



EXCERPT FROM THE
PROCEEDINGS
OF THE
EIGHTEENTH ANNUAL
ACQUISITION RESEARCH SYMPOSIUM

**Assessing the Reliability of the Future Years Defense
Program and Building a Forecast**

May 11–13, 2021

Published: May 10, 2021

Approved for public release; distribution is unlimited.

Prepared for the Naval Postgraduate School, Monterey, CA 93943.

Disclaimer: The views represented in this report are those of the author and do not reflect the official policy position of the Navy, the Department of Defense, or the federal government.



The research presented in this report was supported by the Acquisition Research Program of the Graduate School of Defense Management at the Naval Postgraduate School.

To request defense acquisition research, to become a research sponsor, or to print additional copies of reports, please contact any of the staff listed on the Acquisition Research Program website (www.acquisitionresearch.net).



ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF DEFENSE MANAGEMENT
NAVAL POSTGRADUATE SCHOOL

Assessing the Reliability of the Future Years Defense Program and Building a Forecast

Andrew Hunter—is a Senior Fellow in the International Security Program and Director of the Defense-Industrial Initiatives Group at CSIS. From 2011 to 2014, he served as a Senior Executive in the Department of Defense, serving first as Chief of Staff to Under Secretaries of Defense (AT&L) Ashton B. Carter and Frank Kendall, before directing the Joint Rapid Acquisition Cell. From 2005 to 2011, Hunter served as a Professional Staff Member of the House Armed Services Committee. Hunter holds an MA in applied economics from the Johns Hopkins University and a BA in social studies from Harvard University. [ahunter@csis.org]

Greg Sanders—is a Fellow in the International Security Program and Deputy Director of the Defense-Industrial Initiatives Group at CSIS, where he manages a research team that analyzes data on U.S. government contract spending and other budget and acquisition issues. In support of these goals, he employs SQL Server, as well as the statistical programming language R. Sanders holds an MA in international studies from the University of Denver and a BA in government and politics, as well as a BS in computer science, from the University of Maryland. [gsanders@csis.org]

John Severini—is a graduate student at Columbia University's School of International and Public Affairs, where his research focuses on international security policy and quantitative methods. He currently works for the Arnold A. Saltzman Institute of War and Peace Studies on projects primarily dealing with maritime military issues. Severini holds a BA in political science and in linguistics from The Ohio State University. He can be reached at [john.severini@columbia.edu]

Contributing Authors: **Chitrakshi Bhardwaj, Gabriel Coll, Cuong Nguyen, Gerhard Ottehenning**

Abstract

Discerning, negotiating, and communicating priorities are necessary tasks for the U.S. defense acquisition system to effectively implement its portion of the National Defense Strategy. One of the Department of Defense's central tools for doing so is the Future Years Defense Plan (FYDP), a projection of the cost and composition of the force over the next 5 years. This project created a dataset and employs it to study FYDP reliability, focusing on two sources of uncertainty: differences in approach between military departments and differences in volatility between those line items in the base budget and those that include contingency operation spending.

Introduction

For the U.S. defense acquisition system to properly implement its portion of the National Defense Strategy, it must effectively discern, negotiate, and communicate its priorities. One of the Department of Defense's (DoD's) central tools for this process is the Future Years Defense Plan (FYDP), a projection of the cost and composition of the force over the next 5 years.

Annually updated and submitted as part of the president's budget submission projection, the FYDP provides important insights into the DoD's priorities and projections of the future, both internally and externally. Internally within the DoD, wherein the FYDP is constructed, the process forces the stakeholders involved to debate tradeoffs and outline their visions of the future. Externally, it lays out for Congress a vision of how U.S. national security strategies could be implemented in practice, which the legislature must then choose whether to fund or alter. It helps the U.S. defense industry understand where the DoD plans to invest and thereby allows companies within the industry to align themselves with current priorities. It helps scholars identify trends and do research on major capital-intensive projects, which can be used to inform future projects, both defense and nondefense. It helps U.S. citizens identify how the government plans to spend its taxpayers' dollars. However, the FYDP has a few major drawbacks for these stakeholders that undercut its ability to communicate priorities.



The first drawback is the inherent tension between the FYDP's role expressing the funding amount that the executive branch deems necessary to support the strategy and its role in creating a plan that can be implemented within the funding amount authorized and appropriated by Congress.

A second, related, shortfall is the absence of any measure of reliability or predictive intervals for the projections. Some parts of the DoD budget are easier to predict than others, but the point estimate provided by the FYDP does not differentiate between known quantities, like the purchase of uniforms, and cutting-edge technology, like the development of a next-generation alloy. That said, by design, Overseas Contingency Operations (OCO) budgets operate as a pressure valve for uncertainty by taking some of the most volatile spending out of the base budgets and FYDP and managing them through OCO methods instead. However, the intended functionality of the OCO accounts is muddled when predictable spending is moved to OCO accounts to avoid budget caps.

Third, the unclassified FYDP is released in a form that makes it straightforward to study topline spending or individual line items or programs but challenging to analyze anything in between. This is because the FYDP is released in dozens of PDFs through separate justification books, and not as a centralized database or even in summary documents. Collectively, these limitations present a higher barrier to entry to stakeholders and make it laborious for specialists and unappealing for anyone else to put investment plans in a meaningful context.

The FYDP is a system for planning rather than a forecasting tool, but there are nonetheless multiple benefits to understanding the relationship between its projections and actual spending. First, stakeholders can better employ the system and its results if its strengths and biases are more transparent. Second, this analysis can put common assumptions to the test, for example, the volatility of OCO spending or if the long-term Navy Shipbuilding planning process results in more accurate projections of future needs. Third, for the defense industry, the difference between projected and actual spending can be a key indicator of risk and aid in investment planning. Finally, defense spending must often respond to external changes and updated strategies; large gaps can indicate not just failure of prediction but also speed in adapting the larger defense enterprise to new priorities.

This project has created a dataset to ease research of the FYDP and uses that dataset to address the question, *How reliable are projections within the FYDP as an indicator for actual spending?* It tests the value of the unclassified FYDP for investment spending, RDT&E and Procurement, as a bottom-up indicator of DoD priorities by comparing cumulative projections through fiscal year (FY) 2019 spending from 1, 3, and 5 years in advance with the actual cumulative spending. In addition, the paper examines whether there are differences between military departments and between line items that do and do not include OCO spending.

Literature Review

There are a multitude of challenges in defense planning even within the base budget. The United States, despite its resources and robust analytical staff, faces more difficult challenges than those of many of its peer countries. First, the United States is a presidential system with projections prepared by the executive branch but funding authority resting with Congress. A projection process could be designed that does more to incorporate congressional opinions into the planning process, but the role of two co-equal branches of government means that some degree of uncertainty for both the topline funding and for individual projects of interest to the legislature is irreducible.

Second, the United States is the global leader in defense research and development, and as Light et al. (2017) find, "there is a considerable amount of cost and schedule growth risk



facing all [Major Defense Acquisition Programs] at [Milestone] B” when the DoD commits to significant development spending (p. 44). Even a better estimation approach would be highly unlikely to eliminate uncertainty in defense research and development.

In analyzing the 2020 President’s Budget (PB2020), Matthew Woodward and David Arthur (2019) draw on the FYDP and project specific reporting. Employing historical factors developed from studies by the RAND Corporation and the Institute for Defense Analysis, including those referenced previously, they find that “using the resulting cost estimates instead of DoD’s cost estimates raises total projected acquisition costs by 3.5 percent over the FYDP period and by 6.1 percent over the 2026–2035 period” (p. 16).

The relationship between these project estimates and the larger FYDP projections is complex, with estimated cost influencing budget requests and a sense of total available funds influencing what the budget is able to fund and at what level. For example, the CBO estimates explore the cost implications of the President’s Budget by keeping present plans constant. In practice, MDAPs and other budget lines can be descope, slowed down, or canceled outright.

Topline FYDP projections should not be treated simply as a sum of composite budget lines but as a consequential form of estimating in their own right. As Todd Harrison and Seamus Daniels (2020) note,

Previous inflection points in the defense budget, both up and down, have been influenced by wars, shifts in strategy, changes in the threat environment, and economic conditions. ... While there does not appear to be appetite to cut the defense budget in the remainder of FY 2020 or in FY 2021, as is evident by the inclusion of additional defense funding in stimulus bills, the political environment could shift markedly once an economic recovery is underway in FY 2022 or FY 2023. (pp. 57–58)

While the pandemic makes the present particularly challenging to forecast, the problem has never been an easy one for the DoD. Kevin Lewis (1994) finds that plans routinely expect small incremental changes, but in practice, changes, in topline spending or individual programs, are regularly more dramatic and often cyclical (pp. 110–113). Leland Jordan (2015) goes further and argues that historically most administrations project more funding than materializes, showing “systematic fiscal optimism” (p. 274). Jordan (2015) analyzed budgets from 1975 to 1995 and discovered that 70% of the projections exceeded the appropriated amounts (pp. 282–283).

Effective projection should introduce difficult choices in the present, before expensive commitments have been made, rather than in the future, where cost overruns or budget shortfalls may lead to the termination of programs experiencing difficulty or redirection of funds from programs that are presently successful. Jordan (2015) concludes that “those administrations having demonstrated the greatest bias in their real growth projections also most seriously handicapped program managers” (p. 288).

Enduring Budget in Overseas Contingency Operations

As mentioned in the introduction, OCO budgets acknowledge the existence of uncertainty due to external events such as wars or other cases where events beyond the control of the DoD require rapid responses. However, the distinction between the base budget and OCO has been undercut by budgetary maneuvering since the passage of the Budget Control Act (BCA) of 2011. In an attempt to reduce federal budget deficits, the BCA established spending limits on discretionary budget authority, applying to both defense and non-defense programs. These limits do not apply to the OCO budget. Subsequently, the executive and legislative branches have evaded the caps by transferring some predictable enduring spending



out of the base budget and into OCO spending. This approach has been criticized by some defense experts and government officials from both parties. Katherine Blakeley and Lawrence Korb (2014) from the Center for American Progress voice their concerns that “financially, the free flow of war funding has decimated any pretense of fiscal discipline at the Pentagon. ... Unclear budget guidance and poor financial management have allowed DoD to pay for substantial enduring costs with war funding rather than the base budget, further muddying the waters” (p. 28). Then Acting White House Chief of Staff Mick Mulvaney also criticized in strong terms the “use of OCO funding for base budget requirements” (McGarry & Epstein, 2019, p. 9).

The Congressional Budget Office (CBO) estimates that from 2006 to 2018, more than \$50 billion in OCO funding per year (in 2019 dollars), on average, has gone toward the costs of enduring activities rather than the temporary costs of overseas operations. This is particularly transparent in the FY 2020 President’s Budget. As Harrison and Daniels (2020) report, “the request shifted entire categories of funding, such as Army Ammunition Procurement, from the base budget into OCO” (pp. 4). With the coming expiration of the BCA caps in FY 2021, the FYDP released with the FY 2020 budget request then shifts all the money back to base spending after the expiration.

The insertion of enduring items within the OCO budget does not necessarily undermine the value of OCO in those cases where it is still used as intended. Andrew Hunter (2019) defends OCO by noting that in the last decade, a major portion of OCO funds was used to support the operations of Afghan security forces. As the actual size, operational employment, and equipment of these forces have changed rapidly, OCO funds were extremely useful to sustain the mission. Due to the fact that the levels are not planned out a full 5 years in advance, OCO allows the performance of missions that might be practically impossible otherwise. Moreover, the origin of OCO was an attempt to bring more oversight and transparency to wartime emergency supplemental bills. Senator McCain, objecting to the inclusion of non-emergency procurement in a supplemental bill,

demanded that DoD submit its request for war funding along with the regular budget so that it might receive a similar level of congressional scrutiny as the base budget and so that it would be less easy to embellish with non-war-related funding. (Hunter, 2019)

The Strategic and Communication Role of the Future Years Defense Plan

The FYDP helps the executive branch make strategic choices and communicates them to internal and external audiences. Critically, the FYDP is released as part of the President’s Budget and thus linked to the negotiation between branches of the U.S. government. The numbers are provided to justify the funding requests the executive branch is making to Congress. As Todd Harrison and Seamus Daniels (2020) put it,

The FYDP is therefore best understood as a statement of policy rather than a prediction of where the budget is headed. It is an indication, with considerable detail, of the Defense Department’s priorities and trade-offs among modernization, force structure, and readiness. (p. 11)

This emphasis on policy and strategy provides another set of criteria that can be used to evaluate the effectiveness of the FYDP. In particular, Thomas-Duerrel Young (2018) is critical of long term defense planning as practiced in the United States. While other authors have emphasized the benefits of budgetary stability making efficient choices, Young believes this fails to acknowledge the way adversaries can unpredictably shape choices. Instead, Young highlights two tasks that defense planners can achieve: “to produce costed priorities” and “creating an understanding of future financial projection of current obligations” (p. 366). For Young, the quality of cost estimates matters for costing priorities, but the value of the FYDP is



not to provide reliable predictions. Instead “the utility of these financial projections should be judged by how much flexibility they can provide ministers and senior defense officials to change the way money is being spent to produce relevant defense outcomes” (Young, 2018, p. 370).

While strategic flexibility is inherently appealing, it does still face challenges noted by Kevin Lewis (1994) in his argument for defense planning humility. Lewis (2014) observes a range of cyclical factors in the defense budget and a shrinking portion of the budget going to combatant forces. He cautions, “we should expect inefficiencies, and substantially more negative effects on capability than might be expected from these inefficiencies, because of the increasing role of ‘fixed-cost’ overhead items” (p. 132). That said, the larger emphasis on the FYDP’s role in achieving flexibility harks back to Alain C. Enthoven and K. Wayne Smith (2005), who reject the idea that long-range plans limit the president’s ability to implement a new strategy and argue that “an organization’s flexibility to move in a new direction is greatly reduced if it lacks a clear picture of the direction in which it has been heading” (p. 50).

Young’s (2018) emphasis on the strategic flexibility provided by the FYDP presents a challenge; while there has been extensive research comparing projections to actual costs, measuring the agility of the acquisition enterprise is a less intuitive problem. Will Domke (1984) provided one possible answer by analyzing how the Defense Budget responded to presidential priorities going back to the Eisenhower administration by analyzing the winners and losers among DoD funding accounts. He found that the balance between military departments change most at the start of an administration and more popular president’s have greater influence (p. 389).¹

Picking up on Domke’s (1984) approach, Travis Sharp (2019) argues that what the current debate on gaps between strategy and resources “generally [does] not provide, however, is any objective sense of whether DoD is doing relatively better or worse aligning resources with strategy” (p. 9). Sharp (2019) considers three diagnostic tests including a Winners test that looks at whether spending has shifted into areas identified by the National Defense Strategy as priorities.² He looks at the FYDP’s Major Program Categories and investment areas, with the latter only available through the President’s Budget. In both cases, Sharp (2019) finds that the 2020 President’s Budget failed this test (p. 24).

Data and Methods

Data Sources and Structure

This project focuses on budget lines rather than major programs for both financial and policy reasons. Todd Harrison (2016) reports that as of the FY 2016 President’s Budget, “these smaller programs account for an average of 57 percent of the total acquisition budget over the FYDP” (p. 24). Moreover, during the study period, the DoD is increasingly experimenting with alternate channels, including mid-tier acquisition and other transaction authority, responding to pressure from the Executive Branch and Congress to pursue speed and innovation.

To better understand this era of reduced reliance on the major weapons system pipeline, this report chooses to focus its attention on procurement line items and RDT&E program elements. This is not the most detailed level of analysis available; however, it has the advantage of being available from multiple sources. The first pair of these sources are the P-1s for Procurement and R-1s for RDT&E. These budget documents are provided as spreadsheets that

¹ Domke (1984) does not include FYDP data in his analysis.

² “Instead, a winner is best defined as one that receives the largest increase in its proportional share of DoD’s total spending compared to what DoD’s prior outyear plans forecasted” (Sharp, 2019, p. 45).



cover the entire DoD enterprise, going back to the 1998 President's Budget (DoD Comptroller, 2020). However, these detailed and convenient documents do not include FYDP's out years.

Instead, for detailed future year projections at the procurement line item (P-40) and RDT&E program element (R-2) level, it is necessary to turn to the Justification Books. In these documents, the military departments and agencies lay out their spending request and describe what is being bought as well as providing program management details. A major challenge for open-source researchers is that for investment spending alone, each President's Budget is accompanied by dozens of PDF files, splitting the information based on organization and funding account.

This project overcomes the limitations of the justification books and does so with the benefit of two external sources. First, a predecessor FYDP analysis led by Gabriel Coll bulk laid the foundation for this project by downloading many of the justification books and conducting an initial analysis. A parallel effort by the CSIS's Defense Budget Analysis (DBA) group greatly accelerated this effort with the discovery that the justification books, starting with the FY 2013 FYDP, have included XML encoded spreadsheets that do not require the intensive data cleaning effort necessary when scraping PDFs. This past and parallel work assisted in the creation of the dataset, and, in the DBA case, provided a valuable source for cross checking totals and budget line classifications.

To allow for cross-comparisons, the team has imported R-1 and P-1 budget requests from FY 2011 to FY 2021, using the most recent files and most recent columns within those files. Much of the effort of the dataset focused on the creation of unique identifiers, called CSIS budget line keys, that ease the process of tracking a budget line across different sources and over time. While many of the same columns are available in the R-1s and R-2s as well as in the P-1 and P-40s, there are discrepancies in the labels used, which are more challenging in early years and with procurement data in particular. For example, in some years and for some agencies, the line number, that is to say, the order in which it is presented in that PB, is the same as a line item. This causes problems as P-1s and P-40s include slightly different budget lines, leading to misalignment, and because order regularly changes from year to year. Moreover, line item standards change over time, with many budget lines changing from having a six-character line number to a 10-character one while still having the same broad topical focus. For procurement, for the analysis in this paper, all cost types pertaining to a single program are combined.³

A related challenge is that in a typical year, a few score PEs and LIs will cease to receive funding or be tracked for the first time. This is a natural outflow of changes in strategy, priorities, and technology. However, some of the time, a new PE or LI does mean a genuinely new project but may instead reflect a change in identifiers for an existing budget line of greater magnitude than just an increase in the length of the identifier used. This may be a matter of a change of agency, such as the move from the Defense Health Agency to the Defense Health Program, or a reclassification, such as in the lead-up to the creation of the Space Force.

The ties between the original sources and their unique identifiers are recorded within the dataset's repository for transparency and reproducibility reasons. The study team has taken the additional step of classifying budget lines based on our confidence that potential confounding factors have been removed.

³ In the early years of the dataset, Advanced Procurement LIs sometimes had a different line item than the rest of a program. These divergent advanced procurement budget lines have been combined by the study team with the rest of their program under the same CSIS budget line key.



Figure 3 shows a summary of the R-1 and R-2 portions of the dataset by confidence level in the left and right columns, respectively. The top row shows the total actual spending associated with the projections of each PB. For the R-1, this includes only the actual spending for the budget year in question. The R-2s show substantially more spending because their projection window covers the President's Budget and four out years. For those cases where projections go beyond FY 2019, the spending amount projected is shown in gray.

Focusing on those projections for which we know the actual spending, there are five confidence levels of interest:

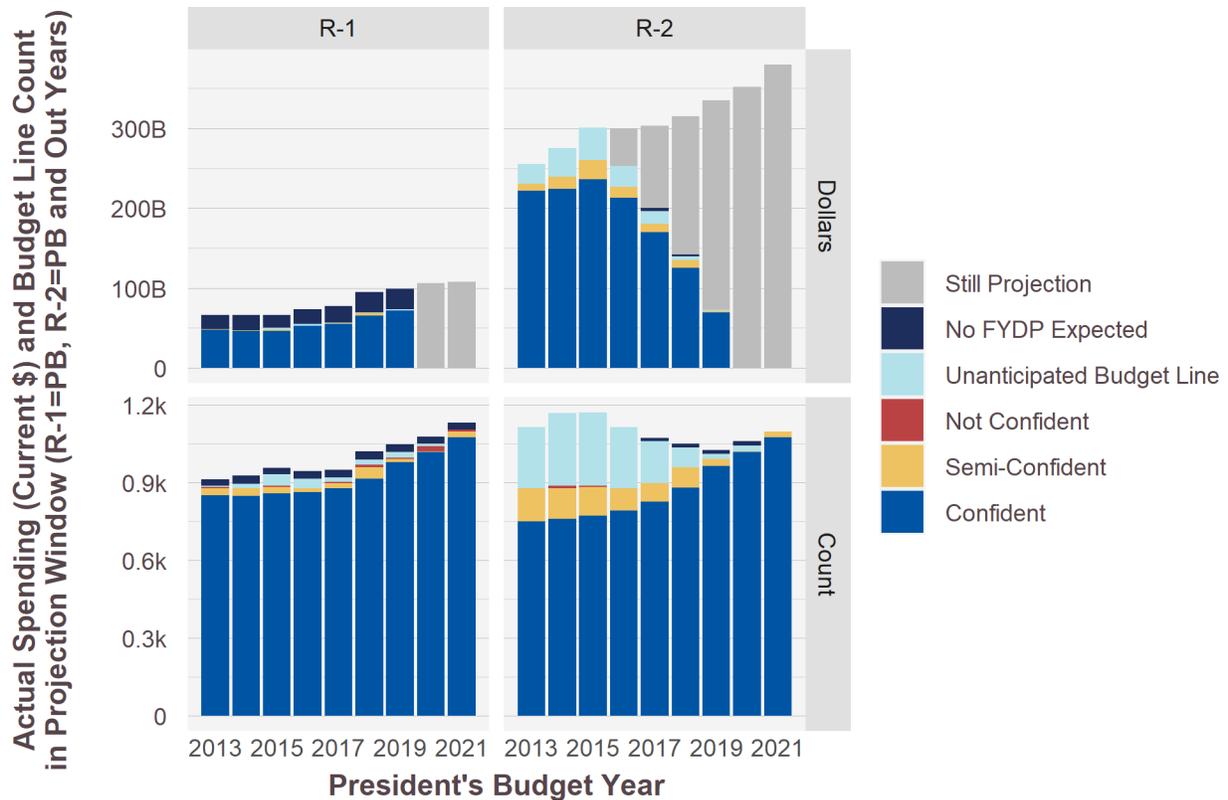
- No FYDP Expected: This covers budget lines, such as classified spending, that are typically not included in the unclassified FYDP as reported by the R-2s or P-40s.
- Unanticipated Budget Line: This covers budget lines that did not yet exist when the PB was published. They may be a genuinely new budget line, or perhaps a transfer whose predecessor was not identified by the study team.⁴
- Not Confident: This category covers budget lines that are present in one source but not in the other.⁵
- Semi-confident: This category covers budget lines that are present in both sources but that have one of a range of known discrepancies. This includes having a difference projection between the two sources of more than \$2,000. In the case where there is a missing cost type, budget lines are reported as semi-confident if there is a net difference between the two projections of \$2,000 or less but the spending occurs in different years depending on the source. Finally, if a budget line projected spending in a future year, but the budget line was not reported in the year in question, it is marked as semi-confident.
- Confident: This is the category that is the basis of most of the charts in this paper. To be classified as confident, both sources must report projections within \$2,000 of one another. A budget line may end during the projection period and still be classified as confident, but only if the ending was anticipated by the President's budget. These strict criteria are intended to limit the sample to those budget lines that are genuinely starting and ending rather than having overlooked connections.

Turning again to Figure 1, the larger light blue blocks present in the R-2 column indicate that there is a substantial amount spent by PEs that the FYDP does not see coming years in advance. The lower row of Figure 1 shows the same budget lines, but using the metric of the count of lines rather than the amount of actual spending.

⁴ Unanticipated Budget Line is smaller in more recent years because it is a lagging indicator and only available once actual spend figures are known.

⁵ For procurement, this also covers cases in which there is a cost type under the budget line, for example, advanced procurement, that is not present in the other source and there is a net difference in projection of more than \$2,000 between the two sources.





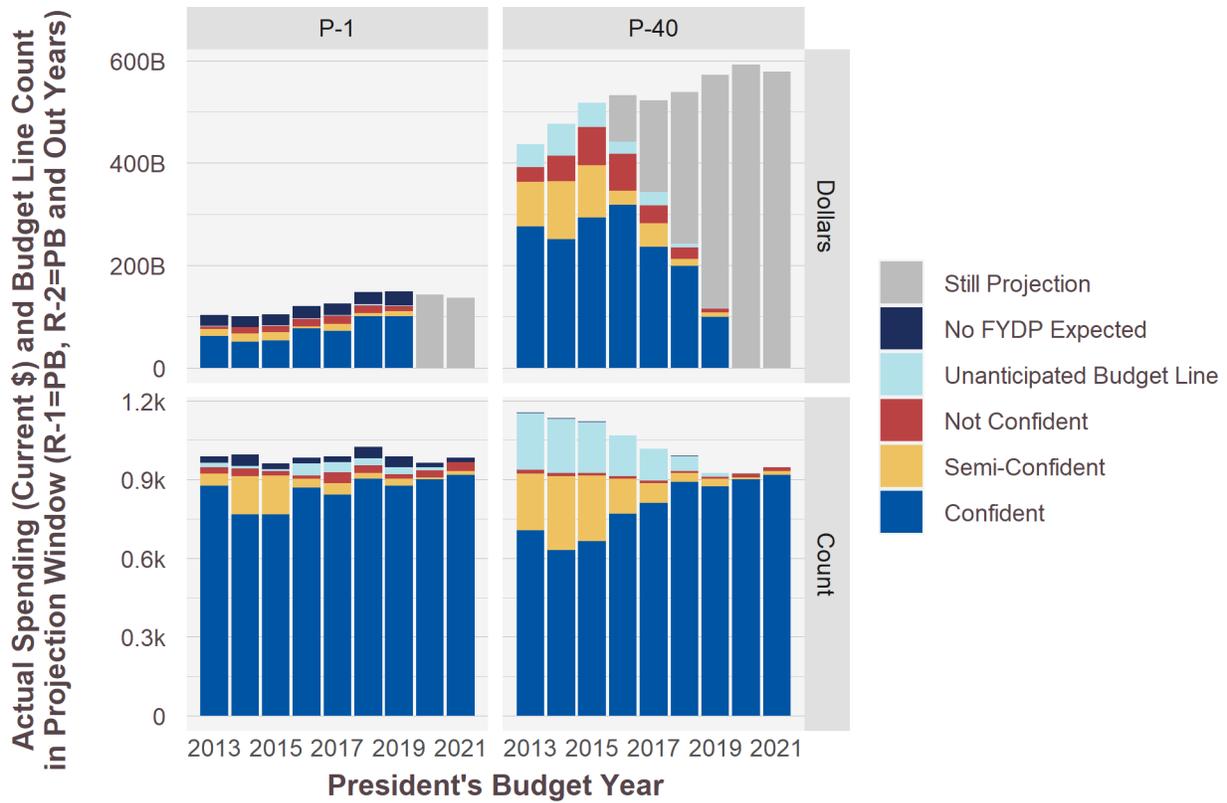
Source: R-1s, R-2s, CSIS Analysis

Figure 1. RDT&E Actual Spending and Budget Line Count by Level of Confidence

The Procurement dataset is summarized in Figure 2, and by the prevalence of red and yellow, shows that there are more disagreements between the P-1s and P-40s than between the R-1s and R-2s. There were multiple contributors to these problems. First, there were gaps in reporting on the P-40 side, though some of this the study team has already overcome by manually transcribing the PDF files that were missing XML files.⁶ Second, cost type categories, particularly reductions to adjust for prior year past procurement, were sometimes missing from P-40 reporting and merit closer examination.

⁶ The most problematic absence in dollar terms had been in the Navy Shipbuilding and Construction account. For some President's Budgets, key columns went unreported, but these were overcome by imputing the values using other available data. For both the R-2s and P-40s PB2014 and PB2015, the justification books did not include OCO spending with a note that those figures were to be released later. The study team imputed these values from the P-1s and R-1s. In PB2016, the total President's Budget spending column was missing from the P-40s, but the study team imputed that number by adding base and OCO spending together for each row. In PB2017, both base and OCO spending were amended after the R-2s and P-40s were published, and the study team again drew from the R-1s and P-1s to impute the amended values.





Source: P-1s, P-40s, CSIS Analysis

Figure 2. Procurement Actual Spending and Budget Line Count by Level of Confidence

All dollar amounts in this report represent current dollars. Although changes in spending timing (delays, shifts, etc.) occur throughout the time period, much of the FYDP estimations themselves have inflation concerns built into their reasoning. For statistical purposes, values across multiple-years have been aggregated into a cumulative expenditure; for example, analyzing the full 5-year window involves comparing the sum of 5 years of projection and 5 years of actual spending.

There are several tradeoffs to this methodology. The analysis is exchanging year-specific sensitivity for a more robust measurement better representative of the discrepancies being assessed over the time periods in question. This sacrifices the ability to directly compare the predictive strength of different FYDP years (i.e., testing the extent to which the second and third out years are more reliable than the fourth and fifth out years). Likewise, this approach makes it harder to account for any single year having an abnormal occurrence, such as sequestration. In return, the variables being compared serve as a much more accurate representation of what is being estimated in total. From an industrial standpoint, this serves as a more natural portrayal of how spending is being looked at overall.



Results

How Reliable Are FYDP Projections?

Figure 3 shows cumulative FY 2019 actual spending plotted against FYDP projections starting in 2015.⁷ The Y-Axis is the Total Spend, the logged set of actual expenditures fully realized as outlays over given years. The X-Axis is the Projected Budget, the logged set of prior expenditure estimates for all budget lines expected to exist in those future actual spending years. The black diagonal lines represent the boundary where the projected budget is exactly equal to the total spend. Points to the upper left of the line have more spending than expected, and points to the lower right projected more spending than actually occurred. Points lying on the axes are the result of projected budget lines that ended up having no actual spend (x-axis) or budget lines appearing with actual spend that projected no spending or were unanticipated (y-axis). The graph's different columns correspond to different comparison periods: the PB2015–PB2019 graph includes 5 years of projections, while the PB2019–PB2019 graph includes only the first year with no out years. The upper row covers all budget lines, while the lower row includes only those budget lines where the study team is confident in the data quality.

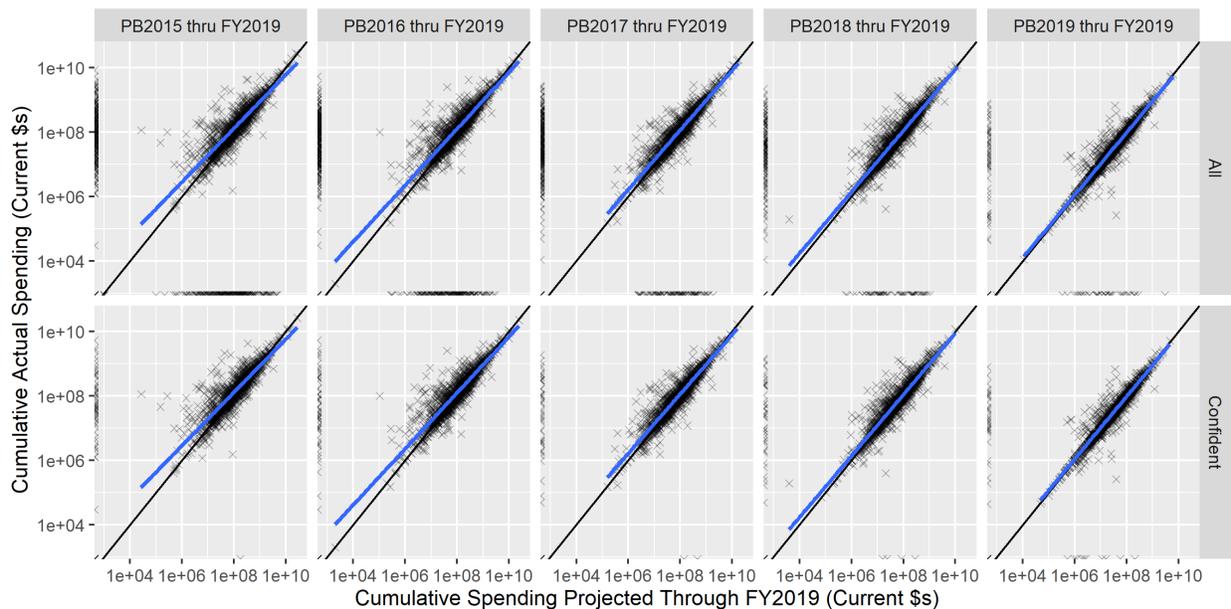


Figure 3. Scatter Plot showing FYDP projections versus Actual Total (Investment)

Each of the plots has a blue regression line of best fit that estimates the relationship between the projections and the actual spend. The closer the line is to the black diagonal line, the better the projections. When limiting the sample to those budget lines where the study team is more confident of the data's quality, as shown in the second row of Figure 3, the quality of the relationship strengthens.

Figure 4 compares the projected and total values shown in Figure 3 but uses a histogram to focus on the differences between the projected and total values. Only the 5-year, 3-year, and 1-year projections are shown in this and subsequent graphs to allow the display of more detail. The X-Axis denotes the relative difference between projected expenditure on any

⁷ Due to the extreme range in spending across different items, all figures will employ logarithmically transformed axes for statistical purposes. Logging both sides results in more normalized sets, rather than most of the points falling into the lower left corner due to the great variation in size of budget lines.



single budget line and the actual spend on that same budget line in FY 2019. Relative difference is used to allow the scale to include cases where the projected or the actual spending figures are zero.⁸ The Y-Axis denotes the count of budget lines with that level of difference.

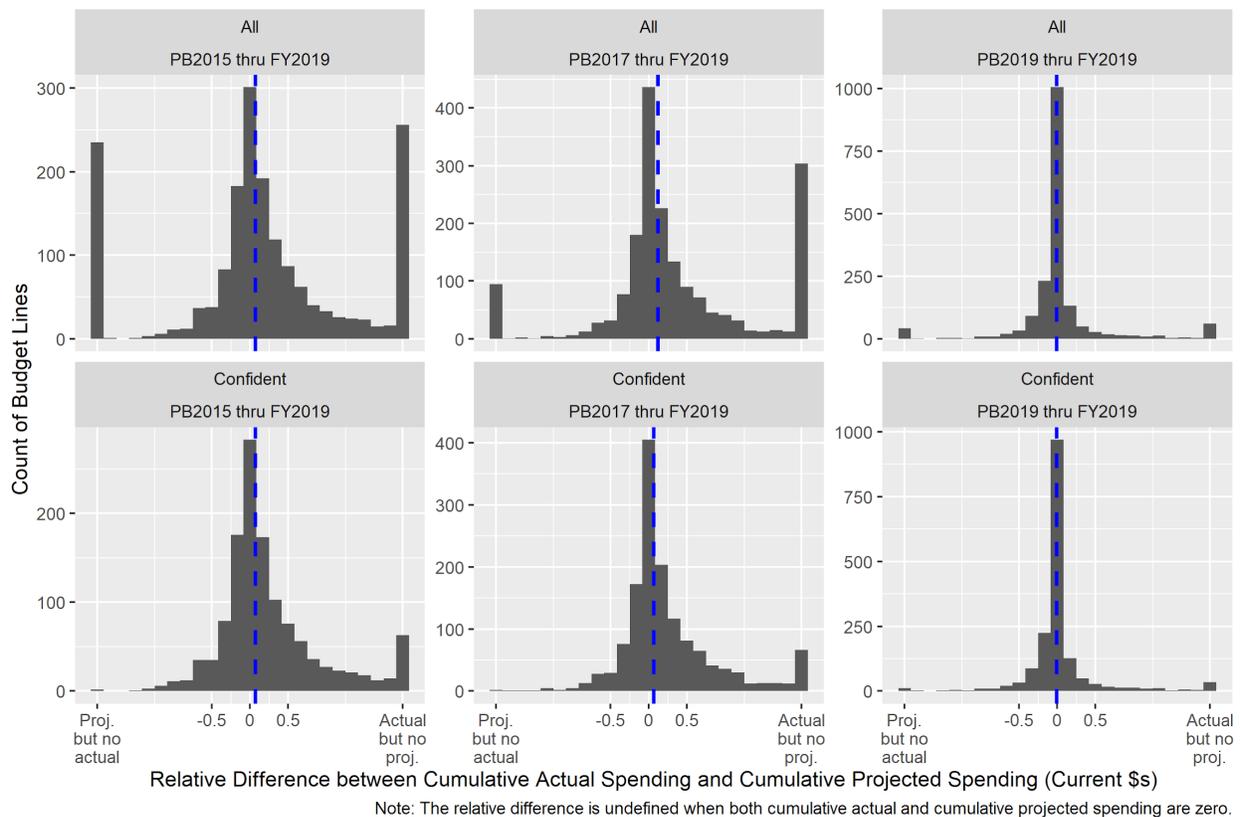


Figure 4. Budget Line Distribution by the Relative Difference Between Cumulative Projection and Actual Spending

For those points in the center of each histogram, the difference between actual and projected values is small. The chart's left side indicates budget line items where the projected value was much higher than what was eventually spent, and the right side of the chart shows budget line items where the estimated value was much lower than what ended up being spent. The groupings all display normal distribution, with the same previously mentioned 0 spend outliers showing on the edges, with the left edge representing cases where spending was projected but none took place and the right edge representing cases where zero dollars were projected, or the budget line was unanticipated by the PB, but spending took place nonetheless. The first row of the graph shows all data, and the second row shows only confident budget lines; note that far fewer budget lines fall at either end of the scale in the confident dataset.

Blue dashed lines indicate the X-Axis median. The closer the estimations are to the actual spending, the more accurate and center-oriented the estimations end up being. There is a distinct drop-off in overall accuracy from the PB year estimation to the out-year estimations, which results in significantly wider tails for the histograms that include more out-years.

⁸ Relative difference, for the purposes of these charts, is calculated by dividing the numerator of (Cumulative Actual Spending - Cumulative Projected Spending) by the denominator of (Cumulative Actual Spending - Cumulative Projected Spending) / 2. Relative difference is not defined in those cases where both the actual spending and the projected spending values are zero.

Which Services and Budget Categories Have the Most and Least Reliable Projections?

This section examines whether the reliability measures discussed previously vary between OCO and base budget lines and when looking across military departments. For context, the count of P-40 and R-2 budget lines for each category is shown in Figure 5. The following analysis is limited to the confident budget lines, shown in dark blue. While most budget lines merit the confident description every year, the semi-confident and not confident categories can account for a reasonable amount of spending. For example, for the Navy, shipbuilding procurement line items, complicated by advanced procurement, have been a regular source of data import problems. While the number of items is small, those budget lines are of high value.

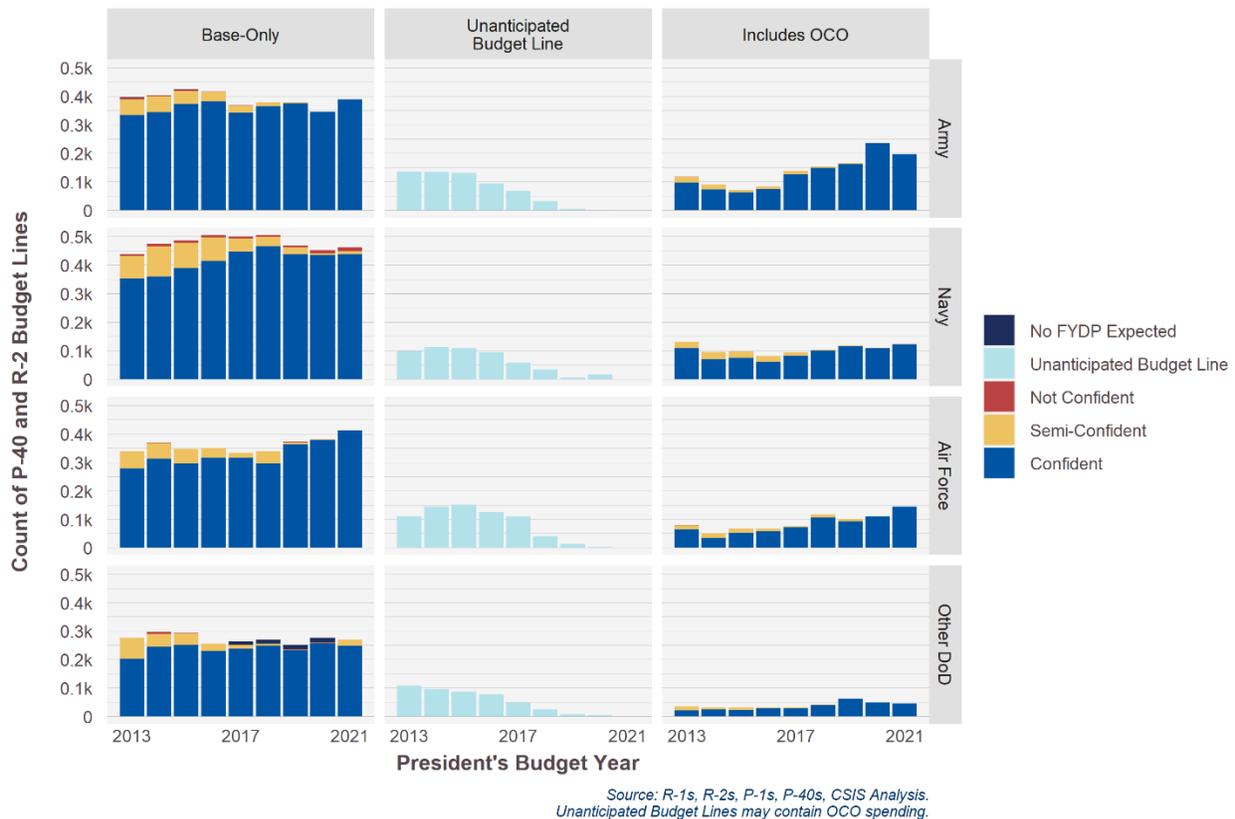


Figure 5. Budget Lines by Military Department and OCO

Influence of Overseas Contingency Operations

When considering the influence of OCO, this paper classifies budget lines based on whether they contained any enacted OCO spending or whether the President's Budget included any OCO spending. This method assumes that budget lines with planned or recent OCO spending may be more likely to have it in the future. A limitation of this method is that sometimes a budget line will add OCO spending even if it had not included it in the past. As is shown in Figure 5, the number of OCO line items experienced a small decline after 2013. However, since PB2016, the number of budget lines including OCO has increased across the military departments.

When examining changes in individual budget lines, as shown in Figure 6, the first year of estimates, PB2019 versus PB2019, shows little difference in distribution between base-only lines and other lines. However, in the 3-year and 5-year projections, shown in the middle and left columns, respectively, the difference is much starker. The first row's base-only budget lines



have peaks centered around 0 proportional difference between projected totals and actual spending. In contrast, that is still a high point for the budget lines including OCO, but the frequency is notably lower, resulting in a more rounded dome shape. Instead, a greater portion of budget lines are spread to the right, indicating growth. This can also be seen in the dashed blue lines, which show the median growth and are shifted to the right, indicating a higher median increase in budget lines. Note that unlike in Figure 5, this graph is simplified by only displaying the confident budget lines due to their superior reliability for analysis.

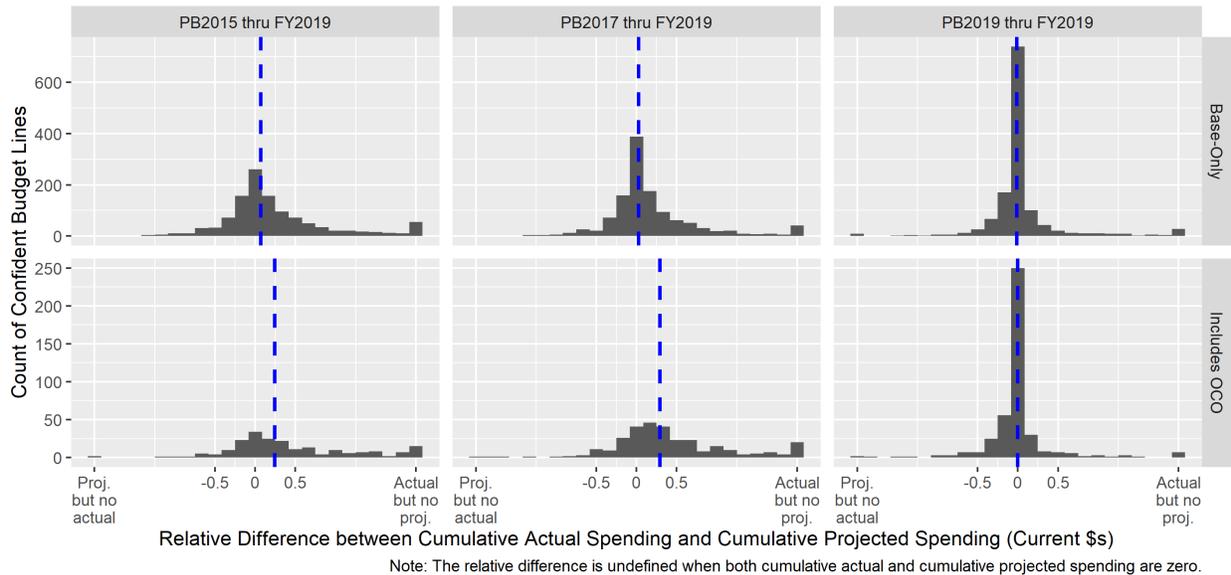
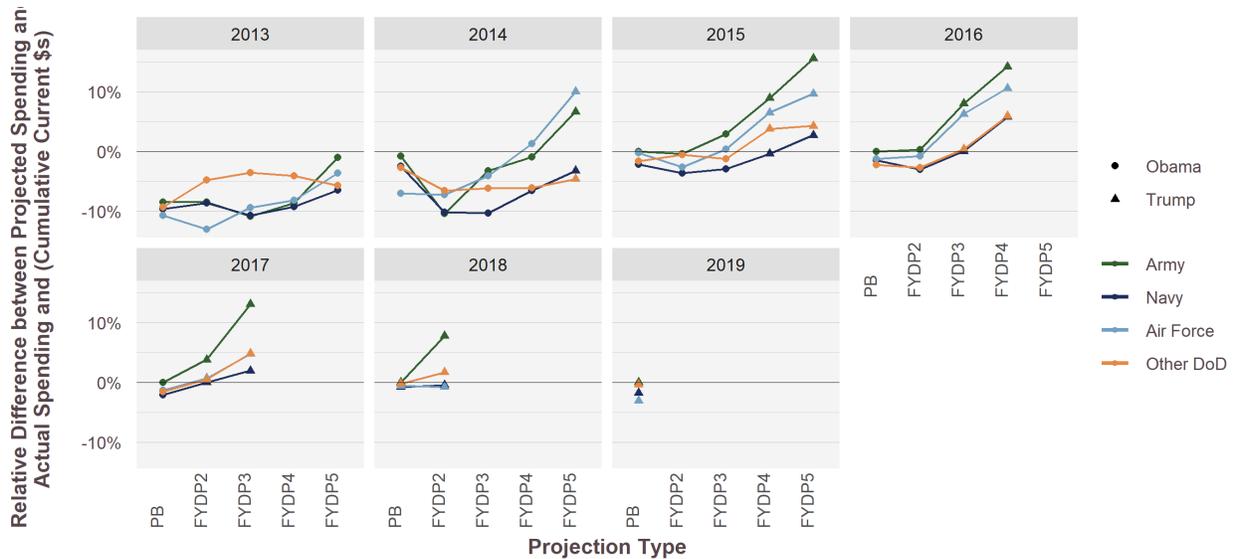


Figure 6. Budget Line Distribution by the Relative Difference Between FY 2019 Actual and FYDP Projection by OCO

Military Services

The differences between Military Services are more subtle than those between base-only and OCO including budget lines, and Figure 7 thus compares the median relative difference between projected and actual spending. Positive values are associated with more spending than projected, while negative values indicate projections exceeded spending. Each of the graphs covers a complete PB year, working from the President's Budget only on the left to the 5-year FYDP on the right. The Military Departments show similar trends within each PB, as assumptions about topline spending levels have widespread influence. PB2013 was devised under the explicit assumption that a budget deal would be reached that would have reduced the strictness of budget caps, a belief that proved optimistic, and as a result, actual spending was consistently below projections. PB2014 still proved optimistic in the early years, though by the fifth out-year, FY2018, Army and Air Force budget lines began to spend more than had been projected. The last three PBs of the Obama administration all proved to underestimate eventual spending. In the latter two budgets, the administration transition for the FY2018 budget leads to particularly stark changes. Finally, the PB2018 projections also proved to underestimate future spending, which may be in part the result of the limited attention top-level leadership in the incoming administration was able to give the FYDP.





Note: The relative difference is undefined when both cumulative actual and cumulative projected spending are zero.

Figure 7. Median Relative Difference by Military Department in Confident Budget Lines

The Army, since PB2015, has had the highest relative difference, and is also the service that most relies on OCO spending. The Air Force has also had relatively higher shifts in median spending, although these Air Force estimates are more reliable in dollar terms, suggesting that Air Force changes are concentrated in lower value budget lines. The pairing of the two services is somewhat surprising, as the Army and Air Force make the most and least use of OCO, respectively.

To better understand the interplay between OCO inclusive budget lines and military department projection reliability, Figure 8 examines both. It shows that across the services, budget lines including OCO, shown in red, tend to underestimate out-year spending to a much greater extent than base budget lines. Indeed, the median relative difference for OCO including lines is repeatedly more than 20%. This unpredictability aligns with expectations, as OCO spending is only reported through the first year of the FYDP. So any OCO spending in subsequent years inherently results in the base budget figures underestimating the eventual funding. There are exceptions to this trend; for example, for the Army and Navy, the PB2013 OCO budget lines consistently overestimated the spending to a greater extent than did base budgets. This spending reduction may be due to the greater variability of OCO-related budget lines, making them a more likely target for cuts once the budget caps arrived in full force.

For both the Army and Air Force, their comparatively high medians shown in Figure 9 can be traced to those years in which OCO inclusive budget lines rocket above 20%. Both the Army and Air Force have been increasing the number of OCO involved budget lines. Interestingly, for the Air Force, this expansion has correlated with a reduction in the degree to which OCO budget lines underestimate actual future spending, suggesting that these new OCO-involved budget lines may be easier to predict. On the other hand, the Army stands out in PB2015–PB2017, as the base-budget lines show larger underestimation than any of the other services.



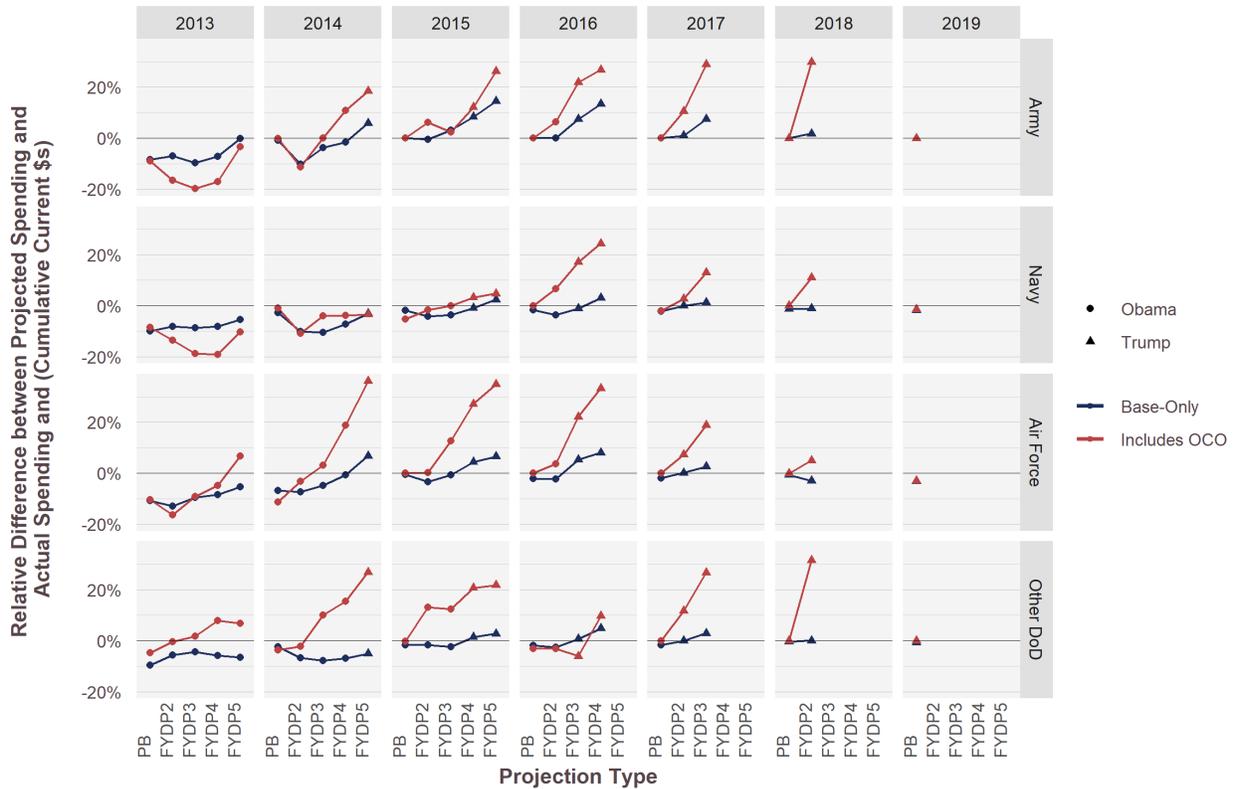


Figure 8. Investment Military Department Topline by OCO Cumulative Percent Difference

Discussion and Conclusions

How Reliable Are Projections Within the FYDP as an Indicator for Actual Spending?

Cumulative investment FYDP projections and actual spending highly correlate, albeit with generally reducing accuracy further out into the out years. Nonetheless, the bottom-up data does replicate previous findings regarding the top-level FYDP: There are notable differences between presidential budgets that with the relationship administration’s approach to the budget caps and the level of support from Congress pushing toward overestimation of future funding in PB2013 and PB2014 switching to underestimation in the later years of the Obama administration.

How Does OCO Spending Relate to Projection Reliability?

In keeping with expectations, budget lines that include OCO had less reliable out-year projections compared to projections by base-budget lines. As Figure 4 shows, OCO budget lines typically underestimate future spending, with greater than projected actual expenditures occurring in the out years for which OCO does not project. That said, in PB2013, Army and Navy OCO budget lines were particularly optimistic and experienced more significant funding shortfalls than other budget lines. In part, this reflects OCO serving its intended purpose of allowing for more rapid changes and thus signifying budget lines that are more difficult to predict.



Which Military Departments Have the Most and Least Reliable Projections?

The Army, across multiple measures, has had the most substantial difference between total projected spending and actual spending, as shown in Figure 7. Separate analysis has confirmed that the correlation between Army projections and actual spending is lower than for the other services. As is shown in Figure 5, the Army is the predominant user of OCO-related budget lines.

What Are the Unclassified FYDP's Biggest Data Quality Problems?

The process of building the dataset highlighted problems that the FYDP presently faces that undercut the applications of data science approaches to develop better forecasts and for analysts to track spending plans to strategy. Most noteworthy is the absence of unique identifiers that make it easy to speak the same language across years and sources. Budget lines regularly appear, disappear, and change labels. This is not inherently a problem; budget lines are sometimes abruptly cut and a new, and perhaps unexpected priority, receives funding. More challenging are the cases where funding may have shifted to a different line or the Justification Books or Comptroller documents have budget lines not present in the other source. The FYDP is a tool intended to aid defense planning, which often means changing priorities, and thus shifts are not just expected, but desirable. However, the data's usability would be enhanced if the newly appearing budget lines reported predecessor lines in a machine-readable manner or indicated if they were a genuinely new initiative.

This project takes a step to increasing the usability of FYDP data in analysis by deriving unique identifiers from labels in the budget documents and managing discrepancies. This dataset's confidence labels are meant to be a stopgap that shows where contradictions are present between sources and track transitions between budget lines where identified.

How Could FYDP Projections be Improved?

The unclassified FYDP has the potential to be used as the basis for better projections. The study team held a workshop with leading practitioners and analysts to present initial results from this dataset. Participants made various valuable suggestions, including separately modeling base and OCO budget lines and looking for known budgeting foibles, such as zeroing out of accounts in the President's Budget that are subsequently restored in congressional enactments. Modeling experiments also found that it was helpful to separately model if a budget line was likely to be funded at all. Unfortunately, the bottom-up approach to modeling is only half of the challenge, as exogenous fiscal factors can matter more than the given budget line's specifics.

Is the FYDP Fulfilling its Purpose?

The period from PB2013 to PB2021, which covers both a significant downturn in spending and a large upswing, offers lessons with relevance beyond the budget caps' expiration. These results demonstrate a critical strategic communication limitation of the FYDP during this period. OCO allows greater flexibility but is not well suited to demonstrate commitment, as the PB2013 experience shows that future funding assumptions may not manifest and budget lines that included OCO spending took more significant cuts than those using the traditional FYDP process.

For those seeking budgetary tools offering both flexibility and fiscal discipline, Light et al. (2017) may offer a better alternative: reserve funds. For example, "funds might be held in reserve to address anticipated further growth within an MDAP portfolio (rather than for any single program)" (p. 44). Their proposal focuses on insulating the overall budget from project uncertainty. Still, if the reserve portfolio's scope matches an area of strategic focus, setting reserves aside could be a powerful priority-setting mechanism. Such dedicated funding could



insulate a key strategic priority from cost increases or topline pressure and leave room for innovation even in the face of fiscal headwinds.

Well-designed reserve funds may offer a way to provide flexibility while mitigating uncertainty but would certainly need support from key congressional committees. A more minor change would be to embrace Travis Sharp's (2019) approach to measuring strategic shifts and to specifying priority areas within the FYDP that align with existing or new publicly available categories of FYDP budget lines. Updated and clearly labeled strategic capability areas or other exhaustive and mutually exclusive categorization systems could overcome many of the data quality limitations documented in this paper and would make it easier to mark a strategy to a budget and serve what Enthoven and Smith (2005) describe as key to the FYDP's value: the way the FYDP forces "the Secretary to make controversial decisions explicitly" (p. 52).

References

- Arena, M. V., Leonard, R. S., Murray, S. E., & Younossi, O. (2006). *Historical cost growth of completed weapon system programs* (Vol. 343). RAND Corporation.
- Blakeley, K., & Korb, L. (2014). *The war chest: War funding and the end of the War in Afghanistan*. Center for American Progress.
- DoD Comptroller. (2020). *DoD budget request*. <https://comptroller.defense.gov/Budget-Materials/>
- Domke, W. K. (1984). Waste, weapons, and resolve: Defense posture and politics in the defense budget. *Policy Sciences*, 16(4), 371–390.
- Enthoven, A. C., & Smith, K. W. (2005). *How much is enough?: Shaping the defense program, 1961-1969*. RAND Corporation.
- Harrison, T. (2016). *Defense modernization plans through the 2020s: Addressing the bow wave*. Center for Strategic and International Studies.
- Harrison, T., & Daniels, S. (2017). *Analysis of the FY 2018 defense budget*. Center for Strategic and International Studies.
- Harrison, T., & Daniels, S. (2020). *Analysis of the FY 2020 defense budget and its implications for FY 2021 and beyond*. Center for Strategic and International Studies.
- Hunter, A. (2019, December 13). *Bad idea: Blaming OCO for our defense budget blunders*. Defense 360. <https://defense360.csis.org/bad-idea-blaming-oco-for-our-defense-budget-blunders/>
- Jordan, L. G. (2015). Systemic fiscal optimism in defense planning. *Defense Acquisition Research Journal*, 22(3), 272–293.
- Lewis, K. N. (1994). The discipline gap and other reasons for humility and realism in defense planning. In *New challenges for defense planning: Rethinking how much is enough* (MR-400-RC, pp. 101–134). RAND Corporation.
- Light, T., Leonard, R. S., Pollak, J., Smith, M. L., & Wallace, A. (2017). *Quantifying cost and schedule uncertainty for major defense acquisition programs (MDAPs)*. RAND Corporation.
- McGarry, B. W., & Epstein, S. B. (2019). *Overseas contingency operations funding: Background and status* (CRS Report No. R44519). Congressional Research Service.
- Sharp, T. (2019). *Did dollars follow strategy? A review of the FY 2020 defense budget*. Center for Strategic and Budgetary Assessments.



Woodward, F. M., & Arthur, D. (2019). *Long-term implications of the 2020 Future Years Defense Program*. Congressional Budget Office.

Young, T. D. (2018). Questioning the “sanctity” of long-term defense planning as practiced in Central and Eastern Europe. *Defence Studies*, 18(3), 357–373.





ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF DEFENSE MANAGEMENT
NAVAL POSTGRADUATE SCHOOL
555 DYER ROAD, INGERSOLL HALL
MONTEREY, CA 93943

WWW.ACQUISITIONRESEARCH.NET