PSJ Policy Studies Journal

The Policy Studies Journal, Vol. 45, No. 2, 2017

Complexity, Capacity, and Budget Punctuations

Derek A. Epp and Frank R. Baumgartner

The budgeting literature has long focused on "institutional friction" as a cause of ubiquitous punctuated equilibrium (PE) findings. A recent wave of scholarship looks to identify specific institutional mechanisms that affect the number of punctuations in policy outputs. We contribute to this growing body of research by focusing on the complexity of the institutional environment surrounding a policy area as well as that of the government as a whole. These factors have opposite effects: the more complex a policy area, the greater the likelihood of extreme spending changes. But, higher institutional capacity in general leads to greater stability. To test these ideas, we develop a novel index of budgetary change that balances the conceptual importance of extreme changes while analyzing the entire distribution of budget changes, not only the tails. In addition, we also demonstrate that findings are robust to a number of important distinctions, such as between series associated with slowly moving demographic trends or quickly moving stochastic events. We, therefore, demonstrate the robustness of important findings from the established literature, add a new measure of the dependent variable, and push the literature forward with a new focus on issue complexity and institutional capacity.

KEY WORDS: budgeting, punctuated equilibrium, policy change, capacity, complexity

This article builds directly on recent theoretical insights presented in the *Politics of Information* by Frank R. Baumgartner and Bryan D. Jones (2015), who note that government greatly expanded its capacity during the post-1947 period. More government agencies were created, often with missions to attack different aspects of the same underlying problems. Not only did government "spread" into new areas, it also dramatically "thickened" by having increased numbers of agencies or institutions dealing with the same issues; this process affected the federal government as well as the states and localities. A major driver of these efforts to grow institutional capacity was an increased focus on solving complex problems, such as poverty and racial discrimination. Complicated problems can be contrasted with what Baumgartner and Jones call "engineering" problems, where solutions are less partisan and more readily apparent. Providing clean water, for example, is an important function of government, but not a conceptually complex one. Prior to the 1940s, the government dealt mostly with engineering problems. As policymakers moved to address more complicated issues, there was a growing need for a thicker government (although by the late 1970s this had led to efforts to cut "waste and redundancy" in

Published by Wiley Periodicals, Inc., 350 Main Street, Malden, MA 02148, USA, and 9600 Garsington Road, Oxford, OX4 2DQ.

government). We focus in this article on documenting the importance of these two elements—institutional capacity and issue complexity—on budget punctuations. We find that they work in countervailing directions. Policies addressing complex issues are much more likely to undergo punctuations than those for simple issues. Periods of higher institutional capacity meanwhile are associated with greater stability than periods of lower capacity.

Recent scholarship by Breunig, Koski, and Mortensen (2010); Ryu (2011); John and Bevan (2012); Robinson, Flink, and King (2014); and Epp (2015) advances punctuated equilibrium (PE) theory by identifying different institutional and governmental forms that condition the levels of friction in policymaking, and therefore affect the magnitude of punctuations in policy outputs. We add to this literature by identifying institutional capacity and issue complexity as powerful predictors of instability in policymaking. Our argument proceeds as follows. Complex issues exacerbate the information problem facing policymakers because there are more streams of policy-relevant information to consider (this is what makes them complex issues in the first place) and less agreement over which streams should receive priority. Attention to complex issues is, therefore, more likely to be distributed unevenly, leading to the stick-slip dynamic of policy change characterized by PE theory. But, the informational capacity of government is not constant; it can be enhanced by developing the tools for policy analysis, that is, by "thickening" government. When capacity is higher, policymakers are better equipped to engage with multiple information streams, and subsequently, attention can be distributed more proportionally. Note that one of our variables differs across issue areas and the other differs across time. Different policy domains feature dramatically different institutional settings, and complexity here is related to punctuation. However, the government as a whole differs over time in the multiplication of institutional venues, or a restriction in this institutional capacity. Controlling for these longitudinal trends is an important element of any modeling exercise, and here we find that the overall capacity of the institutional environment reduces the likelihood of punctuations.

We explain the reasons for these contradictory effects below and introduce empirical measures of institutional capacity and issue complexity. While our first goal is to push the literature forward by directly testing the idea of institutional thickening and issue complexity, we also address two additional elements relating to previous empirical tests of the PE theory. The first is to test if previous findings could be artifacts of combining budget categories that have different statistical characteristics where the pooled distribution reflects more these compositional elements than it does a common data-generating process. We do this by identifying budget categories that are highly correlated with slowly moving demographic trends (e.g., pensions or elderly health care) and those that involve responses to highly unpredictable stochastic shocks (e.g., trade flows, agricultural commodity prices, or international affairs). Budget series such as those associated with Social Security can naturally be expected to vary only slowly with response to slowly moving compositional shifts in the U.S. population. Conversely, farm price supports shift almost automatically to shifting world commodity prices. We provide a direct test of this idea here and demonstrate that the findings are robust. Results are presented in Supporting Information.

The second element in our robustness testing is to use different definitions of what constitutes a punctuation, as the literature has also been somewhat inconsistent in its operational definitions. In fact, rather than focusing only on cases in the tails, we develop an index of budgetary change that uses all the data but scores more highly those cases that are far in the tails. This allows us to avoid any dichotomous coding of cases that are said to be "in the central peak" or "in the tails" but to allow each observation to enter proportionately into our analysis, with cases near the central peak scoring low values and those near either the right or the left tail being weighted proportionately more heavily. Our Supporting Information explores alternative definitions common in the literature, showing highly robust findings but ultimately proposing that our newly developed index of budget change is a more useful measure, as it reflects Breunig and Jones's (2011) admonition to use the entire distribution rather than to focus only on a part of it.

The article advances the study of policy change in three ways. First, we demonstrate the duality of complexity and institutional capacity. Both elements affect the policy process: complexity leads to instability and higher capacity to stability. We note that government thickened so that policymakers had more tools to engage with complicated problems and it does appear that institutional capacity provides at least a partial solution to the challenges that complex issues bring. Furthermore, our measure of complexity provides an observational assessment of the frictions operating in different policy domains. That allows a more direct test of the institutional friction hypothesis, and a more direct one than that typically used in the literature up to now, such as that friction progresses from low at the input stages to higher at the output stages of the policy process (see Baumgartner et al., 2009; Jones, Sulkin, & Larsen, 2003). Second, by identifying and controlling for possible artifactual causes of PE findings, we conduct a critical robustness test; one that finds renewed support for existing scholarship. Third, the newly developed index provides a useful approach to measuring budgetary change, allowing for the analysis of the entire distribution of changes rather than any single part of the distribution.

Background

Jones et al. (2003) were among the first to use a distributional methodology to test theories of policy change. Looking at the outputs from various U.S. political institutions such as congressional hearings, executive orders, media coverage, financial markets, and public budgets, they pooled observations for each institutional output and calculated annual percent changes. The result was a series of histograms, each displaying a great preponderance of cases in the central peak of the distribution, and, at the same time, featuring "weak shoulders" and very wide tails; in short, the distributions showed high kurtosis. The authors argued that the levels of kurtosis they observed across change distributions were a function of the frictions faced by each institution, and they presented empirical evidence to this effect. Later, Jones and Baumgartner (2005) refined these arguments, explaining how the shape of change distributions was consistent with the theoretical expectations they had developed in their earlier collaborative work. More recently, Baumgartner, with various

colleagues (2009), showed similar progression from lower to higher levels of kurtosis in three countries while moving from inputs to lawmaking to budgets, establishing the generalizability of the earlier Jones et al. (2003) finding. Jones et al. (2009) then posited this as a "general law," citing the ubiquity of PE findings in public budgets.

A number of studies support the idea that there are basic differences across policy domains that make punctuations more or less likely. Jordan (2003) looks at local government expenditures in one of the first articles showing differences in levels of stability across budget functions. Another early study by John and Margetts (2003) finds that levels of kurtosis vary across U.K. budget functions. Mortensen (2005) reveals similar variation across policy domains in local Danish budgets as does Jensen (2009) looking at a large number of West European states. More recently, Breunig et al. (2010) compare change distributions across U.S. and Danish budget categories. They demonstrate that changes to certain categories such as interest on the public debt are almost normal, while others, such as Medicare, show significant leptokurtosis. Of particular interest, where international comparisons are possible, they show that the same policy domains tend to show higher or lower kurtosis scores across both countries, suggesting that common attributes of certain policies make them more or less likely to undergo punctuations. Breunig and Koski (2006) provide support for this idea; pooling budget data across the 50 U.S. states they show great variability in kurtosis scores associated with different policy domains. But, while there has been universal consensus that kurtosis varies across policy areas, there has been little agreement as to a common underlying causal factor, beyond the notion of differing levels of friction. Our explanation is a simple one: some policy areas are more complicated than others. The complex ones place greater informational burdens on policymakers, which lead to more punctuations in these areas.

We also note growing concerns about operational definitions in the PE literature. John and Bevan (2012) raise a particularly alarming possibility. They group punctuations in the U.K. national budget according to three causal processes: procedural, low-salience, and high-salience adjustments. Their argument is that punctuations resulting from procedural reclassifications are artifacts of the data classification system used and, therefore, are not appropriate tests of the theory. (For example, changes associated with endangered species protection at the same time as changes in the area of air pollution abatement could be construed as a very large change in the area of the environment, but this is a simple aggregation effect caused because we consider endangered species and air pollution to be part of the same policy domain, although they could be unrelated changes.) Furthermore, they point out that it is difficult to reconcile punctuations occurring in the absence of significant media attention to the causal process identified by PE theory. Their question then is how many of the punctuations they observe can be linked to shifts in attention, rather than these two competing mechanisms. They discover that about half of the punctuations they identify occurred either as part of a procedural adjustment or with an almost complete lack of public attention. In their view, only the high-salience punctuations correspond with what the theory calls for, although their revised definition leads to no significant changes in the theory or the empirical support for it; there are just fewer punctuations by their definition. Similarly, Mortensen

(2009) highlights that legislative attention is only one piece of the budgetary change puzzle; he finds that large spending shifts are also heavily dependent on shifting public opinion.

In summary, scholarship has shown: (i) that greater frictions lead to more punctuations, (ii) variability in the degree to which different policy domains follow a PE pattern, and (iii) concerns over operational definitions of what constitutes a punctuation. We build on each of these elements in developing and testing a model of budgetary change. Our measures of issue complexity and institutional capacity provide empirical approaches to measuring friction, ones that can be replicated internationally. Furthermore, we take John and Bevan's warnings and curiosity about the impact of operational definitions seriously, and push forward to test the robustness of our empirical model when using different definitions of the dependent variable (see Supporting Information for this analysis) and including or excluding cases based on the presence of what might be thought of as artifactual, rather than substantively important, punctuations.

Hypotheses

Attention is one of the scarce resources governments have at their disposal to allocate. Imagine that policymakers have 10 "units of attention" that they can allocate to an issue. If that issue is very simple, it may require only 5 units of attention to comprehend, but if the issue is complex, then it may require 100 units. This is what makes issues simple or complex in the first place; the complex ones are multidimensional and, therefore, place greater informational burdens on policymakers. For a simple issue, policymakers can allocate attention to every relevant dimension of the problem, but for complex issues many dimensions will be ignored (or perhaps even not recognized). Consequently, when faced with complex issues, policymakers will be more likely to latch onto one or a few particular dimensions of a much larger problem. So, for example, rather than trying to meticulously understand and solve poverty, policymakers may choose to address only its most visible elements; homeless children, perhaps. Occasionally, social urgency may build up around another dimension of the same problem—homeless children are still important but the real problem is drug addiction—and public policies will lurch to accommodate the newly in-vogue solution. When issues are simple, lurching from solution to solution is much less common, as the solutions are readily apparent and widely agreed upon. Snow accumulation is a problem faced by many local governments. Democrats and Republicans agree that the solution is to deploy snow plows; an important and basic function of local governments in winter climates. Someday a new technology might improve upon the snow plow necessitating a dramatic update of the laws that govern snow removal, but such technologically driven revolutions are relatively uncommon. In all, our logic follows directly from PE theory, which draws a link between the disproportionate allocation of attention and policy instabilities. We postulate:

Complexity hypothesis: More complex issues will be prone to greater instabilities than simpler issues.

We also note, however, that the number of attention units policymakers have at their disposal is not a constant. A thicker government with more resources for engaging with policy information may allow policymakers to cast their attention more widely than would otherwise be possible. If, for example, policymakers had access to 100 units of attention, then even a very complex issue could be processed comprehensively. Indeed, this was the very justification for the thickening of government that took place in the 1960s and 1970s. Policymakers wanted the tools to engage with and solve the most difficult types of social problems. Our second hypothesis states:

> *Capacity hypothesis: As institutional capacity grows policy instabilities will diminish.*

Our findings and expectations are consistent with two recent arguments put forth by Robinson and various collaborators in an important string of contributions (Robinson, 2004; Robinson & Caver, 2006; Robinson, Caver, Meier, & O'Toole, 2007; Robinson et al., 2014). This is the institutional argument: certain forms of institutional structure, such as centralization, may be more prone to generate punctuations than other structures. The most recent test of this idea is that institutional structures constitute a rival hypothesis to the "error-accumulation" idea inherent in the friction model posited by Jones and Baumgartner (2005). Robinson et al. (2014) note in their study of Texas school district budgets that districts with more punctuations in their history can be predicted to have more in their future. They suggest that this contradicts the friction idea as in that formulation one would expect that the lack of a punctuation would suggest a build-up of pressure, increasing the likelihood of a punctuation in the future. Their data, in any case, show that punctuations beget punctuations. Our hypotheses are also consistent with the idea that, for any given series, punctuations may be more or less likely; they are a characteristic of the build-up of pressure, to be sure, but also of how different sets of organizations deal with changes to their environments. Our findings, then, are consistent with Robinson et al.'s (2014) institutional model although our explanation of the process is different and our empirical points of reference differ as well. In any case, our ideas help explain what could be seen as a contradiction so far apparent in the literature. No single dataset so far used in the literature allows a clear delineation of these different operating mechanisms, so we simply note here that our hypotheses about institutional capacity and issue complexity are consistent with Robinson et al.'s (2014) idea of institutional memory.

Data and Measurement

To test our hypotheses, we use U.S. budget authority (BA) from 1947 to 2012. Figure 1 shows annual changes in these BA pooled across the 66 Office of Management and Budget (OMB) categories (e.g., subfunctions) for that period.[1] This is a simple update of the Jones–Baumgartner figure 4.14, which started the discussion about

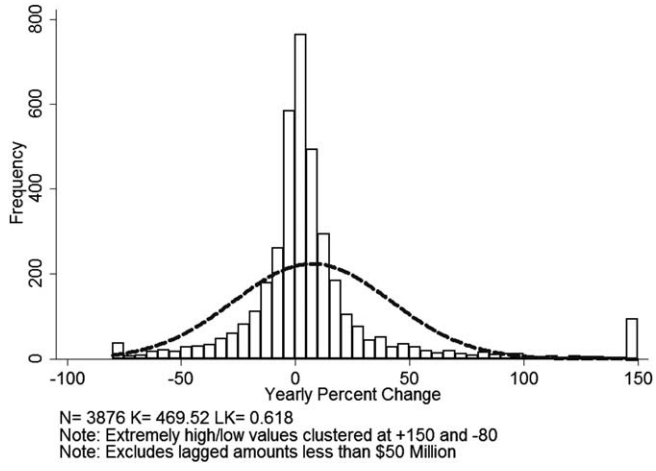

Figure 1. Distribution of Annual Changes in Federal Budget Authority, 1947 to 2012.

punctuations in public budgets (2005, p. 111). Clearly, this distribution is not normal, but instead features a high central peak and extremely wide tails, while the "shoulders," or midrange changes, of the distribution are missing. Note, as well, the l-kurtosis value (a standardized measure of kurtosis), which at 0.618 is well above the 0.123 associated with the normal distribution.

Our interest is identifying factors that predict budgetary instability. From 1947 to 2012, we document 3,876 annual budget changes. Various methods to distinguish between those cases far in the tails and those not considered to be punctuations have been used. One approach is a simple dichotomous classification where changes beyond a certain percentile cutoff are considered punctuations; for example, Jones, Baumgartner, and True (1998) placed this cutoff at +20 and −15 percent. Breunig and Koski (2006) used quantile regression to analyze separately the tails from the center of the distribution. Robinson et al. (2007) compared an observed budget distribution to a normal distribution with the same mean and standard deviation, using the relative mass between distributions to determine thresholds for small, medium, and large changes. More recently, Breunig and Jones (2011) have emphasized the value of exploring the full range of the distribution, not (as is typical) any single part of it. Commonly, scholars have focused on mean values for example, or the variance in values, and all these approaches have in common looking at a single moment or characteristic of the data.

We are less interested in documenting a threshold effect, for example, in distinguishing between "punctuated" and "incremental" changes, and more interested in developing a general indicator based on a continuous measure that allows us to capture the full nuance of budgetary change. Taking the full range of percentage changes, as documented in Figure 1, moving from the 50th to the 51st percentile—a trivial adjustment given the weight of observations in the central peak of the distribution—is a less notable policy change than moving from the 90th to the 91st percentile, where the magnitude of change is much greater.

To balance these considerations, and in concert with the recommendations in Breunig and Jones (2011), we develop an index of budgetary instability. The first step in constructing the index is to group observations by their corresponding percentile rank, so that the largest negative changes are assigned to the 1st percentile and the largest positive changes the 100th. We then transform this standard percentile ranking into a scale where the largest negative changes correspond with the value 50, which decreases toward 0 moving from the left tail to the center of the distribution, and then increases from 0 through 50 moving from the center for the distribution through the right tail. We then take the square of this scale to form our index. So, it is a squared index reflecting a folded percentile ranking.[2]

This transformation of the underlying percentage change values into the index of instability accomplishes two things. First, it measures the magnitude of changes: are they large or small, rather than their direction. (In Supporting Information, we estimate a model that looks separately at the index values corresponding to positive or negative changes, but our primary interest is with instability regardless of the direction it might take.) Second, the index puts an emphasis on punctuations by weighting the changes in the tails of the budget distribution more heavily than those in the central peak. This reflects the idea that punctuations are particularly interesting because they represent a break from the forces of negative feedback that usually operate on the policy process.

Introducing a new measure of an old dependent variable requires some special justification. We do so here in response to perceived uncertainty and dissatisfaction in the extant literature over the measurement of policy punctuations. The Supporting Information demonstrates that the index is robust to alternate specifications: we consider an index based on the cube of the percentile values rather than their square and an index that uses tenths of a percentile to rank the percentage change values. Both approaches preserve the underlying logic of the index construction, and using these specifications does nothing to change the substance of our findings. We also conduct a series of robustness tests to compare results using the index with results that measure the dependent variable in different ways. These include defining punctuations using a cutoff percentile, using the absolute value of the percentage changes, and defining punctuations based on the point at which the observed budget distribution passes a hypothetical normal distribution with the same mean and standard deviation. Each specification provides support for both of our hypotheses, suggesting that there is nothing revolutionary about the index that should call into question previous results. But, we still feel that it offers an operational improvement by emphasizing the most important elements of policy change (punctuations) while providing a continuous measure that uses all of the available data.

The left panel of Figure 2 shows how the index corresponds to a percentile ranking of the data. Budget changes at the 50th percentile (the exact center of the distribution) have an index score of 0, with movement down the left- or right-hand slopes of the distribution resulting in increasing index scores. Note that movement in the tails has a greater effect on the index value than movement in the center of the distribution. For example, moving from the 20th to the 1st percentile increases the index by 1,500, but moving from the 50th to the 30th percentile increases the index by only 400. In

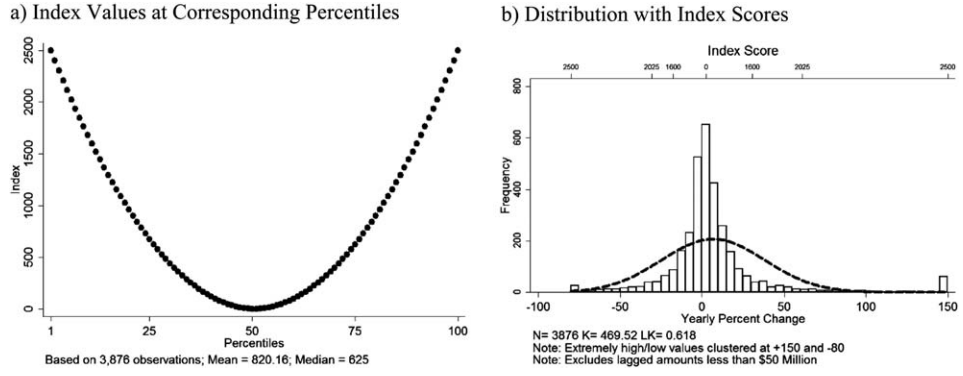

Figure 2. Index of Budget Instability: (a) Index Values at Corresponding Percentiles; (b) Distribution with Index Scores.

this way, changes in the tails of the distribution are weighted more than changes in the center and our index reflects the fact that incremental changes in spending may be relatively straightforward to enact, but that punctuations may require some special attention or legislative effort. Moving further out in the tails indicates greater punctuation in the budget, and this is accurately reflected in our index.

The right panel of Figure 2 displays the distribution from Figure 1, but with an additional *x*-axis at the top displaying index scores. This makes clear how the very largest percent changes, those clustered at −80 and 150 percent, receive the highest possible index score of 2,500. Most observations, however, are grouped in the center of the distribution and receive index scores between 0 and 1,600. This index then is our dependent variable.

Measuring Issue Complexity

A major influence on PE theory was the "garbage can" model of organizational choice, originally developed by Cohen, March, and Olsen (1972). Their insight was that in many organizations there is no iron-clad relationship between problems and solutions. Instead, problems and solutions co-mingle in an organizational garbage can, temporarily pairing off in ways that are not intuitive or definitive (and resulting in policy punctuations as they do so). We would argue that complex issues are much more likely to end up in the garbage can. Simple issues avoid this fate because they often come paired with ready-made solutions; this is what makes them simple. Thus issues that are complex for political or scientific reasons are more likely to attract interest from different organizational agencies. Given the uncertainty involved, each agency chief has a legitimate shot at portraying their agency's specialty as the solution to the problem. Participation can, therefore, be seen as an indicator of latent complexity, with the expectation that more complex issue areas encourage greater agency participation.

Federal spending authority is organized by the "function" or purpose of the spending, but the dollars may be allocated to any number of different federal agencies. Poverty abatement is a "function" of government but that spending may be

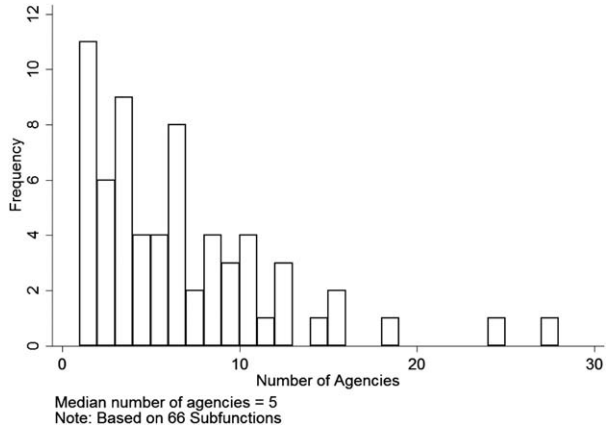

Figure 3. The Distribution of Agencies Across OMB Budget Subfunctions.

related to many different programs, often in different government departments and agencies. Relatively simple policy domains might have just one or a few agencies involved. Social Security spending, for example, is a huge budget category, but provides a solution to a relatively straightforward problem: the inability of elderly and disabled Americans to earn a living. The money is disbursed by just a single agency, the Social Security Administration. So, the budget domain is large, but not complex and there is very little opportunity for other agencies to advocate whatever solution they specialize in as more effective than Society Security payments.

The OMB maintains a dataset linking spending allocations by subfunction to the government agencies charged with implementing them.[3] For example, the National Science Foundation is frequently authorized to spend money allocated to the budget category for "general science and basic research." While the OMB maintains extensive records, the dataset linking subfunctions to agencies is available only from 1976 through 2008. During this period some categories, such as Social Security, have fallen exclusively under the purview of a single agency, while others, such as "advancement of commerce" or "general government," are carried out by more than 20 different agencies.

The complexity measure simply counts the number of distinct agencies linked to each category of spending. Figure 3 shows the number of agencies associated with each of 66 categories of spending. The median value is 5 agencies per category, but 14 categories have been linked to more than 10 agencies and 2 are linked to more than 20.[4]

Measuring Institutional Capacity

Baumgartner and Jones argue that government underwent a "great expansionary period" from the late-1940s through to the late-1980s. During this time, the government's capacity in engage with policy analysis expanded rapidly. Our argument is that policy instabilities should diminish as institutional capacity grows because

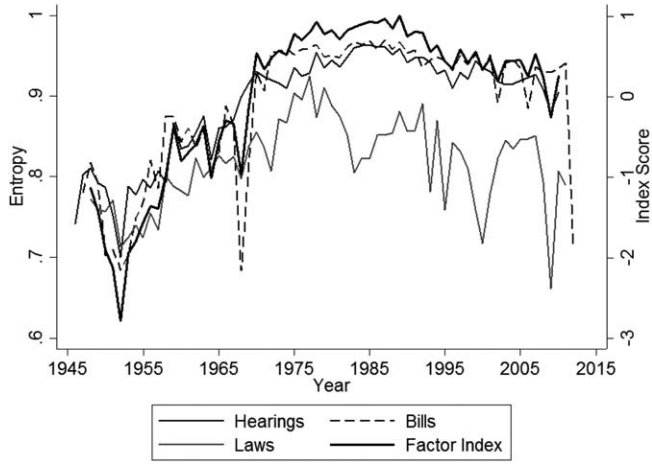

Figure 4. Diversity of the Government's Agenda.

during periods of higher capacity the ability of policymakers to process and respond to information is enhanced. To measure capacity, we turn to the Policy Agendas Project, which records every congressional hearing, bill introduction, and public law from the 1940s to 2013. Altogether this amounts to 95,000 hearings, more than 400,000 bills, and almost 20,000 laws.[5] Each bill, hearing, or law is assigned one of 19 major topic codes, so it is possible to track congressional attention to each of these different areas over the last 70 years.

Our interest is in the spread of attention across topics and we measure this by calculating the entropy of each congressional activity over time. Entropy measures the spread of information (or events) over discreet categories. There are a number of different approaches to calculating entropy; we use Shannon's *H* because it is sensitive to changes in systems where attention is relatively diffuse, as it often is in public organizations (Bodystun, Bevan, & Thomas, 2014). Shannon's *H* takes values between 0 and 1, with 1 being the most entropic possible system and 0 being the most concentrated. So, for example, if we have 10 categories and 100 events, and all 100 events take place in only one category, the resulting entropy score is 0. But if 10 events take place in each of the 10 categories, then the entropy score is 1. Periods when attention falls upon a diverse range of topics are indicative of a government that is more active in seeking out and responding to problems. This logic follows directly from the *Politics of Information*, where Baumgartner and Jones (2015) use a measure of congressional entropy similar to the one we develop here to assess the "thickness" of government. They show that periods of higher entropy are correlated with a government that is better equipped to produce and engage with policy analysis. Congressional entropy can, therefore, be seen as measuring the interest and ability of government to address problems.

Figure 4 shows the entropy of each congressional activity over time. All three activities show a similar trend: the government's attention is most myopic in the mid-1950s but the range of activities policymakers attend to expands rapidly during the 1960s and 1970s. This expansionary period reaches a peak in the 1980s and then

begins a slow but steady decline. Each activity is so closely correlated that it make sense to think of them as compositional elements of a single overarching variable: the diversity of the government's agenda. We can measure this variable using principal-components factor analysis to extract a common factor from the three activities. The resulting factor index is shown in the figure as a dark black line. This factor index is our measure of institutional capacity. Unlike the entropy scores, which vary from 0.6 to 1, the factor index varies between −3 and 1, with lower values indicating less diversity.

Model Specification

Our analysis uses a simple ordinary least squares (OLS) model to predict the magnitude of budgetary changes.[6] The dependent variable is our index measure, described above. Key independent variables include the measures of issue complexity and institutional capacity. We also control for various factors that might affect the likelihood of large shifts in spending: divided government, polarization, presidential popularity, the size of the spending category, and if the category addresses a demographic or a "crisis" area. Before proceeding to the results, we briefly review these variables.

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 (Aldrich, 1995; Coleman, 1999; Cox & McCubbins, 1991). This is represented in the model with a dichotomous variable, coded 1 if allocations were made during a period of unified government. Likewise, congressional polarization might affect the possibility for major policy shifts. When polarization is low, there is more room for cooperation between parties, but during periods of high polarization gridlock can bring even basic functions such as passing a budget to a halt. The model includes a measure of House polarization adopted from Poole and Rosenthal's DW-Nominate scores (Poole & Rosenthal, 1985). From 1947 through 2012, the measure varies between 0.40 and 1.10, with lower values indicating less polarization. A one-unit increase in the House polarization variable is coded as a 0.10 increase in DW-Nominate.

We also include a measure of presidential approval, as presidents who enjoy high approval ratings may be uniquely effective at ushering in large policy changes (Canes-Wrone & de Marchi, 2002). This variable is based on the classic Gallup question asking respondents if they approve or disapprove of the way the president is handling his job, and tracks the percentage of respondents indicating that they approve. The model also controls for the amount of money allocated to each category in each year. A concern is that instabilities are more prevalent 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.[7]

Finally, we note that different spending categories may operate according to very different dynamics. Consider that 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. Of particular concern is that many of the largest—in terms of expenditures—mandatory categories are strongly driven by demographic trends, such as retirements, and should logically have dynamics that perpetuate incremental changes; after all, aging is a predictable and incremental process. Of course, 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 incrementalism than others.

An equal concern would be that much of the instability usually attributed to the rise and fall of political agendas 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 precrisis 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.[8]

To address these concerns, we introduce dichotomous variables for spending that can be said to be driven primarily by demographic trends or by unforeseen crises. Spending on federal employee retirement and disability, general retirement and disability insurance, Medicare, and Social Security are classified as "demographic" categories. In 2012, these four subfunctions accounted for 41 percent of total government allocations. They are far from the only categories affected by demographics, but stand apart in their reliance on formulas that link spending directly to demographic trends. Crises can and sometimes do occur across a wide range of policy domains, but our interest is in identifying only those categories where a primary purpose of the spending program is to address sudden, unforeseen, and dramatic problems. We do so only after a careful reading of the OMB literature relating to each budget category. Altogether, six budget categories are coded as "crisis" categories: unemployment compensation, disaster relief and insurance, farm income stabilization, military-other, international humanitarian assistance, and international security assistance.

Results

Table 1 displays the results of the model, which we estimate in three ways. Recall that the range of index scores is from 0 to 2,500, with a mean of 820 and a median of 625. Model 1 is relatively simple, including issue complexity, institutional capacity, spending levels, and the political control variables; model 2 adds in our measures for demographic and crisis-related spending categories; and model 3 presents the same model with those categories excluded from the data. The last model allows us to test for whether the PE findings in the overall model hold when

Table 1. Predicting Budgetary Instability

Variable	Model 1	Model 2	Model 3
Dollars (billions)	−1.16* (0.20)	−0.64* (0.22)	−0.38 (0.41)
Institutional capacity	−128.25* (15.90)	−126.05* (15.44)	−121.97* (16.76)
Issue complexity	20.18* (2.20)	24.44* (2.16)	25.30* (2.21)
House polarization	−22.09* (6.84)	−24.26* (6.67)	−25.86* (7.18)
Presidential approval	−2.09* (1.00)	−2.05* (0.97)	−2.17* (1.05)
Unified government	−45.16 (26.06)	−45.32 (23.31)	−48.64 (27.38)
Demographic	–	−196.96* (53.90)	Excluded
Crisis	–	613.56* (45.88)	Excluded
Constant	1,021.28* (73.99)	955.93* (72.21)	964.06* (77.94)
N	3,699	3,699	3,167
Adjusted R^2	0.070	0.123	0.075

Note: OLS regression results shown with standard errors in parentheses. The dependent variable is the index of stability (see text for details). Demographic- and crisis-related categories are defined in the text.
*Significant at 0.05 *p*-value.

we exclude those categories where we might expect some artifactual causes for the findings.

Looking at model 2, it is clear that demographic-related spending categories are much less likely to see high percentage changes in spending: a value of −200 suggests spending much closer to the central peak for those cases. Similarly, as we expect, the crisis-related spending categories have much greater likelihoods of observations far out in the tails of the distribution, on average almost 600 points on our scale, which ranges only to 2,500. Looking at the other variables of interest, however, in particular the issue complexity and institutional capacity measures, show that these findings are robust to all three specifications.

As expected, issue complexity has a positive and significant effect on budgetary instability, no matter what control variables are included (in fact, it is slightly stronger with controls). Each additional agency assigned to a budget category increases the index value by about 25, or, converting backward, shifts the magnitude of the percentage change by about 1 percent.[9] Given that our complexity variable ranges from 1 to 27, the model predicts a major effect moving from the least to the most complex budget category. Likewise, the parameter for institutional capacity is statistically significant and in the expected direction. Increases in capacity correspond with major drops in the index scale, indicating that periods of higher capacity can be associated with greater budgetary stability. Moving from the period of lowest to highest capacity is associated with a decrease in the index of 512, as the capacity measure varies from −3 to 1.

In model 2, the variables for demographic and crisis spending show statistically significant and large effects on instability, but, critically, our theoretically motivated variables are robust to their inclusion, offering support for the PE model as an explanation of punctuations in public budgets. Furthermore, when we exclude all cases associated with either the demographic or the crisis-related spending categories, coefficients are robust, as shown in model 3. Thus, the findings are by no means driven by these cases. This reassures us not only about our own study, but those that have been previously published as well.

When it comes to the controls for governing conditions the story is mixed. Presidential approval has the opposite effect as expected: higher approval is associated with

smaller budgetary changes. These effects, however, are only very modest. Also modest (considering that it is a dichotomous variables) is the parameter for unified government, which is nonsignificant. House polarization, conversely, appears to have a large and significant effect. As polarization increases, gridlock settles in and it becomes increasingly difficult for the government to make large budgetary adjustments.

In all, the model shows very clear support for both our hypotheses. Thickening government by developing the capacity for policy analysis and then engaging with that analysis across a wide range of policy topics does appear to reduce instability. But, these informational tools only stretch so far; we also find that factors endogenous to different policy domains are an important part of the story. Areas of high complexity make the information problem faced by policymakers especially acute. So, we add two important and easily identifiable new indicators to the literature: the number of agencies associated with the spending category in question and the diversity of the government's agenda. One differs over time; one across policy domains. Controlling for the complexity of issues being considered during a given historical period, shared control leads to more punctuations. Multiplying the functions and range of government, conversely, leads to fewer punctuations.

Conclusion

Numerous studies have confirmed various elements of the PE theory of public budgeting and we further this research here in several ways. First, we propose a new measure of budget change, one that weighs policy changes far from the central peak of the distribution differently from those closer toward the center and which, therefore, makes use of all the observations in the data series while at the same time focusing more heavily on the unusual punctuations as compared to the more common incremental movements. Having developed a new measure of an old dependent variable, we then propose our key theoretical insight that issue complexity and institutional capacity have opposing effects on the policy process: complexity as a driver of punctuations and capacity as a stability-inducing element. The measures we introduce for complexity and capacity provide useful indicators of "friction," which in the past has not been operationalized for any single institutional process (but only across institutions, each of which has been associated with different levels of friction). Further, we propose controls for possible rival or artifactual processes which could hypothetically drive our findings as well as those in previously published literature: Budget series driven largely by mandatory spending programs associated with slowly shifting demographic trends (e.g., pensions) may have very low levels of punctuation whereas some spending series such as those associated with natural disasters (e.g., "crisis" or "shock" related series) may have very high volatility. We find that these series are indeed significantly different from the baseline, but that even while controlling for them, we support the broader theory. Finally, we include a range of robustness tests based on alternative measures of key variables and including or excluding possibly misleading categories of spending, revealing that the findings demonstrated here, consistent with the previous literature and including our new measure of institutional complexity, are highly robust.

Several authors have suggested explanations for observed differences in the characteristics of different policy domains. For Breunig, Mortensen, Koski, and others, these have been domain-specific or issue-related concerns. For Robinson and various coauthors, it has been institutional design. Recently, May and Jochim (2013) suggested a regime approach: different policy regimes operate with different rules of the game. Our focus on issue complexity and institutional capacity adds to this literature which seeks to understand the different patterns of policy change across different policy issues and institutional forms. We look forward, therefore, to more studies that seek to explain in greater detail the particular sets of institutional structure that promote and inhibit punctuated policy changes, and we believe that issue complexity and institutional capacity will be consistently found to be key drivers in this process.

Derek A. Epp is a postdoctoral research associate at the Nelson A. Rockefeller Center at Dartmouth College.
Frank R. Baumgartner is the Richard J. Richardson Distinguished Professor of Political Science at University of North Carolina-Chapel Hill.

Notes

1. All dollar values in the article are adjusted for inflation and presented as 2012 dollars. As is common in the literature (see Jones & Baumgartner, 2005), we use BA figures, not appropriations. BA refers to the decision to authorize spending, which can differ from actual expenditures, and corresponds better to the decision-making process we study. No large differences in the fundamental shape of the distributions follow from the choice to study BA or appropriations. In fact, because BA figures are not available for states, other countries, and many other budgetary settings, scholars have sometimes used appropriations data instead, with no discernible impact on the findings.
2. Note: This measure is conceptually similar to a *z*-score, but is not based on a standard deviation, which is biased in the case of extreme value distributions.
3. These data are available online through the OMB website.
4. We look at the largest number of agencies associated with each subfunction in the analysis below. That is, we use a static agency count. Numbers change relatively little over time, and the annual count is available only for the period from 1978 through 2008, as noted above. Our Supporting Information explains this in greater detail and shows that our findings are not driven by any of these decisions.
5. Bills are available from 1947 to 2012, hearings from 1946 to 2013, and public laws from 1948 to 2011.
6. The Supporting Information considers 10 alternative specifications. Results are highly consistent.
7. Table 6A in Supporting Information uses a measure of the percentage share of the annual budget rather than amount of spending and shows similar results.
8. The Supporting Information revisits previous findings of high kurtosis in U.S. budget distributions in light of these concerns. We find that even excluding potentially atheoretical budget categories from the analysis, budget distributions still show evidence of leptokurtosis.
9. Here, the index-percentage change crossover in Figure 2 may be helpful. For budget categories with the highest levels of agency participation, the model predicts an increase in the index of around 675, holding all other variables at their means. This corresponds with a percentage change value of around 27 percent.

References

Aldrich, John H. 1995. *Why Parties? The Origin and Transformation of Political Parties in America*. Chicago: University of Chicago Press.

Baumgartner, Frank R., and Bryan D. Jones. 2015. *The Politics of Information*. Chicago: University of Chicago Press.

Baumgartner, Frank R., Christian Breunig, Christoffer Green-Pedersen, Bryan D. Jones, Peter B. Mortensen, Michiel Neytemans, and Stefaan Walgrave. 2009. "Punctuated Equilibrium in Comparative Perspective." *American Journal of Political Science* 53 (3): 602–19.

Boydstun, Amber E., Shaun Bevan, and Herschel F. Thomas III. 2014. "The Importance of Attention Diversity and How to Measure It." *Policy Studies Journal* 42 (2): 173–96.

Breunig, Christian, and Bryan D. Jones. 2011. "Stochastic Process Methods with an Application to Budgetary Data." *Political Analysis* 19 (1): 103–17.

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

Breunig, Christian, Chris Koski, and Peter B. Mortensen. 2010. "Stability and Punctuations in Public Spending: A Comparative Study of Budget Functions." *Journal of Public Administration Research and Theory* 20 (3): 703–22.

Canes-Wrone, Brandice, and Scott de Marchi. 2002. "Presidential Approval and Legislative Success." *Journal of Politics* 64: 491–509.

Cohen, Michael D., James G. March, and Johan P. Olsen. 1972. "A Garbage Can Model of Organizational Choice." *Administrative Science Quarterly* 17 (1): 1–25.

Coleman, John J. 1999. "United Government, Divided Government, and Party Responsiveness." *American Political Science Review* 72: 1243–61.

Cox, Gary W., and Matthew D. McCubbins. 1991. "Divided Control of Fiscal Policy." In *The Politics of Divided Government*, ed. Gary W. Cox, and Samuel Kernell. Boulder, CO: Westview, 155–75.

Epp, Derek A. 2015. "Punctuated Equilibria in the Private Sector and the Stability of Market Systems." *Policy Studies Journal* 43 (4): 417–36.

Jensen, Carsten. 2009. "Policy Punctuations in Mature Welfare States." *Journal of Public Policy* 29 (3): 287–303.

John, Peter, and Shaun Bevan. 2012. "What Are Policy Punctuations? Large Changes in the Legislative Agenda of the UK Government, 1911–2008." *Policy Studies Journal* 40 (2): 89–107.

John, Peter, and Helen Margetts. 2003. "Policy Punctuations in the UK: Fluctuations and Equilibria in Central Government Expenditures since 1951." *Public Administration* 81 (3): 411–32.

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

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.

Jordan, Meagan. 2003. "Punctuations and Agendas: A New Look at Local Government Budget Expenditures." *Journal of Policy Analysis and Management* 22: 345–60.

Jones, Bryan D., Tracy Sulkin, and Heather A. Larsen. 2003. "Policy Punctuations in American Political Institutions." *The American Political Science Review* 97: 151–69.

May, Peter J., and Ashley E. Jochim. 2013. "Policy Regime Perspectives: Policies, Politics, and Governing." *Policy Studies Journal* 41 (3): 426–52.

Mortensen, Peter B. 2005. "Policy Punctuations in Danish Local Budgeting." *Public Administration* 83 (4): 931–50.

———. 2009. "Political Attention and Public Spending in the United States." *Policy Studies Journal* 37 (3): 435–55.

Poole, Keith T., and Howard Rosenthal. 1985. "A Spatial Model for Legislative Roll Call Analysis." *American Journal of Political Science* 29: 357–84.

Robinson, Scott E. 2004. "Punctuated Equilibrium, Bureaucratization, and Budgetary Changes in Schools." *Policy Studies Journal* 32: 25–39.

Robinson, Scott E., and Flounsay R. Caver. 2006. "Punctuated Equilibrium and Congressional Budgeting." *Political Research Quarterly* 59: 161–66.

Robinson, Scott E., Floun'say R. Caver, Kenneth J. Meier, and Laurence J. O'Toole, Jr. 2007. "Explaining Policy Punctuations: Bureaucratization and Budget Change." *American Journal of Political Science* 51: 140–50.

Robinson, Scott E., Carla M. Flink, and Chad M. King. 2014. "Organizational History and Budgetary Punctuation." *Journal of Public Administration Research and Theory* 24: 459–71.

Ryu, Jay Eungha. 2011. *Bounded Bureaucracy and the Budgetary Process in the United States*. New Brunswick, NJ: Transaction Publishers.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

Supporting Information 1—Alternative Measures of Budget Instability
Figure 1A: Distribution of Annual Percentage Changes in U.S. Budget Appropriations, 1947 to 2012: Defining Punctuations as Observations below the 10th and above the 90th Percentiles
Figure 2A: Defining Punctuations as Observations beyond the Point Where the Mass of the Budget Distribution is Greater than the Mass of a Hypothetical Normal Distribution a) Relative Density Plot b) Distribution Cutoffs
Table 1A: Logistic Model Predicting Punctuations (Figure 1A Definition)
Table 2A: Logistic Model Predicting Punctuations (Figure 2A Definition)
Table 3A: OLS Model Predicting Absolute Value of Budgetary Change
Table 4A: OLS Model Predicting Budgetary Instability, with Cubed Index
Table 5A: OLS Model Predicting Budgetary Instability, with Index Based on 10th of Percentiles
Supporting Information 2—Alternative Specifications of Independent Variables
Table 6A: OLS Models Predicting Budgetary Instability, with Alternative Specification of Dollar Amounts
Table 7A: OLS Models Predicting Budgetary Instability for Negative and Positive Changes
Table 8A: OLS Model Predicting Budgetary Instability, with Annual Complexity Measure
Figure 3A: Agencies by Subfunction in 1976 and 2008
Supporting Information 3—Alternative Model Specifications
Table 9A: Cross-sectional Time-series Models Predicting Budgetary Instability, with Random Effects
Table 10A: Predicting Budgetary Instability with Quantile Regression
Supporting Information 4—Are Previous Findings Robust?
Table 11A: Kurtosis of U.S. Budget Distribution with Demographic and Crises Series Excluded, 1947 to 2012
Figure 12A: Tracking the Annual Inter-quartile Range and L-kurtosis of Percent Changes in the US Budget, 1947 to 2012