Measuring Gubernatorial Budgetary Power: A New Approach

Yanna Krupnikov¹ and Charles Shipan²

Abstract
Numerous scholars have considered the relationship between gubernatorial power and political outcomes. In fact, gubernatorial power has been used as a key explanatory factor in analyses of topics such as gubernatorial approval, divided government, regulation, and even individual political behavior. The key to these studies is the precision with which scholars can measure gubernatorial power and many such studies rely on the Formal Powers Index (FPI)—a measure maintained by Beyle. In this article, we reconsider these commonly used power scores in three parts. First, we argue and show that FPI suffers from a key measurement error that is particularly problematic in analyses of time-series data. Second, we present a new approach to estimating gubernatorial power and explain how this approach deals with the measurement errors in the FPI. Finally, we use our new scores to replicate a study that originally relied on the FPI to analyze the effect of gubernatorial power. Given the prevalence of the FPI in the existing literature, our results have key implications for the study of the effects of gubernatorial power.

Keywords
gubernatorial power, measurement

Introduction
Governors vary widely in the amount of formal power that they wield, especially in their dealings with legislatures. In some states, powerful governors can dominate legislatures that are understaffed and that do not meet very often. In other states, however, governors find themselves constrained by the legislature when they attempt to

¹Northwestern University, Evanston, IL, USA
²University of Michigan, Ann Arbor, USA

Corresponding Author:
Yanna Krupnikov, Department of Political Science, Northwestern University, Scott Hall, 601 University Place, Evanston, IL 60208, USA
Email: yanna.krupnikov@northwestern.edu
create public policy. These differences in gubernatorial power are examples of institutional conditions that can affect outcomes at the state level. Indeed, the governor’s formal power—often relative to that of the state legislature—has been identified as a key explanatory factor in numerous behavioral and institutional phenomena, ranging from the way individuals attribute responsibility for political outcomes (Rudolph 2003) to a governor’s influence over state agencies (Dometrius 2002).

Much of this research relies on the same measure of gubernatorial power: the Formal Powers Index (FPI) that was originally created by Schlesinger (1965) and then streamlined and regularly updated by Thad Beyle.¹ The current version of the FPI (Beyle and Ferguson 2008) details a governor’s power over various institutional areas: budget, veto power (VP), separately elected officials (SEP), tenure potential (TP), gubernatorial party control (PC), and appointment power (AP). These ratings, which are widely accepted and used within political science, provide state-level power scores for governors over nearly a 50-year period, from 1960 to 2007. The ratings have appeared in numerous articles over the past 10 years, published in journals such as the American Political Science Review, American Journal of Political Science, The Journal of Politics, State Politics & Policy Quarterly, and Legislative Studies Quarterly. In addition, these scores have been included in countless conference papers and even utilized by the mainstream media in newspapers such as the Chicago Tribune and the New York Times.

The FPI is ubiquitous, and for good reason: These are useful, simple ratings that draw on legitimate differences in gubernatorial power across states. Nonetheless, this ubiquity is a source for concern when scholars rely on these scores to make longitudinal claims. In this article, we consider the underlying issues with using the FPI in a longitudinal manner. Specifically, we will focus our discussion on the measure of budgetary power, which is a key component of gubernatorial power more generally (see, for example, Alt and Lowry 2000; Barrilleaux and Berkman 2003).

Using these ratings to make longitudinal claims is problematic due to the structure of the index. In particular, updates over time to the power scores have included and accounted for different institutional factors. As a result, a state that has an identical power score at different points in time could, in fact, have different institutional structures at those times. Conversely, this may mean that a state whose gubernatorial power structure remained unchanged may receive a different power score in each update due simply to a change in the way the scores are created. This effect is particularly dramatic in the budgetary power scores, as changes in these scores between 1988 and 1994 suggest that a vast majority of states underwent a massive institutional change in gubernatorial power over the budget. Although Beyle has noted these coding shifts and changes to the measure, research has nonetheless continued to treat this measure as longitudinal.²

In this article, we highlight why scholars should exercise caution when measuring budgetary power over time. We proceed as follows. First, we discuss prior scholarship that has relied on the FPI. Second, focusing directly on budgetary power, we explain the dangers of treating this measure as longitudinal. Third, given the analytical problems associated with using the FPI in a longitudinal manner, and given that scholars are
interested in research questions that require the measurement of budgetary power over time, in this section, we introduce a new measure of budgetary power, one better suited to analyses that require a consideration of institutional conditions over time. Finally, we engage previous research that relies on the FPI as an explanatory variable and reestimate the tests using our new measure. In the end, we offer a general caution to scholars who rely on the budget power (BP) scores in longitudinal analyses.

**Existing Literature**

Over the past decade, scholars have incorporated the FPI—either as controls or as key explanatory factors—into investigations of a variety of state-level phenomena. Indeed, the breadth of the analyses relying on the FPI is notable, ranging from studies of individual-level political perceptions to issues of gubernatorial power and state policy. For example, in an innovative analysis of the way voters perceive state-level outcomes, Rudolph (2003) merges individual-level survey data with the FPI to analyze the way citizens attribute responsibility. Relying on a different approach, in her study of gubernatorial success, Ferguson (2003) incorporates the scores in their raw format, although these scores do not reach significance. In addition to these two examples, Table 1 highlights the diversity of this research using a sample of recently published papers.

<table>
<thead>
<tr>
<th>Broad topic</th>
<th>Sample-dependent variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Political behavior</td>
<td>Perception of responsibility for outcomes (Rudolph 2003)</td>
</tr>
<tr>
<td>Bureaucracy</td>
<td>Confidence in state government (Kelleher and Wolak 2007)</td>
</tr>
<tr>
<td>Divided government</td>
<td>Statutory control (Huber, Shipan, and Pfahler 2001)</td>
</tr>
<tr>
<td>Legislative behavior</td>
<td>Change in state revenue (Alt and Lowry 2000)</td>
</tr>
<tr>
<td>State policy</td>
<td>Conflict between governor and legislature (Clarke 1998)</td>
</tr>
<tr>
<td>Gubernatorial politics</td>
<td>Statehouse leadership power level (Clucas 2007)</td>
</tr>
<tr>
<td></td>
<td>Participation in interstate compacts (Bowman and Woods 2007)</td>
</tr>
<tr>
<td></td>
<td>Environmental regulation (Daley, Haider-Markel, and Whitford 2007)</td>
</tr>
</tbody>
</table>

Note: FPI = Formal Powers Index. This table represents only a sample of the articles published that rely on the FPI. A more complete list included in the online Appendix 1. This online appendix focuses on those articles and conference papers that rely on the FPI longitudinally, but for informational purposes includes a fuller listing of articles and conference papers that rely on the FPI cross-sectionally.

Although some of these articles are limited to data from just 1 year, others have relied on gubernatorial power as an explanatory factor in longitudinal analyses. Miller (2006), for example, analyzes incremental and nonincremental policy making covering the years 1980 through 1998, and relies on the index over this time period. Oxley and Fox (2004) include the measure in an analysis of variation in the number of women in state executive offices over the years 1979 to 1998. Similarly, Woods and Bowman (2011) use the measure to consider data from 1960 to 2000 to analyze state compact participation.
In addition to using the FPI to analyze longitudinal data, scholars also rely on it to make judgments about the governor’s power over time. Svoboda (1995), for example, writes “with respect to state executive reforms, regardless of which area one reviews, tenure, appointment power, veto power, or budget making, it is clear that gubernatorial power has increased dramatically within the past twenty years” (137).

Given the frequency with which these scores are used—and more importantly, the frequency with which these scores are used in a longitudinal manner—we turn now to a consideration of the underlying structure of the FPI and how this structure may affect analyses that attempt to trace relationships over time.

**FPI**

The FPI has undergone numerous changes over the past several decades. The earliest version of the index (Schlesinger 1965) included relatively few factors comprising gubernatorial power, focusing on a governor’s TP, VP, AP, and budget creation power. Although this index provided a basis for coding gubernatorial power, scholars questioned its construction (Dometrius 1979). Over the years, this index was revised and updated to include additional facets of a governor’s formal power (Beyle 2007, latest update). These modifications to the FPI offered a new and creative summary of gubernatorial power in the states over time. Currently, the FPI, as coded and maintained by Beyle (2007), measures gubernatorial power in six distinct areas of a governor’s institutional power. Indeed, Beyle’s version of the FPI is the lone gubernatorial power index that spans numerous years. The institutional power areas it includes are as follows:

1. **SEP** assesses the procedures for gubernatorial election.
2. **TP** considers the length of a governor’s term.
3. **PC** looks at the position of the governor’s party in the legislature.
4. **AP** focuses on the governor’s ability to appoint officials in various key areas of governmental administration.
5. **VP** deals directly with the governor’s ability to veto legislation.
6. **BP**—the focus of this article—considers a governor’s role in proposing and implementing the executive budget.

These six factors can either be averaged into a composite index (the FPI) or used separately to account for different types of power. Some scholars, for example, have used these scores to create modified versions of the index (e.g., Dometrius 2002 calculates the average of the components without including PC). Others rely on only one single component of the index (e.g., Rudolph 2003 relies only on the budgetary power score).

Although the information codified by the different facets has shifted over the years, since 1980 the scores within the FPI have ranged along a 1 to 5 scale, where 1 indicates the governor has little power and 5 means the governor is highly powerful. For ease of discussion, in this article, we refer to states with a score of 1 as “low-power” states (states in which the power of the governor is low relative to that of the legislature), states with a score of 5 as “high-power” states (states in which the power of the
Figure 1. FPI factors: Comparison from 1980 to 2007: (a) all FPI scores over time, (b) percentage of “shared-power” states, 1980–2007
Note: FPI = Formal Powers Index.

governor is high relative to that of the legislature), and states with a score of 3 as “shared-power” states (states in which the power of the governor is generally shared with the legislature or balanced with that of the legislature).  

In Figure 1a, we compare the average score, across all states, from 1980 to 2007 on each of the institutional power areas. For clarity, we omit the two most recently
created scores, SEP and PC, because these scores are a more recent addition to the index and because the coding approach to these two factors has remained stable since their initial inclusion in the index. As the figure shows, although most of the governor’s powers remain more or less stable over time, the average governor’s budgetary power appears to decline sharply in 1994 and then remains at this level.

We can also consider these scores in a different manner, shown in Figure 1b. Comparing the percentage of shared-power states in each of the institutional categories within the FPI again shows shifts in institutional structure across states. This approach suggests that the dramatic change in the budget variable shown in Figure 1a was a function of an increase in the number of shared-power states; in the other three categories—appointment, veto, and tenure—again the numbers remain relatively stable.

The change apparent in Figures 1a and 1b speaks to shifts in the creation of the index. Much of the FPI is drawn from data published by The Book of the States (BOTS), and as Beyle’s discussion of the FPI suggests, his definitions of the various categories—and the information used to compile these categories—change over the years. While scholars should heed Beyle’s warning when relying on any component of the index, we demonstrate below that these coding changes have profound implications for the budgetary power variable.

Budget: Shared Power

Beyle’s measure of budgetary power focuses on gubernatorial responsibility for the state budget. As Figure 1 shows, the distribution of gubernatorial budgetary scores shifts dramatically over time. This change is particularly stark when we focus on two categories within this measure: the shared-power category (i.e., BP = 3) and the high-power category (i.e., BP = 5). In particular, between 1988 and 1994 the majority of states shift out of the high-power category into the shared-power category. Overall, in the 1980s the percentage of shared-power states hovers between 0% and 2%, with no states having a shared-power system in 1988 and only North Carolina having one in 1980. In 1994, however, this number jumps to 80%, where it remains in 1998 (and in all more recent updates of the codes). At the same time, we see a sharp decline in high-power states. In 1980 and 1988, nearly 88% of states fall into the high-power category; by 1994, this number is 4%.

This change in scores likely lies in the relationship between the codes and their original source data, the BOTS. In Table 2, we first present the criteria Beyle uses to assign each state a score and subsequently turn our focus to the raw information in the BOTS.

In Beyle’s system, in 1980 and 1988 a high-power state was one where the governor had full responsibility for the budget, whereas a shared-power state was one where the governor shared budgetary responsibility with the legislature. In 1994, however, as Beyle incorporates additional factors into the scores, the coding approach shifts. To begin with, a high-power state is one in which the governor has full control over the budget and the legislature may not increase the budget. In addition, the shared-power state in 1994 is redefined as one where a governor has full responsibility over the budget and the legislature has unlimited power to change the budget. Both these
definitions are quite different from those used earlier. The most recent updates to the FPI retain the 1994 structure.

To consider how these coding changes affected the distribution of states, we turn to the BOTS. In 1980 and 1988, the BOTS reported information about whether the governor had full responsibility over the budget or whether he shared responsibility with another body. Drawing on this information, Beyle used the criteria spelled out in Table 2 to create the budgetary power scores for 1980 and 1988, which show that the vast majority of governors have full responsibility over the budget. Notably, however, during these years the BOTS also contained some information about legislative powers. In 1980, for example, the BOTS directly identified states where the legislature had an unlimited power to change the budget. This information, however, is not a part of Beyle’s (1980) coding scheme (see Table 2), and thus is not taken into account in the coding of each state’s BP score.

In 1994, Beyle began to incorporate the legislature into what he coded as Category 3 (and what we refer to as the “shared-power category”), to evaluate the power of the governor relative to the legislature. This means that the 1994 codes incorporate previously unused (but available) BOTS information to create the BP measure. How would the 1980 scores be affected if Beyle used the same coding rules then as in

| 1 (low power) | Governor shares responsibility with several others with independent sources of strength | Governor shares responsibility with several others with independent sources of strength | Governor shares responsibility with other elected official(s), and legislature has unlimited power to change executive budget |
| 2 | Governor shares responsibility with another popularly elected official | Governor shares responsibility with other elected officials | Governor shares responsibility; legislature has unlimited power to change executive budget |
| 3 (shared power) | Governor shares responsibility with legislature | Governor shares responsibility with legislature | Governor has full responsibility; legislature has unlimited power to change executive budget |
| 4 | Governor shares responsibility with a civil service appointee or with person appointed by someone else | Governor shares responsibility with civil servants or other person appointed by someone else | Governor has full responsibility; legislature can increase by special majority vote or subject it to item veto |
| 5 (high power) | Governor has full responsibility | Governor has full responsibility | Governor has full responsibility; legislature may not increase executive budget |
1994? We consider this and use the 1981 BOTS to trace how many states would have been coded as shared power had Beyle applied the 1994 coding scheme in 1980. Strikingly, we find that by relying on the 1994 coding rules, all the states that were coded as high power in 1980 would have been considered shared power under Beyle’s redefinition.

Considering the budget score more broadly, the shift in categories discussed above is the result of a more general issue, namely, only starting in 1994 does the FPI incorporate legislative activity into the measure of budgetary power. In 1980, for example, a high-power state is simply one where the governor had full responsibility over the budget. In 1994, however, a high-power state is one where the governor has full responsibility over the budget and the legislature cannot increase the budget. In 1980 and 1988, a low-power state is one where the governor shared budget creation responsibility with “other independent sources of strength.” In 1994, however, a low-power state is one where the governor not only has to share budget-making responsibility but the legislature also has an unlimited power to change the budget.

In sum, although these coding changes may seem reasonable, they have consequences for scholars who wish to use the index in longitudinal analyses, further demonstrating that scholars should heed Beyle’s admonition. As our discussion shows, the scores in 1994 and beyond are based on different pieces of information from the scores from 1980 and 1988, and as a result what appear to be changes in institutional structure over time are artifacts of changes to the coding rules and in the amount and use of available information.

Given these problems—and given that numerous research questions require the use of longitudinal data—ideally, we would attempt to make the scores consistent over time by using the most recent coding rules to recode the 1980 and 1988 years. Unfortunately, however, although the BOTS has some information on legislatures (e.g., such as that taken into account when reconsidering the shared-power category), it does not have all the data we would need for all years. Specifically, the BOTS only begins to include extended information about legislative budgetary activities in the early 1990s. Thus, although we can recode the shared-power category, we require more information than the BOTS offers to reconsider other levels of power. In short, correcting the budgetary power scores is not as simple as changing a 5 to a 3. To provide a more consistent portrait of budgetary power in states over time, we reconsider how to code gubernatorial BP and propose a new approach.

A New Measure of Budgetary Power

Beyle obtains much of the data for the budgetary power score from the BOTS, which in turn gathers information about state budgetary activities from the National Association of State Budget Officers (NASBO). What BOTS presents as basic distinctions in budgetary responsibilities thus stem from larger and more detailed NASBO studies called “Budgetary Processes in States.” These NASBO studies were conducted in 1981, 1987, 1989, 1992, 1995, 1997, 1999, 2002, and most recently in 2008, a time period similar to the existing FPI.11 The NASBO studies combine constitutional information with surveys of state budget officers.
Although here we will be using NASBO scores to track gubernatorial power, scholars have put the NASBO data to good use in the past for other purposes, including such topics as budgetary outcomes (Ryu et al. 2008) and state appropriations (Abney and Lauth 1998). For our purposes, NASBO data are beneficial because they allow us to track the following institutional factors: (1) the governor’s ability to spend federal funds without legislative approval, (2) the governor’s line-item VP, (3) the governor’s power to reorganize departments related to the budget without legislative approval, and (4) the governor’s ability to reduce the budget without legislative approval. We combine these four factors with a fifth factor from BOTS: whether the governor has authority for budget preparation. We base our proposed coding of gubernatorial budgetary power on these five factors. We refer to budgetary power scores included in the Beyle’s coding as “FPI budgetary power score” and to the budgetary power score we propose here as the “NASBO score.”

There are several benefits to using the information collected by NASBO. First, the BOTS bases part of its categorization of gubernatorial power on NASBO data. As a result, turning to NASBO will offer a clearer foundation for considering budgetary power. Second, to the extent that the operationalization of abstract concepts benefits from the inclusion of different indicators, our definition of budgetary power will be made stronger by the fact that NASBO data include a number of different areas of power. Most importantly, however, compared with the FPI the NASBO scores present a more longitudinally consistent approach to measuring budgetary power in the states. Given that scholars are interested in tracing the effects of various factors over time, our new score offers a consistent means to account for gubernatorial power.

The Codes

We code each of the NASBO factors on a simple 3-point scale. A code of 1 means that the governor has the power to take the action described in our factor under the conditions described in our factor. This means, for example, that a governor who can spend federal funds without legislative approval receives a score of 1 on this particular factor. Conversely, a score of 0 means that a governor cannot take an action under the conditions described in the factor. A middle score of 0.5 means that the governor has partial or conditional power. This score was used when a governor generally did have the power to take an action under the conditions described in the factor, but with limitations. In 1992, for example, the Pennsylvania governor had the power to reorganize departments related to the budget during a legislative recess, but only within agencies. We code the BOTS budgetary preparation factor on a similar scale: Cases where the governor holds sole power over budgetary preparation are coded as 1, and cases where the governor shares that power are coded as 0.

After the initial coding, we then summed all five factors into an index. In this index, a score ranging from 4.5 to 5 means high power over the budget; a score of 1 indicates low power over the budget. Tracing our results by year shows relative consistency (Figure 2). Although average budgetary power does change over the years, it remains
fairly stable over time, hovering just over a 3. Unlike the FPI budgetary score, we see no sudden and dramatic shifts in budgetary power in the early 1990s.

As a next step, we directly compare our new score with the FPI budgetary score. Given the slight differences in the timing between Beyle’s updates to the FPI and NASBO surveys occurred on slightly different years, we matched the years as follows: FPI 1980 to NASBO 1982, FPI 1988 to NASBO 1987, FPI 1988 to NASBO 1989; FPI 1994 to NASBO 1992, FPI 1994 to NASBO 1995; FPI 1998 to NASBO 1997, FPI 1998 to NASBO 1999; FPI 2001 to NASBO 2002; FPI 2007 to NASBO 2008. Although there was a 2005 Beyle update, including this update rather than the 2004 update does not change the figure, as the average budget score is the same in 2004 and 2005.

Reanalyzing Previous Work

Given that the NASBO scores proposed here are distinctly different from the FPI budgetary score, will relying on these scores lead to different conclusions about

Figure 2. Comparison of the FPI and NASBO scores
Note: FPI = Formal Powers Index; NASBO = National Association of State Budget Officers. As Beyle’s updates to the FPI and NASBO surveys occurred on slightly different years, we matched the years as follows: FPI 1980 to NASBO 1982, FPI 1988 to NASBO 1987, FPI 1988 to NASBO 1989; FPI 1994 to NASBO 1992, FPI 1994 to NASBO 1995; FPI 1998 to NASBO 1997, FPI 1998 to NASBO 1999; FPI 2001 to NASBO 2002; FPI 2007 to NASBO 2008. Although there was a 2005 Beyle update, including this update rather than the 2004 update does not change the figure, as the average budget score is the same in 2004 and 2005.
political relationships? In part, this will best be answered over time, as scholars compare the two measures when conducting new studies or revisiting earlier analyses using the new measure. Here, we start this process by reassessing one of the numerous studies published in *State Politics & Policy Quarterly* (and elsewhere) that uses the FPI as an explanatory variable in a longitudinal analysis: Clucas (2007), which relies on the FPI to analyze data from 1981 to 1995.

We present our replication in three parts. First, we present the original results. Second, we present the results obtained by reestimating the original model with only FPI budget score. Third, we present the results of a model that relies on our new NASBO codes.

### Replication

In “Legislative Professionalism and the Power of State House Leaders,” Clucas (2007) assesses the crucial topic of power in legislatures. More specifically, he analyzes how well three theories of legislative organization—partisan (Cox and McCubbins 1993), distributive (Brady 1988), and career based (Maddox 2005)—can explain state-to-state variation in the power of legislative leaders. To make this comparison, Clucas relies on a novel set of surveys of state legislators, conducted in 1981 (Francis 1989) and 1995 (Carey, Niemi, and Powell 2000), that allow him to measure the extent to which state legislators themselves perceive legislative leaders to be powerful forces within the legislature. As these are two separate surveys, Clucas estimates two models, one for 1981 and one for 1995. His approach provides a useful test for our purposes, as it straddles Beyle’s (1994) coding change of the FPI budgetary variable.

With these measures of relative power serving as dependent variables, Clucas’s (2007) independent variables are designed to tap into the incentive structures in the legislatures and legislators’ electoral motivations. Specifically, his insight is that different

### Table 3. Budgetary Power Comparison: NASBO and FPI

<table>
<thead>
<tr>
<th>NASBO year</th>
<th>FPI year</th>
<th>NASBO (4–5) FPI (5)</th>
<th>NASBO (2–3.5) FPI (3)</th>
<th>NASBO (0–1.5) FPI (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1987</td>
<td>1988</td>
<td>46.94</td>
<td>87.9</td>
<td>48.98</td>
</tr>
<tr>
<td>1989</td>
<td>1988</td>
<td>54.0</td>
<td>87.9</td>
<td>54.0</td>
</tr>
<tr>
<td>1992</td>
<td>1994</td>
<td>44.0</td>
<td>4.0</td>
<td>54.0</td>
</tr>
<tr>
<td>1995</td>
<td>1994</td>
<td>46.0</td>
<td>4.0</td>
<td>52.0</td>
</tr>
<tr>
<td>1997</td>
<td>1998</td>
<td>50.0</td>
<td>4.0</td>
<td>48.0</td>
</tr>
<tr>
<td>1999</td>
<td>1998</td>
<td>50.00</td>
<td>4.0</td>
<td>48.00</td>
</tr>
<tr>
<td>2002</td>
<td>2001</td>
<td>46.00</td>
<td>4.0</td>
<td>54.00</td>
</tr>
<tr>
<td>2008</td>
<td>2007</td>
<td>48.98</td>
<td>4.0</td>
<td>46.94</td>
</tr>
</tbody>
</table>

Note: NASBO = National Association of State Budget Officers; FPI = Formal Powers Index. Cells are percentage per category.
significance patterns among these independent variables would reinforce different theories of legislative organization. First, positive and significant coefficients on the Professionalization and Majority Party Advantage variables would support the partisan theory. Second, negative and significant coefficients on these same variables would suggest the distributive theory. Finally, results where the Majority Party Advantage variable has no effect on the distribution of power would support the career theory. Clucas’s statistical tests show that the Professionalization and Majority Party Advantage variables are positive and significant, leading him to conclude that his analysis “provides support for the partisan theory of legislative organization in explaining the distribution of power in American legislatures” (15).

Clucas (2007) uses the composite version of the FPI—which is the average of all the separate components of the index—as a control. He reports that gubernatorial power does not have a significant effect on the power of legislative leaders, and the inclusion of this index does not affect the relationship between the dependent variable and other independent variables. We reconsider Clucas’s results in several steps, as outlined above. First, we replicate his original findings. Second, we replicate his findings with only the FPI budgetary power variable. Next, we reconsider his results with the new NASBO variable. Finally, we estimate a full model that controls for all the components of the FPI separately.

As a starting point, in Table 4, Models 1 and 7, we perfectly replicate the results reported in Clucas’s article (2007, 13). In his article, Clucas notes that these results continue to hold when the FPI is included in the model. In Table 4, Model 2, that is exactly what we find: The inclusion of the FPI does not have a substantive effect on the results for 1981, nor does the index have a significant effect on the distribution of power. In 1995 (Table 4, Model 8), the results are similar, although not exactly the same: including the index changes the significance of the Professionalization variable, although it has no other effect on the model.

As Clucas (2007) controls for gubernatorial power using the composite version of the FPI, next we split the index into its key components and first estimate a model that relies only on FPI budgetary power scores. The results, shown in Table 4 (Models 3 and 9 for 1981 and 1995, respectively), are consistent with those Clucas’s reports: the previously significant variables remain significant in 1981 (Table 4, Model 3), and the Professionalization variable is once again significant in 1995 (Table 4, Model 9). Finally, we replace the FPI budgetary power variable with the NASBO scores we described in the section “A New Measure of Budgetary Power” and estimate the same models. The results, as revealed in Model 4, show a notable change: In 1981, the coefficient on Majority Party Advantage is no longer significant. Given Clucas’s theoretical setup, the change in this variable is suggestive, but before we consider the full implications we conduct one more check.

In the models discussed above, we control only for budgetary power, yet in his analysis, Clucas (2007) aimed to control for gubernatorial power more generally. Indeed, there is nothing in his theoretical framework to suggest that it is budgetary power in particular that should affect the distribution of power in the legislature. To
Table 4. Reanalysis of Clucas (2007; N = 38)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Professionalization</td>
<td>.437** (.206)</td>
<td>.475** (.218)</td>
<td>.437** (.207)</td>
<td>.551** (.226)</td>
<td>.438* (.231)</td>
<td>.633** (.279)</td>
<td>.243* (.136)</td>
<td>.236 (.140)</td>
<td>.269* (.138)</td>
<td>.243* (.138)</td>
<td>.326** (.141)</td>
<td>.266* (.148)</td>
</tr>
<tr>
<td>Career</td>
<td>−.244*** (.077)</td>
<td>−.256*** (.080)</td>
<td>−.242*** (.078)</td>
<td>−.219*** (.086)</td>
<td>−.269*** (.088)</td>
<td>−.241*** (.099)</td>
<td>−.099*** (.046)</td>
<td>−.094*** (.049)</td>
<td>−.117*** (.049)</td>
<td>−.100*** (.049)</td>
<td>−.123*** (.051)</td>
<td>−.089 (.054)</td>
</tr>
<tr>
<td>Career dead end</td>
<td>−.268*** (.095)</td>
<td>−.283*** (.098)</td>
<td>−.281*** (.096)</td>
<td>−.245*** (.101)</td>
<td>−.316*** (.109)</td>
<td>−.284*** (.115)</td>
<td>−.044 (.050)</td>
<td>.041 (.052)</td>
<td>−.047 (.050)</td>
<td>−.045 (.053)</td>
<td>−.054 (.056)</td>
<td>−.045 (.060)</td>
</tr>
<tr>
<td>Outside options</td>
<td>.821*** (.343)</td>
<td>.835*** (.347)</td>
<td>.795*** (.345)</td>
<td>.798*** (.349)</td>
<td>.920*** (.404)</td>
<td>.877*** (.438)</td>
<td>−.084 (1.71)</td>
<td>.083 (1.73)</td>
<td>−.110 (1.72)</td>
<td>−.085 (1.80)</td>
<td>−.209 (2.00)</td>
<td>−.129 (.218)</td>
</tr>
<tr>
<td>Majority party</td>
<td>.275* (.148)</td>
<td>.282* (.150)</td>
<td>.289* (.149)</td>
<td>.259 (.154)</td>
<td>.295* (.155)</td>
<td>.247 (.165)</td>
<td>.310*** (.104)</td>
<td>.298*** (.113)</td>
<td>.345*** (.109)</td>
<td>.312*** (.115)</td>
<td>.343*** (.120)</td>
<td>.274*** (.130)</td>
</tr>
<tr>
<td>Advantage</td>
<td>FPI</td>
<td>−.038 (.061)</td>
<td>−.171 (.186)</td>
<td>−.182 (.198)</td>
<td>.017 (.054)</td>
<td>−.038 (.035)</td>
<td>.017 (.054)</td>
<td>.001 (.028)</td>
<td>.010 (.039)</td>
<td>.012 (.025)</td>
<td>.010 (.031)</td>
<td>.016 (.032)</td>
</tr>
<tr>
<td>FPI BP</td>
<td>.015 (.046)</td>
<td>.036 (.058)</td>
<td>.013 (.025)</td>
<td>.027 (.026)</td>
<td>.019 (.027)</td>
<td>.011 (.028)</td>
<td>.021 (.021)</td>
<td>.016 (.022)</td>
<td>.026 (.031)</td>
<td>.017 (.037)</td>
<td>.008 (.018)</td>
<td>.006 (.020)</td>
</tr>
<tr>
<td>FPI AP</td>
<td>.0169 (.027)</td>
<td>.0024 (.042)</td>
<td>−.005 (.049)</td>
<td>−.013 (.025)</td>
<td>−.027 (.026)</td>
<td>−.011 (.028)</td>
<td>−.021 (.021)</td>
<td>−.016 (.022)</td>
<td>.026 (.031)</td>
<td>.017 (.037)</td>
<td>.008 (.018)</td>
<td>.006 (.020)</td>
</tr>
<tr>
<td>FPI VP</td>
<td>.012 (.025)</td>
<td>.017 (.037)</td>
<td>.016 (.032)</td>
<td>.016 (.032)</td>
<td>.016 (.032)</td>
<td>.016 (.032)</td>
<td>.016 (.032)</td>
<td>.016 (.032)</td>
<td>.016 (.032)</td>
<td>.016 (.032)</td>
<td>.016 (.032)</td>
<td>.016 (.032)</td>
</tr>
<tr>
<td>FPI TP</td>
<td>.008 (.018)</td>
<td>.006 (.020)</td>
<td>.005 (.020)</td>
<td>.005 (.020)</td>
<td>.005 (.020)</td>
<td>.005 (.020)</td>
<td>.005 (.020)</td>
<td>.005 (.020)</td>
<td>.005 (.020)</td>
<td>.005 (.020)</td>
<td>.005 (.020)</td>
<td>.005 (.020)</td>
</tr>
<tr>
<td>FPI SEP</td>
<td>.035** (.020)</td>
<td>.026 (.021)</td>
<td>.026 (.021)</td>
<td>.026 (.021)</td>
<td>.026 (.021)</td>
<td>.026 (.021)</td>
<td>.026 (.021)</td>
<td>.026 (.021)</td>
<td>.026 (.021)</td>
<td>.026 (.021)</td>
<td>.026 (.021)</td>
<td>.026 (.021)</td>
</tr>
<tr>
<td>Constant</td>
<td>−.35 (0.183)</td>
<td>.828 (0.954)</td>
<td>−.114 (.299)</td>
<td>.847 (1.056)</td>
<td>−.054 (.033)</td>
<td>.688** (0.116)</td>
<td>.632*** (0.217)</td>
<td>.809*** (0.161)</td>
<td>.692*** (0.167)</td>
<td>.764 (0.243)</td>
<td>.548 (0.282)</td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>.475</td>
<td>.483</td>
<td>.484</td>
<td>.495</td>
<td>.505</td>
<td>.071</td>
<td>.297</td>
<td>.320</td>
<td>.295</td>
<td>.446</td>
<td>.374</td>
<td></td>
</tr>
</tbody>
</table>

Note: FPI = Formal Powers Index; BP = budget power; NASBO = National Association of State Budget Officers; AP = appointment power; VP = veto power; TP = tenure potential; SEP = separately elected officials; PC = Party Control.

* p < .1, ** p < .05, *** p < .01.
this extent, controlling only for budgetary power may be limiting, and it is possible
that we obtained the new 1981 results simply because we did not control for other types
of gubernatorial power. To address this potential limitation, we again split the FPI into
its key factors, except in the next series of models we include all the factors when
estimating our models. This allows us to include both the full measure of power and
substitute our NASBO scores as a measure of budgetary power.

Estimating the 1981 and 1995 models with all the factors in the FPI included sepa-
rately leads to little change in Clucas’s (2007) key results (see Table 4, Models 5
and 11). Notably, however, in Model 11 we see that in 1995 the FPI budgetary power
variable has a significant effect on the distribution of power—something that does not
happen in 1981. We do not want to make too much of this change in significance given
the differences in data between 1981 and 1995, but nonetheless this result is sugges-
tive. Does the significant result in 1995 reflect the increased importance of the budget-
ary power in 1995, or is the difference in significance levels due to the change in
coding of the variable?

As a final step (Table 4, Models 6 and 12), we estimate a model that controls for all
the components of the FPI, but replaces the FPI budget score with our NASBO scores.
In 1981 and 1995, the NASBO scores have no significant effect on the distribution
of power. The similar behavior of the coefficients on the NASBO scores in both models
adds to the suggestiveness of the results obtained with the FPI budget factor discussed
above. Furthermore, the inclusion of the NASBO scores in Models 6 and 12 leads to
shifts in several of Clucas’s (2007) variables. In 1981, for example, the coefficient on
Majority Party Advantage is no longer significant once the NASBO scores are
included; this factor remains significant in 1995. In contrast, the Career Springboards
variable remains significant in 1981, but is no longer significant in 1995 once we
use the more accurate and appropriate NASBO scores. In short, the inclusion of the
NASBO variables increases the differences between the 1981 and 1995 models. This
is again a suggestive shift, as Clucas writes that a notable part of his analysis is “how
closely the findings for [1981 and 1995] parallel one another” (14).

Furthermore, these new differences between the 1981 and 1995 analyses affect the
conclusions one draws from the tests. Recall that Clucas (2007) ties his independent
variables directly to competing theories of legislative organization: Significant coeffi-
cients on Professionalization and Majority Party Advantage would signal support for
either distributive or partisan theories, whereas significant coefficients only on
Professionalization would signal a theory of career springboards. Indeed, when Clucas
concludes that the distribution of power is best explained by the partisan theory of leg-
islative organization, he does so because the Professionalization and Majority Party
Advantage variables have significant coefficients. The inclusion of the NASBO vari-
able, however, could point one to a different conclusion. Although Professionalization
is important in 1981 and 1995, the same is no longer true for Majority Party Advantage. Given Clucas’s proposed relationship between independent variables and theories of
legislative organization, the inclusion of NASBO scores may lead one to consider that
the 1981 estimates that suggest that power distributions are actually better explained
by the theory of outside career impact than the partisan theory.
Although we do not wish to overstate these results, they suggest that the choice of measures can shift the types of conclusions one can draw from this analysis, reinforcing our focus on the composition of budgetary power scores. Given that Beyle’s coding scheme changes between 1981 and 1995, his power index in 1981 is different from the one constructed in 1995. When the inconsistent FPI budgetary power factor is replaced with the more consistent NASBO scores, controlling for gubernatorial power suggests alternative conclusions about the distribution of power in state legislatures. Indeed, although Clucas’s (2007) original findings focus on the similarities across years, our results suggest that conditions in 1981 differ from those in 1995.

Conclusion

The FPI has clearly filled a need for a measure of gubernatorial power, as numerous scholars have used both the full index and its various components to analyze a multitude of state-level conditions. Beyle clearly deserves significant credit for recognizing this need and then creating and regularly updating such a measure. Here, we have focused on one of the key factors used in this index: the budgetary power of the governor. By considering the underlying structure of this factor, we have shown that the FPI budgetary power has a measurement difficulty: It is not directly comparable across years, particularly due to coding changes that occurred in 1994. These changes limit scholars’ abilities to use these scores longitudinally.

Given that scholars are interested in longitudinal questions, we offer a new set of gubernatorial power scores, based on NASBO surveys. These scores offer a more consistent portrait of these powers, enabling longitudinal tests. A reanalysis of an existing study suggests that replacing the FPI budgetary power score with our score can lead to different results.

These new scores, of course, have their own limits. We recognize, for example, that the original FPI is appealing because it covers a broad range of gubernatorial powers, whereas our NASBO scores, however, consider only one such power. Nonetheless, we hope that our reconsideration of even one of the factors in the FPI suggests that scholars should thoroughly understand the underlying structure of the measures they are including in their analyses.

Acknowledgments

The authors thank Richard Clucas for sharing his data and Nelson Dometrius for helpful comments on a draft of this paper.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.
Notes

1. Scholars who have used this same power measure have used different titles for the index; for example, the index has been called “Beyle’s index of gubernatorial power” (Clarke 1998), “Beyle’s measure” or “Beyle’s scores” (Dometrius 2002). Although, in this article, we specifically focus on Beyle’s version of the power index, we term it the “Formal Powers Index” (FPI) to illustrate the more general nature of this measure. The most recent update, which occurred in 2007 and which we cite as Beyle (2007) in this article, can be found on Beyle’s website: http://www.unc.edu/~beyle/gubnewpwr.html

2. In describing the data, Beyle provides a series of definitions for the different concepts, noting that definitions may differ by years of data collection.

3. In the online Appendix 1, we provide a more comprehensive list of studies using the FPI, specifically identifying articles and conference papers that rely on the FPI longitudinally. For informational purposes, the appendix also includes a listing of articles and conference papers that rely on the FPI cross-sectionally.

4. Schlesinger’s index was largely administrative; Beyle’s directly incorporates the legislature.

5. Prior to 1980, Beyle used a different coding scale. Given this, and given that there were no updates between 1968 and 1980, our analysis focuses on the period from 1980 to the present. Although others have criticized the very components of Beyle’s composite index and the way these components fit together (Dometrius 1979), we do not address this issue.

6. We use these labels for clarity of discussion, not to suggest that the FPI is an interval measure.

7. In 1980, for example, he relied on information from 46 different offices to estimate appointment power (AP); but in 1988, he relied on information from only 6 offices.

8. Reisinger (2008) details some of the measurement inconsistencies in the other factors.

9. The number tops out at 86% in 2005.

10. The concept of “full responsibility” is from The Book of the States (BOTS), representing cases in which the governor has the role of creating the budget. We return to this concept later on.


12. This is based on a discussion in the BOTS and validated using state guidelines and scholarship on budget preparation. Following the BOTS, when a governor has full responsibility over the budget, the preparation of the budget is in the governor’s hands (rather than in the hands of a special commission or legislature), which means that the “governor [is] in the position of actively presenting the agenda of state spending for the coming fiscal period, while the legislature is for the most part reacting to the governor’s budget initiatives” (Lauth 1986). In contrast, in states where the governor does not have full responsibility over the budget, the legislature or special commission takes the lead in preparing the budget. For example, the initial budget creation in Texas is conducted by the Legislative Budget Board.

13. Although it is difficult to capture the “true” power of the governor, our score provides a more longitudinally consistent snapshot than the FPI.
14. Our factors describe a governor’s ability to take an action under specific conditions (legislative approval, certain points in the legislative term), rather than his ability to take that action at all. We describe our coding procedure in the online Appendix 2.

15. Although a score of 0 is technically possible, no state actually has such a score. Thus, we treat 1—the lowest actual score—as the low-power category.

16. The overall correlation between the NASBO scores and the FPI is .25. This is considered a weak correlation (Cohen 1992). As the FPI coding approach changed in 1994, we also consider the correlation between the score before and after that date. Pre 1994, the correlation between NASBO and FPI is .42, whereas post 1994, the correlation is .20, suggesting that the coding changes do alter the relationship between the two scores.

17. Although Clucas (2007) cites only Beyle (2003), he relies on different FPI by year; these different scores match the FPI updates on Beyle’s website.


References


**Bios**

**Yanna Krupnikov** is an assistant professor of political science at Northwestern University. Her research focuses on American politics and has appeared in the *American Journal of Political Science, Journal of Politics* and *Political Communication*.

**Charles Shipan** is the J. Ira and Nicki Harris Professor of Social Science and Chair of the Department of Political Science at the University of Michigan. He has published extensively on the topics of political institutions and policymaking.