# How Robust are Distributional Findings of Punctuated Equilibrium in Public Budgets?

Derek A. Epp
Ph.D. Candidate in Political Science
UNC-Chapel Hill
derekepp@email.unc.edu

Frank R. Baumgartner
Richard J. Richardson Distinguished Professor of Political Science
UNC-Chapel Hill
Frankb@unc.edu

**Abstract**

Fat-tailed distributions of annual percentage changes in budgets are widely taken as evidence for a punctuated-equilibrium process in government decision-making. We assess the robustness of these findings by exploring some potential artifactual causes of such findings, controlling for these, and evaluating the distributions of data when they are eliminated from the analysis. Using the US federal budget as a test case, we find only small differences in the overall shape of the distribution when various potentially misleading series are excluded. Further, we assess statistical models predicting budget punctuations and find that the predictive power of individual variables are also robust to the inclusion of exclusion of specific data series. Results suggest that the distributional findings relating to fat tails are highly robust.

Keywords: Punctuated equilibrium, budgeting, fat-tailed distribution, public policy, kurtosis

Paper presented at the annual meetings of the Midwest Political Science Association, Chicago IL, April 2–6, 2014.

## Introduction

Punctuated equilibrium in policy studies refers to the idea that political, social, and other forces act to suppress dramatic changes in public policies during periods of equilibrium, but that these negative feedback forces that generate the equilibrium can occasionally be disrupted, creating a surge of self-reinforcing changes that rapidly cause a new equilibrium to be reached. Policies then may appear stable for long periods of time, but suddenly be dramatically transformed when an underlying institutional structure or supporting policy idea is revised. Frank Baumgartner and Bryan Jones developed these ideas in their 1993 treatment of a number of policy issues, demonstrating by looking over many decades of media or congressional consideration of such issues as nuclear power, pesticides, or tobacco regulation that stability is interspersed by punctuation, and that a single theory can explain both the stability as well as the brief periods of change. Scholars have continued to track individual issues over time, and this chronological or historical approach remains an important method for understanding policy dynamics, one issue at a time.

The literature took a dramatic turn in 2005 when Jones and Baumgartner shifted attention to a distributional approach. Looking specifically at the US federal budget over a 50-year period, they pooled observations across 60-some consistently defined spending categories, and calculated inflation-adjusted annual percent changes for each year and each budget category. The result was a histogram that showed the distribution of annual percent changes in budgets (Jones and Baumgartner 2005, Figure 4.14). The shape of this distribution was consistent with the theoretical expectation they had developed in the earlier work: the "normal" period of equilibrium was reflected by a great preponderance of cases in the central peak of the distribution, which they termed a sort of hyper-incrementalism. At the same time, the distribution showed "weak shoulders" and very wide tails. In the time since the publication of

this work, the distributional approach to studying policy change has become widespread. A recent article in which the two original authors were joined by many others looking at several western countries, each showing a similar fat-tailed distributions, was entitled "An Empirical Law of Public Budgeting," reflecting the wide applicability of these findings (Jones et al., 2009).

The movement from historical to distributional treatment of policy changes comes with important costs and benefits. The benefits of the distributional approach include the important fact that such an analysis can literally be comprehensive. Rather than focus on one, a few, or even a great number of individual policy domains, the distributional approach can look across the entire budget, and at whatever level of specificity the categories allow. Further, the approach has allowed international and cross-institutional comparisons, as percentage-change calculations can be made from budgets in virtually any system, and databases other than budgets can also be analyzed from a distributional approach. So there is no wonder why the approach has become widespread in the literature.

At the same time, there are costs to the distributional approach. We can mention three. First, the approach is a-historical—the chronological dimension is lost. Rather than looking at the trace of an individual policy over time, the distribution simply counts up the number of annual changes of various magnitudes. Second, by combining all available policy domains into a single analysis, it removes the details of the underlying policy choices from the analysis. So it removes both history and policy-relevant context from the analysis; these are heavy costs, to be sure. These two issues are the reason why Jones and Baumgartner (2005) were careful to combine the distributional approach with a number of chronological / historical treatments of individual policy issues. What the distributional approach gains in generalizability, it loses in policy specificity and attention to historical context; these are completely absent, in fact.

A third issue in a distributional approach is a concern that different processes could be generating the cases in the central peak of the distribution and those far out in the tails. The normal interpretation has been that the cases in the tails represent punctuations: dramatic shifts in policy direction, or at lease massive changes in budgetary commitments to certain ideas. But what if they are immediately reversed? Rather than a punctuation or a paradigm shift, they might just be a blip. Far from representing a sea-change, they could represent just a temporary dynamic that might immediately be reversed. This paper explores the possibility that the tails and central peaks of budget distributions could be generated by factors inconsistent with the theory of punctuated equilibrium. We begin by exploring those possible issues and then assess the robustness of fat-tailed distributions when we control for relevant factors.

## Background

Our approach is informed by previous research that explores the nature of policy punctuations. In particular, we look to a study by Peter John and Shuan Bevan (2011) that develops a three-tiered typology for punctuations in the U.K. context as an intellectual precursor to the current analysis. They group punctuations according to three causal processes: procedural adjustments, low-salience, and high-salience adjustments. Their argument is that punctuations resulting from procedural reclassifications are a-theoretical and in some cases should be removed from the data. Further, they point out that it is difficult to reconcile punctuations occurring in the absence of any attention to the casual process identified by punctuation equilibrium theory. Their question then is how many of the punctuations they observe can be linked to shifts in attention, rather than the competing mechanisms. They discover that a substantial proportion—about half—of the punctuations they identify occurred either as part of a procedural adjustment or with an almost complete lack of public attention. We engage in a similar process here, identifying mechanisms

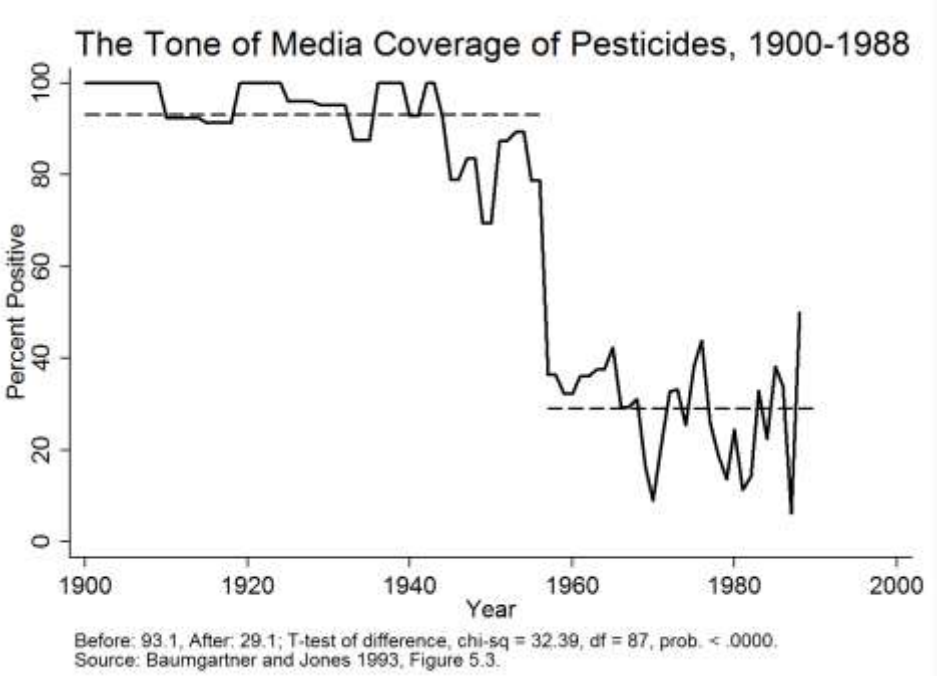
in the data generating process that could potentially produce cases in the tails of a distributional analysis which would not correspond to the data generating process implied in the punctuated equilibrium model.

## Identifying Punctuations

We start by returning to a classic example of a punctuation—coverage of pesticides, as originally documented by Baumgartner and Jones (1993). They describe how from their development around the turn of the 20th century until the late 1950s pesticides were viewed as a marvel of modern technology, a panacea that would usher in a new age of agricultural productivity and public health. Given the positive press surrounding pesticides it seemed logical for the U.S. government to support their liberal application, and indeed large swaths of the continental U.S. and other parts of the world were blanketed in DDT. Then, in the mid-1950s, it began to dawn on people that while very effective at killing insects, pesticides do not discriminate; they kill many other things as well. This idea culminated with the publication of *Silent Spring* in 1962 by Rachel Carson, which documented the disastrous environmental consequences of indiscriminate pesticide use.

Figure 1, which we borrow from Baumgartner and Jones (1993), clearly shows the dramatic reversal of fortunes pesticides saw in the 1950s. The turning point appears to be 1957. Before this year, articles on pesticides had been remarkably supportive—in many years every single article on the topic had a positive tone. After 1957, the majority of articles cast a negative light on pesticides. We have updated the original figure after conducting difference of means tests using 1957 as the dividing point in the data. The dashed horizontal lines show the mean value of support before and after 1957. As the subtext to the figure notes, before 1957 this value was 93% and after it was 29%, representing a major shift in the debate over pesticides.

Figure 1. The Classic Punctuation: Media Coverage of Pesticides



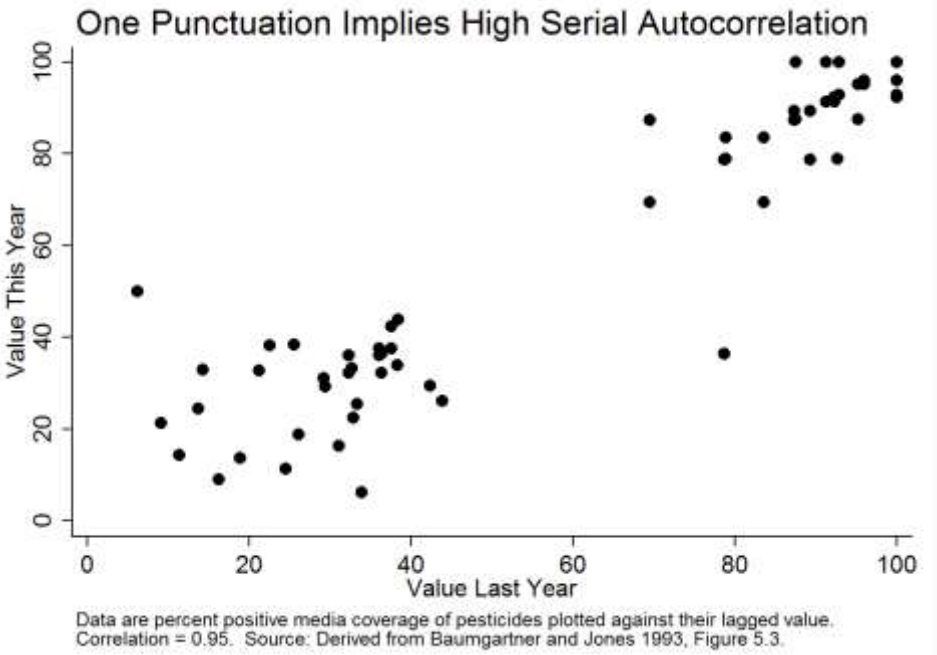

The dynamic on display in Figure 1 exemplifies what is typically thought of as a punctuation in the literature on agenda setting. This is the idea that disequilibria in policy series herald a paradigm shift, where some new approach or solution takes precedent and traditional ways of doing things are rapidly discarded. In the case of pesticides, this shift came when people stopped viewing pesticides as an easy solution to various societal problems and started seeing them as harmful carcinogens. With this type of punctuation we should be able to look at a policy series and draw a clear line denoting the point where perceptions flipped and a new paradigm took hold.

But moving to a stochastic approach allows the possibility that many of the punctuations we observe are 'false' or temporary, in the sense that they are quickly reversed and signify very little about political agendas. How then do we distinguish between 'real' punctuations, the focus of agenda-setting theories, and temporary punctuations? This is the central question of the paper, and we employ various empirical strategies to answer it.

One straightforward approach is to test for serial auto-correlation, with the expectation that volatile policy series where punctuations are quickly reversed will show low correlations between current and previous values. In Figure 1, it is clear that during the early period, values remain consistently high; only in 1957 does the value dramatically shift from the previous value, and remain distinct in its future values from the past ones. Seen in this way, inertia is a key element of the theory. Most issues, most of the time, maintain a certain stickiness, maintained in equilibrium by negative feedback. Serial auto-correlation is a simple way to look at this; series with high inertia should have high correlations between any given value and the previous one. Series with no punctuations, or with high variability around a central value that itself does not change, do not correspond with the theory, and they would show low values of serial auto-correlation. (Note however, that trending series with a no punctuation would have high values on this measure, so it must be combined with other techniques in assessing fit to the theory.)

Figure 2 illustrates the expected relationship by plotting the percent positive media coverage about pesticides against their lagged values (the data are the same as from Figure 1). Here auto-correlation is very high and we see that the data is divided into two groups; high values (stemming from the period before the punctuation in 1957) and consistently low values (coming after the 1957 collapse in the public image of the industry). So, with the one major exception, last year's values are a strong predictor of current year values. As series become more volatile, and punctuations more temporary, we can expect this relationship to break down.

Figure 2. Serial Auto-correlation in Coverage of Pesticides, 1900 to 1988.

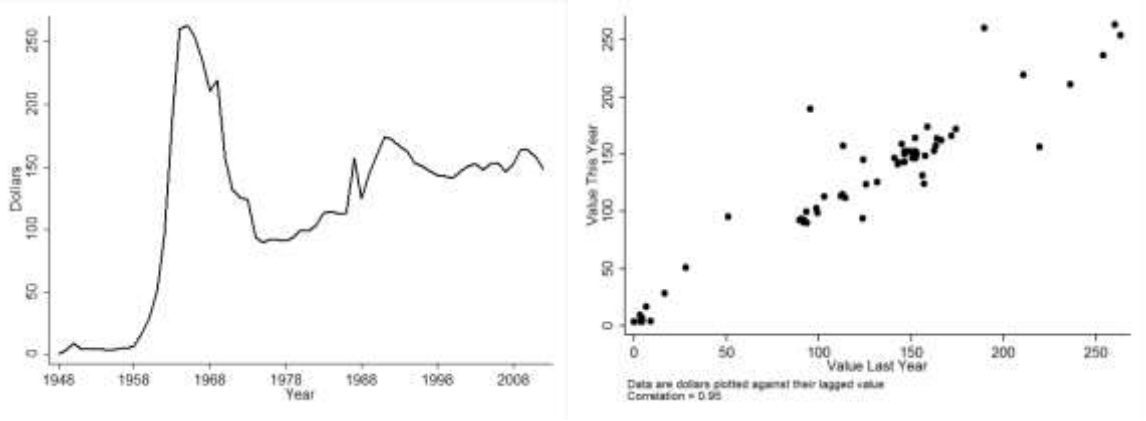



We turn now to spending by the U.S. federal government from 1947 to 2012; a common focus in the agenda setting literature. Our interest is in systematically documenting the occurrence of temporary versus sustained punctuations, but we begin with simple descriptive examples. Remember that the pesticides example represents the classic but also ideal case of a sustained punctuation, so one question is how closely any budgetary series will come to replicating that pattern of change.

Figure 3 shows budget authority toward "space flight, research, and supporting activities", one of the 66 non-financial budget categories (called subfunctions in OMB parlance) that make up the U.S. budget. The left-panel, tracking budget authority in millions of dollars, reveals some dramatic changes in spending. Most notable is the enormous increase that took place in the early-1960s corresponding to the Apollo moon missions. But these high levels of spending were not sustained and after successfully landing a person on the moon the government substantially scaled back spending to this category. Still, spending never returns to its pre-Apollo levels and space flight is certainly a higher priority in the modern era than it was before 1958. So

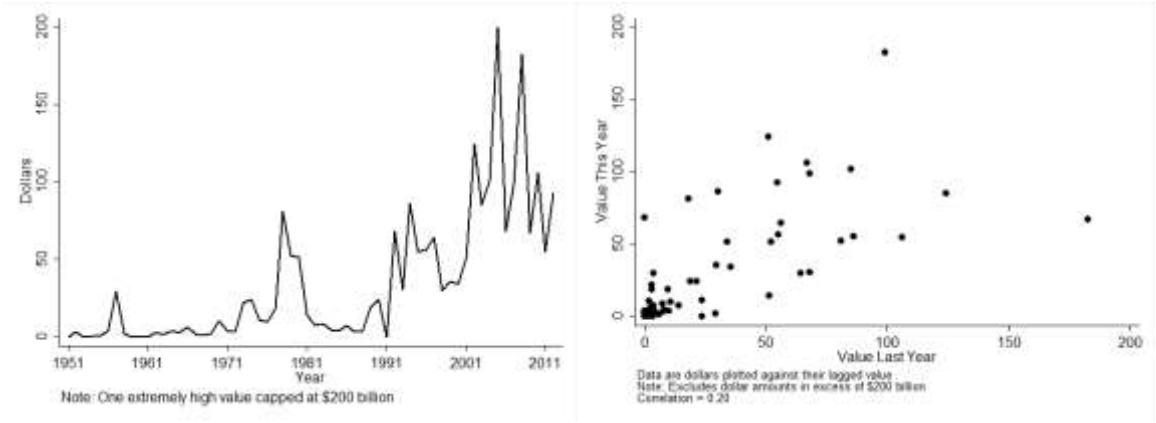
while the major punctuation in this series is not sustained to the same degree as what we saw when looking at pesticides, this is a clear example of 'real', substantively interesting punctuation. The right panel of Figure 3 looks at the serial auto-correlation of spending on space flight, revealing a pattern that is familiar from the pesticides example. Current spending levels are correlated with the previous year's spending at 0.95 and again we see gaps in the coverage corresponding to the occurrence of a punctuation.

Figure 3. U.S. Spending on Space Flight, Research, and Supporting Activities, 1948 to 2012
a) Budget Authority b) Serial Auto-correlation




For an example of a temporary or 'false' punctuation consider Figure 4, which tracks Budget Authority toward "disaster relief and insurance". Again the left-panel of the figure shows annual Budget Authority, where, unlike with spending on space flight, it is difficult to see trends indicative of a larger political agenda beyond the basic need for responsible governments to respond to crises as they occur. The right-panel of Figure 4 supports the assertion that measuring serial auto-correlation can be a powerful tool to distinguish between series that are prone to temporary versus sustained punctuations. Here we see that current values are only correlated at 0.20 with the previous year's spending; exactly what we would expect from a series that is heavily driven by exogenous shocks.

Figure 4. U.S. Budget Authority toward Disaster Relief and Insurance, 1951 to 2012
a) Budget Authority b) Serial Auto-correlation

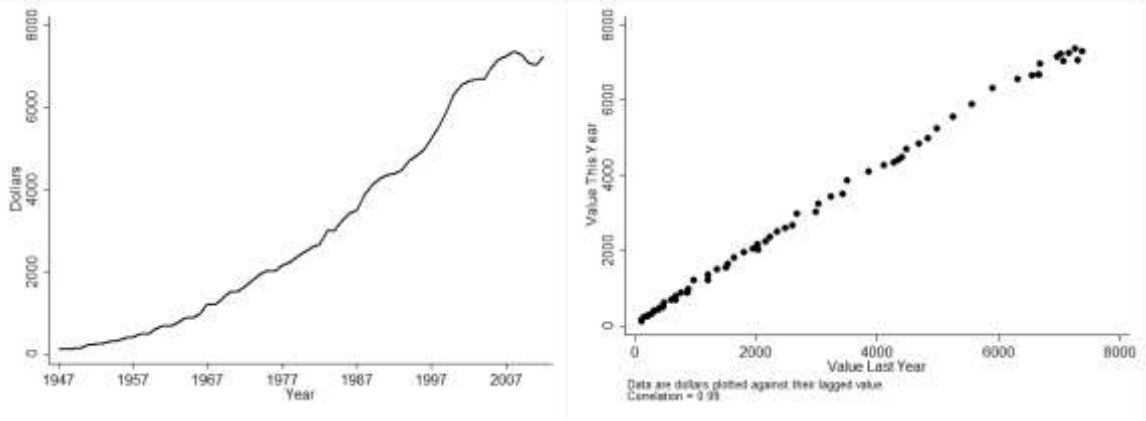



Figures 3 and 4 both look at discretionary spending topics, but the U.S. budget is increasingly devoted to spending on mandatory programs, where spending levels are determined by well-established formulas that are politically difficult to adjust. A substantial part of the budget is therefore largely insulated from the type of agenda setting thought to cause policy punctuations. Further, spending for many of the mandatory categories is often strongly driven by demographic trends, such as retirements, and should logically have dynamics distinct from those domains that are subject to endogenous or exogenous shocks. We cannot state that any particular budget category is driven by a purely demographic logic; even in the case of retirements and pensions, important shifts sometimes occur in the formulae used to determine entitlements. But some budget categories are clearly much more prone to instabilities than others. (Another caveat here is that automatic spending formula may be tied to slow moving demographic trends, but also to highly volatile series such as commodity prices or climatic events such as natural disasters. Our analysis below shows that a simple mandatory v. discretionary distinction is not enough to capture the dynamic that is needed to distinguish theoretically among those series that would be more or less likely to harbor true punctuations.)

Figure 5 considers Social Security, revealing the particular dynamics that appear to govern mandatory spending relating to demographic shifts. While spending on space flight was subject to the whims of political enthusiasm, and in the case of disaster relief the extreme variability of the physical climate, spending on Social Security climbs relentlessly upward regardless of party control of government or historical circumstances. Note that the correlation in the right-panel is at 0.99, emphasizing that there are almost never large shifts in spending.

Figure 5. U.S. Budget Authority toward Social Security, 1947 to 2012

a) Budget Authority b) Serial auto-correlation




The type of change exemplified by spending on space flight, where shifting political ideals determined spending levels, is the best match to the causal process commonly identified in the literature on punctuations. But as this brief review demonstrates, it is far from the only dynamic at work. A fuller understanding of the causes of policy change must take seriously the possibility that policymakers have tied their hands by placing a majority of the budget under automatic spending formulas. While this does not eliminate political agendas as a causal factor—formulas are sometimes updated—it does suggest that much of the budget will not be particularly susceptible to agenda setting dynamics as laid out in the theoretical literature. A larger concern would be that much of the instability usually attributed to the rise and fall of issue frames is actually rooted in a much simpler and politically mundane phenomenon—the need for governments to respond to various military and natural crises. When a crisis occurs spending is

dramatically ramped up in response, but as soon as the emergency dissipates, spending is brought back down to pre-crisis levels. In these circumstances, punctuations in the positive direction would beget major decreases in spending within a few years; we would observe instabilities coming and going. This tidal process could be a powerful source of the instability observed in government budgets, but would have little connection to traditional conceptions of agenda setting.

Given the various concerns and competing causal processes our goals are twofold. First, we make a systematic effort to document the relative frequencies with which temporary and sustained punctuations occur in budgetary time series. Then, informed by that effort, we test the robustness of previous findings after excluding categories prone to temporary punctuations and mandatory spending.

## Measuring Sustained Punctuations in the U.S. Budget

The focus of our efforts is U.S. budget authority from 1947 to 2012 and Figure 5 shows aggregate changes in spending levels across the 66 OMB subfunctions for that period. This is a simple update of the Jones-Baumgartner figure 4.14, which started the discussion about punctuations in budgets (2005, 111). Clearly this distribution is not Normal, but instead features a high central peak and extremely wide tails, while the "shoulders", or mid-range changes, of the distribution are missing. In other words, this is a classic leptokurtic distribution. Empirically we can determine the extent to which a distribution is leptokurtic by looking at that distribution's l-kurtosis value, which for a Normal distribution will be 0.123, with increasing values indicating leptokurtosis and smaller values playkurtosis. Note that the budget distribution in question has an l-kurtosis value of 0.621, indicating that it deviates substantially from the Normal.

Our interest is in the punctuations in Figure 6 and there are various ways to distinguish between those cases far in the tails and those not considered to be punctuations. Analysis of the causes of punctuations proves to be highly robust with respect to where we draw the line between a punctuated change and one that is closer to the bulk of the observations. For simplicity, we draw that line at the top and bottom ten percent of the observed changes, and the figure illustrates this with dashed vertical lines. With 3,831 observations in the overall distribution, 783 then are identified as punctuations, half on the negative side and half on the positive side. Note, as is standard in the literature, we have truncated the presentation of the data by clustering all extremely high positive changes at +150 percent.[1]

[1] More complicated definitions of what constitutes a punctuation, such as those beyond the point where the observed distribution passes the hypothetical Normal distribution with similar variance, or controlling for changes in overall variability across time, generate results highly similar to those we present here, so for simplicity, but with knowledge that our results are robust, we choose a very simple definition of punctuation here. Various authors have drawn these lines differently: Jones, Baumgartner and True (1998) drew them at +20 and -15; Breunig and Koski (2006) have used quintile regression to analyze separately the tails from the center of the distribution.

Figure 6. Identifying Punctuations in the Distribution of Annual Changes in Federal Budget Authority, 1947 to 2012

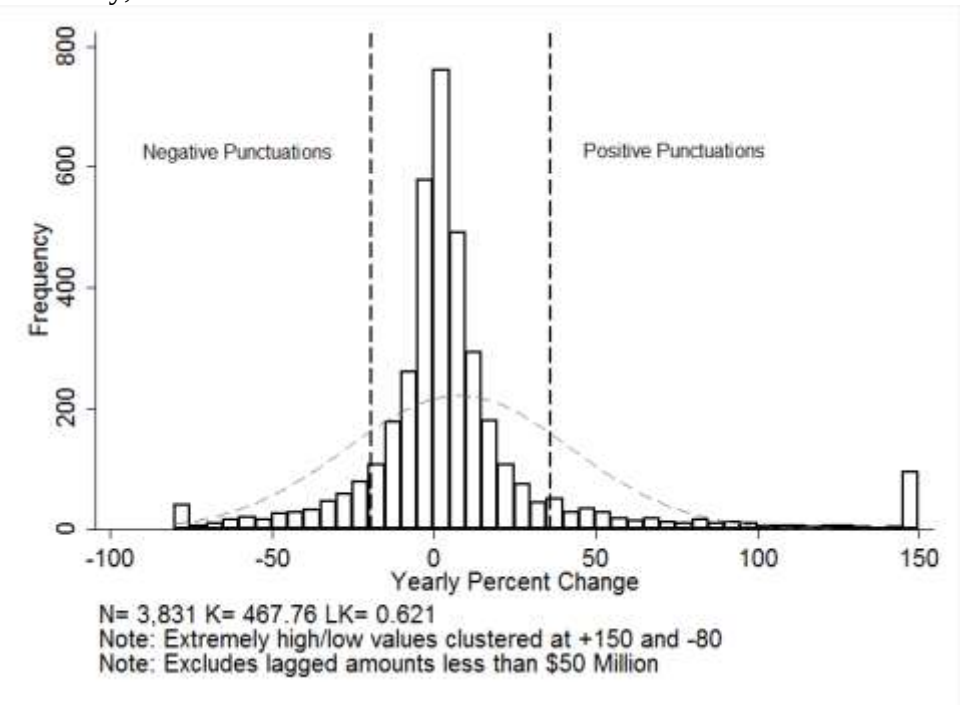



Figure 6 replicates evidence, now commonplace in the literature, that government budgets are prone to instabilities. Many studies have sought to explain the causes of this instability, but for the most part these investigations have rested on broad theoretical arguments about the disproportionately of government information processing. Few attempts have been made to 'drill down' into the data in order to determine how many of the punctuations we observe can be attributed to shifting political agendas versus high volatility in input series that force a government response. Both factors are clearly at work, as the examples looking at Budget Authority toward space flight and disaster relief demonstrated. So we simply want to know how many of the punctuations that we define in Figure 6 are temporary and sustained.

Of course, 'temporary' and 'sustained' are subjective terms, so there are many ways we can go about answering the question. Table 1 provides 20 possible answers. It documents the number of punctuations that are reversed by a certain percent over a certain number of years. Reading the first row of the table from left to right reveals that 376 punctuations, or 49% percent

of the total, were reversed by at least 10% after only 1 year, 394 were reversed by at least 10% after 2 years, and so on. What do we mean by saying that a punctuation was reversed by at least 10%? Consider a punctuation that increased spending to a budget category by 75% over its base value of $100 (so for the year the punctuation took place spending is now at $175). If in the year following that punctuation spending then decreased by at least $7.5 (10% of the $75 increase), we can say that punctuation was reversed by 10% in 1 year[2].

Table 1. How many Punctuations see Reversals between 10 and 90% within 4 years?

| Reversal | 1 Year | 2 Year | 3 Years | 4 Years |
|---|---|---|---|---|
| 10% + | 376 (49%) | 394 (52%) | 400 (54%) | 383 (52%) |
| 25% + | 309 (40%) | 332 (44%) | 342 (46%) | 342 (47%) |
| 50% + | 217 (28%) | 255 (34%) | 278 (38%) | 279 (38%) |
| 75% + | 160 (21%) | 199 (26%) | 224 (30%) | 230 (32%) |
| 90% + | 128 (17%) | 164 (22%) | 189 (26%) | 201 (28%) |

Note: Year 1 Punctuations = 766; Year 2 = 754; Year 3 = 740; Year 4 = 726

Table 1 shows the number of punctuations that saw reversals according to different terms, leaving the remaining punctuations to qualify as sustained[3]. That is, if 49% of punctuations were reversed, then the remaining 51% can be thought of as sustained. In this way, the table presents definition of varying strictness for what constitutes a sustained punctuation. The upper-right cells in the table show definitions that are very strict; here any punctuation that is reversed by 10% over the course of 3 or 4 years is considered temporary. Using this definition would place the majority of the punctuations we document in Figure 5 in the temporary category. The lower-left cell shows the least restrictive definition, where only punctuations that are reversed by more than 90% within 1 year qualify as temporary. Depending on which cell in the table we occupy makes

[2] We calculate reversals separately for positive and negative punctuations. In the case of a negative punctuation, a 75% decrease to a base value of $100 leaves $25. We consider that punctuation reversed by 10% within 1 year, if in the next year spending was increased by at least $7.5.

[3] Note that the total number of punctuations diminishes slightly as we calculate changes multiple years in the future because we lose observations from years 2009 through 2012.

a big difference as to our conclusions regarding the relative frequencies of these punctuations. But regardless of definition, we must conclude that temporary punctuations make up a substantial proportion of the instabilities documented in Figure 5; somewhere between 17% and 54%.

Figure 7 provides a hypothetical example of a punctuation that has decayed by 10% within 4 years (the definition from the upper-right cell in the table). The idea here is to give a visual sense for this type of change. The figure shows two huge increases in spending, which within 4 years have been reversed by exactly 10%, with the vertical lines showing the 4-year window. If we accept the upper-right cell as our definition of a sustained punctuation, then the changes Figure 7 displays would not qualify. This clarifies just how restrictive this definition is, and it should come as no surprise that under this definition fewer than half the total punctuations are considered sustained.

Figure 7. Hypothetical Budget Series with Punctuations that are Reversed by 10% in 4 years

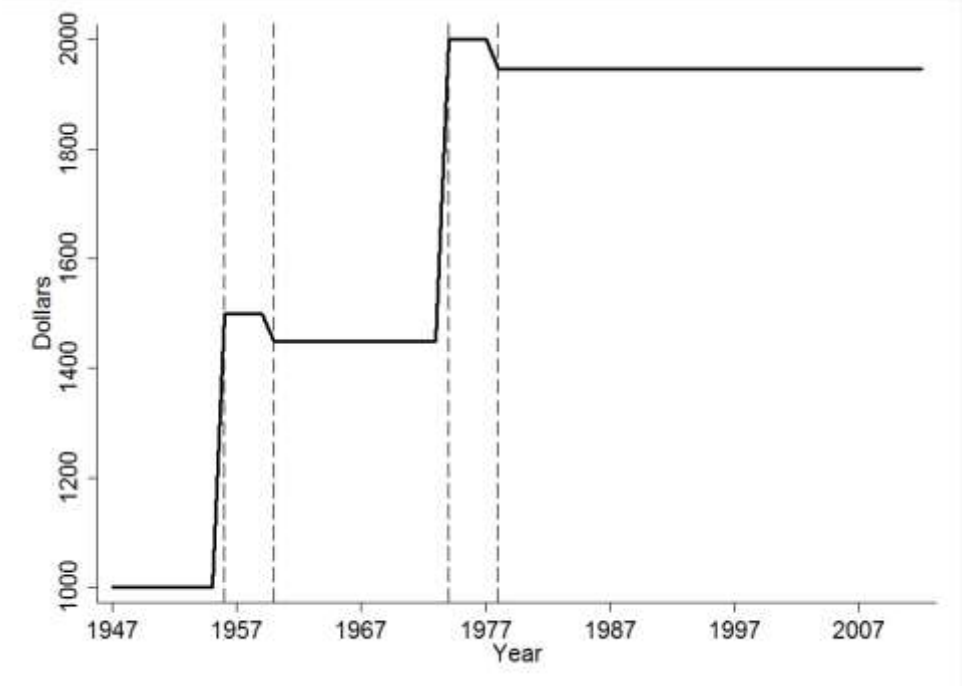
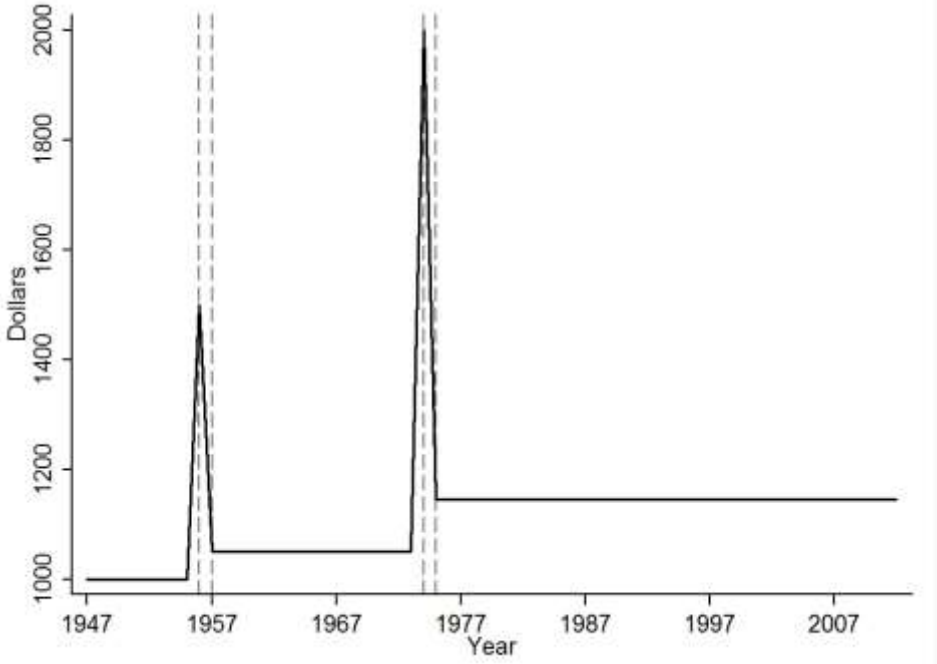


Figure 8 provides a similar hypothetical this time for the cell in the lower-left of the table—punctuations that are reversed by 90% within 1 year. Clearly this is an altogether more dramatic reversal, where within 1 year the original punctuation is almost completely eliminated.

Figure 8. Hypothetical Budget Series with Punctuations that are Reversed by 90% in 1 year




We have tried to show that the number of punctuations that can be considered sustained depends heavily on the parameters involved, but in order to proceed to subsequent analysis we must pick a definition. We define sustained punctuations as those that do not see reversals upward of 50% within 4 years (the middle cell in the right-most column of Table 1). The logic behind this choice is that the process by which new issue frames supplement old ones is thought to play out over many years or decades, so we can reasonably expect punctuations that are brought about by shifting political agendas to last at least 4 years. We pick the 50% reversal rate simply as a conservative, middle-of-the-road option. Based on this definition, we classify 279 punctuations as temporary and 447 as sustained. Table 2 shows how these punctuations are

distributed by OMB subfunction, and the right-most column shows the serial auto-correlation of each budget series.

Table 2. Total Punctuations and Punctuations Sustained by at least 50% over 4 Years, by OMB Subfunction

| OMB Subfunction | Total | Sustained | Temporary | Auto-Corr. |
|---|---|---|---|---|
| Disaster Relief and Insurance | 37 | 20 | 17 | 0.21 |
| Military—Other | 28 | 14 | 14 | 0.23 |
| Farm Income Stabilization | 31 | 18 | 13 | 0.60 |
| Area and Regional Development | 23 | 12 | 11 | 0.52 |
| Community Development | 21 | 10 | 11 | 0.99 |
| General Property and Records Management | 28 | 17 | 11 | 0.48 |
| Other Advancement of Commerce | 28 | 17 | 11 | 0.43 |
| Higher Education | 22 | 12 | 10 | 0.81 |
| Unemployment Compensation | 20 | 10 | 10 | 0.85 |
| Defense-related Activities | 23 | 14 | 9 | 0.45 |
| International Development and Humanitarian Assistance | 23 | 14 | 9 | 0.74 |
| International Security Assistance | 27 | 18 | 9 | 0.79 |
| Housing Assistance | 21 | 13 | 8 | 0.65 |
| Training and Employment | 20 | 12 | 8 | 0.69 |
| Veterans Education, Training, and Rehabilitation | 27 | 19 | 8 | 0.92 |
| Other Income Security | 7 | 0 | 7 | 0.99 |
| Research and General Education Aids | 15 | 8 | 7 | 0.93 |
| Executive Direction and Management | 12 | 6 | 6 | 0.93 |
| Military Construction | 11 | 5 | 6 | 0.75 |
| Other General Government | 23 | 17 | 6 | 0.65 |
| Conservation and Land Management | 12 | 7 | 5 | 0.91 |
| Criminal Justice Assistance | 14 | 9 | 5 | 0.77 |
| Water Resources | 11 | 6 | 5 | 0.55 |
| Central Personnel Management | 10 | 6 | 4 | 0.62 |
| Elementary, Secondary, and Vocational Education | 10 | 6 | 4 | 0.74 |
| Energy Conservation | 6 | 2 | 4 | 0.00 |
| General Purpose Fiscal Assistance | 14 | 10 | 4 | 0.83 |
| General Retirement and Disability | 10 | 6 | 4 | 0.70 |
| Ground Transportation | 12 | 8 | 4 | 0.81 |
| Pollution Control and Abatement | 10 | 6 | 4 | 0.41 |
| Recreational Resources | 10 | 6 | 4 | 0.93 |
| Conduct of Foreign Affairs | 9 | 6 | 3 | 0.98 |
| Health Care Services | 8 | 5 | 3 | 0.99 |
| Legislative Functions | 5 | 2 | 3 | 0.99 |
| Other Labor Services | 7 | 4 | 3 | 0.71 |

| | | | | |
|---|---|---|---|---|
| Space Flight, Research, and Supporting Activities | 13 | 10 | 3 | 0.95 |
| Water Transportation | 6 | 3 | 3 | 0.93 |
| Atomic Energy Defense Activities | 9 | 7 | 2 | 0.68 |
| Emergency Energy Preparedness | 11 | 9 | 2 | 0.73 |
| Energy Information, Policy, and Regulation | 9 | 7 | 2 | 0.84 |
| Federal Correctional Activities | 4 | 2 | 2 | 0.99 |
| General Science and Basic Research | 8 | 6 | 2 | 0.97 |
| Income Security for Veterans | 3 | 1 | 2 | 0.77 |
| Military (1947-1956) | 3 | 1 | 2 | 0.68 |
| Social Services | 7 | 5 | 2 | 0.96 |
| Air Transportation | 9 | 8 | 1 | 0.97 |
| Central Fiscal Operations | 2 | 1 | 1 | 0.97 |
| Consumer and Occupational Health and Safety | 3 | 2 | 1 | 0.99 |
| Federal Employee Retirement and Disability | 5 | 4 | 1 | 0.98 |
| Federal Law Enforcement Activities | 6 | 5 | 1 | 0.99 |
| Food and Nutrition Assistance | 8 | 7 | 1 | 0.99 |
| Military Procurement | 1 | 0 | 1 | 0.87 |
| Other Natural Resources | 4 | 3 | 1 | 0.99 |
| Other Veterans Benefits and Services | 5 | 4 | 1 | 0.95 |
| Agricultural Research and Services | 1 | 1 | 0 | 0.97 |
| Federal Litigative and Judicial Activities | 1 | 1 | 0 | 0.99 |
| Foreign Information and Exchange Activities | 3 | 3 | 0 | 0.88 |
| Health Research and Training | 0 | 0 | 0 | 0.97 |
| Medical Care for Veterans | 2 | 2 | 0 | 0.99 |
| Medicare | 1 | 1 | 0 | 0.99 |
| Military Family Housing | 2 | 2 | 0 | 0.80 |
| Military Operations and Maintenance | 1 | 1 | 0 | 0.97 |
| Military Personnel | 0 | 0 | 0 | 0.92 |
| Military Research, Development, Test, and Evaluation | 2 | 2 | 0 | 0.98 |
| Other Transportation | 3 | 3 | 0 | 0.79 |
| Social Security | 1 | 1 | 0 | 0.99 |
| Total | 726 | 447 | 279 | 0.99 |

Note: Temporary punctuations are correlated with auto-correlation at –0.60. The appendix includes a scatter-plot showing this relationship.

Table 2 is sorted by temporary punctuations, making clear that topics driven by exogenous shocks tend to have the most (disaster relief, farm support), while those topics

associated with mandatory programs have fewer (Social Security, Medicare). The column displaying serial auto-correlation shows the continuation of the pattern established with spending on space flight and disaster relief—frequent temporary punctuations are associated with lower auto-correlation.

## Are Previous Findings Robust?

If much of the instability we observe in budgetary time series can be attributed to policymakers ratcheting up spending to address an unforeseen crisis and then quickly bring it back down to equilibrium or pre-crises levels, this points to a different causal process than is commonly identified in the literature on punctuations. Further, the increasing proportion of the budget that goes to mandatory spending topics suggests that there will be a strong tendency toward incremental adjustments. This raises the possibility that the well-known kurtosis in government Budget Authority is less a function of agenda setting and more attributable to stochastic inputs series and mandatory spending formulas. That is, we observe high leptokurtosis because we combine budget categories that are prone to shocks and incrementalism. Critically, however, the type of change engendered by these categories is not well-explained by traditional ideas about agenda setting and the rise and fall of competing frames. A plausible concern is that by removing these categories from the analysis, isolating the areas of the budget where we do expect agenda setting dynamics to be at work, we can produce a distribution that is much less punctuated, with lower kurtosis.

To investigate this possibility we reproduce the budget distribution from Figure 6 after excluding various budget categories from the analysis. Table 3 displays the kurtosis statistics associated with each modified distribution. The first and second row drop the top 3 and top 9

categories for temporary punctuations as identified in Table 2.[4] Here we see only very marginal differences. The l-kurtosis for the full distribution is 0.62, in the first row it is 0.58, and after dropping the top 9 categories it falls to 0.56. This suggests a decreasing trend to be sure, but in each case the distribution is remains distinctly leptokurtic even after removing the series that could be thought to generate potentially artifactual punctuations. The third row of the table drops mandatory categories, which causes an increase in the kurtosis statistics (contrary to what one might expect if the mandatory categories were all driven by slow-moving demographic trends), and finally the fourth and fifth rows look at the combined effects. In all, there is very little movement in the l-kurtosis statistics across any of the categories. This suggests that the concerns we raised are unfounded; a key finding in literature on punctuations—the high kurtosis of budget distributions—is highly robust. Even when we eliminate categories prone to incrementalism and stochastic shocks in order to focus more directly on areas of the budget where agenda setting dynamics are most applicable, we find the same pattern of change.

Table 3. Kurtosis of U.S. Budget Distribution with Stochastic and Mandatory Series Excluded, 1947 to 2012

| Excluding: | N | Kurtosis | L-kurtosis |
|---|---|---|---|
| Top 3 Categories for Temporary Punctuations | 3,712 | 431.54 | 0.589 |
| Top 9 Categories for Temporary Punctuations | 3,341 | 394.00 | 0.560 |
| Mandatory Spending Categories | 3,118 | 416.93 | 0.632 |
| Mandatory and Top 3 | 3,017 | 400.03 | 0.598 |
| Mandatory and Top 9 | 2,776 | 377.86 | 0.562 |
| Full Distribution | 3,831 | 467.76 | 0.621 |

Note: Excludes lagged values less than $50 million

The finding of high kurtosis in budget series appears ubiquitous, but is it becoming less so over time? Jones, Baumgartner, and True (1998) have demonstrated a general secular decline

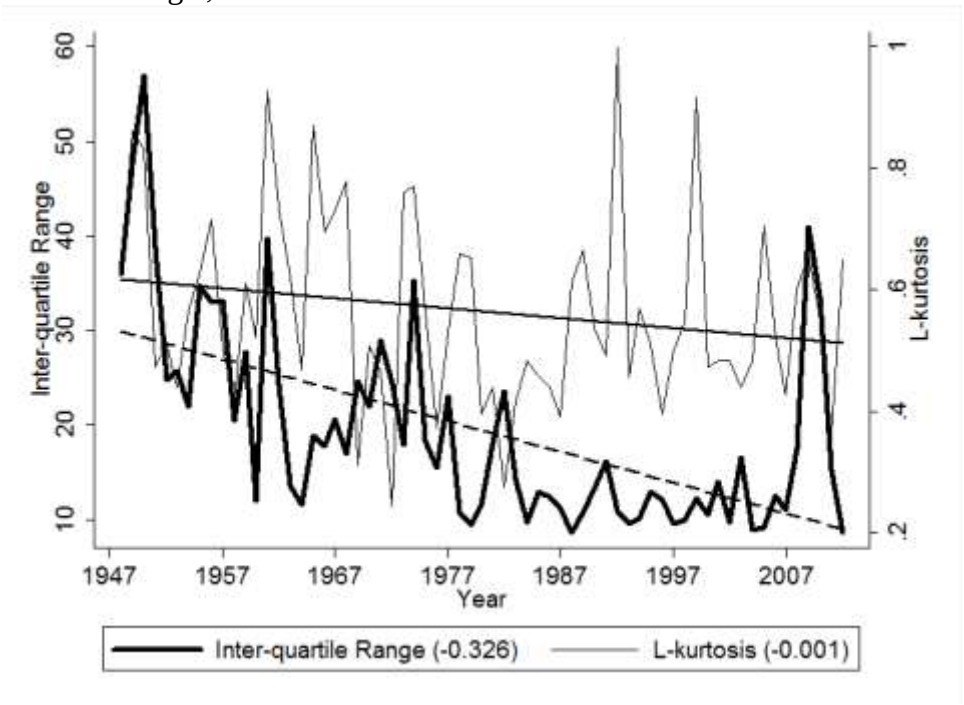
[4] Note: we do not eliminate the punctuations, but the entire series associated with the excess temporary punctuations. That is because we are interested in the full distribution of changes, and eliminating the cases in the tails but leaving all other cases would mathematically generate reduced kurtosis, obviously.

in the volatility of budgets over the decades from 1948 to the recent period. This trend can be attributed to the increasingly large proportion of Budget Authority that are determined by mandatory formulas. As more of the budget becomes insulated from the agenda dynamics thought to cause punctuations, might we observe less kurtosis over time? Does kurtosis track volatility?

Figure 9 plots the inter-quartile range of the percent change values across all 66 budget categories for each year of data on the left-axis, while the right shows annual levels of l-kurtosis[5]. This replicates the general decline in volatility noted by Jones et al. in 1998; note however the surge in volatility corresponding to the 2009 stimulus bill. The l-kurtosis statistic, while also volatile, shows no clear decreasing trend. The estimated best fit lines for both measures support the visual interpretation. As the coefficients (included in the legend to the figure) indicate, l-kurtosis declines only marginal with time (and, given the variability of this measure, the -.001 slope is not significantly different from zero), while volatility decreases at a relatively steep rate. The reassures us that the high kurtosis observed in budget data is not a relic of a previous era, but persists even as a greater proportion of the budget is determined by spending formulas. Kurtosis statistics require many observations to be robust, so authors have shied away from estimating them, for example, on 60 annual series. When we do so as in this figure, we do so with some caution and with a goal of estimating whether the trend is sharply downwards, as is volatility. The answer is that volatility has been declining progressively over time, but kurtosis has remained steady across time.

[5] The figure uses inter-quartile range rather than a direct measure of variance as it is robust against the extreme outliers pervasive in budget data.

Figure 9. Tracking the Annual Inter-quartile Range and L-kurtosis of Percent Changes Values in the U.S. Budget, 1947 to 2009




As a final robustness test we estimate a series of logistic regression models to predict the occurrence of a punctuation. In each model the dependent variable is coded as 1 if a punctuation occurred and 0 otherwise. The first model is for the full dataset, the second after dropping the top 3 categories for temporary punctuations, and the third after dropping the top 9 categories. By keeping the set of independent variables constant across each model, we can assess how the dynamics governing the occurrence of punctuations change as we eliminate those series prone to stochastic shocks. If the findings from the full dataset are robust, we would expect few substantive differences between models. Before proceeding to results we briefly describe the coding and expectations for each independent variable.

A key factor we consider is the complexity associated with each budget category and the expectation is that instabilities will increase with complexity. To operationalize complexity we turn to a dataset available from the Bureau of Economic Analysis (BEA), which links spending

allocations to the government agencies in charge of implementing them. For example, the National Science Foundation is frequently authorized to spend money allocated to the budget category for "general science and basic research". The BEA data is available from 1976 through 2008 and during this period some categories, such as Social Security, have fallen exclusively under the purview of a single agency, while others are carried out by more than 20 different agencies. Our measure of complexity simply counts the number of distinct agencies linked to each subfunction. The idea is that instabilities will be especially acute where multiple agencies are involved, as there will be more room for disagreement over the varying "solutions" that different agencies have to offer. Complexity in organizational design is a possible source of punctuation.

Beyond complexity, one obvious possibility for the presence of dramatic policy shifts is change at the top. New leaders, especially those with a different ideology from their predecessors, might want to make their stamp by dramatically adjusting spending patterns. Further, if presidents benefit from a honeymoon period they may be uniquely effective at ushering in large budgetary changes shortly after taking office. Combined, these factors are a good reason to expect that a president's first budget may feature more punctuations than those coming later in his term. The model accounts for this with a dichotomous variable, coded 1 if the reallocations took place in the first budget of a new president.

We can also look to various governing conditions as a possible cause of budget instability. Periods of unified government may present majority parties with opportunities to pursue major policy initiatives; operating as a "release-valve" on pent-up issues that went unattended through political intractability. Congressional polarization might certainly affect the possibility for major policy shifts. When polarization is low, there is more room for cooperation

between parities, but during periods of high polarization, even basic responsibilities such as funding the government can be sidetracked. Our model includes a measure of House polarization adopted from Keith Poole and Howard Rosenthal's DW-Nominate scores. From 1947 through 2012 the measure varies between 0.40 and 1.10, with lower values indicating less polarization.

Also included are variables for mandatory and defense spending, leaving domestic discretionary as the excluded category. Finally, the model controls for the amount of money allocated to each subfunction in each year. A concern is that punctuations are more likely for small budget categories, as it is comparatively easy to make a large change to a small base value, as compared with budget categories that typically see billions of dollars in spending. Table 3 shows the results of the full model.

Table 3. Logistic Regression Predicting the Occurrence of Punctuations

| Variable | Odds Ratio | Standard Error |
|---|---|---|
| Dollars (Millions) | 0.99* | 0.00 |
| Mandatory Spending | 1.27* | 0.15 |
| Defense Spending | 1.52* | 0.20 |
| United Government | 1.04 | 0.09 |
| House Polarization | 0.20* | 0.05 |
| First Budget Year | 0.93* | 0.11 |
| Subfunction Complexity | 1.04 | 0.00 |

N = 3,289, Pseudo $R^2$ = 0.04, * = significant at 0.05 p-value

The odds-ratios for 5 out of the 7 variables are statistically significant. Punctuations are slightly less likely as budget categories increase in size. Spending for mandatory and defense categories are more likely to see punctuations than spending on domestic discretionary items. This seems counter-intuitive, especially in the case of mandatory spending, but by using a multi-variate regression we have controlled for many of the factors that may lead to punctuations in discretionary categories. That is, we include variables that speak directly to stickiness or friction

of government agendas, whereas the factors that lead to punctuations in mandatory topics may often be exogenous to these governmental controls[6].

As expected polarization is a strong predictor of policy instability. Moving from very low to very high polarization decreases the likelihood of a major policy shift by about 90%. We also see a strong, and highly significant, effect for subfunction complexity. Each additional agency increases the chance for a punctuation by 4% and this variable ranges from 1 to 27, so moving from the least to the most complex budget category has a major effect. The odds-ratio for unified government is above 1 as expected, but not significant, and the odds-ratio for the first budget of a new president is below 1 and also not significant. These effects may simply be drowned out when controlling for other factors. We can imagine, for instance, that the ability of presidents to inflect a budget with their own priorities will be highly contingent on polarization, regardless of any benefits incurred through a honeymoon period, and it may be that we have simply overestimated the ability of presidents to push through dramatic spending adjustments.

How robust are the results in Table 3? Have we identified factors that speak to the dynamics of agenda setting, or are the results contingent on the inclusion of budget categories that are governed by automatic formulas or simply respond to stochastic shocks? Table 4 shows the results of two additional logistic regressions, which run the same model after excluding the top categories for temporary punctuations. In each case, the results are almost identical to those presented in Table 3. The most notable change is that after excluding the top categories for

[6] Consistent with this explanation, when we predict punctuations without the mitigating controls we find that discretionary categories are more likely to see punctuations than mandatory spending.

temporary punctuations, mandatory spending is no longer significantly different than domestic discretionary[7].

Table 4. Logit Regression Predicting the Occurrence of Punctuations Excluding Mandatory Spending and Categories Prone to Temporary Punctuations

| Variable | Odds Ratio | Standard Error |
|---|---|---|
| Excluding Top 3 Categories | | |
| Dollars (Millions) | 0.99* | 0.00 |
| Mandatory Spending | 1.08 | 0.14 |
| Defense Spending | 1.67* | 0.22 |
| United Government | 1.05 | 0.10 |
| House Polarization | 0.15* | 0.04 |
| First Budget Year | 0.90 | 0.11 |
| Subfunction Complexity | 1.05* | 0.00 |
| Excluding Top 9 Categories | | |
| Dollars (Millions) | 0.99 | 0.00 |
| Mandatory Spending | 0.84 | 0.14 |
| Defense Spending | 1.97* | 0.27 |
| United Government | 1.00 | 0.10 |
| House Polarization | 0.10* | 0.03 |
| First Budget Year | 0.85 | 0.12 |
| Subfunction Complexity | 1.04* | 0.00 |
| Model 1: | Model 2: | |
| N = 3,176 | N = 2,819 | |
| Pseudo $R^2$ = 0.05 | Pseudo $R^2$ = 0.05 | |
| * = significant at 0.05 p-value | * = significant at 0.05 p-value | |

The robustness analysis points unequivocally in one direction: a key finding in the agenda setting literature—instabilities in policy series—is highly robust. Even after eliminating budget categories that are least amendable to agenda setting theory, we still see high levels of leptokurtosis in outlay distributions and the same factors predicting the occurrence of punctuations. Still, drawing a distinction between these three types of budgeting is important as they point to very different types of policymaking. Studies in agenda setting must be aware that the idea of dramatic shifts in policy direction as a driver of budget instabilities is only part of a larger, more complex policymaking dynamic.

[7] This is probably attributable to the budget category for "farm income stabilization", which falls under mandatory spending and sees frequent punctuation reversals.

## Conclusion

Punctuated equilibrium is an increasingly popular approach to understanding policy change, not just in a budgetary context, but across a range of organizational outputs. Given the explanatory power of this idea and its wide assimilation through the literature, we want to be sure that its central empirical findings are robust. Here we have identified what could be a major concern: that the dichotomy between incremental and punctuated changes that the theory explains as the result of shifting political commitments, is in fact artifactual. The counter-hypothesis is that we observe incrementalism because much of the budget is tied to slow-moving demographic indicators, and we observe punctuations because certain budget categories are linked to highly stochastic input series. When we combine these factors with a distributional approach we would naturally observe leptokurtosis, but it would actually have little to do with punctuated equilibrium theory. After conducting various robustness tests we can confidently report that this concern is unfounded. Removing potential sources of measurement bias from the data does little to alter the shape of budget distributions and the same predictive elements remain statistically significant. Further, the finding of high kurtosis is robust with respect to time; it remains high even as mandatory spending makes up a larger proportion of the budget.

The findings from the paper contribute beyond the support they lend to punctuated equilibrium theory. Most important is the discovery that many of the punctuations observed in government budgets are short-lived and see reversals within only a few years. Given the large proportion of such cases, and in line with the question that John and Bevan (2011) addressed in their paper, one could wonder whether the distributional approach to the study of punctuations simply has too much error built into it to be worthwhile. Our analysis suggests that the findings remain robust even when we recognize and control for relevant causal processes that could generate significant numbers of "false" or temporary punctuations.

## References


Baumgartner, Frank R., and Bryan D. Jones. 1993. *Agendas and Instability in American Politics*. Chicago: University of Chicago Press.

Breunig, Christian and Chris Koski. 2006. Punctuated Equilibria and Budgets in the American States. *Policy Studies Journal* 34: 363-79.

John, Peter and Shaun Bevan. 2011. What Are Policy Punctuations? Large Changes in the Legislative Agenda of the UK Government, 1911-2008. *Policy Studies Journal* 40: 89-108.

Jones, Bryan D. and Frank R. Baumgartner. 2005. *The Politics of Attention: How the Government Prioritizes Problems*. Chicago: University of Chicago Press.

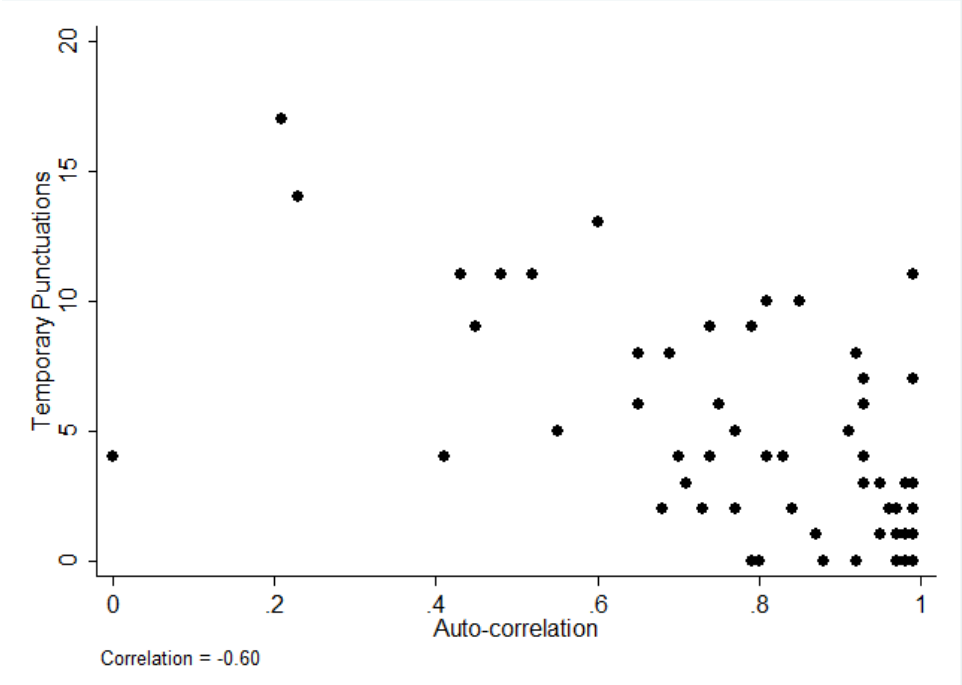
Jones, Bryan D., Frank R. Baumgartner, Christian Breunig, Chris Wlezien, Stuart Soroka, Martial Foucault, Abel François, Christoffer Green-Pedersen, Chris Koske, Peter John, Peter B. Mortensen, Frédéric Varone, and Stefaan Walgrave. 2009. A General Empirical Law of Public Budgets: A Comparative Analysis. *American Journal of Political Science* 53: 855-73.

Jones, Bryan D., Frank R. Baumgartner, and James L. True. 1998. Policy Punctuations: US Budget Authority, 1947–95. *Journal of Politics* 60 (1): 1–33.

## Appendix—Robustness Tests with Serial Auto-correlation

A limitation of the robustness tests we conduct is that they are based on defining a temporary punctuation as one that is reversed by at least 50% within 4 years. As Table 1 made clear, however, there are many different ways we can set this definition, which has a big effect on the number of temporary punctuations we identify. Another approach to robustness testing is to exclude budget categories based on levels of serial auto-correlation. This has the advantage that these measurements are constant and not based on any pre-determined set of parameters. Figure 1A plots the number of temporary punctuations identified in Table 2 against levels of auto-correlation, for each budget category. With a 0.60 correlation, budget categories with lower levels of auto-correlation tend to have more temporary punctuations; a relationship supported by the figures on space flight and disaster relief. Serial auto-correlation appears to be a plausible substitute for direct measures of temporary punctuations.

Figure 1A. Temporary Punctuations and Serial Auto-correlation for 66 Budget Categories




We repeat our logistic regression model, this time excluding budget categories that have levels of auto-correlation below 0.50 and 0.80. Table 1A has the results of both regressions.

Once again, we see that the causes of budgetary punctuations appear robust; there are few substantive differences between the models in Table 1A and the model based on the full dataset from Table 3.

Table 1A. Logit Regression Predicting the Occurrence of Punctuations Excluding Categories with low Serial Auto-correlation

| Variable | Odds Ratio | Standard Error |
|---|---|---|
| Excluding with less than 0.5 | | |
| Dollars (Millions) | 0.99* | 0.00 |
| Mandatory Spending | 1.51* | 0.19 |
| Defense Spending | 1.72* | 0.26 |
| Unified Government | 1.03 | 0.10 |
| House Polarization | 0.14* | 0.04 |
| First Budget Year | 0.92 | 0.12 |
| Subfunction Complexity | 1.04* | 0.00 |
| Excluding with less than 0.8 | | |
| Dollars (Millions) | 0.99* | 0.00 |
| Mandatory Spending | 2.33* | 0.38 |
| Defense Spending | 0.65 | 0.21 |
| Unified Government | 1.14 | 0.15 |
| House Polarization | 0.14* | 0.06 |
| First Budget Year | 1.02 | 0.17 |
| Subfunction Complexity | 1.06* | 0.01 |
| Model 1: | Model 2: | |
| N = 2,995 | N = 2,053 | |
| Pseudo $R^2$ = 0.04 | Pseudo $R^2$ = 0.07 | |
| * = significant at 0.05 p-value | * = significant at 0.05 p-value | |