Program Type and Budget Change: Does it make a Difference?

Jinhai Yu
Hyokyung Kwak*
(University of Kentucky)

Prepared for presentation at the Association for Public Policy Analysis & Management Fall Research Conference, Miami, FL, November 14, 2015

Abstract

This paper applies a public organization theory to the U.S. federal budgeting process to explain changes in federal budgeting. James Q Wilson (1991) divided public organizations into four categories including production, craft, procedural, and coping ones according to whether the agency outputs are easily measured and whether the production processes are visible to outsiders. We attempt to empirically test this typology at the program level by examining its consequences for budget change. The rationale is that agencies can use high levels of output measurement and production transparency as “hard” evidence to bargain for more resources for a program and leave less room for budget cutbacks from President, OMB, or the Congress. The hypothesis is that those programs whose production is readily observable and output is easily measured are more likely to have a stable budget allocation, i.e., smaller variance and less punctuations. We first distinguish discretionary programs from entitlement programs and then focus our analysis on the former category. We code the type of program based on the information provided by the Program Assessment and Rating Tool (PART) in the Bush Administration. As a demonstrative test of the coding rules, we analyze a randomly selected sample of all discretionary programs and find that, compared to production programs, coping programs are more likely to experience big, negative budget change from the actual spending in 2008 to the budget request in 2009. While this preliminary analysis lends tentative support to the hypothesis, the next step of improvement is to refine the coding procedure and to enlarge the sample to include all discretionary programs so as to achieve higher statistical power. By exploring the effects of program characteristics on its budget allocation, this study would contribute to the literature on government budgeting by incorporating theories from other subfields of public administration.

Key words

Federal Program, Output measurability, Production visibility, Budget Allocation

* Hyokyung Kwak is the presenter and contact author: hyokyung.kwak@uky.edu. Hyokyung Kwak and Jinhai Yu are PhD students in Martin School of Public Policy and Administration at University of Kentucky. We would like to thank Prof. Nicolai Petrovsky, Prof. Edward Jennings, and Prof. J.S. Butler for their guidance and help.
Program Type and Budget Change: Does it make a Difference?

1. Introduction

For a long time it has been useful to borrow theories from political science and economics to explain the budgeting process. Examples include incrementalism, the punctuated equilibrium model, and public choice. In contrast, there are relatively fewer public organizational theories applied in such way. This is unfortunate since, by focusing on the characteristics and logic of organizational activities, public organization theory can provide significant insights about the governmental budgeting processes. This paper represents an effort of applying organizational theory to government budgeting issues. Bridging these two fields not only helps test organizational theory empirically but also has the potential to enhance our understanding of the public budgeting process.

In his classic work on American bureaucracy, James Q. Wilson (1989) divided public programs into four types, including production, craft, procedural, and coping ones according to whether the program outputs are easily measured and whether the production processes are readily visible to outsiders. One interesting research question to ask is what implications this framework has for public budgeting. The two dimensions of performance measurability and production visibility are directly related to the level of information asymmetry in budget negotiation and the costs of evaluating and monitoring agency activities from the standpoint of budget bureau. The agencies administering different types of program may have different leverage or bargaining power in the budget process due to the nature of programs they daily run. Consequently, there may be different patterns of budget change for the programs falling into different types.
To empirically test the effects of program type on budget change, we use data on federal programs covered by the Program Assessment and Rating Tool (PART) in the George W. Bush Administration. After identifying the discretionary programs, we try to code their types by the dimensions of output and outcome visibility based on the program information collected by the PART. Due to the uncertainty and costs of the coding, we coded 47 randomly selected discretionary programs as a demonstration of the research design at this stage. The preliminary analysis lends some support to the hypothesis in that coping programs are more likely to face large budget decrease than production programs are between actual budget spending in 2008 and budget request in 2009. We are aware of the limitation of current analysis and aim to improve it in terms of data coding and sample size in the future research. This paper consists of following sections. The next section is a literature review on existing studies on budget change, budget negotiation, and program type. Then the theoretical arguments and hypotheses are proposed. The measurement of program type and data coding are discussed. We lastly report the preliminary empirical test and conclude by discussing possible pathways of improving the study in the next step.

2. Literature Review

Budget Change

There are two theoretical frameworks in the existing literature to explain budget change. On the one hand, the incrementalism budgeting model suggests that government budget changes incrementally over time, with the past year’s budgeting level serving as the major base of the current year’s budget. The main reasons are the bounded rationality of
budget game players, the complexity of budget process, and the fragmentation of budget institutions (Davis, Dempster, and Wildavsky, 1966; White, 1994). On the other hand, the punctuated equilibrium model depicts the budget change as a series of small changes occasionally punctuated by radical ones (Baumgartner and Jones, 2009). The level or frequency of punctuations is positively correlated to the institutional friction embedded in the budgetary process (Jones et al, 2003). These two models agree that budget changes are incremental most of time but differ in their explanations of big budget changes, which are repeatedly found to exist. While the incrementalism budgeting views the big changes as outliers, the punctuated equilibrium model attempts to provide a consistent rational logic for both big and small budget changes. Overall, the punctuated equilibrium model has found empirical support at the federal, state and local budgeting levels (Jones et al, 1998; Jordan, 2003).

While the small and big budget changes are well-documented and generally explained by these two theories, neither is able to specify the conditions under which we observe a particular pattern of budget changes. Taking a further step based on these two theories, Robinson et al. (2007) studied the effects of organization size and centralization on budget change in the setting of school districts. They measured the dependent variable of budget change by defining small, medium, and large changes through comparing the actual leptokurtic distribution of changes of budget spending over time to the assumed normal distribution. By studying the effects of program features on their budget change, this paper may contribute to the literature by exploring the influencing factors of when we observe big or small budget changes.

*Budget Negotiation*
The budget change may result from internal changes in the budget negotiation process. Wildavsky (1964) analyzed the federal budget negotiation process among the agencies, the Office of Management and Budget (OMB), the President and the congressional committees. He described the tricks, norms, and practices of budget bargaining between the agencies and the OMB and showed how this process generated incremental budget changes. However, in these two theoretical models, there is no hypothesis about which agencies or programs are more likely to have incremental changes than others.

There are some studies on the effects of budget negotiation on budget outputs or outcomes in the private sector (Fisher, Frederickson, and Peffer, 2000, 2002, 2006) but not in the public sector. Although it is well recognized that information asymmetry affects the output of budget negotiations, there are limited implications of this research for government budgeting due to the many differences between private and public budgeting.

The effects of budget negotiations on budget change may differ across two distinctive federal programs, i.e., entitlement programs and bureau programs. White (1998) outlined the differences of the budgeting of these two kinds of programs. While the entitlement program spending is typically determined by legal formulas and the budget changes subject to legislative reforms to these formulas, the bureau program spending goes through regular federal budgeting process annually. There is much more room for budget negotiation between federal agencies and the OMB for bureau program spending than that for entitlement spending. This distinction suggests that there must be differences in the effect of program type on budget change for these two kinds of programs as well.
Program Type

While Wilson’s typology of programs makes theoretical sense, there are few empirical tests in the existing literature. Gregory (1995) applied Wilson’s typology to analyze the management culture, accountability and responsibility of public sector reform in New Zealand and pointed out that the particular task an organization faces is critical to determine its type. Though a dozen of government agencies were put into different categories of this typology, this is not a quantitative test by itself. The lack of such test may be due to the difficulties involved in coding programs or agencies in practice, which is a challenge we have begun to address in this paper.

The two dimensions of Wilson’s typology of performance measurability and production visibility receive different level of academic attention. While performance measurability becomes a central concern of performance management and performance budgeting in recent decades, the production visibility is not a similar focus in public administration. According to Jones (1984), “the level of visibility will depend on the extent to which the work context permits the monitoring and evaluation of individual performance. (p.686)” Specifically, the task visibility is affected by such factors as the size, interdependence, routineness, and technology complexity of work (Jones, 1984; Esfahani and Salehi-Isfahani, 1989). This research provides some useful guidelines for specifying the level of production visibility of different federal programs.

3. Theory and Hypothesis

The federal program is the appropriate unit of analysis in this study for several reasons. First, while Wilson’s typology initially target agencies, it can be equally applied at the
program level. Federal budgeting can take place at both the program and agency level, which allows some room of choice between the two options. Nonetheless, either at the agency or program level, the typology can be difficult to quantify because they typically have multiple goals which may differ in the degree of output or outcome visibility (Jennings and Haist, 2004). Second, one federal agency typically administers multiple programs and thus makes it necessary for further aggregation to code agencies into different categories, where the effect of program type on budget change might be more difficult to identify. One potential way to classify agencies is to code them according to the proportion of different types of programs. However, this can be challenging for empirical testing because one agency may manage both entitlement and discretionary programs, while we mainly expect the theory to work for discretionary ones. Hence, even if the agency type could be measured, it is less useful to examine its effects on budget change. Third, the data on the Program Assessment Rating Tool (PART) during the Bush Administration can be readily used to code the program type and contain the information on program budget changes. Therefore, we choose to focus on the program level to reduce the complexities in empirical test of Wilson’s typology.

There are several stages of the federal budgeting process where budget bargaining or negotiation usually happens. First, the agency head proposes their budget request to the OMB, covering the spending of all programs this agency administers. This is usually done under the budget guideline from the OMB in the spring at the beginning of a budget cycle. As a result, the agency has some sense of what level of budget it may get from the OMB this year. Second, the OMB sums up all the budget requests from different agencies and adjust them by taking into account the policy objectives and priorities of the
President and the economic forecasting results. Third, the President presents the total budget proposal to the congress on behalf of his or her administration. Fourth, the two chambers of Congress work on the presidential budget proposals by reviewing them in various congressional committees or subcommittees. The agencies may be called by the congressional committees to testify to defend their budget requests. Fifth, the budget proposal is further adjusted by the congress until a consensus is reached and gets passed by house and senate, sometimes requiring reconciliations between the two chambers. In the cases where there is no consensus being reached, the Congress usually passes a resolution to continue the past year’s budget appropriations to maintain the functioning of core government activities.

Given the scarcity of resources, it is unlikely that all budget requests at either the program or agency level will be met by the Congress in the budget negotiation. Thus agencies are motivated to use some strategies or tactics to win more resources in the budget competition. For instance, agencies could propose budget decrease for some programs valued by the Congress which essentially makes the cutbacks less likely to happen, or they could induce the Congress to pass authorization for a program which turn out in the next few years to be far more expensive than it is originally projected (Wildavsky, 1964). As described by Wildavsky (1964), the mutual expectation between agency and the OMB leads the agency to propose a higher budget request than it actually needs to allow some room for cutback by the OMB or congressional committees. While the focus of Wildavsky (1964) is to argue how this bargaining process generates incremental budget changes, there is no indication of whether or why some agencies or programs may fare better than others by getting the budget share it desires in the end.
Wilson’s typology of programs fit the budget negotiation process by providing different leverage or bargaining power to the agencies that usually face the pressure of budget cutbacks. As found by Fisher, Frederickson, and Peffer (2002), the information asymmetry between agency and budget bureau affects their positions and strategies in budget negotiation. The performance measurability and production visibility of federal programs directly affects the level of information asymmetry between federal agencies and the OMB or the congressional committees. The higher the level of performance measurability and production visibility, the lower the level of information asymmetry, and thus the results of budget negotiation should differ accordingly.

Specifically, the two dimensions of programs may affect the bargaining power and strategies of agencies in three ways. First, the agencies can use quantitative performance measures and transparent production processes of the programs it administers as “hard” evidence to argue for more budget resources in front of the OM and legislators. This is very likely to be the case in the Bush Administration because one goal of the PART is to encourage the linkage of performance measures with budget results. Second, high performance measurability and production visibility of programs reduces the institutional friction or transaction costs of budget bargaining since the discrepancy between feasible budget request and budget cuts are decreased for both the agencies and the OMB or legislators. In other words, it is easier for agencies and the OMB to reach consensus on the “truly appropriate” budget level due to a lower level of information asymmetry among different parties. Third, high performance measurability and production visibility may increase the external autonomy of agencies in the sense that they are less prone to political pressures for cutbacks and easier to defend their positions (Wilson, 1989).
As a result of the different bargaining power of agency, there should be different patterns of budget change both at the agency and program level. The budget change can mean either the difference between the agency budget request and the actual budget allocation it gets or the change of actual budget allocation over time. Since the federal agencies have little discretion of distributing fund across different spending items or programs, the pattern of budget change resulting from program types should be identical at the agency and program level. Nonetheless, the agencies administering different proportion of different types of program may not behave the same as if they only manage one type of program. As long as the budget bargaining between agencies and the OMB or congressional committees focuses on the program spending or the structure of agency spending rather than simply the agency spending as a whole, the program type should matter for program-level budget changes.

We develop four hypotheses about the effects of program type on the program budget change. For the programs with high output measurability and production visibility, there is less room both for the agency to propose higher budget requests than it truly needs and for the OMB or congressional committees to cut down. As a result, the discrepancy between budget request and final budget allocation should be smaller for such programs. The first hypothesis thus is:

**H1a:** The programs with high output measurability and production visibility are more likely to have smaller difference between budget requests and actual budget allocations.

In addition, there should be differences in the direction of budget change for different types of programs in difficult budget times such as the economic recession in 2009.
When there are expectations of reduced revenues and budget cuts, the programs with high output measurability and production visibility should still fare better than those without. In other words, the programs with low visibility of output and outcome allows more room for budget cutbacks because it is easier to propose and more difficult to defend. Thus:

**H1b: The programs with high output measurability and production visibility are less likely to have budget cuts from actual budget allocations to budget requests.**

For the second dimension of budget change, there are two opposite hypothesis about the effects of program type. On the one hand, in the same logic of the first hypothesis, high output measurability and production visibility should make budget bargaining easier to reach agreement due to lower level of information asymmetry and also subject to lower level of institutional friction. According to the punctuated equilibrium model, this should lead to less punctuation and more incremental changes, i.e., more stable budget allocation over time. Therefore:

**H2a: The programs with high output measurability and production visibility are more likely to have smaller budget changes over time.**

On the other hand, low output measurability and production visibility means higher level of information asymmetry and uncertainty in the budget process. The incrementalism budgeting theory suggests that budget actors are more prone to stick to the past-year budgeting level as a base when there is high uncertainty. Since there is no “solid” evidence on program performance or production process at all, the consensus reached in the past negotiations takes more weight for budget actors to simplify the budget
bargaining process since there will be high costs to overthrow it and start over each year. Therefore:

H2a: The programs with low output measurability and production visibility are more likely to have smaller budget changes over time.

Note that the two groups of hypotheses focus on different stages of budget bargaining. The first hypothesis on the difference between budget spending and budget request concentrates on the first half of federal budgeting process, i.e., from the beginning to the President budget proposal. The second hypothesis focuses on the consequences of the whole budgeting process because it is about the changes of actual budget appropriations over time. Furthermore, since the relationship between program type and budget change depends on the intermediate link of budget bargaining, it is understandable that these hypotheses apply to the discretionary program budgeting but not to the entitlement program budgeting. There is simply insufficient room of budget negotiation between agency and the OMB or congressional committees to allow the program type to matter in the entitlement budgeting. Therefore, to better test these hypotheses, we focus only on the discretionary program budgeting and to exclude entitlement programs.

4. Data and Measurement

*General Considerations*

The Program Assessment and Rating Tool (PART) data contains scores on federal program performance and program budget allocation. From 2002 to 2008, 1016 federal programs, or 98% of total federal programs were graded in the PART system (Gallo and Lewis, 2012). PART was used to enhance the budget leverage of presidents and to
reduce inefficiencies and waste in program spending (Dull, 2006). Also, PART was incorporated into the congressional budget process as part of budget justification and federal agency testimony on a limited basis (Frisco and Stalebrink, 2008).

PART has been studied extensively in the existing literature (Gilmour and Lewis, 2006; Frisco and Stalebrink, 2008; Heinrich, 2012; Lavertu and Moynihan, 2012). Most of these studies are concerned about whether the PART reflects true program performance and how the performance information is actually used in the budgeting process. There are mixed findings on both research questions. For example, Heinrich (2012) finds PART ratings reflect some aspects of the quality of evidence of program effectiveness but do not affect the program funding. In contrast, Gilmour and Lewis (2006) find that the PART score do positively affect program funding in traditionally Democratic agencies.

While these studies provide useful guidelines on the structure of PART and the control variables to include in this paper, here the PART is used in a very different way from the approach to the data by these researchers. Particularly, we focus on the measurability of output and outcome of programs, which is only part of the information contained in the PART system and its effects on budget allocation.

There are several possibilities to construct measures for the type of PART programs. One may make use of the PART rating results directly. The overall program evaluation includes results not demonstrated, ineffective, adequate, moderately effective, and effective. A rating of “Results not demonstrated” indicates that the program does not have appropriate performance measures (Gilmour and Lewis, 2006). The first dimension of Wilson’s typology of program, the output measurability of a program could be
measured by differentiating programs which are rated as “Results not demonstrated” from others. The rationale is that the result of PART review stating that a program was not able to demonstrate its performance is an indicator for the difficulty involved in performance measurement.

However, a counterargument is that the grade “Results not demonstrated” can tell us very different information other than the difficulty of performance measurement. For example, a program receiving such a rating may be due to the reluctance or slow response of the agencies to provide sufficient performance information to the OMB while it is inherently easy to do so. Moreover, agency ideology is found to affect the grades the programs obtained in the PART system. Gallo and Lewis (2012) find the programs hosted by the agencies with conservative ideology received higher PART scores. Furthermore, we do not know whether a program receiving such a grade is due to the lack of output or outcome measure, which is critical to the Wilson’s typology. Therefore, “Results not demonstrated” is not a quality measurement of output measurability and we turn to the qualitative coding instead.

There is no option of direct measurement of the other dimension of Wilson’s typology--production visibility. There is no readily available measure in this dataset. To measure production visibility, the two authors have coded the federal programs covered in the PART system as high or low level of production visibility independently according to some rules of threshold. This same coding process also applies to the dimension of output measurability. At the minimal level, this coding needs an operational definition of “production visibility” and “output measurability” to be feasible, which is also a
challenge remaining to be met. Ideally, taking these two dimensions together, any program in the dataset can be assigned to the two by two matrix of Wilson (1991).

The complete PART score dataset is available on David Lewis’s website (https://my.vanderbilt.edu/davidlewis/data/). Considering all the programs covered in PART, this is essentially cross-sectional data in which the PART score is coded once in different years. The key independent variable in the present study is the program type based on performance measurability and production observability, which arguably did not change too much over time, so a cross-sectional analysis may be sufficient for the purpose to explore the effects of program feature on the budget change. In other words, this relationship varies across different program types but not over time.

The dependent variable is not the absolute level of budget allocation of a program since it is difficult to argue why different types of program should lead to high or low level of budget shares. Instead, the dependent variable is some measure of changes in budget allocation of a program, either compared to the program budget request or the actual budget allocation historically. There are four possible measurements of the dependent variable of budget change. First, the level difference between the budget request and actual budget allocation of a program each year may reflect how the program fares in the budget negotiation. Second, the annual percentage change of actual budget allocation for a program over time is a simple measurement of budget change. Third, the variance of program budget allocation over time is an overall measurement of the stability of budget change. Fourth, the kurtosis of program budget allocation over time is a measurement of fluctuation of budget change (Robinson et al, 2007).
The first and the last three measurements are built upon slightly different theoretical arguments. While the first one measures to what extent a program is able to get what it requests, the other three measures how stable the program budget it is over time. Moreover, to make the last three measures and the test of third and fourth hypotheses feasible, we need to know the federal budget appropriations at the program level between 2002 and 2008. Unfortunately, we are unable to obtain such data at this stage and thus we focus on testing the first two hypotheses. In the PART score data set mentioned above, there are variables of budget spending for each program in 2007 and 2008, and the budget request for each program in 2009. Therefore, we construct our dependent variable based on this data.

**Entitlement vs. Discretionary Programs**

The first step of our data analysis is to identify which program in the PART system is discretionary. There is no such variable in the PART data set we use. This means we need to find other information source on the entitlement/discretionary categories of federal program. It turns out there seems to be no such information at the federal program level. What we managed to find is a data set of federal budget outlay from the Office of Management and Budget (OMB) website [https://www.whitehouse.gov/omb/budget/Supplemental](https://www.whitehouse.gov/omb/budget/Supplemental) which contains such categories at the account level. The critical variable of interest in this data is a column indicating the Budget Enforcement Act (BEA) category. There are three categories of federal spending according to the BEA, discretionary, mandatory, and net interest. Specifically, “Discretionary spending is controlled through annual appropriations acts”. “Mandatory spending generally operates under permanent authority. Statutes generally specify what
must be paid and who is eligible to receive payments.” Net interest refers to “Interest payments to and from the public and intragovernmental payments of interest, primarily to trust funds, are not included as either mandatory or discretionary spending.” (User’s Guide, the OMB, https://www.whitehouse.gov/omb/budget/Supplemental). The terminology of “account” is explained by the OMB as “assigned by the Treasury Department and are used for budget presentation as well as financial management. In some instances, groups of small accounts (usually miscellaneous receipt accounts) have been consolidated into single accounts” (User’s Guide, the OMB).

There are both positive and negative sides of this data set, which is also the closest to the ideal data we would like to have. The positive side is that we can use the BEA category to identify which federal spending is discretionary. The negative one is that this information is only available at the account level, which may or may not match the PART programs. As indicated by the OMB, the account is at higher aggregate level than the programs. One additional complicating factor is that while the PART data set we use contain the program names from 2002 to 2008, the federal budget outlay data shows the same account name over time. Therefore, we may not be able to match the programs to the account because of the name change of the program.

In spite of all these challenges, we set out to match the account name in the budget outlay data to the PART programs. There are more than half of the cases where we can exactly match the account names and the program names. Nonetheless, we cannot match the others either because there are more than one account name corresponds to the PART program name or the account name and program name are simply different. We code the PART programs into three categories including discretionary, entitlement, and mixed.
The “mixed” category contains the programs that we cannot identify its BEA category with this method. We end up with about 243 discretionary programs, 61 mandatory programs, and 155 mixed programs. Due to the compounding factors discussed above, this number should be an under-estimate of the discretionary PART programs because there is some proportion of “mixed” programs which actually is discretionary but we cannot identify. In any event, this is the fraction of PART programs that we expect to use to test our hypothesis on the relation between program type and budget change.

**Data coding**

The most critical and challenging part of implementing this empirical test is to conduct a proper coding of the PART programs according to Wilson’s typology. While Wilson’s typology is insightful, there is little guideline on how one could possibly put it into empirical tests. There are no such guidelines either from Wilson or other scholars on how one could decide the output measurability and production visibility of a program or agency. The first step to simplify the task is to re-focus the two dimensions of the Wilson’s typology to output and outcome measurability according to existing studies such as Jennings and Haist (2004). Essentially, the slight change is made so that the production visibility equals outcome measurability. Then coding task is to discern whether the output and outcome of a program is measurable.

In principle, there are two possible approaches. Wilson’s typology emphasizes whether it is possible to measure the output or outcome of a program or agency and his viewpoint is that in some cases the program output or outcome are inherently difficult to measure. However, this creates a challenge for empirical testing. It is difficult to know for sure
whether a program is “measurable” because it may depend on how many efforts or resources one would like to invest to measure it and other factors such as technology of measurement. It seems there are no direct ways to determine whether the program output or outcome is measurable in theory given the varieties of federal programs and huge amount of information to take into account. Therefore, we turn to the second, indirect way of measurement. What we observe in practice is that some programs have output or outcome measures while others do not. We can infer from such facts that the programs with more output or outcome measures are easier to measure than those without. The quality of output and outcome measures certainly deserves equal attention. We can safely conclude that the output or outcome of program is measurable only when there is certain number of such measurements of certain “acceptable” quality. In this study, we base our coding on the performance measures contained in the PART system.

Although still having room for ambiguity, the second approach is what we currently adopt to code the PART programs. We conduct the coding according to the two dimensions of Wilson’s theoretical framework—to what extent the output and outcome of a program is measurable. If the output or outcome of a program is measurable, we code the variable of output or outcome measurability as 1; otherwise, we code it as 0. We try to assess these two characteristics of a program from the perspective of the political principals, which include the OMB, the President, and the congressional committees. This is because how the output or outcome measurability affects information asymmetry between these principals and federal agencies as agents is at the core of our theoretical argument.
Even though “output” or “outcome” is commonly used concept in public administration, it is still challenging to come up with operational definitions in a specific context like the PART program. The OMB has a general definition of these two concepts. Specifically, outcomes are “events or conditions external to the program and of direct importance to the public, beneficiaries and/or customers. They relate to the program’s mission, purpose and strategic goals.” Outputs are “internal program activities; products and services delivered to the public, beneficiaries”. We try to follow this conceptual distinction offered by the OMB.

To proceed, we attempt to make use of the information generated in the PART review process. For each PART program, there are assessment details including the questions, answers, and scores which are available online at the Expectmore.gov website (http://georgewbush-whitehouse.archives.gov/omb/expectmore/summary/10003230.2006.html). Typically, the answer to the first question on the PART questionnaire that “Is the Program Purpose Clear?” provides a brief summary of the objectives of a program. The four questions under the “Strategic Planning” section concern the performance measures in terms of output or outcome of a program. Although they appear to focus on outcome measures, the answers by most programs include both output and outcome measures. Specifically, these questions are shown in the Table 1 below.
Table 1: The Questions on Performance Measures in the PART Evaluation

<table>
<thead>
<tr>
<th>Number</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Does the program have a limited number of specific long-term performance measures that focus on outcomes and meaningfully reflect the purpose of the program?</td>
</tr>
<tr>
<td>2.2</td>
<td>Does the program have ambitious targets and timeframes for its long-term measures?</td>
</tr>
<tr>
<td>2.3</td>
<td>Does the program have a limited number of specific annual performance measures that can demonstrate progress toward achieving the program's long-term goals?</td>
</tr>
<tr>
<td>2.4</td>
<td>Does the program have baselines and ambitious targets for its annual measures?</td>
</tr>
</tbody>
</table>

Even though the OMB tries to define the concepts of output and outcome, as shown in the training document provided by the OMB (http://georgewbush-whitehouse.archives.gov/omb/performance/), they did not make an explicit effort to distinguish the output, outcome or efficiency measures when asking the agencies for performance information. However, for our purpose, the key is to figuring out whether a program has output or outcome measures of decent quality.

As directed by the OMB training document, the answers to the four questions in Table 1 provide information on the output or outcome measures associated with any given programs. We make the assumption that the output or outcome measures of a program answered by the agencies in the PART review process are nearly exclusive. In other words, there should not be any major undocumented output or outcome measures for any program during the PART evaluation. The assumption is reasonable because, given the incentives of the agencies to improve their PART scores, it make senses for them to provide as many performance measures as possible. In addition, there are no clear guidelines from the OMB on whether output or outcome measure is more desirable. Even if one agency wants to hide some performance measures, there is no reason to
systematically hide one type or another. Moreover, previous studies show that it is easier
to use output measures than outcome measures as the former is usually more ready to
quantify and more controllable by the agency. Hence, we would expect many of the
performance measures provided by an agency are output measures, even though the
questions seem to emphasize outcome measures. Thus, the key to our coding becomes
separating output measures from outcome measures among all the performance measures
of any PART program as reported by the hosting agencies or bureaus.

More specifically, when we evaluate the output or outcome measures of a program, we
do not totally rely on the number of output or outcome measures. It is presumably true
that some performance measures are of higher quality than others. Though there are no
operational guidelines on what constitutes a good performance measure, we would like to
make qualitative judgments based on such general principles as relevance to program
goals, feasibility, and timeliness, among others. Therefore, we take account into both the
number and quality of output or outcome measures associated with a program to
determine its type by the Wilson’s framework.

Due to time constraints and the uncertainty on our coding methods, we decide to code
part of all the discretionary programs as a demonstration. We randomly sample 47
programs from the 243 discretionary programs and code them according to the
procedures above. Both of the authors conduct the coding independently and then we
come together to solve the inconsistency of our coding results. We reconcile the
inconsistencies by stating our reasons of our own coding and re-examine the performance
measures. It turns out that our coding results are the same for about 60 percent of the
programs. This inter-rater consistency rate is not promising compared to 80 or 90 percent
in the existing literature. It implies that either our coding rules are still ambiguous or the coders are not familiar enough with the programs. This suggests room for improvement of our coding rules in the future study.

5. A Preliminary Test

Model

At this stage, we make use of the coding results for these 47 programs to conduct a preliminary analysis. The first step is to measure the dependent variable of budget change between actual budget spending in 2008 and budget request 2009, the actual dollar amount of which is available in our data set. Robinson et al. (2007) developed a method to distinguish budget punctuations from incremental changes of the instructional spending per pupil in school districts in Texas. They imposed a normal distribution on the top of the aggregate distribution of changes in the instructional spending from 1989 to 2001. The big budget changes are defined as those that fall on both tails of the distribution and whose frequencies exceed the normal distribution. Small budget changes are those that fall in the middle and whose frequencies exceed the normal distribution. Medium budget changes are those that fall between far tails and the middle and whose frequencies are below the normal distribution.

We adapted this approach to measure the size of budget change in our case. Given the small sample size, we instead distinguish only big and small budget change. A big budget change is coded 1 if the budget change falls on both tails and its frequency exceeds the normal distribution. All other budget changes are defined as small change and the variable of big budget change is coded 0. In addition to the size of budget change, we are
also interested in the direction of budget change from the actual budget spending in 2008 to the budget request in 2009. The direction of budget change is an index which equals 1 if there is a budget increase, 0 if there is no change, and -1 if there is a budget decrease. In the sample, there is high correlation between the actual spending in 2008 and budget request in 2009, which is 99.45%. A closer look at the data reveals that most of the budget changes are very small, while a few of them are quite large. The two dependent variables capture the occurrence of such big changes and their directions.

The key independent variable of interest is program type. This variable is constructed based on our data coding according to Wilson’s typology. If both output and outcome of a program are measurable, the program is defined as a production program and the program type variable is coded 0. If only output of a program is measurable, the program is defined as a procedural program and the program type variable is coded 1. If only outcome of a program is visible, the program is defined as a craft program and the program type is coded 2. Lastly, if neither output nor outcome of a program is visible, the program is defined as a coping program and the program type variable is coded 3. Thus, there are 21 production programs, 17 procedural programs, 4 craft programs, and 5 coping programs in our sample.

In consistency with the studies on the PART programs such as Gallo and Lewis (2012), Lavertu and Moynihan (2012), we also control the categories of PART programs assigned by the OMB, the PART rating results of each program, the ideology of the agency managing the program, and a dummy variable indicating the year when the program was assessed in the PART system. According to the OMB, there are seven categories of PART programs, including block/formula grant, capital assets and service
acquisition, competitive grant, credit, direct federal, research and development, and regulatory programs. Each of these categories is indicated by a dummy variable in our model where 1 means the program is of a particular category.

For the PART rating results, we include an index ranging from 0 to 4. Specifically, 0 means ineffective, 1 means results not demonstrated, 2 means adequate, 3 means moderately effective, and 4 means effective. Thus, the higher the value of this variable is, the more positive PART evaluation a program has. According to Gilmour and Lewis (2006), this variable should be positive associated with budget increase. Agency ideology is a dummy variable which equals 1 if the agency managing the program is conservative and 0 if the agency is liberal. The data on agency ideology is from the estimation of federal agencies by Clinton and Lewis (2008). Since the PART covered only a fraction of federal programs every year from 2002 to 2008, the year when a program is evaluated may capture some heterogeneity among them. We tried to include all the six year dummies and found only the year dummy for 2008 matters. Given our focus of dependent variable on the budget change from 2008 to 2009, this makes sense and we include only the year dummy indicating whether a program is assessed by the PART in 2008. Table 2 shows the descriptive statistics for all the dependent and independent variables in our model.
Table 2 Descriptive Statistics for the Dependent and Independent Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs.</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big budget change</td>
<td>47</td>
<td>.2979</td>
<td>.4629</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Direction of budget change</td>
<td>47</td>
<td>.1064</td>
<td>.8656</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>Program type</td>
<td>47</td>
<td>.8511</td>
<td>.9775</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Block/formula grant</td>
<td>47</td>
<td>.2553</td>
<td>.4408</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Capital assets and service</td>
<td>47</td>
<td>.0851</td>
<td>.2821</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>acquisition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competitive grant</td>
<td>47</td>
<td>.1277</td>
<td>.3373</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Credit</td>
<td>47</td>
<td>.0213</td>
<td>.1459</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Direct federal</td>
<td>47</td>
<td>.3617</td>
<td>.4857</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Research and development</td>
<td>47</td>
<td>.0213</td>
<td>.1459</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Regulatory</td>
<td>47</td>
<td>.1277</td>
<td>.3373</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Assessed in 2008</td>
<td>47</td>
<td>.1064</td>
<td>.3117</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Agency ideology</td>
<td>47</td>
<td>.4681</td>
<td>.5044</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>PART rating</td>
<td>47</td>
<td>2.3830</td>
<td>1.0744</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

Results

To explore the effects of program type on the size and direction of budget change from the actual spending to budget request between 2008 and 2009, we run an OLS regression and an ordered probit regression, respectively. Table 2 shows the results of the first model. The overall model is statistically significant. Coping program shows a positive and statistically significant (p=0.1) effect on big budget change. Since the base category for program type is production program, this indicates that coping program is more likely to experience big budget change from actual budget spending in 2008 to budget request in 2009. This lends preliminary support to our hypothesis that production programs are more likely to have stable budget change, which is at least the case compared to the coping programs.
Table 3: The Effects of Program Type on Size of Budget Change

<table>
<thead>
<tr>
<th>Explanatory variables</th>
<th>Big budget change (0/1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procedural program (0/1)</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>(0.74)</td>
</tr>
<tr>
<td>Craft program(0/1)</td>
<td>-0.13</td>
</tr>
<tr>
<td></td>
<td>(-0.54)</td>
</tr>
<tr>
<td>Coping program(0/1)</td>
<td>0.47*</td>
</tr>
<tr>
<td></td>
<td>(2.00)</td>
</tr>
<tr>
<td>Block/formula grant(0/1)</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>(1.23)</td>
</tr>
<tr>
<td>Capital assets and service acquisition(0/1)</td>
<td>0.89*</td>
</tr>
<tr>
<td></td>
<td>(1.93)</td>
</tr>
<tr>
<td>Competitive grant(0/1)</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>(1.68)</td>
</tr>
<tr>
<td>Direct federal(0/1)</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>(1.01)</td>
</tr>
<tr>
<td>Research and development(0/1)</td>
<td>-0.45</td>
</tr>
<tr>
<td></td>
<td>(-0.74)</td>
</tr>
<tr>
<td>Regulatory(0/1)</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>(0.35)</td>
</tr>
<tr>
<td>Agency ideology(0/1)</td>
<td>0.45***</td>
</tr>
<tr>
<td></td>
<td>(3.11)</td>
</tr>
<tr>
<td>Assessed in 2008(0/1)</td>
<td>-0.26</td>
</tr>
<tr>
<td></td>
<td>(-1.24)</td>
</tr>
<tr>
<td>PART rating</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>(0.46)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.51</td>
</tr>
<tr>
<td></td>
<td>(-1.17)</td>
</tr>
</tbody>
</table>

Observations      47  
R-squared          0.468

Note: OLS regression, adjusted R-squared: 25.04%, test of the regression: F (12, 34) = 2.28, t-statistics in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Two of the control variables also reach statistical significance. For the variables of program categories assigned by the OMB, credit program is used as a base category. At p=0.1 level, capital asset and service acquisition programs are more likely to face big budget change from actual spending in 2008 to budget request in 2009 than the credit program. Agency ideology has a positive and statistically significant effect (p=0.05) on
big budget change for a program. This implies that programs administered by conservative agencies are more likely to have a big budget change than those managed by liberal agencies are.

Table 3 shows the results of the ordered probit model. The overall model is statistically significant. Note that two observations are completely determined in the sample and thus the standard errors are questionable. Specifically, two agencies report requested amounts of 0 in 2009, which is unlikely and probably due to data reporting error. The results are not strongly affected by collinearity, as the highest collinearity observed, about 65% value of R squared for an explanatory variable on the others, is not particularly high.
<table>
<thead>
<tr>
<th>Explanatory variables</th>
<th>Direction of budget change (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procedural program (0/1)</td>
<td>-0.72 (-1.49)</td>
</tr>
<tr>
<td>Craft program (0/1)</td>
<td>-0.95 (-1.27)</td>
</tr>
<tr>
<td>Coping program (0/1)</td>
<td>-1.94** (-2.54)</td>
</tr>
<tr>
<td>Block/formula grant (0/1)</td>
<td>-1.15* (-1.94)</td>
</tr>
<tr>
<td>Capital assets and service acquisition (0/1)</td>
<td>-1.80** (-2.18)</td>
</tr>
<tr>
<td>Competitive grant (0/1)</td>
<td>-1.73** (-2.44)</td>
</tr>
<tr>
<td>Credit</td>
<td>4.88 (0.00)</td>
</tr>
<tr>
<td>Research and development (0/1)</td>
<td>6.64 (0.01)</td>
</tr>
<tr>
<td>Regulatory (0/1)</td>
<td>1.19 (1.31)</td>
</tr>
<tr>
<td>Agency ideology</td>
<td>-0.73 (-1.45)</td>
</tr>
<tr>
<td>Assessed in 2008 (0/1)</td>
<td>-1.06* (-1.71)</td>
</tr>
<tr>
<td>PART rating</td>
<td>-0.18 (-0.95)</td>
</tr>
<tr>
<td>Constant cut 1</td>
<td>-1.63* (-1.95)</td>
</tr>
<tr>
<td>Constant cut 2</td>
<td>-2.56*** (-2.91)</td>
</tr>
</tbody>
</table>

Observations: 47

Note: Ordered probit regression, LR chi2(12)=23.07, Prob > chi2=0.0271, z-statistics in parentheses, *** p<0.01, ** p<0.05, * p<0.1

The production agency is again chosen as a base category. We find that coping program has a negative and statistically significant effect (p=0.05). This means that coping program is less likely to experience a budget increase from the actual spending in 2008 to budget request in 2009 than the production program is. While the first model shows that the coping program is more likely to have big budget change than the production program is.
programs, this result adds more information on the direction of the difference. Taken both results together, we can tentatively conclude that coping programs are more likely to have big, negative budget change from actual spending in 2008 to budget request in 2009 than the production programs are.

Four of the control variables have statistically significant effects on the sign of budget change. Compared to the direct federal programs, the block/formula grant, capital asset and service acquisition, and competitive grant are more likely to have negative budget change from the actual spending in 2008 to budget request in 2009. These effects are statistically significant at p=0.05 level. Note that the dependent variables in both models measure the relative size and direction of the budget change from actual spending in 2008 to the budget request rather than the actual spending in 2009. In the budget cycles, the budget request in 2009 was made by Bush Administration and was approved by the Congress in June, 2008. This is well before the American Recovery and Reinvestment Act of 2009 (ARRA) took place. Particularly, combined with the result from the first model, it shows that federal capital spending was projected to have a big decrease before the ARRA was signed into law in February 2009. In addition, the programs assessed by the PART in 2008 are more likely to have budget decrease and the effect is statistically significant at p=0.1 level.

6. Discussion and Outlook

The main objective of this study is to empirically test Wilson’s typology by linking it to one of the potential significant consequences-budget allocation. While it seems to make theoretical sense, it is challenging to implement in practice. The key difficulty is to
develop a decent method to classify the federal programs according to Wilson’s theory framework. We currently use qualitative coding to put PART programs into different categories and conduct a preliminary analysis with the coding results. The present coding rules are not sufficiently clear as indicated by the relative low inter-rater consistency of the coding results. Nonetheless, the preliminary results look promising, as it seems to indicate that there are differences of the size and direction of budget change from actual spending to budget request as predicted by the theoretical arguments. We have to interpret the results cautiously because of the coding rules and small sample results. This is also an incomplete test since we do not detect any difference between production programs and procedural programs or craft programs, which will be subject to future examination.

The second practical challenge is to discern discretionary programs from all the 1016 PART programs. While our current method of matching account names to program names allows us to make some progress, we may need to seek for cleaner ways to do so. There are a fairly large proportion of PART programs which we have to put it in the “mixed” categories but can be actually discretionary programs. The third limitation is that we have not located the budget appropriations for each of PART program between 2002 and 2007. Had this data been available, we could have tested the third and fourth hypotheses on the relation between program type and patterns of budget change over time.

Therefore, there are four ways to improve the results of the current preliminary analysis. The first and foremost one is to clarify and refine the coding rules on the basis of our present coding method. For example, there are some guidelines from the OMB on what
research designs are considered as higher quality evidence of program performance—experiment or quasi-experiment research is weighted more than correlation or case studies. This may help us to distinguish performance measures of different quality. We may also specify, conditional on quality, the number of performance measures as threshold for “high” or “low” level. Once we have done this, we can at least enlarge the sample size to the more than 200 discretionary PART programs we have already identified. Second, the direction information on which PART programs is discretionary, or the BEA category of these programs would be extremely helpful for us to expand our sample size. Third, the data on historical budget spending for the PART programs between 2002 and 2007 would be critical to the empirical tests of the last two hypotheses. Last but not least, radically different from the former three ways, an empirical test of Wilson’s typology may be still feasible if we re-focus on its other consequences rather than budget allocation. This might be a future direction to be explored if all the former three ways do not work.
References


Robinson, Caver, Meier, and O’Toole. 2007. “explaining policy punctuations” American Journal of Political Science


