goods (Kirkpatrick, 2003; Hove and Lillekvelland, 2015). Rival nations competing for military victory face similar incentives as employees competing for promotions. Nations must decide how much defense equipment to invest in to improve their chances of victory; the greater the gap between a nation and its rivals, the greater its chance at victory. In turn, the more significant the victory, the more resources the nation will commit to increasing the probability of victory (Hove and Lillekvelland, 2015). Planners must therefore analyze requirements both in the context of the utility that defense forces provide as well as the evolution of the threats that they are expected to face. This complicates the work of planners, who must optimize around a constantly shifting definition of readiness.

Planners also lack a direct avenue of communication with voters as consumers of national defense. Voters, with their limited knowledge of defense matters, must entrust politicians with the minutiae of defense policy. A classic principal-agent problem emerges, with the voters as principals and the politicians as agents (Hartley, 2012). Because the principal's preferences must be expressed through the agent, the agent may pursue his own benefit at the expense of the principal's. A self-interested agent may choose to promote populist defense policies that are more beneficial to his polling numbers than to actual national defense goals (Hartley, 2012; Twight, 1990).

In practice, this means that defense planners may find undue opposition against base closures, but undue support for defense pay increases (Hartley, 2012; Carrell and Hauge, 2009; Twight, 1990). Ultimately, the planner executes a version of voter preferences that is filtered by the political bureaucracy.

These challenges, discussed at length by economists since the socialist calculation debates of the early 20th century, tell us that that we, as a society, are not likely to generate the optimal amount of defense (Hartley, 2012; Coyne Lucas, 2016). The marginal value of more or less defense is neither known to decision makers, nor by a citizenry which must monitor its decision makers. All of this bodes poorly for the defense planner who must determine a common marginal valuation of military preparedness and allocate resources accordingly.

Analyzing budgets strictly in terms of their raw magnitude therefore ignores the central role that planners fulfill in the defense budgeting. Instead, it is necessary to examine budgets in the context of how greatly they deviated from the amount outlined by their planners. Considering the impacts of gaps not only evaluates the marginal impact of expenditures on readiness across different programs, but also evaluates the effectiveness of the planner's ability to plan for the budgetary needs of the military. If the gaps have no impact, the implication is that planners may be overestimating their needs during the planning process.

2.4 Budget Planning in the Air Force

In the active duty Air Force, the budgetary planning of depot purchased equipment maintenance (DPEM) has recently been centralized away from the MAJCOM and consolidated at the directorate level (Fry, 2010; AFMAN 63-143). The task of submitting the annual Project Objective Memorandum (POM) and spreading funds across MAJCOMS now falls under the Central Asset Management office (CAM), a subsidiary of the Air Force Logistics, Engineering and Force Protection directorate (A4). The ostensible advantage of this setup is that planners may move funds more freely across requirements at the enterprise level, as opposed to having consult multiple requirements owners at the MAJCOM level (Fry, 2010). Figure 3 outlines the funding process of depot level expenditures via CAM.

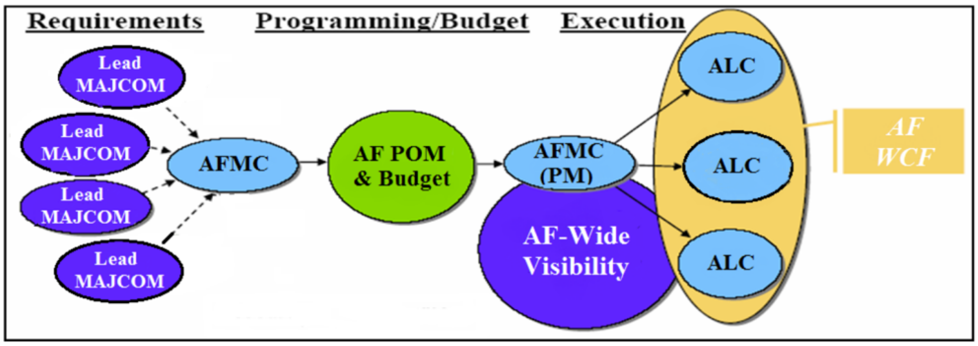

Figure 3. The CAM Budget Allocation Process (AFMAN 63-143)

It is incumbent upon these planners at CAM to efficiently program resources such that readiness is provided at the optimal level. However, just as the economic framework of the defense sector complicates the generation of the optimal level of national defense, the conceptually broad nature of readiness adds layers of complexity to the planner's task. Readiness is not objectively defined, and must be rigorously operationalized to have meaning.

2.5 The Components of Readiness

Readiness is conceptually comprised of two qualities: utilization and availability (Advanced Technology Incorporated ,1980; Harrison, 2014). Utilization indicates the amount of time that a resource is being actively employed. Meanwhile, availability in the modern military context denotes the amount of time a resource is ready for use. A study commissioned by the Chief of Naval Operations in 1980 expresses the two facets of readiness mathematically within an example of radars:

$$R_1(utilization) = \frac{Radar\ Uptime}{Radar\ Uptime + Radar\ Downtime} \tag{3}$$

$$R_2(availability) = \frac{Time\ in\ Period - Radar\ Downtime}{Time\ in\ Period} \tag{4}$$

For the purposes of equations 3 and 4, the researchers define uptime as the time that the radar is in active use. The definition of downtime is more broad, and counts all of the time that the system is broken, regardless of whether or not there is demand for the system.

An important takeaway is that utilization metrics do not distinguish between inactivity and inoperability. A system that is never utilized will have the same R_1 as a system that is totally broken: its readiness is 0 in both cases (Advanced Technology Incorporated, 1980). Availability metrics, meanwhile, do not distinguish between operability and warfighting proficiency (Harrison, 2014). An available system is assumed to *ipso facto* be able to successfully execute the mission it is designed for.

In aircraft terms, flying hours and sorties are examples of metrics that track utilization (AFLM, 2009). The sortie rate (SR), as seen in equation 6, is an example of a utilization metric. SR measures utilization by approximating the maximum quantity of sorties possible given average maintenance turnaround times and flight durations (Stillion Orletsky, 1999). Availability, on the other hand, is expressed through metrics that indicate the quantity of operable aircraft in the fleet (AFLM, 2009). Aircraft Availability (AA), as mentioned in the previous chapter, is the prime method through which the Air Force identifies fleet availability (Fry, 2010; Ritschel et al., 2019).

$$Inventory\ Hours = \#\ of\ Planes\ in\ Fleet * 24 * Days\ in\ Period \tag{5}$$

$$Utilization = SR = \frac{24\ Hours}{Time\ Required\ for\ Repairs + Time\ to\ Accomplish\ Pre-Flight\ Checklist + Flight\ Duration} \tag{6}$$

$$Availability = AA = \frac{MC\ Hours}{Total\ Active\ Aerospace\ Vehicle\ Inventory\ Hours} \tag{7}$$

Utilization metrics tend to be less stable than availability metrics (Boito, Keating, Wallace, DeBlois, and Blum, 2015). This is because utilization rates are at the mercy of a variety of factors, including unit taskings, unit training requirements, and weather (Boito et al., 2015). Flying hour policies must be flexible in order to accommodate all of these contingencies. Normalizing the unique policy factors influencing flying hour goals at each unit is therefore a difficult task.

Changes to aircraft inventory conversely involve high level decision makers and require considerable amounts of planning (Boito et al., 2015). Decisions made at such high levels invariably impact entire fleets. Consequently, the policy factors influencing aircraft inventory are more uniform across units (Snyder, Kim, Carrilo, and Hildebrandt, 2012). Comparisons based on availability, as opposed to utilization, lessens the complexity of the data normalization required to compare readiness across units or platforms. It is for this reason that examining the impact of budgets across a wide variety of aircraft is best undertaken in terms of availability and not the end utilization of the aircraft. Availability acts as an effective control for the variety of utilization policies across different aircraft.

2.6 Measuring Readiness: Utilization vs. Availability

Readiness reporting in the Air Force is done primarily through metrics that indicate availability (Harrison, 2014). Per AFI 10-201, units report readiness via two interrelated systems: The Status of Resources and Training System (SORTS) and the Defense Readiness Reporting System (DRRS). Commanders identify the percentage of unit resources, both materiel and personnel, that are available to deploy. Although AA is not explicitly reported in these systems, the manner in which availability is reported in SORTS and DRRS is congruous with the concept of availability as calculated in AA.

Units are assigned a readiness rating based on their ability to meet target availability levels established by the Air Force (AFI 10-201; Harrison, 2014). The result is that the resources of a unit are used as a proxy for its ability to carry out its mission (Harrison, 2014). The information from SORTS then carries over to DRRS, where commanders have the opportunity to make a self-assessment of the unit's readiness (AFI 10-201; Harrison, 2014).

Critics of Air Force readiness reporting question the validity of using availability as a proxy for performance (Moore et al. 1991; Harrison, 2014). Todd Harrison, in his commentary "Rethinking Readiness", observes that Air Force readiness reporting operates under the flawed assumptions that: (1) any resources that are available are also fully capable; (2) availability targets correctly encapsulate the requirements of the mission; (3) availability targets properly correlate to mission needs; (4) the commander will accurately state his subjective readiness assessment in DRRS (Harrison, 2014). Harrison argues that the conflation of availability with performance, as well as the inevitable bias that seeps into any self-assessment, makes accurate readiness reporting impossible through current Air Force systems (Harrison, 2014).

Harrison, echoing previous criticisms of defense readiness reporting, proposes the implementation of metrics that directly measure mission-related skills (Moore et al. 1991; Harrison, 2014). The goal of these new "strategy-based metrics" is to shift the focus from readiness inputs, such as inventory availability, to readiness outputs, which would change based on the mission. Using a fighter wing as an example, Harrison identifies the need for tailored metrics that would measure low-altitude bombing accuracy and air-to-air combat skills. Data for these metrics would be collected over the course of already scheduled training exercises. Strategy-based metrics would thereby serve as a more accurate approximation of performance than traditional input-based metrics (Harrison, 2014).

The accuracy of strategy-based metrics comes at analytical cost. The most glaring issue is that the Air Force does not presently aggregate metrics for mission performance in any of its force-wide systems. Availability metrics, meanwhile, are directly reported in the Logistics, Installations, and Mission Support – Enterprise View (LIMS-EV) system. LIMS-EV has been well received by the defense community as a particularly accurate repository for fleet health metrics (Petcoff, 2010). Hence, while compelling arguments can be made for the fidelity of readiness outputs to wartime performance, the pure fact is that they are not yet available for analytical purposes.

Even if they were available, the specificity of strategy-based metrics makes them unsuitable for readiness analysis across different conflicts. In addition to measuring warfighting capability, readiness most broadly encompasses the ability of the military to achieve national defense goals (Snyder et al., 2012). National defense goals are not homogenous; conflicts involving near-peer adversaries are strategically separate from contingencies involving insurgencies, or operations containing a significant humanitarian component (Snyder et al., 2012).

Each contingency accordingly requires the projection of different capabilities. As political scientist Stephen Biddle observes, "There is. . . no single, underlying quality of generic 'capability' to which all specific mission capacities are epiphenomenal" (Biddle, 2004). Biddle's remark identifies a key weakness of capability-based metrics: they are not generic enough to provide an accurate picture of Air Force readiness across the gamut of threats it must face. AA, on the other hand, is a metric that is easily understood regardless of mission set or utilization; AA communicates how many aircraft are ready to take to the skies in the event of a conflict.

Comparing the operability of aircraft across airframes is thus a task best undertaken through an availability-based metric, such as AA. Using a utilization-based metric, such as SR, would muddy the waters between inoperability and inactivity; given that the Air Force is not presently engaged in a total war, it is best to capture readiness in terms of the force that *could* be employed as opposed to the force that is actively employed. Likewise, more refined strategy-based metrics are both unavailable and ill-suited to the snapshot perspective needed for a generalized view of Air Force readiness (Snyder, 2012).

2.7 The Aircraft Availability Metric

AA is the most appropriate method to quantify readiness in a cross-platform analysis of Air Force platforms. What follows is deeper exploration of the AA equation and the factors that previous research has identified as drivers of availability in a manner similar to Ritschel et al. (2019). As established in the introduction, AA expands upon the MC by including aircraft awaiting depot-level repairs in its calculation of availability (Fry, 2010). AA is then expressed as a ratio of the total hours aircraft in the fleet are in mission capable over the inventory hours in the period (Equation 9).

$$MC = \frac{MC\ Hours}{PAI\ Hours} \tag{8}$$

$$AA = \frac{MC\ Hours}{TAI Hours} \tag{9}$$

Figure 4 illustrates the sub-metrics that comprises AA (Ritschel et al., 2019; Fry, 2010; AFLM, 2009). Units must account for degradations in AA by applying one of five root causes: (1) depot, (2) not mission capable maintenance, (3) not mission capable supply, (4) not mission capable both, and (5) unit possessed not reported. Uptime occurs when an aircraft is not affected by a down time label.

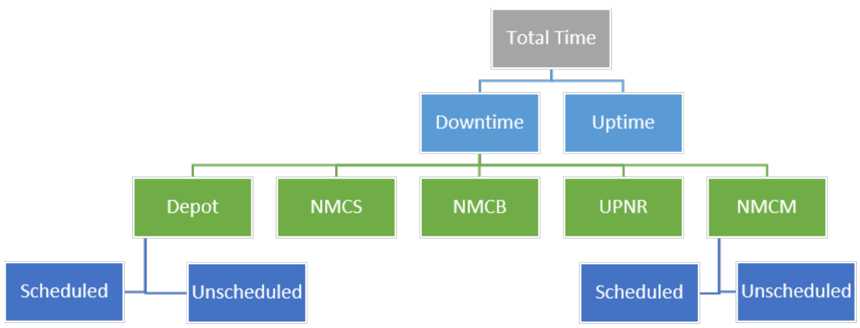

Figure 4. Taxonomy of Aircraft Availability (Ritschel et al., 2019)

The drivers of downtime each cover a unique mission-impacting status (Fry, 2010; Ritschel et al., 2019). Depot aircraft are those aircraft whose repairs or maintenance needs are too serious to be taken care of at the unit level. Not Mission Capable Maintenance (NMCM) refers to aircraft being repaired at the unit level. Not Mission Capable Supply (NMCS) refer to aircraft who are unavailable due to an absence of repair parts. Not Mission Capable Both (NMCB) is applied when an aircraft is awaiting both repair and parts. Lastly, Unit Possessed Not Reported (UPNR) is a special status applied to aircraft whose repairs are complex enough so as to require input from an outside agency to determine the way forward. The UPNR status is applied during the waiting period for this determination.

Downtime is not split equally among the five statuses (Ritschel et al., 2019).

Figure 5 shows the taxonomy of non-availability over the last five years; NMCM and depot repairs constitute the largest proportion of non-available hours. The overwhelming majority of NMCM, 70%, is unscheduled (Ritschel et al.,2019). Meanwhile, the inverse is true of depot level maintenance, 70% of which is *scheduled*.

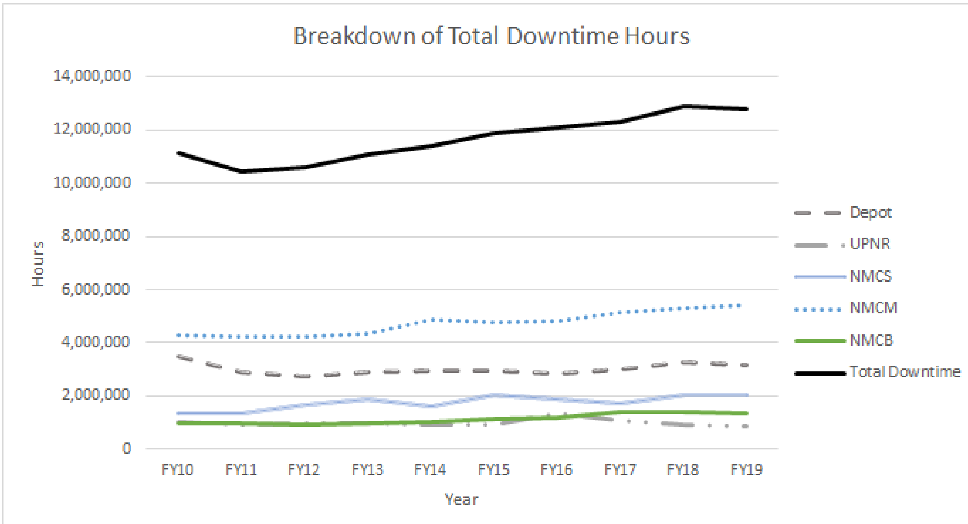

Figure 5. Total Downtime Hours by Label

Accounting for the drivers behind NMCMU is the main focus of Ritschel et al. (2019). The research found that total unscheduled maintenance hours did not decrease proportionally with decreases in TAI; spreading resources across a smaller pool of aircraft did not result in efficiency gains in maintenance needs. The research subsequently identified platform type, airframe age, repair cannibalization rates, breakdown incidence, fleet size, and pilot reported discrepancy incidence as significant drivers of NMCMU.

NMCM and depot time are conceptually corelated metrics due to the fact that any repairs that cannot be resolved at the unit level must flow to the depot level. As a result, the driving variables of breakdowns requiring unit-level repair, as identified by Ritschel et al. (2019), can be theoretically extended to an analysis of the drivers of breakdowns requiring depot-level repair. These variables ultimately fall in line with variables previously identified as theoretically pertinent to AA as a whole (Oliver, 2001; Fry, 2010). Fry (2010) provides a comprehensive list of the researched inputs to AA, reproduced here in Table 1.

Table 1. Pertinent Factors to Aircraft Availability (Fry, 2010)

Category	Variable	Correlation	Author
Personnel	Ratio of 3-levels to 5-levels	Negative	Oliver, 2001
	Ratio of 3-levels to 7-levels	Negative	Oliver, 2001
	Total # of Inexperienced Maintainers by Rank or Skill Level	Negative	Oliver, 2001
	Maintainers Per Aircraft	Positive	Oliver, 2001
	Total # of Maintainers	Positive	Oliver, 2001
	Overall Reenlistment Rate	Positive	Oliver, 2001
	Reenlistment Rate of First-Term Airmen	Positive	Oliver, 2001
	Reenlistment Rate of Career Airmen	Positive	Oliver, 2001
	Reenlistment Rate of Eligible Crew Chiefs	Positive	Oliver, 2001
	Crew Chief Manning Levels	Positive	Huscroft, 2004
	Percentage of 7-level Maintainers	Positive	Chimka and Nachtmann, 2007
	Percentage of 9-level Maintainers	Positive	Chimka and Nachtmann, 2007
Environment	Average # of Possessed Aircraft	Positive	Gilliland, 1990
	Aircraft Age	Mixed (Bathtub Curve)	GAO, 2003
	Transition to Combat Wing Structure in 2002	Positive	Barthol, 2005
Reliability & Maintainability	Cannibalization Rate	Negative	Gilliland, 1990; Moore, 1998; Oliver, 2001
	Awaiting Maintenance Discrepancies	Negative	Gilliland, 1990
	8-Hour Fix Rate	Positive	Oliver, 2001
Funding	CLS supported	Positive	RAND, 2009
Aircraft Operations	Sorties	Mixed	Moore, 1998;
Logistics Operations	Awaiting Parts Discrepancies	Negative	Gilliland, 1990
	% of Requests for Consumables Filled in 1-2 days	Positive	Moore, 1998

In addition to the largely environmental factors explored by Ritschel et al. (2019), Fry's table underscores the significance of personnel matters in determining availability. The theoretical framework tying manning to operability in the modern Air Force is chiefly laid out in the research of Oliver, Johnson, White, and Arostegui (2001). It should be noted that, technically, Oliver et al. (2001) focused on MC, and not

AA. However, the similarity between the two metrics is close enough that the insights provided by older research on MC can still be applied to contemporary explorations of AA (Fry, 2010).

The researchers found that the ratio of 3-level maintainers to 5 and 7-level maintainers, a proxy for the overall level of experience of the unit, was a significant determinant of the MC rate. Moreover, metrics measuring the total amount of O-3s and 9-levels assigned to the unit, as well as the reenlistment rate, were also found to be significant. Oliver's conclusions are upheld by Dahlman and Thaler (2002) as well as Chimka and Nachtmann (2007), who found that the presence of *experienced maintainers* in the unit was critical to high operability rates.

The literature has thus coalesced around three main potential drivers of availability: empirical characteristics of the aircraft (Ritschel et al., 2019), the allocation of personnel to the aircraft (Oliver et al., 2001; Chimka Nachtmann, 2007), and the allocation of budgets to the programs (Fry, 2010). Although the factors identified in Ritschel et al. (2019) were not explicitly linked to depot repairs, unit level maintenance ultimately flows into the depot level sooner or later. The empirical characteristics and and manning levels of the aircraft will need to be controlled for in order to properly isolate for the impact that budgets, as allocated by CAM, have on depot downtime.

2.8 Measuring the Costs of Availability

Planners can deduce the marginal *costs* of readiness by examining the per-unit costs of readiness inputs. One of the more widespread examples of per-unit estimates is Costs per Flying Hour (CPFH), which is often used to compare the marginal cost of utilizing the aircraft (Boito et al., 2015). Costs can similarly be expressed in terms of availability by dividing costs by Primary Aircraft Inventory (PAI). The

Direct Cost Per Aircraft (DCPA) indicates the marginal cost of making an aircraft available, independent of how many hours it flies (Boito et al. 2015). Crucially, both measures are obtained *a posteriori*; the total costs must be known *before* the analyst can determine the marginal cost.

$$CPFH = \frac{O\&S\ Costs}{FH} \tag{10}$$

$$DCPA = \frac{O\&S\ Costs}{PAI} \tag{11}$$

Building predictive models of readiness is famously difficult (Hildebrandt Sze, 1990; Higgs, 1990; Boito et al., 2016). The Aircraft Availability Model (AAM) was one of the Air Force's first attempts to build a predictive model of availability (Fry, 2010). The AAM focused exclusively on availability as a function of spares inventory, and excluded declines in availability driven by maintenance events (O'Malley, 1983). Required part supply levels were modelled against varying squadron sizes to determine the optimal level of spares to acquire at different budgets (O'Malley, 1983). Consequently, the model is effective at mitigating NMCS rates, but does little in the way of helping mitigate NMCMU or depot rates (Fry, 2010).

Recent research at RAND has focused on an approach that treats budget as a fixed known quantity that must be optimally allocated across different readiness priorities (Snyder et al., 2012). Optimization was evaluated based on the marginal impact of increased budget allocation on the AA of the aircraft in question. The researchers selected AA on the basis of its ability to provide insight into readiness across multiple platforms, a merit that was highlighted earlier on in this chapter.

Broader examinations of the impact of CAM-associated budgets on AA conducted by Fry (2010) yielded inconclusive results. However, Fry's research was limited by the fact that it only considered CAM budgets that were part of the cost per flying hour (CPFH) program. CPFH budgets are a subsection of the monies managed by CAM that are entirely separate from the DPEM funding outlined at the top of the chapter (AFMAN 64-143). Despite the name, FHP funding is not an exhaustive grouping of funding categories that promote sortie generation; one of the main takeaways from Fry (2010) is that DPEM funding categories should be included in future analysis of the interaction between budgets and readiness.

The broad list of sub-categorizations of AA offer many avenues of analysis of the various driving factors behind each. However, the models just outlined do not consider the role of the planner in the formation of budgets. Budgets are treated as either known constraints or as a freely adjusted variable in the mix between cost and readiness level. None consider that the planner has already theoretically identified the cost of optimal amount of readiness in the budget requirement, and the readiness consequences that result from deviating from the planned amount.

2.9 Conclusion

The defense economy operates under constraints that are not entirely analogous to those found in the private sector. The invisible hand of Adam Smith is supplanted by the defense planner, who must wade through the entanglement of politics, uninformed citizens, and unknown threats to provide budgets for the optimal level of readiness. In turn, readiness may be viewed in terms of utilization or in terms of availability. For the purposes of a quantifiable analysis across platforms, availability is the preferred perspective.

Depot level maintenance constitutes a significant proportion of aircraft unavailability. The centralization of depot level budgets under a single entity, CAM, offers the opportunity to evaluate how deviations from planned budgets impact depot downtime, and readiness as a whole in turn. Further analysis will need to be conducted in order to understand how budget gaps impact downtime differently across platforms and expenditure type at the enterprise level.

III. Methodology

3.1 Model Overview

The previous chapter identified the role that the empirical characteristics of the aircraft, manning levels, and budgets play on aircraft downtime (Ritschel et al., 2019; Chimka Nachtmann, 2007; Fry, 2010). However, none of the identified research specifically focused on the impact of these factors on downtime at the depot level. Figure Data from the last 10 years indicate that depot possession is the second most common downtime status impacting availability. The purpose of this chapter will be to identify a model that applies the empirical, personnel, and budgetary factors identified as conceptually relevant to aircraft downtime and apply them to the specific context of depot level downtime, thereby filling the gap in the current body of literature.

Much of the methodology expands upon that employed by Fry (2010). Many of the same considerations involving modelling cohort and analytical technique were made accordingly, but with key differences in the databases and specific variables involved. The relationship between depot downtime and budgets is investigated through the analysis of two distinct cohorts of aircraft. The first is a cohort of 23 separate platforms analyzed on an annual period from FY2010 to FY2019. The purpose of this first analysis is to provide a general look at the top-level relationships between funding gaps and depot downtime on an annual level. The analysis of the first cohort then gives way to analysis of a smaller, more concentrated, cohort of 7 different platforms that are analyzed at quarterly intervals from FY2015 to FY2019. The analysis of this second cohort serves to provide a more focused examination of how particular types of funding, down to quarterly movements, impact depot downtime.

3.2 Data Sources and Variable Collection

The empirical characteristics of the aircraft, manning levels, and budget allocation are sourced from three separate databases. Empirical characteristics of the aircraft, such as flying hours, age, and pilot reported discrepancies (PRD), are found in the Logistics Installations and Mission Support – Enterprise View (LIMS-EV) database. Personnel allocation and skill level, as identified through Air Force Specialty Code (DAFSC), are provided by the Air Force Total Ownership Cost (AFTOC) database. Finally, OS budgets allocated for depot level repairs are provided through the Air Force Logistics, Engineering and Force Protection directorate (A4) Funding Requirements Management (FRM) system. What follows is a description of each of the three databases and an outline of the data and variables collected therein.

LIMS-EV is comprised of a series of subsystems that provide enterprise-level analytics for different Air Force competencies (Headquarters Air Force [HAF]/A4 Public Affairs [PA], 2020). Metrics regarding aircraft utilization, availability, and fleet characteristics are contained within the LIMS-EV Weapons System View (WSV) panel (HAF/A4 PA, 2019). The data are updated directly from the Air Force's Core Automated Maintenance System (CAMS) and the Reliability and Maintainability Information System (REMIS) on a daily basis. The empirical aircraft characteristics, with the exception of the dependent variable, mirrors those initially explored in Ritschel et al. (2019) and are reproduced in Table 2. All of the variables in Table 2 are found using the LIMS-EV WSV. Where reasonable, the metrics collected from LIMS-EV are divided by Total Available Inventory (TAI) to provide a per-aircraft average of the metric. Providing the data in terms of an average facilitates the comparison of platforms with widely disparate fleet sizes (Fry, 2010; Ritschel et al., 2019).

Table 2. Empirical Variables

LIMS-EV Variables		
Variable Name	**Type of Variable**	**Description**
Depot hours (dpt.tai)	Continuous	Average number of depot hours per aircraft (Dependent Variable)
Year	Categorical	Fiscal year
Quarter	Categorical	Fiscal quarter
Total Active Inventory (TAI)	Continuous	Number of aircraft in the inventory
Mission Design	Categorical	Aircraft platform
Mission type	Binary	Mission type, such as bomber or tanker, as categorized in LIMS-EV
Aircraft age	Continuous	Average age of all aircraft in the corresponding platform
Hours flown per aircraft	Continuous	Average hours flown per aircraft
Sorties per aircraft	Continuous	Average sorties flown per aircraft
Pilot Reported Discrepancies (PRD) per aircraft	Continuous	Average number of times a pilot reported a discrepancy per aircraft
Breaks per aircraft (breaks.tai)	Continuous	Average number of PRDs that resulted in the aircraft becoming non-mission capable
Cannibalization (cann.tai)	Continuous	Average removals of a serviceable part from an aircraft or engine to replace an unserviceable part on another aircraft or engine
Repeat/recur per aircraft (rep.rec.tai)	Continuous	The average times a discrepancy occurs on the same system or subsystem on the first through fourth sortie attempts after originally reported
Mean Time Between Maintenance (MTBM)	Continuous	The average time, in hours, between all maintenance actions, both corrective and preventative

Depot possession hours serve as the model's dependent variable. Depot hours are sourced from LIMS-EV and are similarly divided by TAI to provide a per-aircraft average. Both the dependent and independent variables found within LIMS-EV can be further aggregated on a quarterly and annual level and sorted by platform.

AFTOC is a cost reporting system that primarily provides information on the costs of Air Force programs broken down by cost element structure (CES), but also provides insight into manning levels of each program (AFTOC portal website SAF/FMC heading). AFTOC calculates manning levels by providing a headcount of the quantity of personnel attached to a weapons system by DAFSC. DAFSC and weapons system are linked by the Program Element Code (PEC), an accounting identifier associated with each Air Force mission. The data variables collected from AFTOC are found in Table 3.

Table 3. Personnel Variables

Types of Budgetary Gap		
Variable Name	**Short Name**	**Description**
5 & 7 Levels (sup.rat)	Continuous	The quantity of enlisted 5 and 7-level maintainers per aircraft
3 Levels (e3.rat)	Continuous	The quantity of enlisted 3-level maintainers per aircraft
5 Levels (e5.rat)	Continuous	The quantity of enlisted 5-level maintainers per aircraft
7 Levels (e7.rat)	Continuous	The quantity of enlisted 7-level maintainers per aircraft
9 Levels (e9.rat)	Continuous	The quantity of enlisted 9-level maintainers per aircraft
1 Levels (o1.rat)	Continuous	The quantity of officer 1-level maintainers per aircraft
3 Levels (o3.rat)	Continuous	The quantity of officer 3- level maintainers per aircraft

Manning data from AFTOC were filtered to include DAFSCs identified as maintenance-oriented in the respective Air Force Officer/Enlisted Classification Directory. For enlisted airmen, only DAFSCs with the 2A prefix were included. For officers, only DAFSCs with the 21A prefix were included. Further filters were imposed to exclude officers in command billets. Out of all of these personnel, only enlisted 5, 7, and 9 levels are permitted to perform unsupervised maintenance on an aircraft (GAO, 2019). Moreover, enlisted 9 levels serve in a predominately supervisory capacity (GAO, 2019). Enlisted 5 and 7 levels are aggregated into one number, "sup.rat", to reflect the proportion of manning accomplishing technical work on the flightline.

Conversations with AFTOC database administrators revealed that the calculation of manning through AFTOC is not without caveats. Headcounts are averaged across the fiscal year such that a change in PEC or duty AFSC results in non-integer headcount values; if an airman changes station 6 months into the fiscal year, she will contribute 0.5 to her losing unit's headcount and 0.5 to her gaining unit's headcount. Moreover, one PEC can support multiple weapons systems. Headcounts are thus averaged across the weapons system just as they are for changes in PEC or DAFSC. This averaging occurs regardless of whatever the operational reality is at the squadron level. Additionally, the PEC does not distinguish between Overseas Contingency Operations (OCO) and peacetime funding. The manning data provided within AFTOC should therefore be only used to draw insights into the broad force-structure for each weapon system, not specific information as to the exact man-hours used to support each weapons system.

FRM contains data on budgetary execution levels as communicated in the execution plan and found at end of quarter and end of year positions. Per discussions with FRM analysts, the budget data reported within FRM are divided into three positions: requirement, plan, and actual. The required budget is the amount formulated at the unit-level that represents the ideal amount of funding that the unit would receive for an expenditure. The planned budget reflects the requested amount that is modified to reflect the reality that requirements cannot expect to be fully funded. Both the required and planned amount are found in the annual execution plan. Finally, the actual amount is the amount of funding that the unit has successfully obligated by the end of the fiscal period. All three of these budgetary positions are presented in thousands of dollars and escalated to FY2019 dollars using the appropriate inflation index from the office of the Deputy Assistant Secretary for Cost and Economics. Table 4 provides an outline of the different ways to measure the gaps between expected and received budgets.

Table 4. Methods of Expressing Budgetary Gaps

Types of Budgetary Gap		
Variable Name	**Short Name**	**Description**
Absolute Gap from Requirement	rqmt.delt	The amount, in dollars, that the actual budget deviated from the requirement budget. Positive numbers indicate a surplus. Negative numbers indicate a shortfall.
Gap as % of Requirement	rqmt.perc	The percentage that the actual budget constituted of the required budget. 100% indicates a fully funded requirement.
Absolute Gap from Planned	plan.delt	The amount, in dollars, that the actual budget deviated from the planned budget. Positive numbers indicate a surplus. Negative numbers indicate a shortfall.
Gap as % of Planned	plan.perc	The percentage that the actual budget constituted of the planned budget. 100% indicates a fully funded quarterly budget.
Quarterly Aggregated Gap	delt	The amount, in dollars, that the total budget summed across all AFEEICs deviated from the quarterly planned budget. Positive numbers indicate a surplus. Negative numbers indicate a shortfall.
Quarterly Aggregated Percentage	per.actual	The percentage the the total budget summed across all AFEEICs deviated from the quarterly planned budget

Budgets in FRM can be further sorted by Air Force Element of Expense Investment Code (AFEEIC), year, quarter, and mission design. Table 5 provides a legend to the different types of AFEEIC.

Table 5. Legend to Types of AFEEIC

AFEEIC Codes		
AFEEIC	**Title**	**Description**
54001 (SOF)	Software Depot Maintenance-Organic	Accounts for the purchase by the organic depots from Depot Maintenance Activity Group (DMAG) of the production effort required to design, code, test, and produce embedded weapon system and associated test system software after establishment of an initial software production baseline.
54101 (ACR)	Aircraft Maintenance - Organic	Accounts for purchase by organic depot of aircraft maintenance from DMAG.
54301 (ENG)	Engine Maintenance DPEM - Organic	Accounts for purchase by organic depot of engine maintenance from DMAG.
54401 (OMEI)	Major Item Maintenance DPEM - Organic	Accounts for purchase by organic depot of maintenance on other major items from DMAG.
54501 (EXC)	Exchange Item Maintenance DPEM – Organic	Accounts for purchase by organic depot of maintenance on exchangeable items from DMAG.
54601 (ABM)	Area Base Support Equipment DPEM - Organic	Accounts for purchase by organic depot of maintenance on area or base support equipment from the DMAG.
54801 (STOR)	AMARC DPEM – Organic	Identifies payments to Aerospace Maintenance and Regeneration Center (AMARC) or services provided to process, re-preserve, reclaim and withdraw aircraft.

Once again, the data are not provided without caveats. Of key import is the fact that the methodology used to categorize expenditures into different AFEEICs is not directly comparable between organic and contractor logistics supported (CLS) functions (Fry, 2010). Organic expenditures are categorized in accordance with the guidance provided by AFMAN 63-143. Contractors, meanwhile, report costs in accordance with their own accounting classifications that may differ across contactors (Fry, 2010). The result is that the same expenditure may be classified differently depending on whether it is organically supported or CLS. It is for this reason that only the definitions of organic EEICS are included in Table 5.

Furthermore, although CAM was formed in 2007, the FRM analysts cautioned that data prior to FY 2010 are unreliable. The analysts further advised that units only began reporting execution plan projections at the quarterly level in FY 2015. Prior to 2015, planned budgets are only reported at the annual level and quarterly spend plans were left blank. Finally, the advent of the 5-digit AFEEIC as a replacement of the 3-digit EEIC took hold in FRM in 2013. Distinguishing between organic and CLS expenditures is only possible using the full 5-digit AFEEIC.

3.3 Platform Selection and Data Aggregation

The limitations of the data create the need for two different modelling cohorts: one mixed cohort of CLS and organic aircraft and one cohort of strictly organic aircraft. A key deficiency of CLS data is that CLS aircraft definitionally lack any information on organic manning; maintenance on CLS aircraft is done by non-military contractors. Acquiring manning data that corresponds to Air Force skill level would require reaching out to each contracting support function individually, which is beyond the scope of this research. Similarly, detailed empirical data, such as break or cannibalization rates, are not available in LIMS-EV for many CLS platforms. Most importantly, the differing accounting classifications between CLS and organically supported systems means that budgetary analysis of a cohort involving both types of aircraft must remain at the aggregate level and cannot be broken out by AFEEIC.

Data on CLS aircraft are still analytically useful. Top-level requirement, planning, and actual budget data are still available for CLS aircraft. Data on average fleet age, flying hours, and number of sorties are also still available. These data are sufficient to create an initial, exploratory, cohort containing both CLS and organically support aircraft that are compared on the basis of fleet age, flying hours, number of sorties, and aggregate budget gaps. Likewise, aggregating CLS and organic airframes allows the usage of the full 10 years of FRM top-level budget data.

However, the selection of the platforms, regardless of support type, is further complicated by the need to for each platform to have robust data across *all three* of the databases. The list of platforms recorded in FRM from FY2010 to FY2019 serves as the starting point for possible aircraft platforms to be used in the model. These aircraft are then cross-checked against the lists of platforms available in AFTOC and LIMS-EV for the same period. Any aircraft that were phased out or phased in during this period were not included. Platforms with multiple Mission Design Series (MDS) suffixes, such as the A-10A and A-10C, were condensed to their base platform (MD). The exception to this is the F-15, which was split to distinguish between the F-15 and the F-15E.

Aggregating aircraft by platform, as opposed to the more granular MDS, is necessary as not all three databases provide the same level of granularity at the MDS level. As an example, the F-15C and F-15D are aggregated under a single heading in LIMS-EV, but broken out separately in AFTOC. Platform level nomenclature is more uniform than MDS nomenclature across the databases. Analyzing at the platform level ensures that the same aircraft are being compared. Data are thereby aligned by platform, year, and, for the second cohort, quarter for each of the three systems.

Nevertheless, initial analysis revealed that the C-130 platform, and its derivatives, is referred to inconsistently across the databases. Within FRM, the C-130 existed both as "C-130" and "C-130(SOF)", with no way to determine which MDS these platform titles encompass. Within AFTOC, certain C-130 MDS suffixes were included at the platform level, adding to the confusion. The C-130 family of platforms was therefore excluded from the dataset to prevent misalignment of data across the three databases.

Platforms were further restricted to those for which depot maintenance intervals were clearly outlined within Air Force Technical Order 00-25-4: Depot Maintenance of Aerospace Vehicles and Training Equipment. The U-2 was a notable absence. Furthermore, rotary-wing and unmanned aircraft were also excluded from analysis to preserve the homogeneity of the cohort. Lastly, because CAM only manages active duty funding, only active duty platforms and manning levels were analyzed. The full list of excluded aircraft is found alongside the reason for exclusion in Table 6. The remaining included aircraft are found in Table 7. Aircraft that were kept on for further analysis as part of the second cohort are bolded.

Table 6. Platforms Excluded from Analysis

Platform Name	Reason for Exclusion	Platform Name	Reason for Exclusion
AC-130J - Ghostrider	C-130 family	HH-60G - Pave Hawk	Not fixed wing
C/KC - 135	Not present in all years	KC-46	Not present in all years
C-130 - Hercules	C-130 family	MC-12	Not present in all years
C-130J - Super Hercules	C-130 family	MC-130J - Commando II	C-130 family
C-20	Not present in all years	MQ-1 - Predator	Unmanned
CV-22 - Osprey	Not fixed wing	MQ-9 - Reaper	Unmanned
E-8 Joint STARS Aircraft	Not present in all years	RQ-4 - Global Hawk	Unmanned
E-9	Not present in all databases	SOF (C130)	C-130 family
F015	Refers to phased out MDS	SOF (C-130J)	C-130 family
F015 (F-15A-D)	Refers to phased out MDS	TH-1H	Not fixed wing
F-117 - Nighthawk	Not present in all years	U2 Systems	Not included in TO 00-25-4
F-35	Not present in all years	UH-1H/TH-1H	Not fixed wing
HC-130 - King	C-130 family	UH-1N	Not fixed wing
HC-130J - Combat King II	C-130 family	VC-25	Not present in all databases

Table 7. Platforms Included in Analysis

Included Platforms	
A-10 Thunderbolt II	E-3 Airborne Warning and Control System
B-1 - Lancer	E-4
B-2 - Spirit	**F-15-C/D - Eagle**
B-52 - Stratofortress	**F-15-E - Strike Eagle**
C-17 Globemaster III	**F-16 - Fighting Falcon**
C-12 - Huron	F-22 - Raptor
C-21	KC-10 - Extender
C-32	KC-135 - Stratotanker
C-37	RC-135 - Manned Reconnaissance Syster
C-40	T-1 - Jayhawk
C-5 - Galaxy/Super Galaxy	T-38 - Talon
	T-6 - Texan II

Table 7 shows that the platforms in the second, smaller, cohort are strictly organically supported attack/fighter and bomber aircraft, as categorized in LIMS-EV. A fully organic cohort was necessary in order to include the full compliment of empirical data, as well as data on manning, that are otherwise unavailable in a set that contains CLS platforms. Attack/fighters and bombers were selected due to their unified fundamental mission purpose of deploying munitions against the enemy. Other platform types, such as tankers and ISR, are left as areas for future research. The F-22 was also excluded from the second cohort due to a lack of expenditures in organic AFEEIC categories.

The data for the second cohort are examined at the quarterly, as opposed to annual, level in order to increase the quantity of data points for the smaller set of platforms. FY2010 to FY2019 spans 10 years of data, whereas Q1 FY2015 to Q4 2019 spans 20 quarters. The additional time periods provide critical degrees of freedom needed to ensure a more robust analysis. The switch to quarters also requires that budgets are expressed in terms of planned, not required, budgets. Spend plans for the full requirement budget do not currently exist at the quarterly level.

3.4 Initial Model Construction

The many aircraft characteristic variables found in the literature review focus on the proximal events that would trigger the need for a repair (Ritschel et al., 2019). Previous research specifically found that age, breaks, MD, cannibalization, PRD, and fleet size were the most relevant aircraft characteristics relevant to unscheduled *unit* level repair events (Ritschel et al., 2019). However, as previously discussed, depot maintenance crucially differs from unit level repairs in that the overwhelming majority of depot level maintenance (DLM) is planned. Variable selection for depot level model therefore focuses less on the metrics that correlate with proximal causes of downtime, and instead focuses more on factors that are likely to increase the complexity and duration of scheduled DLM events.

With this in mind, the pertinent variables proposed by the body of literature on aircraft availability add up quickly. Between Tables 2, 3, and 5, there are 28 potential independent variables that have been pulled from the various databases. Many of the variables, such as sorties and flying hours, overlap significantly on even a conceptual level. Such variables that are so closely related conceptually are likely to introduce significant multicollinearity to the model. Paring down these variables to a theoretically cohesive list is the first step of the analysis. The theoretical models for the first and second modelling cohorts are outlined here respectively in Equation 12 and Equation 15:

$$Annual\,Total\,Depot\,Hours\,per\,Aircraft = f(Aircraft\,Empirical\,Variables)+ (Top\,Level\,Budgetary\,Variables) \quad (12)$$

$$Quarterly\,Total\,Depot\,Hours\,per\,Aircraft = f(Aircraft\,Empirical\,Variables)+ (Manning\,Variables) + (Top\,Level\,Budgetary\,Variables)+ (AFEEIC\,Level\,Budgetary\,Variables) \quad (13)$$

Following the lead of Fry (2010), significant relationships between independent variables are explored with the aid of correlation tables. The correlation table in Figure 6 illustrates the relationships between the possible variables to be used in the first, exploratory, cohort of aircraft. The wide variety of possible MDs, 23 in total, led to the aggregation of MDs into LIMS-EV mission type, such as bomber or tanker, in order to improve readability of the table. As expected, the strictly mathematical relationships between the different methods used to measure budgetary gaps resulted in high correlation coefficients between budgetary variables. Furthermore, sorties and flying hours shared a correlation coefficient of 0.42, indicating a moderate positive relationship between the two.

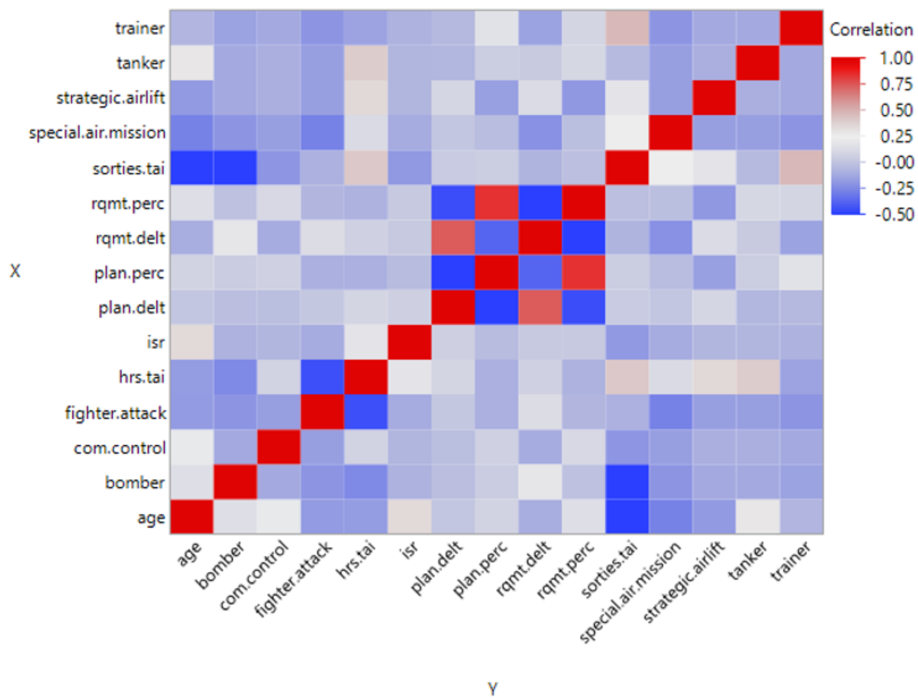

Figure 6. Correlation Matrix for Cohort 1

The demonstrated correlation between sorties and flying hours indicates that including both would introduce redundancy in the model. Importantly, sorties effectively track the amount of times aircraft systems are cycled on as they leave the hangar. Sorties are therefore a better measurement of the wear placed on components, such as the starter, that are used every time the aircraft leaves the hangar, but not necessarily for every hour of flight.

Meanwhile, the correlation table in Figure **??** , illustrates the correlations between the multitude of variables used for the second, fully organic, modelling cohort. Once again, the airframes have been aggregated by mission type to improve graph readability:

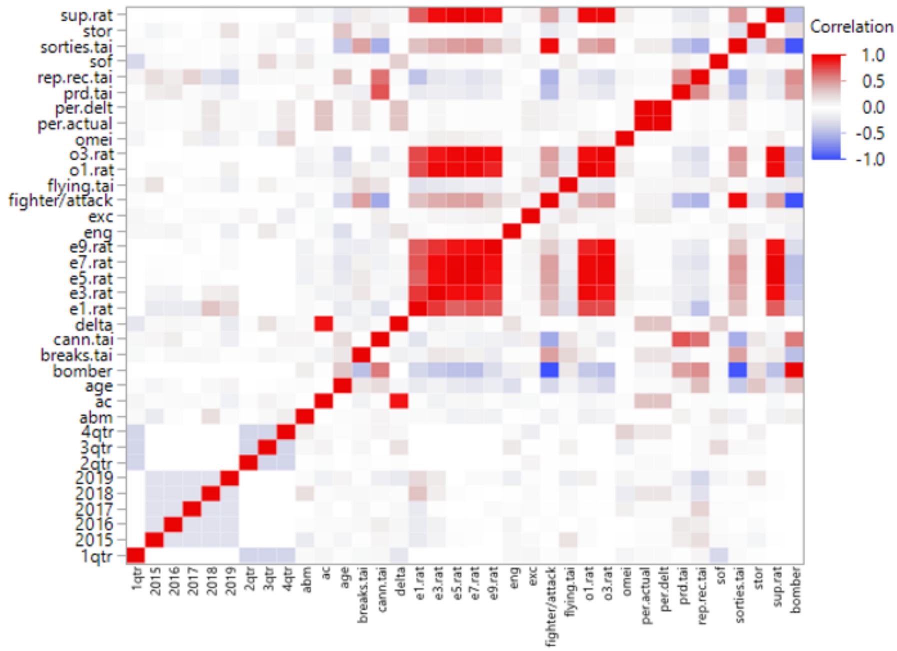

Figure 7. Correlation Matrix for Cohort 2
belfig:corr2

Of note are the significant correlations between all of the measurements of manning. The ratios of the quantity of maintainers of all skill levels, both officer and enlisted, to aircraft are highly interrelated. The high correlation is logically consistent with the fact that units will invariably seek to maintain specific ratios of skill levels; a 1-level officer is expected to supervise a set range of enlisted counterparts. The result is that it is possible to extrapolate from any one manning quantity of a specific skill level the number of superiors, peers, and subordinates that would support that quantity; the presence of one officer implies the presence of, say, ten enlisted. It is therefore unnecessary to include multiple measures of manning within the model. Instead, only the number of enlisted 5 and 7 levels, "sup.rat", who perform unsupervised work on the aircraft are included.

Aircraft empirical characteristics are also somewhat interrelated. The cannibalization rate is correlated both with the quantity of pilot reported discrepancies (PRDs),

=0.73, as well as the discrepancy repeat and recur rate, =0.64 . Indeed, bothersome PRDs may be corrected in the short run by cannibalizing parts from another aircraft in the hangar. Cannibalization rate is selected to remain in the model as a proxy for aircraft with recurring, but not ultimately mission compromising, issues. The logic is that aircraft with recurring issues at the unit level are more likely to require more involved maintenance at the depot level.

Budget variables, whether aggregated at the top level or broken out by AFEEIC, do not tend to correlate with any of the other variables. The only exception to this is the very strong correlation, =0.91, between the absolute magnitude of the budgetary gap in dollars, “delta”, and expenditures in the 54101 aircraft, “ac”, AFEEIC category. This correlation indicates a strong relationship between the magnitude of the overall budgetary gap and the magnitude of 54101 dollars as a subsection of the overall budget.

3.5 Balancing the Dataset

Breaking out the data by AFEEIC in the second cohort caused the dataset to initially be unbalanced. The issue is that requests for budgets in each AFEEIC are not made for all aircraft in all quarters. Typically, units have needs for expenditures that fit in two or three of the seven AFEEICs in a given fiscal year. Consequently, there were blank spaces in the raw dataset whenever a unit did not formulate an organic requirement for a specific AFEEIC in a given period. Additionally, there were 59 instances in which no requirement was planned, but fallout funds were still received and executed later in the fiscal year.

AFEEICS with no requested or received funding in a given period were assumed to be \$0. Crucially, the \$0 value is not assumed out of a lack of data. The budgets are not *missing* values; the values are simply blank. The \$0 assumption hinges on the interpretation that a lack of articulate need is conceptually identical to requesting $0. Future research may opt to explore the significance of this interpretation in greater detail.

Balancing the dataset in this way created a dataset for the quarterly cohort comprised of 7 platforms with an entry for each of the 7 AFEEICS in all 20 periods, totaling 140 rows of data. The annual dataset is comprised of 23 airframes over the course of 10 periods, totaling 230 rows of data. The data are presented in separate Microsoft Excel workbooks.

3.6 Regression Assumptions

The model is constructed using the ordinary least squares (OLS) method of estimation. The first three requirements of running an unbiased OLS estimation requires that the model be linear in parameters, the data be collected independently, and that there is no perfect collinearity between variables (Hilmer and Hilmer, 2014). These three requirements are met in the very creation of the model: the model is specified to be linear in parameters, the data that are collected are exhaustive of the populations that they examine, and none of the variables are perfectly collinear.

The next three assumptions concern the probability distribution of the error term, , and are required to ensure the OLS method provides the best linear unbiased estimates. (Hilmer and Hilmer, 2014) The three error terms requirements are that: 1) the error term has zero mean; 2) the error term is uncorrelated with each independent variable and all functions of each independent variable; 3) the error term has constant variance.

The first assumption is resolved by including an intercept variable within the equation that will account for non-zero bias in the residuals (Hilmer and Hilmer, 2014). The second assumption is violated when residuals from one period correlate with the residuals from a previous period. This issue, known as autocorrelation, is frequently found in data featuring a time component (Hilmer and Hilmer, 2014). Oliver (2001) indicated that time-series of aircraft operability are susceptible to this error, an observation that is upheld by Fry (2010). Autocorrelation is tested for using the Breusch-Godfrey/Woolridge test and is corrected for by using standard errors that are robust to autocorrelation through the Arellano method using the "sandwich" package in R (Torres-Reyna, 2010). Autocorrelation can also be controlled for by time-lagging the dependent variable and introducing it into the model as an independent variable (Hilmer and Hilmer, 2014). The dataset is reduced in size by the amount of time lags (Hilmer and Hilmer, 2014). The Arellano method is ultimately preferred to the lag method as it does not reduce the size of the dataset.

The third assumption is tested using the Breusch-Pagan test for heteroskedasticity. Low-p values reject the null hypothesis of constant variance (Fry, 2010). Heteroskedasticity is also mitigated through the usage of the aforementioned Arellano robust standard errors, as they are robust to both autocorrelation and heteroskedasticity (Torres-Reyna, 2010).

3.7 The Fixed Effects Model

Panel data are datasets that present the data both over a cross-section of units as well as over the course of a period of time (Croissant and Millo, 2008). In the context of this research, aircraft platforms are the units that are being observed longitudinally by year and quarter. The presence of different sub-groups of units within the dataset create the risk of a correlation between the independent variables and the error term (Hilmer and Hilmer, 2014). Left unchecked, this correlation will bias parameters generated through an ordinary least squares analysis (OLS).

Fixed effects models control for unwanted correlation by controlling for each subgroup within the dataset. Isolating each subgroup ensures that any error specific to a sub-group is included within the regression (Hilmer and Hilmer, 2014). Fixed effects models can be run in R by creating entity specific intercepts using the “plm” package. The method by which “plm” isolates subgroups within the panel data is mathematically equivalent to creating a dummy variable for each platform in the dataset. The choice to use a fixed-effects model over a random-effects model is tested through a Hausman test, which examines whether the error terms of the individual subgroups are correlated with the regressors. A significant p-value indicates the primacy of a fixed-effects model. Additionally, fixed effects models are preferred when the dataset constitutes an exhaustive sample of the population it is taken from, and especially when the data contain fewer than 8 sub-groups (Fox, Negrete-Yankelevich, and Sosa, 2014).

Fixed-effects models that control for a time component in addition to the sub-group component are known as two-way fixed-effects models. In the first, exploratory, modelling cohort, the periods are measured in consecutive fiscal years. Ten fiscal years correspond to 10 periods. Modelling the first cohort using a two-way model is once again identical to creating dummy variables for each time period. The second cohort, however, uses five years and four quarters per year. A two-way model will specify 20 different time periods. Including years and quarters as dummy variables will yield only 9 periods.

The nature of budgetary flows means that the second cohort is best modelled without the time effect. The typical longitudinal dataset tracks the observational units as they naturally evolve with the passing time; one period of time flows into the next and the effects of the independent variables accumulate (Hilmer and Hilmer, 2014). Quarterly budgetary data will flow from one quarter to the next as leftover funds rollover, but these funds reset at the beginning of each fiscal year. Thus, when faced with the choice between treating the quarterly data as one continuous stream or as separate clusters of time, the latter option is more appropriate.

3.8 Summary of Analytical Process

The first, larger, cohort of 23 aircraft is analyzed using a two-way fixed effects model. Data on age, numbers of sorties, number of flying hours, platform, and top-level budget gaps are included within the model that is refined over a few iterations of testing. The Hausman test is used to verify the suitability of the fixed effects format. Examinations of variation inflation factor (VIF) scores are used to check for multicollinearity in the model. The Breusch-Godfrey/Woolridge test and Breusch-Pagan test are used to test for autocorrelation and heteroskedasticity. Arellano robust standard errors are used to then mitigate any eventual autocorrelation and heteroskedasticity. The general form of the model is shown in Equation 14.

The second, more focused, cohort of 7 fighter/attack and bomber aircraft is analyzed using an individual fixed effects model. The inclusion of only organic aircraft allows for the inclusion of a larger quantity of variables relevant to the underlying characteristics of the aircraft, as well as variables that describe the manning levels of the aircraft. Budgets are further broken out by AFEEIC to investigate the relationship between specific categories of expenditure and hours of depot downtime. As with the first cohort, the model is validated using the Hausman, Breusch-Godfrey/Woolridge, and Breusch-Pagan tests. The model is similarly refined over multiple iterations. The general form of the focused cohort is shown in Equation 15.

$$
\begin{aligned}
&Annual\,Total\;Depot\;Hours\;per\;Aircraft\\
&\quad = f(Year) + (MD) + (Aircraft\;Age) + (\#\;of\;Sorties\;) + (Flying\;Hours)\\
&\qquad + (Total\;Enlisted\,5\;\&\;7\;levels) + (Total\;Budgetary\;Gap) \qquad (14)
\end{aligned}
$$

$$
\begin{aligned}
&Quarterly\,Total\;Depot\;Hours\,per\;Aircraft\\
&\quad = f(Year) + (Quarter) + (MD) + (Aircraft\;Age) + (Repeat/Recur)\\
&\quad + (\#\;of\;Breaks) + (\#\;of\;Sorties) + (Total\;Enlisted\,5\;\&\;7\;levels)\\
&\quad + (Budget\;Gap\;Software\;) + (Budget\;Gap\;Aircraft) + (Budget\;Gap\;ABM)\\
&\quad + (Budget\;Gap\;Engines) + (Budget\;Gap\;Exchangeables) + (Budget\;Gap\,OMEI)\\
&\qquad + (Budget\;Gap\;Storage) + (Total\;Budgetary\;Gap) \qquad (15)
\end{aligned}
$$

IV. Results

4.1 Results

The first iteration of the model, as seen in Table 8, includes all measurements of budgetary gaps as well as both sorties and flying hours alongside age in order to facilitate deeper analysis of the inclusion of each variable. The results of the Hausman test yielded a p-value of 0.0175, rejecting the null hypothesis that the unique errors of the model are not correlated with the regressors. This result indicates that the model is correctly specified using a fixed effects approach as opposed to a random effects approach. The Breusch-Pagan test for heteroskedasticity produced a p-value < 0.01. Furthermore, the Breusch-Godfrey-Woolridge test for serial correlation yielded a similar result, $p < 0.01$. The results of these tests respectively indicate the presence of both heteroskedasticity and autocorrelation in the model. Finally, Table 9 shows the VIF scores of the model.

The models are specified so as to mitigate the bias and imprecision brought about by the heteroscedasticity, autocorrelation, and multicollinearity revealed in the first model. Heteroscedasticity and autocorrelation are corrected using Arellano robust estimators (HAC estimator), as previously discussed in the methodology section (Torres-Reyna, 2010). The Arellano method is employed in all iterations of the model to ensure robust estimators.

Table 8. Annual Model 1

	Dependent variable	
	Total Annual Depot Hours per Aircraft	HAC Standard Errors
	(1)	(2)
Age	319.056***	319.056**
	p = 0.010	p = 0.030
Sorties	-6.788**	-6.788**
	p = 0.014	p = 0.030
Flying Hours	-0.795	-0.795
	p = 0.221	p = 0.212
Gap Requirement vs Actual	0.002	0.002
	p = 0.263	p = 0.227
Gap Planned vs Actual	-0.002	-0.002
	p = 0.233	p = 0.188
% Actual of Requirement	-1.097	-1.097
	p = 0.713	p = 0.579
% Actual of Planned	1.364	1.364
	p = 0.603	p = 0.580
Observations	230	
R^2	0.132	
Adjusted R^2	-0.040	
F Statistic	4.166*** (df = 7; 191)	
Note:	Significance Codes: "*" 0.05, "**" 0.01, "***" 0	

Table 9. VIF Scores of Annual Model 1

VIF Scores	
Age	1.599
Sorties	1.987
Flying Hours	1.266
Gap Requirement vs Actual	3.403
Gap Planned vs Actual	3.406
% Actual of Requirement	4.780
% Actual of Planned	4.950

Even when using Arellano robust estimators to correct for heteroskedasticity and autocorrelation, the output of the first iteration of the model yields very little explanatory power. It should be noted at this juncture that the interpretation of coefficients created by 2-way fixed effects models is not as straightforward as for individual fixed effects models. Coefficients in a two-way model are expressed relative to the individual's own average over time, but also relative to the averages of the other individuals in the model (Kropko & Kubinec, 2018). Kropko and Kubinec (2018) find that the coefficients can, at best, be interpreted as "generalizations of the effect of deviations from the case-means at a particular point in time... (or) for each particular case." Hence, the significance of sorties, $p<0.05$, in the model should not be taken as an indicator that every additional sortie decreases depot time by 6.788 hours. Instead, the results express the potential for a loosely negative relationship between sorties and depot maintenance downtime hours.

Additionally, the fixed effects methodology employed by the "plm" package reports the R^2 strictly of the within estimators, not of the model as a whole; the explanatory power of the fixed individual and time effects are not included in the adjusted R2 output of Table 8. Although the "within" R^2 is expected to be lower than the overall R2, the results of this model are so low so as to make it impossible to draw any meaningful inferences from the relationship between the size of a gap in depot level budgets and the amount of hours aircraft spend undergoing depot level maintenance.

The concern of possible collinearity between sorties and flying hours was not ultimately born out in the model. Table 9 shows that the VIF scores are both well under 5, indicating minimal overlap. Indeed, the VIF scores for all of the models fell comfortably below the threshold score of 5. However, while age does not have an explicitly high VIF score theoretical considerations make age's inclusion in the model inappropriate. From a conceptual standpoint, average age of the fleet can conceivably decrease from one year to the next if particularly old aircraft are retired. In practice however, the average change in fleet age from one year to the next was only 0.97 years. The absolute proximity of 0.97 to 1 indicates that the impact of age is already effectively accounted for within the time component of the fixed-effects model.

The next set of models built off of the first model by examining deviations from required budgets and deviations from planned budgets separately. Furthermore, the choice remained between measuring budgets in terms of absolute deviation in dollars or in terms of the percentage the received amount constituted of the initially required or planned amount. Table 10 compares the two approaches using required budgets and Table 11 does the same using planned budgets. The HAC estimators are also presented within the table.

Table 10. Annual Model 2

	Budget from Requirement Standpoint			
	Annual Depot Hours vs. Budget in $	HAC Standard Errors	Annual Depot Hours vs. Budget as %	HAC Standard Errors
	(1)	(2)	(3)	(4)
Sorties	-5.748**	-5.748*	-5.740**	-5.740*
	p = 0.039	p = 0.052	p = 0.038	p = 0.055
Flying Hours	-0.502	-0.502	-0.496	-0.496
	p = 0.442	p = 0.475	p = 0.447	p = 0.479
Gap Requirement vs Actual	-0.00004	-0.00004		
	p = 0.944	p = 0.917		
% Actual of Requirement			0.512	0.512
			p = 0.697	p = 0.665
Observations	230		230	
R^2	0.080		0.081	
Adjusted R^2	-0.080		-0.079	
F Statistic	5.672*** (df = 3; 195)		5.725*** (df = 3; 195)	
Note:		Significance Codes: "*" 0.05, "**" 0.01, "***" 0		

Table 11. Annual Model 3

	Budget from Planned Standpoint			
	Annual Depot Hours vs. Budget in $	HAC Standard Errors	Annual Depot Hours vs. Budget as %	HAC Standard Errors
	(1)	(2)	(3)	(4)
Sorties	-6.116**	-6.116**	-5.872**	-5.872*
	p = 0.028	p = 0.046	p = 0.034	p = 0.051
Flying Hours	-0.502	-0.502	-0.485	-0.485
	p = 0.440	p = 0.467	p = 0.456	p = 0.482
Gap Planned vs Actual	-0.001	-0.001		
	p = 0.276	p = 0.316		
% Actual of Planned			1.232	1.232
			p = 0.276	p = 0.376
Observations	230		230	
R^2	0.086		0.086	
Adjusted R^2	-0.074		-0.074	
F Statistic	6.102*** (df = 3; 195)		6.103*** (df = 3; 195)	
Note:			Significance Codes: "*" 0.05, "**" 0.01, "***" 0	

The results of the tables indicate that budgets gaps continue to have low explanatory power over total depot hours at the annual level. Although sorties are statistically significant across all of the models, the overall "within" R^2 and adjusted R^2 do not change greatly between models. Measuring budgets in terms of their deviations from the required amount or the planned amount does not meaningfully change the explanatory power of the model. Similarly, alternating between examining budget as an absolute deviation, versus as a percentage does not appear to make a difference. The result is that it is not reasonable to make inferences about the relationship between budgets and depot level downtime at this level of analysis.

The focus of the analysis then shifts from a large cohort of aircraft at the annual level to a smaller subsection of organically supported attack/fighter and bomber aircraft at the quarterly level. The examination of the smaller, fully organic, cohort allows the inclusion of more aircraft-specific metrics, such as break-rates, as well as variables related to the manning levels of each MD. Budgets are also divided into AFEEIC category for a more granular investigation of which types of budget dollars have a measurable relationship with depot hours.

The smaller cohort is run without a fixed time effects component to reflect the aforementioned non-continuous nature of budgets from fiscal year to the next. Each year and quarter are assigned a dummy variable for inclusion in the model. 2015 and quarter 1 serve as the base cases. Moreover, the absence of a fixed time effect allows for the reintroduction of the age variable into the model. The individual effects of each platform still remain in the model. Table 12 shows the first iteration of the model featuring the smaller cohort.

Table 12. Quarterly Model 1

	Dependent variable:	
	Total Quarterly Depot Hours per Aircraft	HAC Standard Errors
	(1)	(2)
Age	-0.569	-0.569
	p = 0.928	p = 0.860
Sorties	-16.037^{***}	-16.037^{**}
	p = 0.001	p = 0.042
Flying Hours	0.625	0.625
	p = 0.347	p = 0.518
Breaks	9.301	9.301
	p = 0.576	p = 0.496
Repeat/Recur	13.603	13.603
	p = 0.658	p = 0.521
PRD	0.891	0.891^{**}
	p = 0.147	p = 0.021
Cannibalization	14.319	14.319
	p = 0.103	p = 0.284
% Actual of Planned	0.004	0.004^{*}
	p = 0.229	p = 0.055
Gap SOF AFEEIC	-0.001	-0.001
	p = 0.451	p = 0.519
Gap ACR AFEEIC	0.00002	0.00002
	p = 0.931	p = 0.929
Gap ENG AFEEIC	-0.0004	-0.0004
	p = 0.712	p = 0.526
Gap OMEI AFEEIC	0.017	0.017
	p = 0.510	p = 0.417
Gap EXC AFEEIC	-0.003	-0.003
	p = 0.937	p = 0.847
Gap ABM AFEEIC	0.032	0.032
	p = 0.347	p = 0.231
Gap STOR AFEEIC	-0.011	-0.011^{**}
	p = 0.230	p = 0.016
5&7 Levels	-4.450	-4.450
	p = 0.802	p = 0.836
2016	-27.030	-27.030
	p = 0.106	p = 0.246
2017	-4.528	-4.528
	p = 0.824	p = 0.889
2018	22.586	22.586
	p = 0.430	p = 0.582
2019	51.981^{*}	51.981^{***}
	p = 0.085	p = 0.006
Qtr 2	21.077	21.077^{**}
	p = 0.170	p = 0.022
Qtr 3	38.220^{**}	38.220^{**}
	p = 0.038	p = 0.048
Qtr 4	13.687	13.687
	p = 0.418	p = 0.498
Observations	140	
R^2	0.293	
Adjusted R^2	0.107	
F Statistic	1.982^{**} (df = 23; 110)	

Note: Significance Codes: '*' 0.05, '**' 0.01, '***' 0

Table 13. VIF Scores of Quarterly Model 1

VIF Scores	
Age	1.845
Sorties	6.872
Flying Hours	1.369
Breaks	3.149
Repeat/Recur	4.118
PRD	2.832
Cannibalization	4.598
% Actual of Planned	1.328
SOF	1.266
ACR	1.332
ENG	1.230
OMEI	1.172
EXC	1.157
ABM	1.147
STOR	1.279
5&7 Levels	1.923
2016	1.692
2017	1.706
2018	2.390
2019	2.593
Qtr 2	1.689
Qtr 3	1.959
Qtr 4	1.893

Breusch-Pagan and Breusch-Godfrey/Woolridge tests reveal the presence of heteroskedasticity and autocorrelation, $p>0.05$, in the initial model. These issues are once again mitigated using Arellano robust standard estimators. VIF scores for the model are within the acceptable range of $VIF < 5$, with the exception of sorties, $VIF = 6.872$. The inclusion of the sorties variable is conceptually supported by the way that sorties act as a proxy measure for the wear and tear on aircraft systems that occur every time the aircraft goes through the power cycle. Moreover, the results of the models have consistently indicated that sorties are statistically significant. For these reasons, flying hours are removed from the next iteration of the model despite having the lower VIF score.

The overall fit of the quarterly model is greatly improved over the annual model.

The adjusted "within" R^2 indicates that 10.7% of the model's variation is explained by the model when excluding the predictive power of platform. Moreover, budgetary variables appear as statistically significant for the first time. The overall percentage that the actual budget received constitutes of the initially planned budget is significant, p=0.055, with a positive coefficient of 0.04. This result is counterintuitive, as it suggests that higher amounts of funding relative to the planned amount coincide with greater depot downtime hours, even if only just slightly. The negative and significant, p=0.016, value for storage expenditures is more intuitive. The result indicates that the more money spent within this category, the fewer hours of depot maintenance. The magnitude of the coefficient is nevertheless very small. Making great inferences into the model's output remains ill-advised.

In the final quarterly model, flying hours, breaks, and the repeat/recur rate are excluded. None of these variables approached statistical significance in the first passthrough. Moreover, the cannibalization rate has previously been determined to be the most appropriate indicator for problematic aircraft in the methodology section. The final iteration of the quarterly model is found in Table 14.

The results of the final quarterly model do not differ significantly from the first iteration of the quarterly model. As before, storage expenditures are statistically significant, p=0.003, and slightly negative. Sorties remain a significant variable, p=0.042, and negative. This result suggests that aircraft that fly more sorties experience fewer depot downtime hours. The quantity of PRDs is also significant, p=0.021, and positive. This is an intuitive outcome, as it suggests that an increased quantity of PRDs results in more depot downtime hours down the line.

Table 14. Quarterly Model 2

	Dependent variable:	
	Total Quarterly Depot Hours per Aircraft	HAC Standard Errors
	(1)	(2)
Age	0.199	0.199
	p = 0.975	p = 0.938
Sorties	-12.957***	-12.957***
	p = 0.003	p = 0.006
PRD	1.159**	1.159**
	p = 0.045	p = 0.050
Cannibalization	14.490*	14.490
	p = 0.092	p = 0.244
% Actual of Planned	0.004	0.004*
	p = 0.248	p = 0.095
Gap SOF AFEEIC	-0.001	-0.001
	p = 0.517	p = 0.496
Gap ACR AFEEIC	0.00003	0.00003
	p = 0.904	p = 0.917
Gap ENG AFEEIC	-0.0002	-0.0002
	p = 0.834	p = 0.703
Gap OMEI AFEEIC	0.017	0.017
	p = 0.515	p = 0.451
Gap EXC AFEEIC	-0.005	-0.005
	p = 0.886	p = 0.676
Gap ABM AFEEIC	0.038	0.038
	p = 0.264	p = 0.244
Gap STOR AFEEIC	-0.013	-0.013***
	p = 0.164	p = 0.003
5&7 Levels	-4.172	-4.172
	p = 0.811	p = 0.854
2016	-31.328*	-31.328
	p = 0.055	p = 0.146
2017	-5.370	-5.370
	p = 0.786	p = 0.862
2018	14.792	14.792
	p = 0.553	p = 0.655
2019	45.781*	45.781***
	p = 0.090	p = 0.0002
Qtr 2	17.737	17.737***
	p = 0.228	p = 0.003
Qtr 3	35.257**	35.257**
	p = 0.041	p = 0.024
Qtr 4	12.443	12.443
	p = 0.448	p = 0.460
Observations	140	
R^2	0.279	
Adjusted R^2	0.114	
F Statistic	2.192*** (df = 20; 113)	
Note:		Significance Codes: "*" 0.05, "**" 0.01, "***" 0

V. Conclusions and Future Research

5.1 Conclusion

The results of this study, like many of those that preceded it, do not provide definitive conclusions. No matter how budgets were examined in concert with other variables, budgets were rarely significantly associated with depot downtime hours. Deviation from top level budgets, in terms of percentage, and storage budgets were found to be significant, but with coefficients of fairly low magnitude and in models with little overall explanatory power. Examining the results in the context of the initial research questions will shed light into the potential policy implications of these results, as well as provide an opportunity to highlight the limitations of the study and future avenues of study.

How do shortfalls or surpluses in depot-allocated budgets impact depot downtime hours?

The results of the annual model indicate that neither shortfalls nor surpluses impact the amount of depot downtime hours. Digging down to the quarterly level indicated the potential for a relationship between overall budgets, expressed as a percentage, and downtime hours. The results, although statistically significant, fundamentally did not hold large amounts of explanatory power.

The ability to uncover robust relationships is ultimately limited by the availability of data within FRM. Annual data were only available over the past 10 fiscal years. 10 years is not a large window in economic terms. Moreover, the 5 years for which quarterly data represent an even smaller piece of the puzzle. In comparison, Spinney had access to over 30 years of data at the time of his initial report. Identifying large relationships, such as that between budgets and depot hours, require decades of data to better capture the ebbs and flows in spending trends within the DoD. Indeed, the last 10 years have seen sequestration as well as drawdowns in both Iraq and Afghanistan. The 10 years before that saw the start of the Global War on Terror and associated increase in forces. The bottom line is that 5 to 10 years of data may not be sufficient to view the full range of funding levels that any given requirement can expect to receive.

Investigating the relationship between depot hours and budgets is also limited by the fact that depot level maintenance is scheduled along rigorously predetermined timeframes that limit the volume of truly unexpected costs. TO 00-25-4 explicitly states that "depot maintenance will be accomplished on a planned basis to facilitate the programming of funds, material, manpower, facilities and other resources." A clear nexus is drawn on the doctrinal level between formalized depot schedules and easing the task of budget allocation. Furthermore, depots have 90 days of flexibility when conducting program depot maintenance. Hence, depot maintenance is simultaneously predictable, but also flexible in the event of contingencies. It is possible that the margins of error and high level of information available to maintenance planners allows depot managers to balance depot down hours within a certain range for each aircraft. The absence of a counterfactual, in which an airframe is consistently and significantly defunded, limits the ability to investigate the impacts of budgets on downtime hours at the extremes.

In a world of perfect planning, any deviation from a perfect plan will cause imperfect outcomes. The current combination of variables within the model suggest that deviations from planned budgets do not have immediate effects on availability. This is not proof that downtime will be the same regardless of surplus or shortfall. Instead, the research indicates that there may be enough flexibility in the current process of budget allocation that linear analyses of budgets and downtime may not uncover the interaction between the two; The readiness “receipt” remains more complex than its name would imply.

Which types of expenditures, by AFEEIC, have the greatest impact on depot maintenance hours?

Only the AFEEIC corresponding to storage expenditures was found to be statistically significant. Even still, within the broader context of the explanatory power of the model, the role of storage expenditures should not be overstated to imply a causal relationship. The remaining AFEEIC categories were not found to be significant, and therefore conclusions regarding their relative impact on depot maintenance hours cannot be made.

Investigating the impact of budgets by category only examines a piece of the resources managed by defense planners. In addition to funds, the above quote from the TO 00-25-4 also makes reference to the programming of “material... and other resources”. This research ultimately does not go into detail into the mechanisms behind the allocation of manpower or the logistical issues that depots face when accomplishing depot maintenance. Indeed, LIMS-EV does not categorize depot maintenance to the same degree that unit-level maintenance is. The NMCS, NMCM, and NMCB labels all offer insight that is unavailable for depot level repairs.

5.2 Future Research

Autocorrelation within panel data additionally presents unique difficulties. The cross-section of different sub-groups makes lagging variables a complex and intensive process; each variable must be lagged for each time period, for each subgroup. Future methods of analysis may want to use methods that are more favorable to time-series data, such as Auto Regressive Integrated Moving Average (ARIMA), as opposed to traditional OLS methods.

From a theoretical standpoint, the planned nature of depot maintenance implies the existence of a base timeline for depot down hours. Future researchers may opt to investigate actual depot hours relative to their deviation from the amount of depot hours planned for the platform in the given year. Focusing future research to a single MDS, as opposed to a cohort of multiple platforms, may be the optimal choice for future researchers seeking to create a scalable model of the budget and depot hour relationship.

Bibliography

Advanced Technology Incorporated. (1980). *Sensitivity of System Readiness to Resource Allocation*. Defense Technical Information Center. https://doi.org/10.21236/ADA090290

Air Force Logistics Management Agency. (2009). *Maintenance Metrics U.S. Air Force* (J. Rainey & C. Young, Eds.).

Biddle, S. (2004). *Military Power*. Princeton University Press; JSTOR. https://doi.org/10.2307/j.ctt7s19h

Boito, M., Keating, E., Wallace, J., DeBlois, B., & Blum, I. (2015). *Metrics to Compare Aircraft Operating and Support Costs in the Department of Defense*. RAND Corporation. https://doi.org/10.7249/RR1178

Carrell, S. E., & Hauge, J. A. (2009). Politics and the Implementation of Public Policy: The Case of the US Military Housing Allowance Program. *Public Choice*, *138*(3/4), 367–386. JSTOR.

Chimka, J. R., & Nachtmann, H. (2007). Operational Readiness as a Function of Maintenance Personnel Skill Level. *Air Force Journal of Logistics*, *31*(3), 8.

Coyne, C. J., & Lucas, D. S. (2016). Economists Have No Defense: A Critical Review of National Defense in Economics Textbooks. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2742391

Coyne, C. J., Michaluk, C., & Reese, R. (2016). Unproductive entrepreneurship in US military contracting. *Journal of Entrepreneurship and Public Policy*, *5*(2), 221–239. https://doi.org/10.1108/JEPP-12-2015-0037

Croissant, Y., & Millo, G. (2008). Panel Data Econometrics in R: The plm Package. *Journal of Statistical Software; Vol 1, Issue 2 (2008)*. https://doi.org/10.18637/jss.v027.i02

Dahlman, C., & Thaler, D. (2002). Ready for War but not for Peace: The Apparent Paradox of Military Preparedness. In Z. Khalilzad & J. Shapiro (Eds.), *Strategic appraisal: United States air and space power in the 21st century* (pp. 437–481). Rand.

Dunne, J. P. (1995). The Defense Industrial Base. In K. Hartley & T. Sandler (Eds.), *Handbook of Defense Economics* (pp. 400–427). Elsevier.

Eisenhower, D. (1961, January 17). *Transcript of President Dwight D. Eisenhower's Farewell Address (1961)*. https://www.ourdocuments.gov/doc.php?flash=false&doc=90&page=transcript

Fox, G. A., Negrete-Yankelevich, S., & Sosa, V. J. (Eds.). (2015). *Ecological Statistics: Contemporary Theory and Application*. Oxford University Press.

Fry, F. G. (2010). *Optimizing Aircraft Availability: Where to Spend Your Next O&M Dollar* [Master's Thesis]. Air Force Institute of Technology.

Goldfein, D., & Donovan, M. (2019, September 12). *Receipt to Congress—Air Force readiness on the rise*. The Air Force Times. https://www.airforcetimes.com/news/your-air-force/2019/09/12/receipt-to-congress-air-force-readiness-on-the-rise/

Government Accountability Office. (n.d.). *Air Force Readiness: Actions Needed to Rebuild Readiness and Prepare for the Future* (Testimony GAO-19-120T). Government Acountability Office. https://www.gao.gov/assets/700/694923.pdf

Government Accountability Office. (2019). *Military Personnel: Strategy Needed to Improve Retention of Experienced Air Force Aircraft Maintainers* (Report to Congressional Committees GAO-19-160). https://www.gao.gov/assets/700/696747.pdf

HAF/A4PA. (2019, August). *LIMS-EV Playbook.*

HAF/A4PA. (2020, January 22). *LIMS-EV 101 Briefing*.

Hankins, M. W. (2018). *The Cult of the Lightweight Fighter: Culture and Technology in the U.S. Air Force, 1964-199*. 481.

Harrison, T. (2014). Rethinking Readiness. *Strategic Studies Quarterly*, *8*(3), 38–68. JSTOR.

Hartley, K. (2012). Conflict and Defence Output: An Economic Perspective. *Revue d'économie Politique*, *122*(2), 171. https://doi.org/10.3917/redp.218.0171

Hartley, K., & Solomon, B. (2016). Special Issue: Defence Inflation. *Defence and Peace Economics*, *27*(2), 172–175. https://doi.org/10.1080/10242694.2015.1123464

Hartung, W. D. (1999). Ready for What?: The New Politics of Pentagon Spending. *World Policy Journal*, *16*(1), 19–24. JSTOR.

Higgs, R. (1990). Fifty Years of Arms, Politics, and the Economy. In *Arms, Politics, and the Economy: Historical and Contemporary Perspectives* (pp. xv–1). Holmes & Meier Publishers.

Hildebrandt, G. G., & Sze, M.-B. (1990). *An Estimation of USAF Aircraft Operating and Support Cost Relations*. RAND Corporation. http://www.rand.org/pubs/notes/N3062/

Hilmer, C. E., & Hilmer, M. J. (2014). *Practical Econometrics: Data Collection, Analysis, and Application* (First edition). McGraw-Hill Education.

Hlavac, M. (2018). Stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.1. *Central European Labour Studies Institute (CELSI)*. https://CRAN.R-project.org/package=stargazer

Hove, K. H., & Lillekvelland, T. (2019). On Growing Operating Costs in the Armed Forces. *Defence and Peace Economics*, *30*(4), 438–453. https://doi.org/10.1080/10242694.2017.1413621

Hummel, J., & Lavoie, D. (1990). National Defense and the Public Goods Problem. In *Arms, Politics, and the Economy: Historical and Contemporary Perspectives* (pp. 37–60). Holmes & Meier Publishers.

Kaldor, M. (1986). The Weapons Succession Process. *World Politics*, *38*(4), 577–595. https://doi.org/10.2307/2010167

Kirkpatrick, D. (2003). A UK Perspective on Defence Equipment Acquisition. *Institute of Defence and Strategic Studies*, *41*, 23.

Kropko, J., & Kubinec, R. (2018). Why the Two-Way Fixed Effects Model is Difficult to Interpret, and What to Do About It. *SSRN Electronic Journal*, 27.

Lazear, E. P., & Rosen, S. (1981). Rank-Order Tournaments as Optimum Labor Contracts. *Journal of Political Economy*, *89*(5), 841–864. JSTOR.

Losey, S. (2018, March 6). *Fewer planes are ready to fly: Air Force mission-capable rates decline amid pilot crisis*. Air Force Times. https://www.airforcetimes.com/news/your-air-force/2018/03/05/fewer-planes-are-ready-to-fly-air-force-mission-capable-rates-decline-amid-pilot-crisis/

Losey, S. (2019, August 1). *Aircraft mission-capable rates hit new low in Air Force, despite efforts to improve*. Air Force Times. https://www.airforcetimes.com/news/your-air-force/2019/07/26/aircraft-mission-capable-rates-hit-new-low-in-air-force-despite-efforts-to-improve/

Mattis, J. (2018). *Memorandum for All Members of the Department of Defense*. The United States Department of Defense. https://media.defense.gov/2018/Mar/27/2001894968/-1/-1/1/Be-Peerless-Stewards-of-Taxpayers-Dollars.PDF

Mehta, A. (2017, December 2). *Air Force officials: Return of sequestration would 'break' service*. Air Force Times. https://www.defensenews.com/air/2017/11/10/air-force-officials-return-of-sequestration-would-break-service/

Meserve, J. (2010, October 26). *USAF Maintenance Metrics: Looking Forward with Aircraft Availability*.

Mintzberg, H. (2017, March 31). The U.S. Cannot Be Run Like a Business. *Harvard Business Review*. https://hbr.org/2017/03/the-u-s-cannot-be-run-like-a-business

Moore, S. C., Stockfisch, J. A., Goldberg, M. S., Holyord, S. M., & Hildebrandt, G. G. (Eds.). (1991). *Measuring Military Readiness and Sustainability*. Rand.

Oliver, S., White, E., & Arostegui, M. (2001). Forecasting Readiness. *Air Force Journal of Logistics*, *25*(3), 31–42.

O'Malley, T. J. (1983). *The Aircraft Availability Model: Conceptual Framework and Mathematics* (p. 112) [Model Documentation]. Logistics Management Institute.

Petcoff, R. (2010, May 5). *Air Force Program Recognized for Excellence in Government*. U.S. Air Force. https://www.af.mil/News/Article-Display/Article/116767/air-force-program-recognized-for-excellence-in-government/

R Foundation for Statistical Computing. (2019). *R: A Language and Environment for Statistical Computing.* (Version 3.5.3) [Computer software]. http://www.R-project.org

Ritschel, J. D., Ritschel, T. L., & York, N. B. (2019). Providing a piece of the puzzle: Insights into the aircraft availability conundrum. *Journal of Defense Analytics and Logistics*, *3*(1), 29–40. https://doi.org/10.1108/JDAL-09-2018-0015

Snyder, D., Kim, J., Carrilo, M., & Hildebrandt, G. (2012). *Improving Air Force depot programming by linking resources to capabilities*. RAND.

Spinney, F. C. (1985). *Defense facts of life: The plans/reality mismatch*. Westview Press.

Stillion, J., & Orletsky, D. T. (1999). *Airbase Vulnerability to Conventional Cruise-Missile and Ballistic-Missile Attacks*. RAND.

The Department of Defense. (2019). *DOD Dictionary of Military and Associated Terms*. United States Department of Defense.

Twight, C. (1990). Department of Defense Attempts to Close Military Bases: The Political Economy of Congressional Resistance. In *Arms, Politics, and the Economy: Historical and Contemporary Perspectives* (pp. 37–60). Holmes & Meier Publishers.

Von Mises, L. (1958). *Liberty & property*. https://mises.org/library/liberty-and-property

Zeileis, A. (2004). Econometric Computing with HC and HAC Covariance Matrix Estimators. *Journal of Statistical Software*, *11*(10), 1–17. https://doi.org/10.18637/jss.v011.i10

Zeileis, A. (2006). Object-Oriented Computation of Sandwich Estimators. *Journal of Statistical Software*, *16*(9), 1–16. https://doi.org/10.18637/jss.v016.i09

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
26-03-2020	Master's Thesis	SEP 2018 - MAR 2020

4. TITLE AND SUBTITLE

The Effects of Funding Gaps on Depot Maintenance Hours

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

D'Amato, Carlo, S. 1st Lt, U.S. Air Force

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENV-MS-20-M-196

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Undisclosed

10. SPONSOR/MONITOR'S ACRONYM(S)

Undisclosed

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

12. DISTRIBUTION / AVAILABILITY STATEMENT

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

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The relationship between expenditures and readiness level is a topic of interest to military senior leaders, defense resource planners, and the American taxpayer alike. Senior leaders within the Air Force (AF) justify increased defense budgets by pointing to the potential adverse effects that decreased funding could have on military readiness. Resource planners within the AF are then tasked with the responsibility of ensuring that budgets are allocated most effectively to maximize the AF's ability to project airpower across a variety of contingency operations. This thesis investigates the relationship between budgets and readiness by examining the relationship between depot level funding and hours of aircraft downtime spent at the depot. Funding is analyzed in terms of the magnitude that the amount of funding receives deviates from the amount of funding requested by the planner. The analysis ultimately did not find any conclusive relationship between deviations from requested depot budget levels and the number of hours of downtime spent at the depot.

15. SUBJECT TERMS

Budget Allocation, Defense Planning, Resource Allocation

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Lt Col Scott Drylie, Ph.D., AFIT/ENV
					19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	64	(312) 785-3636, x4441; scott.drylie@afit.edu