FLORIDA STATE UNIVERSITY

COLLEGE OF SOCIAL SCIENCES

STRATEGIC BUDGETING AND BUREAUCRATIC CONTROL

By

JUSTIN ESAREY

A Dissertation submitted to the
Department of Political Science
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Summer Semester, 2008

The members of the Committee approve the dissertation of Justin Esarey defended on June 25th, 2008.

William D. Berry
Professor Directing Dissertation

Tim Salmon
Outside Committee Member

John T. Scholz
Committee Member

Bumba Mukherjee
Committee Member

T.K. Ahn
Committee Member

The Office of Graduate Studies has verified and approved the above named committee members.

I dedicate this dissertation to my wife, Betsy Barre, whose love and support sustained me through the trials of graduate school.

ACKNOWLEDGMENTS

I would like to thank my committee for their advice throughout the creation of this dissertation. In particular, Bill Berry has been an invaluable resource to me and dedicated countless hours to the dissertation's improvement and to my career more generally.

I also thank the FSU Department of Political Science, and especially Dale Smith, Tom Carsey, and Charles Barrilleaux, for their support. It is their commitment to graduate education and research that made this dissertation possible.

Finally, I thank Will Moore for suggesting that I look into principal-agent models for a dissertation project.

This research was supported in part by the National Science Foundation (#0720055) and by Florida State University's University Fellowship.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Who, if anyone, controls the massive government bureaucracy responsible for the elaboration, implementation, and enforcement of laws and executive orders? Typical managerial techniques—screening, reward, and punishment—are hampered by the rules of the civil service system. All but the most senior bureaucrats are hired and promoted according a non-political examination and review system, and most are protected from punishment or termination by tenure. I argue that the executive can control the bureaucracy by creating competition for budget allocations within and between agencies, a process I call *strategic budgeting*. These incentives work under realistic assumptions: highly imperfect monitoring, bounded rationality of executives and bureaucrats, ideological motivations, and professional norms are all a part of the model. I test the predictions of my theory in the laboratory and in data from the American states, finding evidence to confirm that strategic budgeting is an effective strategy for bureaucratic management.

CHAPTER 1

INTRODUCTION

Are bureaucrats independent policy makers, or are they controlled by political masters in the elected branches of government? How is political control accomplished when bureaucrats enjoy extensive career protections against political retaliation? The parts of a typical employment contract—unambiguous goals an employee must meet, objective observation of those goals, and application of rewards or penalties for meeting for failing to meet those goals—are substantially weaker compared to the private sector. All but the most senior civil servants are hired under a non-political career system, protected by a tenure system that makes them difficult to discipline or fire, and have jobs where the quality and quantity of work performed are difficult to objectively or quantitatively observe.

Barriers to control of middle and lower level bureaucrats have led many to conclude that these bureaucrats possess substantial discretion over policy making. Recent literature finds boundaries on this discretion set by the institutionalized power of elected officials to appoint bureaucrats with similar preferences and oversee or reverse their decisions. But inside these boundaries, the literature is unclear about whether and how the bureaucracy will exercise its discretion.

In this dissertation, I argue that the executive's power over the career bureaucracy extends well beyond the power to appoint like-minded agency heads. Even with limited information and weak managerial authority to punish or reward, an executive can control the policies set by an administrative agency by creating competition for budget allocations within and between these agencies.

The intuition behind this idea is simple. Bureaucrats desire larger budgets and greater slack and discretion within those budgets, perks that the executive is uniquely placed to provide due to the central role of the executive in the budgeting process. By strategically manipulating which bureaucratic units—individuals, projects, agency subunits, or even entire agencies—receive budget increases, the executive can gain control over the output of these agencies. I refer to this process as *strategic budgeting*.

The process is easy to understand and implement, and does not require a great deal of fine-grained information about bureaucratic performance in order to be effective. Even if the executive can only state goals in an inexact, qualitative way, and even if compliance with these goals is inexactly measured, the incentives created for compliance can still be effective.

1.1 Strategic Budgeting in a Nutshell

What is strategic budgeting, in straightforward political terms? Strategic budgeting means pitting government agencies (or subunits of an agency) against each other in achievement of a government goal. The goal (or goals) can be as well-defined as increasing revenue, or as ill-defined as improving educational outcome, as long as (i) an elected official (or appointee) has enough information to qualitatively judge performance, even imperfectly, and (ii) competing agencies know the goal that the elected official wants to pursue. Competition winners get budget increases and/or avoid budget cuts. Competition losers face budget cuts.

I expect strategic budgeting to be effective because bureaucrats want bigger budgets, and tournament-style competition to gain desired rewards is known to be an effective motivator of effort (Bolton and Dewatripont, 2005, pp. 316-326). Ironically, bureaucratic desire for ever-expanding budgets (and greater discretion to spend their budgets) is typically cited as an indictment of bureaucracy, a reason why government agencies are less-effective than their market counterparts (e.g., Niskanen, 1971). When budgets are competitively awarded, however, bureaucrats' desire for bigger budgets becomes an asset; it moves them to become more efficient and more responsive to elected authority in order to avoid budget cuts and receive budget bonuses.

The goal of strategic budgeting is to control the policy made by the hundreds of thousands of career bureaucrats who, under the direction of political appointees, put legislation into practice and fulfill executive orders. Though control over appointees is by no means guaranteed, the president can fire appointees at his/her pleasure, and can hire new ones with Senate approval (for the highest positions) or at will (for lower-level appointees).[1] This power allows the president (and sometimes the Senate) to screen a candidate for ideological agreement, and allows the president to terminate those he/she deems rebellious.

For career bureaucrats, screening mechanisms and the reward of promotion (or punishment of termination) is unavailable. Strategic budgeting fills the gap: loyal appointees harness the effort of careerists by dividing them into competing units or projects, then rewarding those careerists who best-fulfill the president's agenda with larger budgets and punishing the laggards with budget cuts. The appointee, using a small appointed staff, is responsible for judging a bureau's performance relative to its competitors.

The same process of strategic budgeting can, if desired, be repeated at a higher level. Appointees in the Executive Office of the President (or the president him or herself) setting up competing agencies and judge their relative performance. The best agencies are given larger budgets, while the worst have their budgets cut.

The technique is feasible because redundancy is already common in government, with multiple agencies often serving similar or even identical goals.[2] Law enforcement agencies,

[1] For more information on government positions appointed at-will or in consultation with the Senate, see the Plum Book (U.S. House of Representatives Committee on Government Reform, 2004).

[2] Where redundancy does not exist, it can be created at low cost without hiring new staff by simply taking an agency of size N and creating k internal subunits of size N/k. Where tasks are divisible (education, law enforcement, etc.), this form of redundancy does not create potentially wasteful duplication though it may hamper the government's ability to take advantage of any economies of scale that might exist. Setting k to balance economies of scale against competitive pressure will be necessary in these cases.

for example, often have overlapping or nested jurisdictions (Russell, Conser and Paynich, 2005). Different branches of the military (the Army, Air Force, and Navy/Marines) often run separate, parallel projects that attempt to solve the same problem (Bendor, 1985, pp. 3-9). This redundancy makes it possible to pit agencies against each other on the same playing field in a competition refereed by an elected official or his/her appointee.

The competition I envision is not the same as other forms of competition, such as the typical bidding process used in procurement. In the procurement process, the incentive to work hard and meet government goals often disappears once a contract is awarded and the competition is over (Burnett, 1987, pp. 29-30). Rather, competition must be ongoing, with performance consistently compared and the victors consistently rewarded.

As I show in this dissertation, the form of competition provided by strategic budgeting is highly robust to the complex environments faced by the bureaucracy. Strategic budgeting is an effective motivator even in cases where bureaucratic performance is difficult to assess, such as when policy outputs and outcomes are not easily quantified (Wilson, 1991, pp. 158-175). Nor does strategic budgeting require bureaucrats to be exactingly hyper-rational calculators or myopically self-interested to work: bounded rationality and realistic (even ideological) preferences suffice and are built into the model.

1.2 Applications for Strategic Budgeting

Strategic budgeting may be of interest to policy makers for normative reasons. As this dissertation will show, the technique appears to be an effective means for elected officials to exert closer control over bureaucratic policy-making. By increasing effort levels, it can also make the bureaucracy a more economically efficient tool of rule-making and implementation. A more representative and efficient government is presumably desirable for democracy.

But this work should also be of interest to political scientists for positive, descriptive reasons. A great deal of scholarly interest has focused on how different branches of government in a divided system work together to make policy (Ferejohn and Shipan, 1990; Hammond and Knott, 1996; Shipan, 2004). Typically, these studies have focused on the institutional powers possessed by each branch and how these powers influence bargaining over policy. But the executive's power to influence the bureaucracy via strategic budgeting is heretofore unrecognized in this literature.

It is therefore both normatively and positively important to understand how strategic budgeting techniques can be employed in a government. The potential application of strategic budgeting techniques to service-oriented agencies, which are in continuous operation and can be constantly compared, should be apparent. Primary and secondary schools are one example of this kind of agency; regulatory and enforcement bureaus like the Justice Department and Environmental Protection Agency are another.

1.2.1 Example: Public Schools

The No Child Left Behind program put into place in 2002 compares schools to a rising baseline of performance goals, rewarding successful schools with larger budgets and punishing unsuccessful schools by firing staff, allowing students to leave (undermining the school's funding), and ultimately closing the school.

A strategic budgeting approach, by contrast, would attempt to compare a school's performance to other schools in a similar position (e.g., working with students from similar backgrounds at the same grade level). Schools that did better than their peers would receive budget increases, while schools that did worse would suffer budget cuts. Much of the work of the program would be focused on ensuring that comparisons were truly fair, in the sense that compared schools faced similar problems with similar financial resources to solve them.

A strategic budgeting approach to education would avoid many of the difficulties and criticisms leveled at No Child Left Behind because it relies on comparison of schools to each other rather than to a target or baseline. For example, NCLB has been criticized for allowing teachers and principals to manipulate the system to artificially inflate their scores.[3] In a strategic budgeting system, these techniques are ineffective, as these techniques have no net effect on relative performance if everyone uses them. Nor can watering down a test help a school system appear to perform better (as long as all schools take the same test) because performance relative to other systems will stay the same.

1.2.2 Example: Government Litigation in Civil Rights and Environmental Protection

During the Reagan administration, the President and his appointees made an effort to bring bureaucratic actions into closer accord with the President's policy agenda. Marissa Golden's in-depth study of Reagan's effort includes profiles of the Justice Department's Civil Rights Division and the Environmental Protection Agency (Golden, 2000).

In both agencies, the strategy was generally to transfer responsibilities away from career bureaucrats and toward political appointees, while increasing the size of the appointee staff. In the Civil Rights Division, the process was comparatively conciliatory toward careerists. While appointees did take especially sensitive cases away from careerists (Golden, 2000, p. 105), it was more typical for Reagan appointees in the division to exhaustively vet (and, if necessary, revise) careerist work so that it was consistent with the president's agenda (Golden, 2000, pp. 88-92).

In the Environmental Protection Agency (and in particular under Anne Gorsuch[4]), a similar but more confrontational approach was used. Gorsuch employed a variety of techniques, including (as listed, verbatim, by Golden):

1. The use of hit lists to target career employees for firing, transfer, or demotion
2. Three reorganizations of the Enforcement Division
3. Secrecy and the exclusion of careerists from the decision-making process
4. The replacement of careerists with appointees to produce the agency's work product
5. Centralization of the rule-making process so that all work was extensively reviewed by appointees

[3] See Section 2.6.2 for more details.

[4] Anne Gorsuch is also known as Anne Burford, due to a subsequent remarriage after joining the EPA (Martin, 2004).

6. Disparagement of the role of career civil servants (Golden, 2000, pp. 121-122)

A strategic budgeting approach would rely less on expanding appointee ranks and shifting effort toward them, and more on establishing career-staffed sub-agency units that must compete for budget funding under the supervision of a small appointee staff. For example, it would be straightforward to create multiple litigation units within the Office of Civil Rights and to compare their work product annually to determine budget cuts and increases.

The strategic budgeting approach has at least two advantages over the direct control techniques that Reagan's appointees employed. First, properly motivated careerists are likely to be better at their jobs than appointees. Careerists are hired on the basis of their skill and dedication, while appointees are valued for their loyalty. The incompetence and ineffective management of the patronage bureaucracy in the United States in the late 19th century was a key factor that led to the institution of the current civil service system (Johnson and Libecap, 1994, Chapter 2), and Golden's study provides some evidence that appointees were often less competent than careerists [5].

Second, making full use of the career staff eliminates the need for a large appointed staff to duplicate or replace careerist work. Appointees are only necessary for monitoring careerist output and making decisions about budget expansions or cuts. These judgments need not be perfect for strategic budgeting to be effective (as will be shown in Chapter 3), and hence the need for appointed staff is small. As a side effect, a smaller appointed staff should be easier for the president and his/her appointees to supervise.

1.3 Proving that Strategic Budgeting Works

I hope that the story I have told about strategic budgeting and its potential applications is interesting and provocative. The task of this dissertation is to make the story scientifically convincing, to prove that it has merit as a management strategy. I will proceed as follows.

1.3.1 Formalizing a Theory

I start by specifying a theory of bureaucrat behavior in an agency under conditions of weak supervision in the production of ideological goods. If competition over budgets induces bureaucrats to comply with an executive's preferences, I argue that researchers should expect to see executive control within legislative constraints, rather than bureaucratic independence within executive and legislative constraints. The theory I present is novel in that it is both mathematically rigorous and allows for the relaxation of many restrictive assumptions imposed by prior theories in the field. In my model, both bureaucrats and supervisors are ideologically-charged, imperfectly or boundedly rational, and have a sense of professional

[5]For example, when the National Highway Traffic Safety Administrator Ray Peck decided to strike the rule requiring automobiles to have airbags, [a]lthough they disliked the decision, careerists felt that decision-making responsibility fell appropriately within Peck's purview. The telling aspect of this episode is the comment that consistently followed the recounting of the episode, the gist of which was that if Peck had let the careerists write the order rescinding the regulation, it would not have been dismissed by the Supreme Court as 'arbitrary and capricious"' (Golden, 2000, p. 50). For other examples, see pp. 123-124, 129-130, and 133.

loyalty to the agency. I am able to show that the effectiveness of strategic budgeting is robust to these factors.

1.3.2 Testing the Theory with Multiple Methods

The strength of the dissertation lies in combining field and experimental tests, which have complementary design strengths. I aim to convince the reader that strategic use of budget incentives by the executive causes bureaucrats to comply with the wishes of executives despite those executives' lack of managerial authority and their inability to closely monitor bureaucratic output. In field data, observing a relationship between budget authority and bureaucratic behavior suggests a substantively significant connection in the relevant political world, but almost never definitively establishes a causal relationship between the two even under the most careful designs (Holland, 1986). Experimental designs, by contrast, require comparatively less-demanding assumptions in order to draw causal inferences, but face difficulties in matching the precise social and political environment of the external world. By combining the two approaches, evidence from the field supports the contention that causal relationships observed in the laboratory are relevant in live bureaucracies.

As noted, one essential problem with bureaucratic management is that good measures of bureaucrats' work product are difficult to obtain; the same problem exists with measuring their ideology. I therefore test my theory's predictions using experimental data from the laboratory. By taking an experimental approach, I am able to causally identify the power of budget competition on administrative behavior in a way that would be very hard to do outside the laboratory.

I also test implications of my theory in observational field data, specifically in state-level enforcement data from state environmental protection agencies, by investigating the relationship between a governor's power to control an agency's budget and the measurable work product of those agencies. My test distinguishes bureaucratic independence from executive control by examining differences behavior across institutions, where some states allow the governor a great deal of control over the budget and some allow less control. Controlling for other influences, a greater degree of sensitivity to gubernatorial preferences when gubernatorial budget control was higher would indicate less bureaucratic independence.

1.4 A Summary of Things to Come

The rest of the dissertation proceeds as follows. Chapter 2 presents the overall argument of the dissertation in greater detail, including the state of the literature in the field and how strategic budgeting fits into that literature. Chapter 3 mathematically presents my theory of strategic budgeting, deriving testable predictions from realistic assumptions about the complex nature of bureaucratic behavior and preferences. Chapters 4 and 5 test these predictions with complementary approaches, using data from a laboratory experiment in Chapter 4 and observational data from state environmental protection agencies in Chapter 5. In Chapter 6, I conclude with what I think political scientists should learn from the dissertation and suggest directions for future research in this field.

CHAPTER 2

COMPLIANCE AND DISCRETION IN GOVERNMENT BUREAUCRACIES

Before I develop and test my strategic budgeting technique in-depth, it is important to carefully demonstrate why it is necessary at all. Hence, in this chapter I discuss what we know about bureaucratic management and independence. To summarize, bureaucrats should enjoy discretion within policy boundaries established by the preferences and institutional powers of elected officials and the courts. Within the boundaries of discretion, however, there is not a clear story about what bureaucrats will do.

I argue that the executive will be able to control bureaucratic behavior within these boundaries, but standard techniques will not work due to the protections of the civil service system. Lower and middle level bureaucrats cannot be screened for political compliance, nor can they be fired at will if found to be non-compliant.

Instead, executives can exploit their privileged place in the budgeting process to establish competition for budgets within and between bureaucracies. Prior work in other fields has shown that tournament-style competition creates powerful incentives for the compliance with managerial directives, even in cases where monitoring is difficult and error-prone. This form of control enjoys advantages over other techniques, including ideological screening and alternative forms of budget management.

2.1 The Principal-Agent Problem

How do officials high in a hierarchy induce their subordinates to follow directives rather than pursuing their own interests? This basic question, typically called the *principal-agent problem*, has received such an extensive examination that a comprehensive review of the literature is well beyond the scope of this chapter. In economics alone, where the question is primarily studied in the context of the firm, a vast array of managerial scenarios have been studied. These include (but are not limited to) the case of costly but available information on individual worker productivity (Alchian and Demsetz, 1972), hidden differences in worker quality (Spence, 1973; Mirrlees, 1976), workers who possess payoff-critical information that the manager lacks (Crawford and Sobel, 1982), production in teams (Holmstrom, 1982), workers whose product is stochastically related to their effort (Mirrlees, 1976; Grossman and Hart, 1983), and countless others. The majority of these models turn on the writing of an optimal, incentive-compatible contract that can induce high effort from a worker despite the

fact that the worker's characteristics or effort levels are hidden.[1]

While much can be learned by comparing the principal-agent problem presented by government bureaucracies to the one faced by a typical private firm, there are at least three relevant differences to be considered. First, public agencies usually have multiple principals with different policy goals. In particular, an administrative bureaucrat typically answers both to the executive and his/her appointees and to the legislature. Previous work has shown that the existence of multiple principals with conflicting goals tends to dilute managerial authority and enable both shirking and greater policy discretion by agents (Bernheim and Whinston, 1986; Dixit, 1997; Dixit, Grossman and Helpman, 1997).

Second, though top-level positions are appointed, mid- and low-level civil servants in the administrative bureaucracy typically enjoy legal protections that obstruct their supervisors' ability to reward or punish them based upon their performance, protections that their private sector counterparts do not share (Wilson, 1991, Chapter 8). Promotion is based on a combination of seniority and competitive examination with little role for supervisory review,[2] while disciplining tenured employees (those with more than 2 years service) is a slow, ponderous process involving a great deal of effort and opportunities for appeal (Johnson and Libecap, 1989*b*, pp. 59-60; see also Johnson and Libecap, 1989*a*). These protections make it difficult for an executive to exercise his/her nominal supervisory authority over the agency. Of course, just as in many private firms, the work effort or output of many employees is very costly or impossible to measure directly for a variety of reasons, further weakening a principal's ability to control his/her subordinates.

Third, work effort in public agencies often carries an ideological charge that does not exist in a private firm. In addition to the financial, moral, social, and professional incentives to work or shirk that private workers face at their workplace, civil servants may derive benefit (or suffer harm) from accomplishing policy objectives that are consistent with (or inconsistent with) their own ideological convictions.

2.2 Management Techniques in a Government Bureaucracy

The executive's weak managerial authority and the difficulty with which outputs are measured rules out a number of standard approaches for management. For example, in public bureaucracies, executives cannot usually offer piece-rate compensation because most career bureaucrats are salaried and their jobs involve inputs and/or outputs that are not easily measured (Wilson, 1991, pp. 158-175). As a result, a great deal of scholarly effort has gone into exploring alternative ways that legislatures and executives can turn the bureaucracy into their willing instrument. These efforts generally fall into two categories: efforts to eliminate the information disparity between principal and agent, and efforts to eliminate the conflict of interest between principal and agent.

[1] Useful textbook overviews of the vast literature in economics can be found in Laffont and Martimort (2001) and Bolton and Dewatripont (2005).

[2] Although supervisory feedback is supposed to be a part of the promotions process, "[i]n practice, however, once an employee has received tenure, performance evaluations do not serve as a means of distinguishing among those employees who do or do not perform" (Johnson and Libecap, 1989*b*, p. 59)

2.2.1 Instituting Self-Monitoring Processes

Investigating the activities of any bureaucratic agency, let alone the entire administrative apparatus of a government, for political compliance would be a prohibitively costly venture. For a legislative body like Congress, which must use its own staff or membership to engage in oversight activities, these costs are especially acute. The fact that Congress rarely engages in such active oversight led many scholars to conclude that Congress had abrogated its oversight responsibility entirely,(McCubbins and Schwartz, 1984, footnote 1) leaving the bureaucracy to respond either to its own preferences or those of the executive.

However, the absence of active investigations and dramatic hearings does not *ipso facto* imply a lack of oversight: indeed, the presence of truly effective, low-cost monitoring would presumably deter bureaucratic overreaching that would necessitate oversight. Several avenues have been proposed by which this monitoring could take place. For example, legislators may rely on interest groups, lobbyists, corporations, and others with acute interest in bureaucratic outcomes to alert them whenever the bureaucracy takes an action counter to the legislators' preferences, effectively outsourcing the monitoring duty to a willing outside group whose interests mirror the legislators' own. This "fire alarm oversight" (McCubbins and Schwartz, 1984, p. 166) technique is at once less costly to Congress, and equally or more effective than constant investigation. Congress can make the technique even more effective by establishing rule-making procedures for the bureaucracy (e.g., requiring public notice and comment on new rules) that maximize transparency and ensure that relevant interests will be aware of rule changes that may require Congressional intervention (McCubbins, Noll and Weingast, 1987). The same fire alarm oversight technique can be used by supervisors within the executive (Downs, 1967).

2.2.2 Leveraging Institutional Powers

While the bureaucracy serves as the front line of implementation and enforcement for the government, elected officials are possessed of institutionalized powers with which they can shape and reverse bureaucratic decisions if (i) they are aware of the decision, and (ii) they are able to leverage their institutional powers. Examples are easily illustrated at the U.S. federal level. Congress has the power to make laws that can overturn bureaucratic rules, and can also revise or eliminate operating budgets to fund or defund programs, as long as it can gather support from 60 senators and 218 representatives for a particular course of action. The president can veto legislation, reorganize the bureaucracy, and has a powerful influence over bureaucratic budgets. The courts, and particularly the Supreme Court of the United States, can overturn administrative rulings and issue orders to the bureaucracy.

However, the institutional powers of government are, at least in the United States, often explicitly designed (Madison, 1999) to offset and stymie one another so that no branch of government exercises independent power over policy making. Combined action by multiple branches of government, or at least tacit approval from other branches of action by one branch, may be necessary to reverse bureaucratic actions. At any rate, leveraging institutional power against every objectionable action of the bureaucracy would be extremely costly if not impossible; indeed, delegation to a bureaucracy occurs precisely because legislatures do not have the time or expertise to directly administer the operations

of a modern government (Bendor, Glaser and Hammond, 2001). Bureaucratic discretion can lie in the the space of disagreement between and apathy from elected officials.

2.2.3 Screening for Loyalty

One of the executive's most important institutional powers is the power to appoint and remove high-level bureaucrats. A great deal of the literature has focused on the setting up screening mechanisms, so that those who are hired share the interests or policy views of the executive and thus will presumably do that executive's bidding loyally (Golden, 2000, p. 6). The strategy has been actively employed by U.S. presidential administrations, particularly since the Reagan era (Warshaw, 1995; Lewis, 2008). Much of the empirical literature on presidential control over the bureaucracy has cited the presidential appointment power as a preeminent causal mechanism relating presidential administration to bureaucratic outcomes (Moe, 1985; Chubb, 1985; Moe, 1987). Some studies have explicitly included high-level appointments as a key variable in determining agency activity (Wood, 1988; Wood and Waterman, 1991; Wood and Anderson, 1993).

While the loyalty of top-level bureaucrats may well be explained by their selection for ideological adherence (and the fact that their job depends on executive favor), the compliance of those below the uppermost level is not as easily explained by this model. In most cases, including the U.S. federal government, the executive does not have the authority to control the hiring, promotion, and firing of career bureaucrats. Indeed, multiple case studies have shown that the president and political appointees are deeply suspicious of career bureaucrats' loyalties—even when this suspicion is unwarranted, as it usually is (Rosen, 1989, Chapter 9; Wilson, 1991, pp. 274-275; Brehm and Gates, 1999; Golden, 2000). This suspicion sometimes prompts executives to create parallel agencies within an agency, staffed with appointees, that cut career bureaucrats out of decision-making and implementation (Golden, 2000, Chapters 5 and 6).

Even in the case of top-level bureaucrats, many positions require legislative confirmation. This power is not necessarily exercised to the fullest for all positions; as Moe points out, "Senate committees have traditionally adopted the view that—except in geographically 'sensitive' cases, like the federal district courts—the president has a right to build his own administrative team in his own way" (Moe, 1987, p. 489; see also Gerhardt, 2003, p. 162). But the absence of overt conflict[3] does not necessarily indicate the absence of legislative power over appointments. Rather, the president typically avoids conflict by nominating persons who will be acceptable to the Senate.

> Knowing the Senate has the power to reject his nominations, and wanting to avoid the embarrassment and political agony of that, the President normally makes an effort to check the acceptability of his nominees with key Senators before announcing them. If strong opposition appears likely, he may well decide not to make a planned nomination and to reopen the search for a more acceptable candidate" (Mackenzie, 1981, p. 174).

[3] "The formal outcomes of the confirmation process can be characterized very simply: Nominations to federal offices are almost always approved" (Mackenzie, 1981, p. 174).

In addition, the modern Senate is not as deferential to presidential prerogatives as in the past. Once the honeymoon period of a president's initial election has passed, both the rejection rate and wait time for high-level executive appointees rises (Loomis, 2001). Congress has forced the withdrawal of a number of bureaucratic appointees, even below the cabinet level, particularly during the Clinton administration (Gerhardt, 2003, pp. 162-174).

2.2.4 Acculturating Professionalism

At the end of the nineteenth century, changing political realities caused the bureaucracy to morph from an openly partisan instrument, and most importantly a reservoir of patronage positions, into a professional institution designed to impartially implement laws passed by Congress (Johnson and Libecap, 1994; Gerhardt, 2003). Along with this change came a fundamental change in the culture of the bureaucracy, a culture of professionalism and service to elected that some regard as central to bureaucratic performance (Wilson, 1991; Brehm and Gates, 1999). Hence, developing and fostering a culture of service and non-political professionalism may make the bureaucracy loyal to the authority, if not personally loyal, to elected officials.

But again, the bureaucracy serves multiple masters, and it may be difficult to determine from whom a professional bureaucrat should take direction. At the federal level, both Congress and president would seem to have competing claims on bureaucratic loyalty. As noted above, the existence of multiple principals tends to dilute hierarchical authority and promote shirking and allow for greater policy discretion (Bernheim and Whinston, 1986; Dixit, 1997; Dixit, Grossman and Helpman, 1997). In addition, some bureaucrats believe that the American public is their ultimate principal (Golden, 2000, pp. 124-125); Downs calls these bureaucrats "statesmen" (Downs, 1967, 110-111). While surely defensible, this belief presumably gives a bureaucrat considerable discretion to define what the public interest is.

2.3 Policy-making in a Bureaucracy

Given the dearth of effective managerial techniques available to any of the principal branches of government, one may expect to see a bureaucracy unaccountable to the public living by its own rules, seeking to implement its own policy preferences, shirk its duties to pursue personal diversions, and provide itself with perks and other benefits at taxpayer expense. Indeed, accounts of extreme bureaucratic discretion do exist (Niskanen, 1971). But in-depth studies more commonly find that bureaucrats work hard and attempt to serve the public interest, even if this attempt is not entirely effective (e.g., Downs, 1967; Wilson, 1991; Brehm and Gates, 1999; Golden, 2000). The story of quantitative studies is more mixed, but generally do not support the hypothesis of an out-of-control bureaucracy: evidence exists to suggest that federal bureaucracies are responsive to the President (Moe, 1987), Congress (Weingast and Moran, 1983), both (Wood and Waterman, 1993), or neither (Eisner and Meier, 1990).

The theoretical literature in this area, responding to studies of the management techniques above, finds that bureaucracies are capable of exercising bounded discretion in policy-making with the boundaries established by the preferences and powers of elected branches of government. However, these theories are largely silent about what happens or who is

in control within this zone of discretion. I argue that a technique of managerial control, known to social science but heretofore unrecognized in the bureaucratic literature, allows the executive to control bureaucratic behavior inside this zone.

The accumulated empirical and theoretical study of the bureaucracy, techniques available to manage it, and the effectiveness of these techniques has culminated in a series of papers attempting to comprehensively model the process of policy-making in a bureaucracy, including the interactions between elected branches of government and the bureaucracy (Ferejohn and Shipan, 1990; Hammond and Knott, 1996; Shipan, 2004). All of these papers are built on knowledge of the management techniques above, and assume that (i) the bureaucracy moves first to set policy in an area, (ii) other branches of government are aware of bureaucratic action, and (ii) these branches have institutional powers that can alter or reverse bureaucratic action or the action of other branches. In all cases, the conflicting interests and countervailing powers of these separate branches allow zones of policy-making within which the bureaucracy can safely exercise a great deal of discretion.

In Ferejohn and Shipan's (1990) model, the bureaucracy can sometimes set a policy consonant with its own preferences that a gatekeeping legislative committee would be unwilling to overturn: opening the gates to new legislation would require allowing the floor to set a policy that could be even worse. In Hammond and Knott's (1996) model, the agency is able to set the status quo point for policy in a bargaining game among legislative and executive actors: by placing the status quo at the most favorable point in the bargaining core, the bureaucracy can maximize its own utility under the constraint of the core boundaries without fear of being overturned by later bargaining among its principals.

In neither of these papers is the source of bureaucratic preferences clarified. This is a critical point, as a great deal of policy-making action takes place within this zone of discretion. The degree of control that elected officials exercise over the bureaucracy, particularly during periods of divided government, will depend on how the bureaucracy exercises this discretion. One of the papers does provide a hint: Shipan's (2004) model includes an empirical test, and therefore must specify a source for bureaucratic preferences. While explicitly stating that he is not assuming that the agency's preferences are identical to the president's, Shipan says:

> I assume that the president is able to set the general location of the agency's ideal point. ...This is a strong assumption, one that essentially means that the president is in a privileged position, relative to other political actors, with respect to the agency. It is, however, an assumption that is based on convincing empirical and descriptive accounts of the powers of presidents—or, more generally, chief executives.... Among other things, the president can, for example, reorganize agencies, alter jurisdictions, and, in many cases, make appointments to key positions, and can do all these things without needing the cooperation of Congress (Shipan, 2004, p. 471).

The question is: how good is this assumption? As Shipan points out, ready measures of bureaucratic preferences are unavailable, making a direct test untenable. But as I have shown above, though certainly defensible as an assumption, executive control over bureaucratic action is by no means assured by the present state of theory.

2.4 Executive Budget Power

There is a reason to think that executives may exert disproportionate control over administrative preferences, even if their ability to select or terminate the administrators is limited: executives typically play a privileged role in the budgeting process. Though the legislature almost always has the final say over appropriations, the executive usually moves first by presenting the legislature with a comprehensive budget proposal. In the federal government, this process is managed by the Office of Management and Budget (OMB) housed in the Executive Office of the President, who works in concert with agencies to develop a budget in line with economic conditions and the president's policy goals (McCaffery and Jones, 2001, pp. 97-108; Meyer, 2002, pp. 55-71). Unlike the Congressional budgeting process, which operates through committees and ultimately by majority consensus, the president can formulate this budget unilaterally. Furthermore, as with any bill, the president can veto budgets that are not to his/her liking, or can unilaterally reorganize agencies and their responsibilities to control allocation of budget resources.

At the state level in the United States, governors often enjoy even more far-reaching powers over the budget. The National Council on State Legislatures effectively summarizes the degree of gubernatorial budget power in the states:

> In most, the proposed executive budget dominates the legislative budgeting process; it establishes the agenda for budget discussion and negotiation. In some of these states, it is adopted with relatively few changes. However, strong legislative budget processes dominate in a few states–Arizona, Colorado and Texas, for example. In those states the executive budget may be largely disregarded. ...Generally, far more cooperation exists between the legislative and executive branches than is typical for Congress and the president. The agenda-setting role of the executive budget in many states is one reason. The executive line-item veto is another reason in the 43 states where it exists. A final reason is the political expectation–that is almost always met–that a state will enact a balanced budget before the beginning of its fiscal year. Late budgets and vetoes are most likely when a state is experiencing serious fiscal difficulties. Otherwise, executive and legislative negotiations are intended to resolve budget disagreements and avoid vetoes and late budgets (National Council of State Legislatures, 1997).

Governors may therefore be exceptionally able to implement budget-based strategies for administrative control.

2.5 Creating Competition for Budgets

I argue that the executive can leverage his/her control over the budget to create competition for budget allocations within the bureaucracy, a process that I call *strategic budgeting*. Suppose that an administrative employee is fully rational and has preferences over three dimensions: salary, policy outcome, and the size of his/her agency's budget. The employee always prefers to be paid more, to set policy closer to his/her ideal point,

and to have control over a larger budget (and the attendant prestige, slack resources, and political influence). It is clear that a bureaucrat will refuse to exert additional effort toward a politically unpleasant task when budgets and salaries are fixed. But this bureaucrat might well exert a great deal of effort when this effort is being compared to others when determining which agencies will receive a budget expansion (or a budget cut). The effect is to bring bureaucrats into compliance with the policy wishes of the executive.

Such a system has already been systematically explored in the industrial organization literature in the context of the firm, i.e. in an organization scheme whose tasks have no political relevance for an employee, in a field of study called *tournament theory* (Lazear and Rosen, 1981; Green and Stokey, 1983). The notion of applying tournament-style incentives to the public sector has already gained some scholarly attention, though toward a slightly different end than the one I suggest. Whitford (2006) argues that using tournaments for promotion in the hierarchy may be a more efficient way of boosting productivity in the public sector than alternatives, such as piece-rate pay or extensive screening of applicants. His argument is bolstered by findings from the economics literature, which show that tournament-style incentive schemes are effective at inducing high effort, and are as efficient as piece-rate schemes for firms in a competitive marketplace when agents are risk-neutral (Bolton and Dewatripont, 2005, pp. 316-326).[4] The informational requirements of a tournament are much lower than the information requirements of screening or piece-rate pay, each of which require detailed and reasonably accurate observation of (potential) employees; hence, tournaments should be easier and cheaper to implement. But in the bureaucracy, extensive career protections fostered by the Pendleton Act and subsequent legislation largely rule out the possibility of running a politically-motivated promotion tournament, at least explicitly, though political motivations may still act at the margins (Johnson and Libecap, 1994).

I argue that using larger budget allocations as tournament rewards is a more easily implemented solution. As William Niskanen (1971) points out, bureaucrats are interested in larger budgets as a means of building up their own resources, authority, and career advancement. Control over a larger budget bestows more power to a bureaucrat: more money can hire more subordinate staff, pay for more work-related trips abroad, and afford the bureaucrat an overall more comfortable work lifestyle and enhanced opportunity for advancement. Niskanen intended his book as an indictment of the wastefulness of administrative government, but with appropriate supervisory techniques, a bureaucrat's desire for bigger budgets can be leveraged to control bureaucratic output, a surprising and unexpected result of the application of tournament theory.

2.6 Advantages of Strategic Budgeting

There are reasons to expect strategic budgeting techniques to perform better than alternative management strategies, including other strategies that involve the use of budget power. Some of the disadvantages of alternative strategies were touched upon in subsection 2.2. Here, I make the comparison with tournament-style competitive budgeting explicit.

[4]Tournaments are less efficient than piece rates when agents are risk-averse, and more efficient when common shocks to productivity exist.

2.6.1 Minimize Congressional and Clientele Pushback

Setting up a competition for budget expansion is not the only way to leverage executive budget authority. One might also consider monitoring bureaus for compliance, then using direct budget cuts, agency reorganization, or outright eradication against recalcitrant agencies. These tactics were employed by President Reagan, who announced in his 1982 State of the Union address that "The budget plan I submit to [Congress] on February 8th will realize major savings by dismantling the Departments of Energy and Education and by eliminating ineffective subsidies for business" (Reagan, 1982).

The effectiveness of agency elimination as a political tactic is questionable: both departments mentioned in Reagan's speech are still in existence 26 years after that address as of this writing, while new agencies (e.g., Homeland Security) have been created by his successors. Indeed, the Department of Education's budget rose in every year but one during Reagan's term, though staying nearly constant relative to inflation (Verstegen, 1990). Perhaps, by this inflation-adjusted metric, Reagan's campaign against the agency was somewhat successful.

There is reason to believe that establishing competition between bureaucracies will be a more effective strategy. Taking overtly punitive actions against an agency, particularly by cutting or eliminating its budget, is very likely to alarm and mobilize clientele groups of the agency and their allies in Congress. However, with internal competition, policy actions by the bureaucracy in any given area can be shaped more subtly without overt cuts in overall funding: while some sub-agencies win and some lose, the overall level of funding to a field does not change. Agency competition encourages compliance with executive policies while creating new clientele groups (of the successful agencies) that can counter the influence of the aggrieved agency's clientele.

2.6.2 Lower Monitoring Requirements

Monitoring agencies for compliance with the president's agenda, and setting budgets accordingly, is a task that requires a great deal of time, effort, and personnel to execute effectively. The Office of Management and Budget must work with an agency in order to develop its budget, as the agency knows far more about its requirements and operations than OMB. Agencies very typically exploit their information advantage to thwart OMB objectives and maximize their budgets (Meyers, 1994), and consequently it can be difficult for the executive to know precisely what budget figures are appropriate for meeting his/her policy objectives in an agency.

In addition, as has been mentioned above, performance can be difficult to objectively assess in an agency where outputs and/or policy outcomes are unobservable (Wilson, 1991). Just setting a performance goal can be confusing and controversial, as has proved to be the case in the No Child Left Behind program. The program is designed to require all schools' students to meet increasing standardized test performance requirements, with attention to ensuring that disadvantaged subgroups (minorities, special education students, etc.) meet these targets. However, the nature and effectiveness of these tests has been hotly disputed. States, which have the discretion to set their own standardized tests, have considerably different assessments that require differing levels of performance (Lewin, 2007). This

allows states to surreptitiously lower standards while demonstrating apparently increasing performance. Individual schools can also inflate their performance by a variety of means (see, e.g., Ortiz, 2008). For example, schools in Texas have inflated their performance by encouraging nonperforming students to drop out; the increased dropouts do not count against the school as long as the students declare intent to obtain a GED (McNeil et al., 2008). Some education scholars claim that the performance standards do not address the inequalities of secondary education, as schools teaching financially and culturally advantaged students are compared to those serving disadvantaged communities (Darling-Hammond, 2007*a*; Hursh, 2007; Darling-Hammond, 2007*b*).

Creating competition for budgets facilitates monitoring and lowers its cost. Competition allows comparison of performance to other agencies, rather than comparison to a difficult-to-establish baseline. The gimmicks that agencies use to inflate their performance or their budget become meaningless in this relative context, as the relative effect of the gimmicks is zero if all the competing agencies are using them. Special care must be taken to ensure that agencies (e.g., schools) are compared to agencies that face similar problems and resource constraints—otherwise, competitions simply reflect and perpetuate resource inequality—but if comparable agencies can be identified, then relative performance is easier to assess than absolute performance.

2.6.3 Ease the Contracting Problem

The idea of injecting competitive forces into the bureaucracy is not a new one. The idea of moving work from a government agency to a private firm is predicated on the idea that private firms must be more efficient in order to survive market competition (Niskanen, 1971; Miranda and Lerner, 1995). If privatization results in increasing policy bang for the taxpayer buck, this can be an effective strategy.

But experience with military procurement surely shows that private firms do not always provide low-cost, on-budget, timely, and effective solutions for a government bureaucracy, even when competitive bidding processes are held. If only one supplier is selected, as is more common, most competitive pressure evaporates once a contract is awarded to a government supplier. The process creates an incentive for private firms to underbid on price and create overoptimistic delivery schedules to obtain a contract, with little incentive to hold to these guarantees once the contract is awarded. This particular problem is one of several contracting problems extensively studied by Williamson (Williamson, 1998, pp. 169-179). To cite just one of many examples of this phenomenon, the U.S. Navy suspended the Littoral Combat Ship due to massive cost overruns and multiple delays in delivery (Taubman, 2008). Indeed, these problems materialized despite that program's use of two parallel shipbuilding programs up through the procurement phase that should have increased competitive pressure to deliver.

As Williamson argued more generally (Williamson, 1998), there are sometimes monitoring and commitment problems associated with inter-firm contracting that are better solved inside a firm. While contracts between firms and the government may be difficult to legally enforce because outputs are qualitative and difficult to measure—and many products of the bureaucracy are qualitative and difficult to measure—inside the government monitoring is cheaper and authority unambiguous. While the potential for a hold-up problem may make contracting costly and inefficient, inside the government such problems can be avoided

through exercise of hierarchical authority. For these reasons, and because alternative mechanisms of executive authority are weak, I argue that budget competition inside the government can be a more effective solution.

2.6.4 Increase Redundancy

By their very nature, competitive budget tournaments require a degree of duplication in the bureaucracy. There have to be multiple bureaucrats, sub-agencies, and/or agencies performing similar work in order to establish a competition. On its face, this may seem to be a disadvantage: duplicating services could increase the cost of government. In particular, if economies of scale exist for administration, then the establishment of two smaller, competing bureaus might be less efficient than one larger bureau if the larger bureau can somehow be motivated to produce efficiently.

But redundancy also has advantages. Several scholars have argued that duplication and redundancy allow the government to ensure the success of projects (Landau, 1969; Bendor, 1985; Ting, 2003). The argument, in brief, is that all projects have some probability of failure, and hence having several bureaucracies simultaneously (and separately) working on variants of a project[5] increases the chances of success. Depending on the project's importance and any individual project's likelihood of success, redundancy may therefore be cost-effective and valuable. Redundancy may be especially effective at increasing production in cases where the bureaucracy is indifferent or hostile to the success of the project, particularly if the executive has the power to observe outcomes (though not inputs) and terminate bureaucrats accordingly (Ting, 2003).

2.7 Disadvantages to be Anticipated and Questions to be Answered

Competitive incentives like these do present special challenges. First, when a bureaucrat or agency believes that gaining budget increases (or avoiding cuts) is impossible because of the superior abilities of other agents, that person or organization may give up and behave according to its own preferences (e.g., by shirking). This behavior has been observed in previous experiments testing the power of tournament-based incentives (Schotter and Weigelt, 1992). Therefore, in a repeated production setting, performance may separate into a class of career-motivated, high-performing bureaucrats (the young and the successful) and another class of stagnant, lower-performing bureaucrats who have realized that they will never win promotion. This outcome was observed in the bureaucracy years ago by Downs Downs, 1967, who labeled these classes climbers and conservers respectively.

Second, the effectiveness of tournaments depends on minimizing opportunities for collusion or sabotage. Tournament models are based on the assumption that agents can choose to exert effort toward the desired goal, or shirk on that effort, or even undermine the goal of the principal, but cannot attack or directly sabotage the effort of other agents toward the end of increasing one's own probability of winning the tournament. If this option is available, a tournament's efficiency can be compromised (see Bolton and Dewatripont, 2005, pp. 26-

[5]E.g., different military weapons systems that accomplish the same goal (Bendor, 1985, pp. 3-9).

38). This collusion need not even be overt: if agencies know that they are working toward a common goal where only one success is required and punishment/reward is uncertain or nonexistent, then the collective action problem may discourage individual bureaus from exerting effort in order to free ride on other bureaus' output (Ting, 2003).

Finally, very little of the existing theoretical or empirical work on tournaments has been directed at public agencies in particular. Some of the differences may not be relevant. For example, while political executives have weaker managerial authority compared to their private sector counterparts, budget competition does not rely on expansive authority. Though the bureaucracy faces multiple principals, if the executive can exercise control over budget incentives the bureaucracy may act as though the executive is the sole (or the most important) principal to obey.

But these are empirical questions and must be resolved empirically. Private sector tournaments typically rely on salary bonuses or job promotions for the winners. Will winning a larger budget be as effective an incentive as direct financial reward, or at least effective enough for government work? The executive plays an important role in budgeting, but the legislature holds final authority—and many other authorities besides. Will the bureaucracy obey this alternative master and ignore the executive? Finally, private sector jobs rarely involve work with a political or ideological charge. When bureaucrats are hostile to a policy goal, when they do not share the values of the executive will budget competition motivate them to produce nonetheless? These are the questions that I strive to answer in the remaining chapters of the dissertation.

CHAPTER 3

A THEORY OF BUREAUCRATIC CONTROL

In the first two chapters, I laid out a heuristic argument for how strategic budgeting techniques would work, why they fill a void in our knowledge about bureaucratic management, and how they could be applied to real political problems. In this chapter, I formally prove the logic of my claims by mathematically modeling bureaucratic behavior under a strategic budgeting system.

The substantive political value of this chapter is in demonstrating that strategic budgeting techniques work to increase efficiency and secure compliance in complex, realistic environments. My model does not rely on unrealistic assumptions about human behavior. People are imperfectly rational and have multiple motivations for their behavior. As a result, all actors in my model—appointed supervisors and career bureaucrats—are boundedly rational. While my model assumes that bureaucrats prefer larger budgets to smaller, it also assumes that they have norms of professionalism and their own policy preferences that shape their behavior.

None of these assumptions change the core finding: strategic budgeting is an effective way of motivating bureaucratic efficiency and compliance with elected officials. As a result, we can be more confident that strategic budgeting techniques will work in the real world, not just on paper.

3.1 Setting Up a Model

A supervisor, either an elected official or a political appointee, is charged with overseeing a government agency on behalf of the executive (in the American context, the governor or president). Career civil servants who work at the agency influence policy outcomes by exerting effort toward tasks assigned to them by their supervisor. Tasks may differ across agencies and from day to day for any individual, and include new rules and regulations, enforcement actions, investigations, and the like. A bureaucrat may choose to exert effort towards the executive's tasks, neglect work in favor of time-wasting activities (socializing, web surfing, and the like), or exert effort that undermines accomplishment of the task. These correspond to the working, shirking, and sabotage actions studied by Brehm and Gates (1999).

All theories, formal and otherwise, are sustained by assumptions, and this one is no different. While examining their initial plausibility is a valid exercise and one that I engage in here, the ultimate test of their appropriateness will come from the empirical

tests conducted in the later chapters. For now, it is important to carefully and completely lay these assumptions bare in order to fully understand the operation of the model and the conclusions that are derived from it.

3.1.1 Individual Behavior

First, and most importantly, I make assumptions about how people make decisions. Many formal models of politics rest on the foundation of rational choice theory, the basic premise of which is that individuals have well-defined and consistent preferences and take actions that are consistent with satisfying those preferences. The assumption of well-defined preferences motivates the creation of a utility function mapping outcomes into preferences; the assumption of rationality motivates the idea that actors maximize this utility subject to resource constraints and the strategic reactions of other actors.

However, particularly in the study of the bureaucracy, there are compelling reasons to question this foundation. Herbert Simon's aptly-titled study of administrative behavior is summarized nicely in the introduction to the second edition:

> In one sentence, the thesis... is this: The central concern of administrative theory is with the boundary between the rational and non-rational aspects of human social behavior. Administrative theory is peculiarly the theory of intended and bounded rationality—of the behavior of human beings who satisfice because they have not the wits to maximize (Simon, 1997, p. xxiv).

This work on satisficing behavior, with its reasoning but not fully rational civil servants, has been extremely influential on the development of political science and public administration. Hence, we would do well to incorporate the ideas of bounded rationality and satisficing into a theory of compliance with the executive.

Fortunately, a formal framework that incorporates these ideas already exists in the concept of Quantal Response equilibrium, or QRE.[1] While a comparatively new concept—QRE was introduced by McKelvey and Palfrey in a series of papers in the 1990s (McKelvey and Palfrey, 1995, 1996, 1998)—it has already been fruitfully applied in a variety of empirical settings in political science (Signorino, 1999; Guarnaschelli, McKelvey and Palfrey, 2000; Aragones and Palfrey, 2004; Goeree and Holt, 2005). Most importantly, it melds the idea that people attempt to maximize their utility with the idea that their ability to maximize is mitigated by psychological factors and plain error.

Suppose that an individual i faces a normal form game with n possible actions in the set A^n. These actions map to payoffs by way of a utility function, u, where utility is contingent both on a player's own actions and on the actions of others. How will the individual decide which action to take? According to Quantal Response equilibrium, the person compares the expected utility of each of his possible actions and chooses that with the highest utility. However, the comparison is subject to error from a variety of sources. For example, a bureaucrat may be cognitively ill-equipped to distinguish fine differences

[1] There are other mathematical approaches to bounded rationality, and comparatively evaluating them is beyond the scope of this project. I employ the QRE approach to bounded rationality because it is well-developed and widely used in the literature, and works well in the bureaucratic context.

in utility among multiple options. Alternatively, the bureaucrat simply lacks the time to undertake the analysis necessary to make a fully-rational decision. A supervisor's information about bureaucrat productivity may be limited or erroneous, leading to an error in judgment. Psychological factors (such as mood) that have a random component might subtly influence decision making. For all these reasons and others, a person chooses action a_{ix} over all other actions if:

$$u_i(a_{ix}) + \varepsilon_{ix} > \max_{y \in A^n \setminus x} u_i(a_{iy}) + \varepsilon_{iy}$$

Hence, actions are taken probabilistically. The frequency with which an action a_{ix} is taken, $\Pr(a_{ix})$, is contingent both on systematic components of utility for all actions in A^n, and on the distribution of the stochastic component of utility ε. The distribution of $\Pr(a_{ix})$ depends on the distribution of the vector of error terms $\vec{\varepsilon_i} = [\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{in}]$. It can be shown[2] that if the components of $\vec{\varepsilon_i}$ are independently and identically distributed according to the Type I Extreme Value Distribution,[3] $\Pr(a_{ix})$ is given by the multinomial logistic distribution:

$$\Pr(a_{ix}) = \frac{\exp[\lambda u_i(a_{ix})]}{\sum_{y=1}^{n} \exp[\lambda u_i(a_{iy})]} \tag{3.1}$$

If a_i is defined over a continuous action space $[\underline{a}, \overline{a}]$, the definition can be revised to reflect this continuity (Anderson, Goeree and Holt, 2002; Yi, 2005) :

$$\Pr(a_{ix}) = \frac{\exp[\lambda u(a_{ix})]}{\int_{\underline{a}}^{\overline{a}} \exp[\lambda u_i(a_i)]\, da_i} \tag{3.2}$$

A *logistic quantal response equilibrium* is located by calculating a fixed point of the best response functions in 3.1 for all actions x and all players i. That is, at equilibrium, all persons should take actions with probability given by 3.1, and none should wish to revise their strategies in light of other people's strategies. Existence of such an equilibrium is guaranteed for all games (McKelvey and Palfrey, 1995, p. 10).

The λ parameter measures an agent's *propensity to best respond*. As $\lambda \to 0$, an agent's actions become increasingly random, until at the limit the agent simply selects a random action from A^n regardless of its systematic utility component. As $\lambda \to \infty$, an agent becomes increasingly likely to pick the utility-maximizing action (i.e., the best response). A unique branch of the QRE correspondence connects the completely random outcome at $\lambda = 0$ to a single Nash equilibrium at the limit $\lambda \to 0$, an equilibrium McKelvey and Palfrey refer to as the *limiting logit equilibrium* (McKelvey and Palfrey, 1995, p. 12). This branch can be interpreted as a path taken by those gradually learning the equilibrium. Players start out with nearly random strategies due to their unfamiliarity with the game and with other players' strategies, then gradually converge to some distribution of play as the error components of their play are eliminated and they become better able to respond to others'

[2]The proof, omitted here, is somewhat involved; see Maddala (1986) and Long (1997).

[3]Other distributions can be derived by assuming different distributions for $\vec{\varepsilon}$; for example, a form like the multivariate probit can be derived by assuming that $\vec{\varepsilon} \sim \Phi(0, \Sigma)$, with Σ giving the variance and covariance of the error terms. I will assume the multinomial logistic distribution throughout this chapter, primarily because the multinomial logistic has a closed form (unlike the multivariate probit).

strategies (McKelvey and Palfrey, 1995, pp. 17-25; see also Nyarko, 2001 and Turocy, 2005). A logistic quantal response equilibrium can also be shown to be the steady state outcome of a learning process whereby actors adjust their strategic play according to whether marginal payoffs are increasing or decreasing in play of the strategy (Anderson, Goeree and Holt, 1999).

3.1.2 Policy Output Space

I assume that a bureaucrat can choose only the degree of effort to exert toward (or against) accomplishment of a policy task. This setup supposes that bureaucrats do not have control over the mix of policy outputs to be produced or the tasks they can work on; these are decided by the supervisor. Hence, bureaucrats do *not* choose policy in an area along some ideological dimension: in my model, they lack the authority and discretion to directly move policy in this fine-grained fashion. Rather, a bureaucrat (or agency) is assigned some task—inspect facilities, write and enforce citations, prosecute cases, et cetera—with an ideological valence.

This model is obviously inappropriate for all settings. For example, when a bureaucrat has the authority to effectively set policy, modeling the bureaucrat's decision as the mere decision to exert productive or counter-productive effort is probably inaccurate. This inaccuracy is most pronounced when civil servants engage in rule-making, a task that determines the way in which a law written by Congress will be applied. To model these cases, a different (and more complex) model would need to be applied.

3.1.3 Supervisor's Ability to Monitor Bureaucrat Output

As is consistent with a scenario of tournament incentives, I assume that the supervisor will award a budget bonus to the single bureaucrat that she perceives to be the best in terms of net effort exerted toward the executive's desired goal. Consistent with the tradition of the principal-agent literature and with the prevailing wisdom in public administration, I assume that the supervisor's ability to monitor work effort or work product is severely restricted. More specifically, I assume that the supervisor is only able to *ordinally* (or relatively) rank the individuals with respect to their output, without any knowledge of either employee's *cardinal* (or absolute) work effort. That is, the supervisor can come to a judgment about the bureaucrat exerting greatest effort, second-greatest effort, and so on, without knowing precisely how much effort any bureaucrat actually exerts.

Furthermore, I assume that this ranking is subject to error. A number of reasons why this decision is error-prone have already been listed. In an information-poor environment, the supervisor may simply misjudge the top producer as a result of erroneous observation. Transient psychological states (good and bad moods) may cause over/underestimation of some bureaucrats' output. Whatever the reason for error, its practical upshot is that winning the bonus is stochastic and can be treated probabilistically.

This point can be restated mathematically. The supervisor awards a higher budget to bureaucrat i if:

$$e_i + \varepsilon_i > \max_{j \neq i} e_k + \varepsilon_j \tag{3.3}$$

In this equation, e_i is the effort exerted by bureaucrat i, and ε_i is the error associated with observation of that output. Note that bureaucratic sabotage is permitted in this framework: if a bureaucrat chooses to sabotage the work of the agency, this is modeled as negative effort e_i that subtracts from the total output of the agency.

The probability that condition 3.3 is met will be abbreviated $\pi_i(e_i, e_{-i})$. This probability depends on the distribution of the vector of error terms $\vec{\varepsilon} = [\varepsilon_1, \varepsilon_2, ..., \varepsilon_n]$. The logistic form of quantal response equilibrium dictates that $\vec{\varepsilon}$ is distributed according to the Type I extreme value distribution and π_i is given by the multinomial logistic distribution:

$$\pi_i(e_i, e_{-i}) = \frac{\exp[\lambda_s e_i]}{\sum_{j=1}^{n} \exp[\lambda_s e_j]} \tag{3.4}$$

Here, the λ_s parameter corresponds to the degree of error in the supervisor's judgment. As this parameter rises, the supervisor's ability to distinguish the most productive bureaucrat from the others also rises. As $\lambda_s \to 0$, the supervisor's choice of winner is random; as $\%\lambda_s \to \infty$, her choice of a winner always corresponds to the highest producer.

3.1.4 Bureaucrats' Utility Function

I assume that the goal of a civil servant is to set a policy that maximizes the combined value of winning a budget increase, fulfilling his professional norms, and doing work that comports with his own ideological goals; these benefits are traded off against the cost of exerting effort. Mathematically, the relationship can be expressed as:

$$u_i(e_i, e_{-i}) = w + B \cdot \pi_i(e_i, e_{-i}) + r_i \cdot \alpha_s \cdot e_i + (1 - r_i) \cdot \alpha_i \cdot e_i - c_i(e_i) \tag{3.5}$$

The bureaucrat chooses e_i, the effort to exert toward productivity or sabotage. The parameter r_i indicates the strength of a person's professional obligation to conform to the wishes of the administration in office. Utility is determined by a combination of the wage w, the budgetary bonus B received with probability π_i, the ideological (dis)utility associated with the policy $\alpha_i * e_i$, the supervisor's ideological utility $\alpha_s \cdot e_i$, and the cost of effort given by the function $c_i(e_i)$.

I assume that $c(0) = 0$, and that the function is convex and symmetric about the status quo. Substantively, this form means that exerting no effort is costless, that sabotage and productivity are equally costly, and that each additional unit of effort becomes more costly to exert. While some information can be cleaned with these general assumptions, more precise predictions will sometimes necessitate the assumption of a functional form for cost. I assume the following form, with $\beta_i > 0$:

$$c_i(e_i) = \beta_i e_i^2 \tag{3.6}$$

One aspect of the utility of bureaucrats, the professionalism parameter r, deserves special attention. It receives this attention in the next section.

3.1.5 Professionalism

Professionalism is shorthand notation for a complicated influence on behavior. The term encapsulates the combined effect of sociological and normative influences on a bureaucrat's behavior that come from the community of which the bureaucrat is a part. While excluding it entirely would deprive the model of substantive richness, it is also important to avoid overly complicating the model or building in conclusions too transparently. Hence, my goal is to model the effect of professionalism and how it interacts with tournament-style incentives without losing the parsimony of the theory.

Some aspects of professionalism will already be captured in the term α_i. If the training and professionalization process changes a person's policy preferences, this change should be reflected in the non-monetary utility they derive from setting a certain policy. For example, civil rights lawyers often vigorously prosecute their cases because law school inculcates the desire to litigate into its graduates. Liberals might also wish to vigorously prosecute these cases, albeit for other reasons. We would expect members of either group to prefer more litigation to the average person, and would expect someone who is a member of both groups to prefer it even more. This preference is represented in the utility that a civil servant assigns to production of the policy output.

However, one aspect of professionalism requires an additional parameter. The tradition of public administration in the United States seeks to separate politics from administration; elected officials set the policies of the land, and administrators pursue these policies in the most technically efficient way possible. While a great deal of research has shown this ideal to be imperfect, it may influence the behavior of civil servants. That is, bureaucrats may feel an obligation to internalize the wishes of their elected superiors. Hence, I assume that the utility that an administrator assigns to an outcome is a linear combination of his own utility function and that of the supervisor:

$$r_i \cdot \alpha_s \cdot e_i + (1 - r_i) \cdot \alpha_i \cdot e_i$$

The r parameter, which ranges between 0 and 1, indicates the weight that an administrator assigns to conformity with the supervisor's preferences. A civil servant for whom $r_i = 1$ would be completely sympathetic with the supervisor's wishes and ignore his own, while a person for whom $r_i = 0$ would disregard the supervisor's preferences entirely and weigh only his own feelings about the policy in making a decision.

3.2 Deriving Predictions from the Model

With the necessary assumptions in place, I can now derive predictions from the model. Unfortunately, closed-form analytic expressions for the equilibrium probability of play cannot be derived. Therefore, I begin by presenting the Quantal Response equilibrium best response functions for bureaucrats and supervisors. I then present comparative static predictions on the quantal response functions of bureaucrats; that is, I predict how bureaucrat effort levels will change with changes in the environment *ceteris paribus*. Finally, I present numerical solutions for equilibrium play using a technique implemented in freely-available software. These numerical examples serve to establish that budget competitions induce productive effort from bureaucrats, even in cases of an ideologically heterogeneous bureaucracy. All

these techniques present a number of behavioral predictions that can be confirmed or falsified via empirical analysis.

3.2.1 Quantal Response Functions

The supervisor's choice of which bureaucrat to award the bonus has already been described in equation 3.4:

$$\pi_i(e_i, e_{-i}) = \frac{\exp\left[\lambda_s e_i\right]}{\sum_{j=1}^{n} \exp\left[\lambda_s e_j\right]}$$

As described above, a bureaucrat will choose the effort level that maximizes his expected utility, subject to the proviso that his judgment of expected utility is error-prone. The only uncertain element of 3.5 is the effort level exerted by other players, e_{-i}. Hence, the expression for expected probability of victory is:

$$E\left[\pi_i(e_{ia}, e_{-i})\right] = \tilde{\pi}_i = \frac{\exp(\lambda_s e_{ia})}{\exp(\lambda_s e_{ia}) + E\left[\sum_{j \neq i} \exp(\lambda_s e_{jc})\right]} \tag{3.7}$$

The bureaucrat's expected utility can therefore be written as:

$$\tilde{u}_i = w + \tilde{\pi}_i \cdot B + r_i \cdot \alpha_s \cdot e_i + (1 - r_i) \cdot \alpha_i \cdot e_i - c_i(e_i) \tag{3.8}$$

Substituting 3.8 into 3.2, the probability that bureaucrat i exerts effort level e_{ia} is:

$$\Pr(e_i = e_{ia}) = \frac{\exp(\lambda_b \tilde{u}_i(e_{ia}))}{\int \exp(\lambda_b \tilde{u}_i(e_i)) de_i} \tag{3.9}$$

At equilibrium, all bureaucrats exert effort consistent with equation 3.9. Note that the $\tilde{u}_i(e_{ia})$ term in 3.9 is contingent on an expectation of other agents' play which, if players follow Bayes' rule, should correspond to the actual expected play of others in the long run. The existence of at least one equilibrium is guaranteed (McKelvey and Palfrey, 1995, p. 10).

3.2.2 Quantal Response Comparative Statics

Although analytic derivation of closed-form probabilities of play is intractable, we can use the bureaucrat's quantal response function in 3.9 to determine how her response dynamics will change as aspects of the environment change. The relationship between a bureaucrat's effort choice e_i and several of these environmental factors is of particular interest; these factors are rivals' effort, the size of the budget bonus, the cost of effort, a bureaucrat's professionalism, the degree of error present in supervision, and a bureaucrat's ideological congruence with the assigned task.

As a preliminary step, I examine the relationship between a bureaucrat's expected utility and the effort level exerted. As I will show, many of the comparative static relationships above rely on the utility earned at some e_i. To understand these relationships, it will therefore be helpful to understand how utility and effort relate. Taking the derivative of bureaucrat i's expected utility 3.8 with respect to own effort e_i, we find:

$$\frac{\partial \tilde{u}_i}{\partial e_i} = \frac{\partial \tilde{\pi}_i}{\partial e_i} * B + r_i * \alpha_s + (1 - r_i) * \alpha_i - c_i'(e_i) \tag{3.10}$$

In order to come to a closed form solution, we will assume a cost function as in 3.6; hence,

$$c_i'(e_i) = 2\beta_i e_i \tag{3.11}$$

Taking derivatives of 3.7 with respect to e_i:

$$\frac{\partial \tilde{\pi}_i}{\partial e_i} = \lambda_s * \tilde{\pi}_i * (1 - \tilde{\pi}_i) \tag{3.12}$$

Substituting 3.11 and 3.12 into 3.10:

$$\frac{\partial \tilde{u}_i}{\partial e_i} = \left[\lambda_s * \tilde{\pi}_i * (1 - \tilde{\pi}_i)\right] B + r_i * \alpha_s + (1 - r_i) * \alpha_i - 2\beta_i e_i$$

Hence, the relationship between expected utility and effort depends primarily on the relationship between cost and the value of the budgetary bonus adjusted for probability of victory. When the expected value of the bonus at some effort level exceeds the cost of that effort level, utility increases as effort increases; otherwise, utility decreases as effort decreases. The influences of professionalism and ideological congruence add to (or subtract from) the value of the budget bonus. A completely professional bureaucrat adds the supervisor's marginal utility from effort α_s to his own $\tilde{u}_i'$. If a bureaucrat tends to assign the same ideological value to a task as the superior, then a less-professional bureaucrat does the same. If a bureaucrat tends to overvalue (or undervalue) a task relative to the superior, then $\tilde{u}_i'$ rises (or falls) as the bureaucrat becomes less professional.

Relationship between Productivity and Rivals' Effort

How will a bureaucrat's productivity change as rivals increase their productivity? To abbreviate notation, let:

$$Q = E\left[\sum_{j \neq i} \exp(\lambda_s e_{jc})\right]$$

Taking the derivative of 3.9with respect to expected utility, we find:

$$\begin{aligned}\frac{\partial \Pr(e_i = e_{ia})}{\partial Q} &= \left[\int \exp(\lambda_b \tilde{u}_i(e_i)) de_i\right]^{-1} \lambda_b \exp(\lambda_b \tilde{u}_i(e_{ia})) \frac{\partial \tilde{u}_i(e_{ia})}{\partial Q} \\ &\quad + \exp(\lambda_b \tilde{u}_i(e_{ia})) \left[-\frac{\int \left(\lambda_b \exp(\lambda_b \tilde{u}_i(e_i)) \frac{\partial \tilde{u}_i(e_i)}{\partial Q}\right) de_i}{\left(\int \exp(\lambda_b \tilde{u}_i(e_i)) de_i\right)^2}\right]\end{aligned}$$

$$
\begin{aligned}
&= \lambda_b \Pr(e_i = e_{ia}) \frac{\partial \tilde{u}_i(e_{ia})}{\partial EQ} - \lambda_b \frac{\Pr(e_i = e_{ia})}{\int \exp(\tilde{u}_i(e_i))de_i} \int \left(\exp(\tilde{u}_i(e_i)) \frac{\partial \tilde{u}_i(e_i)}{\partial Q} \right) de_i \\
&= \lambda_b \Pr(e_i = e_{ia}) \left[\frac{\partial \tilde{u}_i(e_{ia})}{\partial Q} - \frac{\int \left(\exp(\lambda_b \tilde{u}_i(e_i)) \frac{\partial \tilde{u}_i(e_i)}{\partial Q} \right) de_i}{\int \exp(\lambda_b \tilde{u}_i(e_i))de_i} \right] \qquad (3.13)
\end{aligned}
$$

The first term in brackets in equation 3.13 is the change in utility from effort level e_{ia} as the sum of others' (exponentiated) effort increases. The second term is the weighted average change in utility for all effort levels, where the weight is $\exp(\tilde{u}_i(e_i))$, the exponentiated utility from effort level e_i. It is easily shown that the change in utility from an increase in others' output is negative:

$$
\frac{\partial \tilde{u}_i(e_{ij})}{\partial E \left[\sum_{j \neq i} \exp(\lambda_s e_{jc}) \right]} = -B \left[\frac{\exp(\lambda_s e_{ia})}{\left(\exp(\lambda_s e_{ia}) + E \left[\sum_{j \neq i} \exp(\lambda_s e_{jc}) \right] \right)^2} \right] \qquad (3.14)
$$

Expected utility is strictly decreasing in increased output by rivals. Hence, as other agents' output increases, the direction of effort change is contingent on two factors: the current probability that the effort level will be exerted, and the extent to which the speed of utility change deviates from the average. Effort levels whose expected utility decreases faster than average as others' output rises will be exerted less often by a bureaucrat. Effort levels whose expected utility decreases slower than average will be exerted more often. The speed of utility decline is given by 3.14, which can be immediately recognized as an upside-down probability density with minimum value at:

$$
\exp(\lambda_s e_{ia}) = E \left[\sum_{j \neq i} \exp(\lambda_s e_{jc}) \right] \qquad (3.15)
$$

Hence, the probability of exerting effort levels far above and below the average output value of other agents will tend to be least-affected by changes in rivals' effort. Those closest to meeting the condition 3.15 will tend to be most affected.

Relationship between Productivity and Budget Incentives

Intuitively, one expects effort levels to rise as the rewards for effort increase. We can analytically confirm this intuition in a straightforward manner. Taking the derivative of 3.9 with respect to budget rewards B:

$$
\begin{aligned}
\frac{\partial \Pr(e_i = e_{ia})}{\partial B} &= \left[\int \exp(\lambda_b \tilde{u}_i(e_i))de_i \right]^{-1} \lambda_b \exp(\lambda_b \tilde{u}_i(e_{ia})) \frac{\partial \tilde{u}_i(e_{ia})}{\partial B} \\
&\quad + \exp(\lambda_b \tilde{u}_i(e_{ia})) \left[- \frac{\int \left(\lambda_b \exp(\lambda_b \tilde{u}_i(e_i)) \frac{\partial \tilde{u}_i(e_i)}{\partial B} \right) de_i}{\left(\int \exp(\lambda_b \tilde{u}_i(e_i))de_i \right)^2} \right]
\end{aligned}
$$

$$
\begin{aligned}
&= \lambda_b \Pr(e_i = e_{ia}) \frac{\partial \tilde{u}_i(e_{ia})}{\partial B} - \lambda_b \Pr(e_i = e_{ia}) \frac{\int \left(\exp(\tilde{u}_i(e_i)) \frac{\partial \tilde{u}_i(e_i)}{\partial B} \right) de_i}{\int \exp(\tilde{u}_i(e_i)) de_i} \\
&= \lambda_b \Pr(e_i = e_{ia}) \left[\frac{\partial \tilde{u}_i(e_{ia})}{\partial B} - \frac{\int \left(\exp(\tilde{u}_i(e_i)) \frac{\partial \tilde{u}_i(e_i)}{\partial B} \right) de_i}{\int \exp(\tilde{u}_i(e_i)) de_i} \right] \qquad (3.16)
\end{aligned}
$$

Again, we have a relationship that is primarily determined by the difference between the change in utility with respect to the environmental variable (B) and the weighted average change in utility with respect to that variable. The change in utility with respect to B is:

$$
\frac{\partial \tilde{u}_i(e_{ia})}{\partial B} = \tilde{\pi}_i(e_{ia}, e_{-i})
$$

This expression is strictly positive. Hence, effort levels with a probability of winning higher than the utility-weighted average probability of winning are more likely to be played as the budget reward rises. Effort levels with a probability of winning less than the utility-weighted average are less likely to be played as the budget reward rises.

Relationship between Productivity and Cost Levels

The relationship between productivity and cost levels is also analytically revealed to agree with our intuition: if effort becomes more costly, then less effort will be exerted. It is necessary to assume a cost function as in 3.6 in order to easily relate changes in effort with the slope of the cost function, as equation 3.6 allows us to summarize this slope in the β_i parameter. Taking the derivative of 3.9 with respect to β_i:

$$
\begin{aligned}
\frac{\partial \Pr(e_i = e_{ia})}{\partial \beta_i} &= \left[\int \exp(\lambda_b \tilde{u}_i(e_i)) de_i \right]^{-1} \lambda_b \exp(\lambda_b \tilde{u}_i(e_{ia})) \frac{\partial \tilde{u}_i(e_{ia})}{\partial \beta_i} \\
&\quad + \exp(\lambda_b \tilde{u}_i(e_{ia})) \left[- \frac{\int \left(\lambda_b \exp(\lambda_b \tilde{u}_i(e_i)) \frac{\partial \tilde{u}_i(e_i)}{\partial \beta_i} \right) de_i}{\left(\int \exp(\lambda_b \tilde{u}_i(e_i)) de_i \right)^2} \right]
\end{aligned}
$$

$$
\begin{aligned}
&= \lambda_b \Pr(e_i = e_{ia}) \frac{\partial \tilde{u}_i(e_{ia})}{\partial \beta_i} - \lambda_b \Pr(e_i = e_{ia}) \frac{\int \left(\exp(\lambda_b \tilde{u}_i(e_i)) \frac{\partial \tilde{u}_i(e_i)}{\partial \beta_i} \right) de_i}{\int \exp(\lambda_b \tilde{u}_i(e_i)) de_i} \\
&= \lambda_b \Pr(e_i = e_{ia}) \left[\frac{\partial \tilde{u}_i(e_{ia})}{\partial \beta_i} - \frac{\int \left(\exp(\lambda_b \tilde{u}_i(e_i)) \frac{\partial \tilde{u}_i(e_i)}{\partial \beta_i} \right) de_i}{\int \exp(\lambda_b \tilde{u}_i(e_i)) de_i} \right] \qquad (3.17)
\end{aligned}
$$

By now, the pattern is familiar: the relationship between costs and effort is determined by whether that effort level's change in utility with respect to β_i is greater or lesser than the

exponentiated-utility-weighted average of this value. Taking the derivative of utility with respect to β_i, we find:

$$\frac{\partial \tilde{u}_i(e_{ia})}{\partial \beta_i} = -e_i^2$$

Utility is strictly declining in increasing β_i. It may be helpful to substitute the above into 3.17 and rearrange terms:

$$\frac{\partial \Pr(e_i = e_{ia})}{\partial c(e_i)} = \lambda_b \Pr(e_i = e_{ia}) \left[\frac{\int (\exp(\lambda_b \tilde{u}_i(e_i)) e_i^2)\, de_i}{\int \exp(\lambda_b \tilde{u}_i(e_i)) de_i} - e_i^2 \right] \tag{3.18}$$

As β_i rises, generally speaking, the probability of playing higher effort levels falls and the probability of playing lower effort levels rises. More specifically, $\Pr(e_i > e_{i0})$ will fall and $\Pr(e_i < e_{i0})$ will rise, where:

$$e_{i0} = \sqrt{\frac{\int (\exp(\lambda_b \tilde{u}_i(e_i)) e_i^2)\, de_i}{\int \exp(\lambda_b \tilde{u}_i(e_i)) de_i}}$$

Relationship between Productivity and Professionalism

What of the relationship between productivity and a bureaucrat's degree of professionalism, parameterized by r_i? Based on intuition and the relationship between utility and effort above, we might expect the influence of professionalism to be contingent. That is, when the supervisor and the bureaucrat agree, professionalism does not affect productivity. When the bureaucrat prefers more (or less) effort than the supervisor, the degree of professionalism dictates just how much more (or less) effort the bureaucrat will actually exert. Taking the derivative of 3.9 with respect to r_i:

$$\begin{aligned} \frac{\partial \Pr(e_i = e_{ia})}{\partial r_i} &= \left[\int \exp(\lambda_b \tilde{u}_i(e_i)) de_i \right]^{-1} \lambda_b \exp(\lambda_b \tilde{u}_i(e_{ia})) \frac{\partial \tilde{u}_i(e_{ia})}{\partial r_i} \\ &\quad + \exp(\lambda_b \tilde{u}_i(e_{ia})) \left[- \frac{\int \left(\lambda_b \exp(\lambda_b \tilde{u}_i(e_i)) \frac{\partial \tilde{u}_i(e_i)}{\partial r_i} \right) de_i}{\left(\int \exp(\lambda_b \tilde{u}_i(e_i)) de_i \right)^2} \right] \\ &= \lambda_b \Pr(e_i = e_{ia}) \frac{\partial \tilde{u}_i(e_{ia})}{\partial r_i} - \lambda_b \Pr(e_i = e_{ia}) \frac{\int \left(\exp(\lambda_b \tilde{u}_i(e_i)) \frac{\partial \tilde{u}_i(e_i)}{\partial r_i} \right) de_i}{\int \exp(\lambda_b \tilde{u}_i(e_i)) de_i} \\ &= \lambda_b \Pr(e_i = e_{ia}) \left[\frac{\partial \tilde{u}_i(e_{ia})}{\partial r_i} - \frac{\int \left(\exp(\lambda_b \tilde{u}_i(e_i)) \frac{\partial \tilde{u}_i(e_i)}{\partial r_i} \right) de_i}{\int \exp(\lambda_b \tilde{u}_i(e_i)) de_i} \right] \end{aligned} \tag{3.19}$$

The pattern presents itself again. Taking the derivative of utility with respect to r_i:

$$\frac{\partial \tilde{u}_i(e_{ia})}{\partial r_i} = (\alpha_s - \alpha_i)\, e_{ia}$$

Substituting the above into 3.19 and collecting terms:

$$\frac{\partial \Pr(e_i = e_{ia})}{\partial r_i} = \lambda_b \left(\alpha_s - \alpha_i\right) \Pr(e_i = e_{ia}) \left(e_{ia} - \frac{\int \left(\exp(\lambda_b \tilde{u}_i(e_i)) e_i\right) de_i}{\int \exp(\lambda_b \tilde{u}_i(e_i)) de_i} \right) \tag{3.20}$$

Equation 3.20 confirms our intuition. If the gap between the supervisor's preferences and bureaucrat's preferences $\alpha_s - \alpha_i$ is zero, then changes in professionalism are inconsequential; they grow in consequence as the gap grows. More effort (or less effort) tends to be exerted with increasing professionalism if the supervisor values effort more (or less) than the bureaucrat.

Relationship between Productivity and Error of Supervision

In examining the relationship between bureaucratic effort levels and the level of error in supervision, intuition is less-useful as a predictor of behavior. It is therefore best to move directly to analysis; fortunately, the pattern of derivatives has become familiar enough that intermediate steps can be omitted from the exposition. Hence, we can move straight to the following expression:

$$\frac{\partial \Pr(e_i = e_{ia})}{\partial \lambda_s} = \lambda_b \Pr(e_i = e_{ia}) \left[\frac{\partial \tilde{u}_i(e_{ia})}{\partial \lambda_s} - \frac{\int \left(\exp(\tilde{u}_i(e_i)) \frac{\partial \tilde{u}_i(e_i)}{\partial \lambda_s} \right) de_i}{\int \exp(\tilde{u}_i(e_i)) de_i} \right] \tag{3.21}$$

The derivative of utility with respect to error in supervision is somewhat more involved than usual:

$$\begin{aligned} \frac{\partial \tilde{u}_i(e_{ia})}{\partial \lambda_s} &= \left(\exp(\lambda_s e_{ia}) + E\left[\sum_{j \neq i} \exp(\lambda_s e_{jc}) \right] \right)^{-1} \exp(\lambda_s e_{ia}) e_{ia} \\ &\quad - \frac{\left(\exp(\lambda_s e_{ia}) e_{ia} + E\left[\sum_{j \neq i} \exp(\lambda_s e_{jc}) e_{jc} \right] \right)}{\left(\exp(\lambda_s e_{ia}) + E\left[\sum_{j \neq i} \exp(\lambda_s e_{jc}) \right] \right)^2} \exp(\lambda_s e_{ia}) \\ &= \tilde{\pi}_i(e_{ia}, e_{-i}) \left(e_{ia} - \frac{\exp(\lambda_s e_{ia}) e_{ia} + E\left[\sum_{j \neq i} \exp(\lambda_s e_{jc}) e_{jc} \right]}{\exp(\lambda_s e_{ia}) + E\left[\sum_{j \neq i} \exp(\lambda_s e_{jc}) \right]} \right) \end{aligned} \tag{3.22}$$

Note that3.21 looks very much like 3.22; both are weighted averages. In the case of 3.22, effort levels are averaged across bureaucrats with exponentiated effort level as the weight. Hence, the change in utility with respect to λ_s is positive or negative depending on whether

the specific effort level is above or below the weighted average effort of all bureaucrats. This is unsurprising: as the supervisor becomes better at spotting higher effort levels, the value of effort above the average level to a bureaucrat rises because the probability of receiving the bonus rises. Whether an effort level gets played more is a question of whether this change in utility in equation 3.22 is greater than its weighted average, with exponentiated utility as the weight. Effort levels with a greater than average increase in utility are more likely to be exerted, with the converse effect for effort levels with a smaller than average increase in utility.

Relationship between Productivity and Ideological Congruence with the Task

Finally, we examine how a bureaucrat's ideological agreement (or disagreement) with the task affects the effort level he exerts. Intuitively, a bureaucrat might be expected to exert greater effort at tasks that he ideologically agrees with, and less effort at those he does not. My analysis largely confirms this intuition. The derivative of a bureaucrat's probability of exerting effort level e_i with respect to α_i is:

$$\frac{\partial \Pr(e_i = e_{ia})}{\partial \alpha_i} = \lambda_b \Pr(e_i = e_{ia}) \left[\frac{\partial \tilde{u}_i(e_{ia})}{\partial \alpha_i} - \frac{\int \left(\exp(\tilde{u}_i(e_i)) \frac{\partial \tilde{u}_i(e_i)}{\partial \alpha_i} \right) de_i}{\int \exp(\tilde{u}_i(e_i)) de_i} \right] \tag{3.23}$$

And the derivative of bureaucrat utility with respect to α_i is:

$$\frac{\partial \tilde{u}_i(e_{ia})}{\partial \alpha_i} = (1 - r_i) * e_{ia} \tag{3.24}$$

Substituting 3.23 into 3.24:

$$\frac{\partial \Pr(e_i = e_{ia})}{\partial \alpha_i} = \lambda_b (1 - r_i) \Pr(e_i = e_{ia}) \left[e_{ia} - \frac{\int (\exp(\tilde{u}_i(e_i)) e_i) \, de_i}{\int \exp(\tilde{u}_i(e_i)) de_i} \right] \tag{3.25}$$

Equation 3.25 reveals that the change in effort related to a change in one's own ideological congruence with the task depends, as usual, on a weighted average. If e_{ia} is larger than the average e_i weighted by exponentiated utility, then the probability of exerting this effort level rises; otherwise, this probability declines. This relationship weakens as professionalism rises.

An expression similar to3.25 can be created for the relationship between effort and the supervisor's ideological congruence with the task:

$$\frac{\partial \Pr(e_i = e_{ia})}{\partial \alpha_s} = \lambda_b r_i \Pr(e_i = e_{ia}) \left[e_{ia} - \frac{\int (\exp(\tilde{u}_i(e_i)) e_i) \, de_i}{\int \exp(\tilde{u}_i(e_i)) de_i} \right]$$

Here, the probability of playing e_{ia} rises with a rise in α_s if e_{ia} is greater than its utility-weighted average, and falls otherwise. Again, the relationship is related to professionalism: it strengthens as professionalism rises.

3.2.3 Numerically Computing Exact Equilibria

The expression determining bureaucratic output, equation 3.9, is difficult to work with analytically. A closed form solution for the exact distribution of play at equilibrium, $f^*(e_i)$, would entail the analytical solution of a system of n equations with the form in 3.9. However, numerical solutions to such a system are obtained with relative ease using existing software. In this section, I will present the technique used to obtain these solutions and present some sample solved equilibria.

To compute equilibria, I will use a discrete approximate to equation 3.9 of the following form:

$$\Pr(e_i = e_{ia}) = \frac{\exp(\lambda_b \tilde{u}_i(e_{ia})])}{\sum_{w \in S_i} \exp(\lambda_b \tilde{u}_i(e_{iw}))} \tag{3.26}$$

Here, S_i is a bounded grid of points in one-dimensional unbounded effort space. The fineness of this grid can be arbitrarily increased to improve the accuracy of the solution, up to the limits of computability. With this approximation in hand, we can calculate equilibrium predicted probabilities for any given constellation of parameters. The technique I employ was introduced by Theodore Turocy (2005). In the article, he describes the calculation of a limiting logistic quantal response equilibrium using the *homotopy principle*. As Turocy describes it (quoting Govindan and Wilson (2003)), the homotopy principle recommends the following approach to computing an equilibrium:

> Given a system of equations whose zeroes one wants to compute, first deform the system to one with a unique easily-computed solution, then reverse the deformation to trace (a selection of) solutions of the associated systems along the way to find a solution of the original system at the terminus (Govindan and Wilson, 2003, p. 66).

In this case, the system with the easily computed solution is the quantal response equilibrium for the case where $\lambda_b = 0$: at this point, all efforts are played with equal probability (choice is random). Suppose that $y(s)$ is $(e(s), \lambda(s))$, where effort and the error parameter are parameterized by a new variable s. $y(s)$ has $J + 1$ elements, where $J = \sum_{i=1}^{n} Q_i$ and Q_I is the cardinality of the set of i's strategies S_i (the extra element in y is λ_b). Turocy proposes to compute QREs at other values of λ_b by computing $dy(s)/ds$ using a system of differential equations, up to a point where λ_b reaches a sufficiently high value that e approximates the asymptote as $\lambda_b \to \infty$. The details of Turocy's approach, which are beyond the scope of this chapter, are discussed thoroughly in Turocy (2005) and Judd (1998).

This technique for computing Quantal Response equilibria in normal form games is implemented in the GAMBIT software of McKelvey, McLennan, and Turocy (N.d.). An example of the output from this program is given in Figure 3.1, which shows a computed profile of QRE strategies at $\lambda_b = 2.6$. This equilibrium was computed for two identical bureaucrats for whom $r = 0.5$, $\alpha_s = 2$, α_i =1, $\beta_i = 1$, $\lambda_s = 1$, and $B = 5$. As the Figure shows, both players exert an effort generally between $e_i = 0$ and 2, with a peak slightly above one. In fact, if we allow $\lambda_b \to \infty$, $e_i = 1.35$ is played with probability ≈ 1.

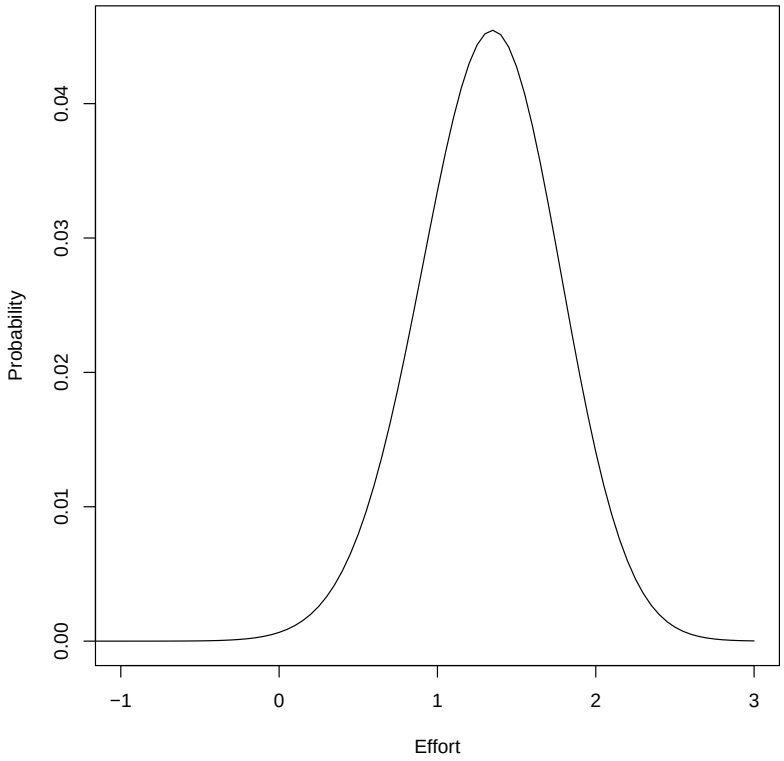


Figure 3.1: A Computed Quantal Response Equilibrium, $\lambda_b = 2.6$, $\lambda_s = 1$

One can also generate numerical comparative static predictions using the same technique. The technique is straightforward: compute the limiting logit equilibrium[4] for a variety of settings, where all parameters but one are held constant and one is allowed to vary over a certain range. Figure 3.2 demonstrates a few of these predictions.[5] For all these graphs, there are two bureaucrats, $r = 0.5$, $\alpha_s = 2$, $\alpha_i = 1$, $\beta_i = 1$, $\lambda_s = 1$, $B = 5$, and the bureaucrat's action space is confined to a grid between -3 and 3 in increments of 0.5. However, each of these graphs allows one of these parameters to vary between the values shown on the x-axis. For example, in the upper-left panel, bureaucrat 1's ideology varies between -1 and 1 in increments of 0.1.

As the top two panels in the figure reveal, increases in bureaucrat 1's ideology tend to increase her effort level. Eventually, sufficiently large increases (or decreases) can prompt strategic responses from other bureaucrats: in the top right panel, bureaucrat 1's production

[4]In the cases examined here, the limiting logit equilibrium is always a pure strategy. However, this will not always be true: mixed strategy equilibria can be limiting logit equilibria.

[5]The computations for Figure 3.2 were conducted by calling the `gambit-logit` program of McKelvey, McLennan, and Turocy (2007) within `R 2.4.1` using game data files generated by `R`. The `snow` 0.2-2 networking package (Tierney et al. 2007) was used to distribute calls to `gambit-logit` for different parameter values to a 4-CPU workstation for faster processing.

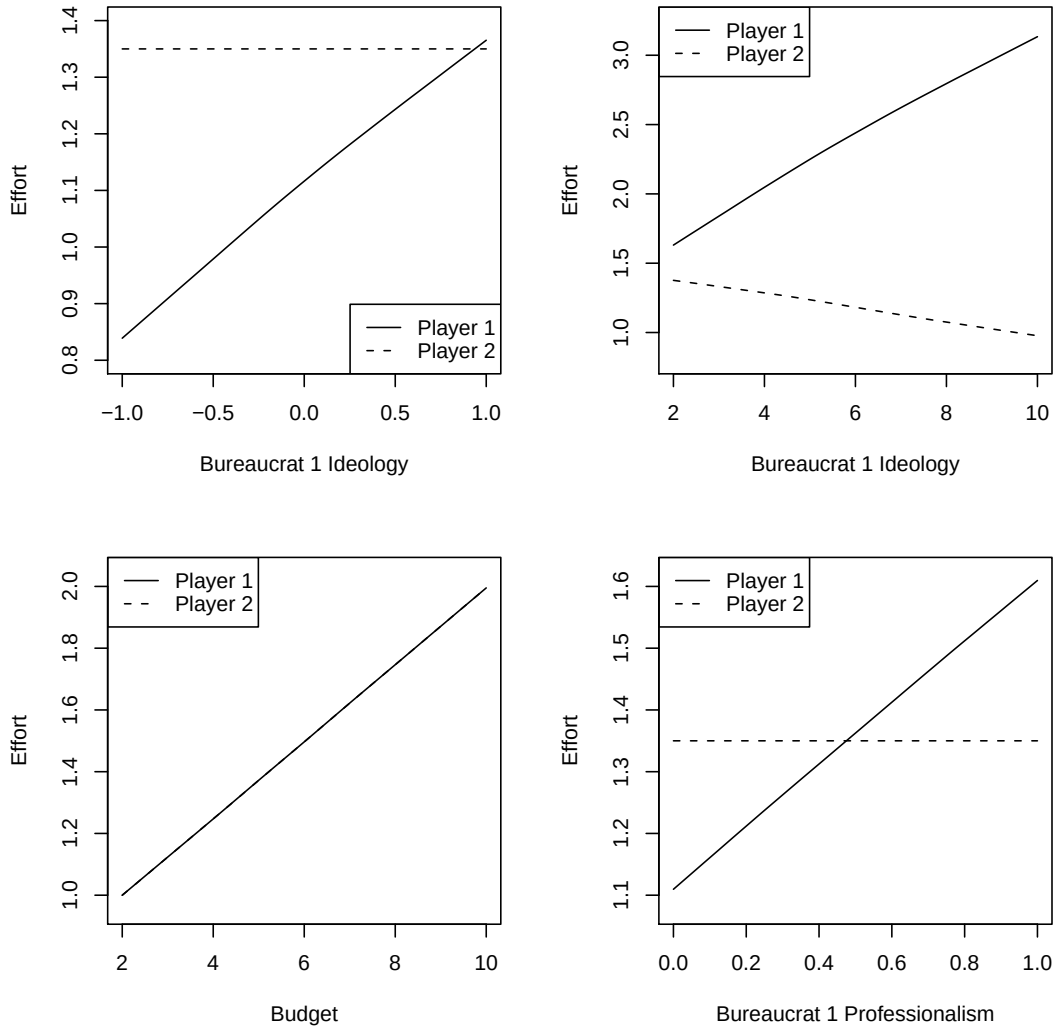


Figure 3.2: Changes in the Limiting Logistic Equilibrium, $\lambda_s = 1$

eventually grows large enough that bureaucrat 2 lowers her output (because she has so little chance of winning the budget bonus that it provides no incentive for effort.) The lower left panel reveals (unsurprisingly) that effort linearly increases in the size of the budget bonus. Similarly, increases in Player 1's professionalism linearly increase her effort, without impact on Player 2 (at least at these effort levels; with sufficiently high or low effort from Player 1, eventually Player 2 will lower her effort in response to lesser budget incentives to produce).

3.3 Conclusion

This chapter has offered a theory to explain how, if the executive can control budgets inside administrative agencies, tournament-style incentives suffice to ensure that civil servants will make policy outputs consistent with the executive's goals, even under conditions of limited information. The solution concept used to derive predictions from the model incorporates the assumption that bureaucrats and supervisors are boundedly rational. Comparative static predictions were generated, both analytical and computationally derived.

With these predictions in hand, I can proceed to the task of empirical testing. In the next chapter, I operationalize a bureaucratic budget tournament in the laboratory and test my theory's predictions. Most importantly, I test two aspects of the model. First, I determine the willingness of subjects to engage in tasks they find ideologically distasteful in order to receive expansions in their budget (cash payments in the experiment.) Secondly, I investigate whether subjects are willing to obey executive mandates without financial compensation (i.e., whether norms of subordination bind them.) It is to these tests, and tests of the other propositions above, that I now turn.

CHAPTER 4

LABORATORY EVIDENCE FOR THE POWER OF STRATEGIC BUDGETING IN AN IDEOLOGICAL ENVIRONMENT[1]

In order for strategic budgeting to be normatively and scientifically noteworthy to policy makers and political scientists, it is important to demonstrate that using this technique causes bureaucrats to comply with a supervisor's policy agenda even when the bureaucrats ideologically disagree with that agenda. A test could be run in the field, with the cooperation of a field agency, but such a test would be expensive and politically difficult to secure without existing evidence for the technique's effectiveness. In addition, that test could not definitively rule out rival causes for any change in bureaucratic responsiveness.

In this chapter, I run a laboratory experiment designed to assess the marginal impact of ideologically-charged tasks on strategic budgeting's effectiveness while ruling out all rival causes through experimental control. Subjects are asked to contribute to ideologically-formed groups at a cost to themselves. Subjects compete against each other in pairs, with a cash penalty probabalistically assigned to one of the pair according to their relative contribution. Contributions are compared to a condition where group assignment is random (non-ideological).

Some of the findings are surprising. As expected, I find that strategic budgeting remains effective as a management technique in an ideological environment, and is in fact more effective than theory predicts. Ideological congruence with the task tends to boost output (and ideological incongruence tends to depress output). However, I also find that competition against a partner with similar ideology facilitates collusion against the supervisor: both subject lower their output, so that they both need to contribute less for the same probability of receiving the penalty. This collusion only occurs when group assignment is ideological and when both competing subjects share the same ideology.

4.1 Experimental Design

In the experiment, 6 subjects are first broken into two groups, depending on the treatment. In the ideological treatment, subjects are asked about their general political ideology before any of the other aspects of the experiment are described; the three most

[1]This research was supported by National Science Foundation Grant #0720055.

conservative subjects are then placed into "the conservative group," while the three most liberal subjects are placed into "the liberal group." In the non-ideological treatment, the survey question is not asked; instead, group assignment occurs randomly into either "Group A" or "Group B."

Once assigned to a group, subjects are sorted into pairs with a counterpart. Two pairs contain members of the same group, with one conservative/conservative or A/A pairing and one liberal/liberal or B/B pairing, while the third pair contains a mixed conservative/liberal or A/B pairing. Each subject is given 10 tokens and 250 ECUs, or Economic Currency Units ($1.00 = 200 ECUs). Subjects may then choose whether to contribute tokens to one of two groups, or keep the tokens for themselves. Each token contributed to a group creates 2 ECUs for every member of that group, but costs the subject money: x tokens given to a group costs a subject x^2ECUs.

The tournament aspect of the experiment is similar to the one in described Bull, Schotter, and Weigelt (1987, see also Schotter and Weigelt, 1992). Subjects are informed that a penalty[2] of 150 ECUs will be assessed to themselves or their counterpart, with the penalty assessed according to the number of tokens that they and their counterpart give to one of the two groups, which I will call the *target group*.[3] The probability of a subject receiving the penalty is given by the cumulative distribution function of the uniform distribution: if subject X gives x tokens to the target group while subject Y gives y tokens, then subject X's probability of receiving the penalty is:

$$\Pr(X) = \frac{y}{x+y}$$

The penalty selection process is explained to the subjects as having the computer simulate placing all contributed tokens into a bag, then randomly drawing one token from the bag. If the token drawn is one's own, the counterpart received the penalty; if the token drawn is the counterpart's, than oneself received the penalty. If $x + y = 0$, as explained to the subjects, the penalty is determined by a simulated coin flip. Note that, by design, subjects cannot lose money in the experiment: the lowest possible payoff is 0 ECUs.

The principal-agent game embedded in this experiment is between the experimenter (or the computer program), who represents the collective interest of the target group, and the subjects. Subjects have the choice between contributing to the target group (at a cost to themselves) and avoiding the penalty, shirking by contributing to neither group (and avoiding the cost of contribution), or subverting the principal's wishes by contributing to the non-target group.

The experiment lasts 20 periods. For the first set of 10 periods, the target group remains the same. The second set of 10 periods switches the target group to the other group. A subject's pairing with the counterpart remains the same throughout all these periods. Hence, there are potential ordering effects in the experiment depending on which of the two groups

[2]A similar experiment could be performed using rewards instead of penalties, but this experiment uses penalties because in applied situations governments may find it easier to cut budgets than add to them. Hence, this experiment should more closely reflect the effects of strategic budgeting in the bureaucracy.

[3]Subjects were not given the label of "target group", but merely told that "A penalty of 150 ECUs will be assessed to you or your Counterpart. This penalty will be assessed on the basis of the number of tokens that you and your Counterpart gave to the <name of target> group."

was the target group first. I therefore systematically vary the ordering of the target group, with half of the sessions using the conservative group/Group A as the first target group and the other half using the liberal group/Group B as the first target group. Combined with the ideological/non-ideological group treatment variation, the overall design of the experiment is a 2x2 block. At the conclusion of the experiment, a risk-assessment task and demographic survey were administered; these treatments do not enter into the present analysis and took place strictly after the treatments analyzed here; subjects did not know about these future elements of the experiment when participating in the tournament.

4.2 Expectations from Prior Experiments

There have been several experimental studies of competition for a prize that help guide design decisions and enable the formation of expectations for the present experiment. Though none take place in an ideologically-charged environment, these studies have examined a wide variety of settings and help to shape expectations for the experiment presented in this paper. First, I expect the competitive incentives provided by the penalty tournament to be effective at motivating token contributions (i.e., effort) and equilibrium predictions to be accurate, as they have been in previous experiments. Second, I expect there to be considerable heterogeneity in the response that subjects have to these incentives. Finally, in ideological treatments, equilibrium predictions about the direction of response should be accurate, but (again) I expect a considerable degree of heterogeneity on this front.

4.2.1 Effectiveness

Past experiments on competitive incentives suggest that, on average, these incentives are effective at motivating effort. This point is demonstrated by the experiments of Bull, Schotter, and Weigelt (1987) the first attempt to examine the effectiveness of tournament incentives in the laboratory and one of the first to examine tournaments empirically in any form. The authors ran a wide variety of experiments designed to investigate a number of different influences on tournament outcomes.

The basic procedure of Bull, Schotter, and Weigelt is as follows. Subjects are randomly paired and allowed to choose an effort level e between 0 and 100, inclusive. Higher decision numbers result in quadratically increasing costs to the subject ($c = e^2/10000$). After a random number between -40 and 40 is added to this decision number to form that subject's total effort, it is compared with the total effort of his/her partner; the person with the larger effort level wins a prize of \$1.45, with the other subject earning \$0.86.[4] The procedure is repeated for 12 rounds, with the same subject pairing in each round. The authors find that, on average, subjects in this treatment make decisions that are consistent with the equilibrium decision $e = 37$. Similar results are confirmed in a replication and extension by Drago and Heywood (1989).

As noted, the paper also explores a wide variety of alternative treatments to test the robustness of tournament incentives to different environments. Convergence to equilibrium (on average) is robust to changes in the cost function that raise the equilibrium decision

[4] A coin was flipped to determine the winner in the event of a tie.

to $e = 74$, to longer experiments with more rounds of play, to play against robots, and to widening of the distribution of the random number that makes outcomes less-related to one's decision.

4.2.2 Heterogeneity in Response

Although competitive incentives are effective on average, individual response to these incentives can be heterogeneous. While the Bull, Schotter, and Weigelt experiments do confirm convergence to predicted equilibrium effort levels on average, there is a great deal of variance around this average, even in the late stages of an extended experiment where play lasts for 25 rounds. Thus, individual response to the treatment is homogeneous and can be inconsistent with the equilibrium prediction.

Bull, Schotter, and Weigelt believe that variance in tournament play, which is greater than variance in a comparable piece rate pay institution run by the authors, can be explained by the fact "that a tournament, unlike the piece rate, is a game and so requires strategic, as opposed to simply maximizing, behavior" (1987, p. 3). Note that they do not believe that the phenomenon is solely driven by information asymmetry (for example, by the difficulty of inferring what the other subject's strategy actually is). The authors conducted experiments wherein subjects competed with robots that always played a fixed decision number known to the subjects and found that substantial variance in play remained, though some was eliminated. Instead, they concluded, "one reason for the higher variance of choices in tournaments as opposed to piece rates is that the computational aspects of the maximizations in the tournaments are harder than those in the piece rates" (1987, p. 26).

This explanation is rejected by Drago and Heywood (1989), who believe that the variance observed by Bull, Schotter, and Weigelt was due to the relatively flat payoff function around the optimum. By conducting a new experiment with a more dramatically peaked payoff function around the optimum,[5] Drago and Heywood are able to reduce the level of variance considerably, so much that the variance around effort is no different than the variance around the effort outcomes of a comparable linear piece rate mechanism. Their result seems to confirm the idea that tournaments are no more computationally difficult than piece rate mechanisms.

4.2.3 Asymmetric Incentives

Finally, when subjects have unequal motivation to win the tournament, the less-motivated subject's response to competition with a more-motivated subject is itself heterogeneous. Asymmetric tournaments, where subjects face different parameters in the tournament, are of particular interest: bureaucrats are likely to be a heterogeneous population with respect to ideology, and thus may perceive different costs of exerting effort when competing for budgets.

One of the variations of the Bull, Schotter, and Weigelt experiments exposes each member of a subject pair to a different cost function, $c_{low} = e^2/25000$ and $c_{high} = 2e^2/25000$, and

[5]In this new treatment, the cost function is manipulated so that payoffs at every effort choice are equal to the payoffs of the piece rate mechanism in Bull, Schotter, and Weigelt's paper.

prizes of \$1.60 and \$0.80. Both subjects know their own cost function and the cost function of the person with whom they are matched. In this experiment, the cost-disadvantaged subjects (those with the steeper cost function c_2) choose decision numbers of 56.45 on average, far higher than their equilibrium choice of $e = 37$, while cost-advantaged subjects choose decision numbers of 75.55 on average, close to their equilibrium choice of 70. The authors speculate that the greater effort exerted by cost-disadvantaged subjects could be due to a preference for winning beyond its pecuniary value. However, as the authors point out, this explanation cannot explain why the cost-advantaged subjects do not play a best response to the cost-disadvantaged subjects' strategy: although 70 is an equilibrium, the best response to $e = 56$ for a cost-advantaged subject is $e = 86$ (Bull, Schotter and Weigelt, 1987, p. 24).

In a later paper, Schotter and Weigelt (1992) further extend the exploration of asymmetric tournaments. The basic design of their experiments is similar to the Bull, Schotter, and Weigelt experiments described above, but with the introduction of two types of asymmetric tournament, *unfair* and *uneven*. In an *unfair* tournament, subjects' total effort must be k units greater than that of his counterpart in order to win the larger prize, where both subjects know the value of k. In an *uneven* tournament, subjects have different cost functions, one steeper than the other, so that $c_i = \alpha c_j$, $\alpha > 1$. In *unfair* tournaments, the authors find that subjects exert greater effort than predicted in equilibrium, although less effort than the equilibrium of a symmetric tournament. In *uneven* tournaments, on average subjects exerted the level of effort predicted in equilibrium, but at the individual level tended to follow one of two patterns: either the cost-disadvantaged subject dropped out completely (set $e = 0$), or contributed more than would be predicted in equilibrium.

4.3 Theoretical Predictions

One advantage of the experimental environment is that the situation can be carefully designed to produce firm theoretical expectations for testing. As long as the parameters involved are numerically observable—such as the size of the penalty, the shape of the cost function, and the direct reward of contribution to a group—it is straightforward to calculate point predictions for token contribution to the target group. When parameters involved are only indirectly observable—such as the degree of ideological affinity that a subject has for the task—point predictions will not be meaningful, but comparative static predictions can still be derived by looking at the calculated change in token contributions as the partially-observable factor changes.

In terms of the theory described in Chapter 3, subjects in the experiment play the role of bureaucratic agents. The principal makes no decisions in this experiment: penalty assignment is conducted by the computer, implementing a common-knowledge rule. Hence, all tested predictions pertain to bureaucratic responses to the incentives created by strategic budgeting.

Note that Chapter 3 describes a one-shot game, not the repeated game implemented in the experiment. However, given the relatively short and obviously finite period of the experiment, expectations for behavior should not be different than those for a one-period game.

4.3.1 Benchmark Prediction

When subjects do not derive any utility from the task beyond that of direct payment, utility that might (for example) derive from minimal group identification, ideological affinity or professional regard for the assigned task, all elements of their utility function are numerically observable and a point prediction is meaningful. Let e_k be subject k's contribution to the target group, i denote a subject, j denote the subject's counterpart, and C denote contributions to the target group by all other subjects in the session. Subject i's utility in the case where the subject is a member of the target group should be:

$$u_i\left(e_i, e_j, C\right) = 250 - 150\frac{e_j}{e_i + e_j} + 2\left(e_i + e_j + C\right) - e_i^2 \tag{4.1}$$

How many tokens will subject i contribute in equilibrium? Because the tournament occurs only in pairs in this experiment, the game is simple enough to solve using traditional techniques to find a Nash equilibrium, which in this case will be identical to the limiting logit QRE (see below). Taking derivatives with respect to e_i:

$$\frac{\partial u_i}{\partial e_i} = 150\frac{e_j}{\left(e_i + e_j\right)^2} - 2e_i + 2 \tag{4.2}$$

When i maximizes his/her utility, equation 4.2 will equal zero. Note that the decisions of all other subjects, as captured in C, have dropped out of this expression. A symmetric solution can be derived for subject j:

$$\frac{\partial u_i}{\partial e_j} = 150\frac{e_i}{\left(e_i + e_j\right)^2} - 2e_j + 2 \tag{4.3}$$

Setting equations 4.2 and 4.3 equal to zero forms a system of two equations in two unknowns, which can be solved using a computer algebra system. In this case, the equilibrium prediction is $e_i = e_j \approx 4.86$. This prediction is identical to the limiting logit QRE prediction for this case, which I calculated using the Gambit software package (McKelvey, McLennan and Turocy, N.d.). Because subjects can contribute only a discrete number of tokens in the experiment, I round this prediction up to 5 tokens.

The same process can be used to generate a benchmark prediction for cases where both subjects are *not* members of the target group. In this case, the subject's utility is equation 4.1 with the $2\left(e_i + e_j + C\right)$ term removed. In this case, the calculated equilibrium is $e_i = e_j \approx 4.33$, or a discrete prediction of 4 tokens contributed to the target group in the experiment.

Finally, a benchmark prediction can be calculated for the case where one subject is a member of the target group, while the other subject is not a member of the target group. Here, one subject will have utility as given in equation 4.1, while the other will have the $2e_i + 2e_j + C$ term removed. Here, the equilibrium prediction for the subject who is a member of the target group is ≈ 4.85 or 5 tokens, while the equilibrium prediction for the non-member is ≈ 4.32 or 4 tokens. These predictions are similar (though not precisely identical) to the behavioral predictions inside of matched pairings.

Note that the difference in predicted contribution for target and non-target group members *without* ideological affinity is merely the product of the small direct benefit (2 ECUs

per contributed token) that target group members receive and non-target members do not. The experiment is not designed to produce a difference in these two levels of contribution, as this difference would be hard to statistically detect. Rather, the experiment is designed to detect a difference between ideological and non-ideological subjects, and the predictions here are benchmarks to which ideologically-charged subjects will be compared.

4.3.2 Comparative Statics: Ideological Affinity

The experimental design systematically varies whether group membership is determined by ideological affiliation or random assignment. Hence, if subjects do assign a value to group contributions over its direct monetary payoff value to them, subject behavior in this treatment will reflect that component of value. In addition, the design is able to distinguish value assignment due to minimal group affiliation, as this effect (if any) will be present in the treatment that uses random group assignment to arbitrarily labeled groups.

With ideological affinity for the task added, the utility function for the subject becomes:

$$u_i\left(e_i, e_j, C\right) = 250 - 150\frac{e_j}{e_i + e_j} + \left(2 + \alpha_i\right)\left(e_i + e_j + C\right) - e_i^2$$

The degree of ideological affinity for the task is given by α_i. Taking derivatives as before:

$$\frac{\partial u_i}{\partial e_i} = 150\frac{e_j}{\left(e_i + e_j\right)^2} - 2e_i + 2 + \alpha_i$$

Hence, by solving the system of equations as before, I can generate predictions for the number of tokens I expect subjects to contribute. Point predictions about the difference are not directly helpful, as I do not measure the precise degree of ideological affiliation with the task and theory does not provide direct guidance without such a measurement. These calculated predictions can, however, be leveraged to make a comparative static prediction about how the number of tokens played will change as the degree of ideological affiliation changes. Figure 4.1 displays a contour plot of the number of tokens played by subjects i and j as their ideological affinities change. The contours are plotted over a range of plausible values[6] for ideological affinity; the extremes of 15 and -15 depict situations where every token played is worth the ideological equivalent of 15 or -15 ECUs, respectively. As the plots depict, when holding the ideological affiliation of the other player constant, the number of tokens played by a subject in equilibrium is monotonically increasing in his or her own ideological affiliation. Translating this prediction to the experiment, positive ideological affiliation with the target group should increase the number of tokens a subject contributes to that group, while negative ideological affiliation should decrease the contribution.

The plot in Figure 4.1 raises an important point: the number of tokens played in equilibrium will vary not only according to one's own ideological affiliation, but also according to that of the counterpart. The mechanism is straightforward to understand: if one's partner is motivated by ideology to play more tokens (or fewer tokens), then my probability of winning is decreased (or increased) and my strategy may change accordingly. The contour

[6]Plots ranging even wider—for example, from -50 to 50—yield the same result, but predict token allocations outside the boundaries of 0 and 10.

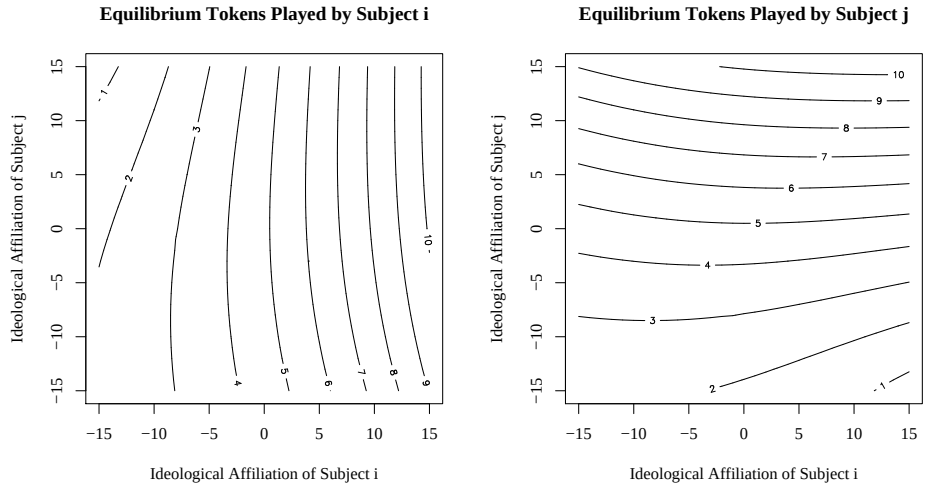


Figure 4.1: Tokens Played in Equilibrium Under Different Ideological Affinity Pairings

plot does not yield firm predictions of ideologically like vs. unlike pairings, as there are regions in the plot where increasing counterpart affinity first decreases, then increases one's own play in equilibrium. However, the plot does show that significant differences between like and unlike groups may exist if ideological valence is a factor, despite the almost-indistinguishable benchmark predictions without ideology. Hence, it will be prudent to compare like pairings to other like pairings, and unlike pairings to unlike pairings, to ensure a fair comparison.

4.4 Data and Statistical Analysis

The experiments were conducted in the xs/fs laboratory for experimental social science at Florida State University, using 102 undergraduate students as subjects (54 in the ideological treatment, 48 in the non-ideological treatment). The laboratory provides a large number of private, networked computing carrels that allow subjects to make their decisions anonymously. Each session of the experiment experiment was run using 12 subjects in front of private computer terminals, with the 12 subjects randomly assigned to two subsessions of 6 subjects simultaneously taking place in the same room. The experiments were programmed using zTree (2007), a software toolbox for economic experiments. All subjects received a $10 show-up fee in addition to what they earned in the experiment; subjects earned a total of $27.13 on average.

4.4.1 Descriptive Statistics

The vast majority (92%) of subject contributions in the data set are made to the target group. In later rounds (period 15 and later), this percentage rises to 95.7%. This outcome is

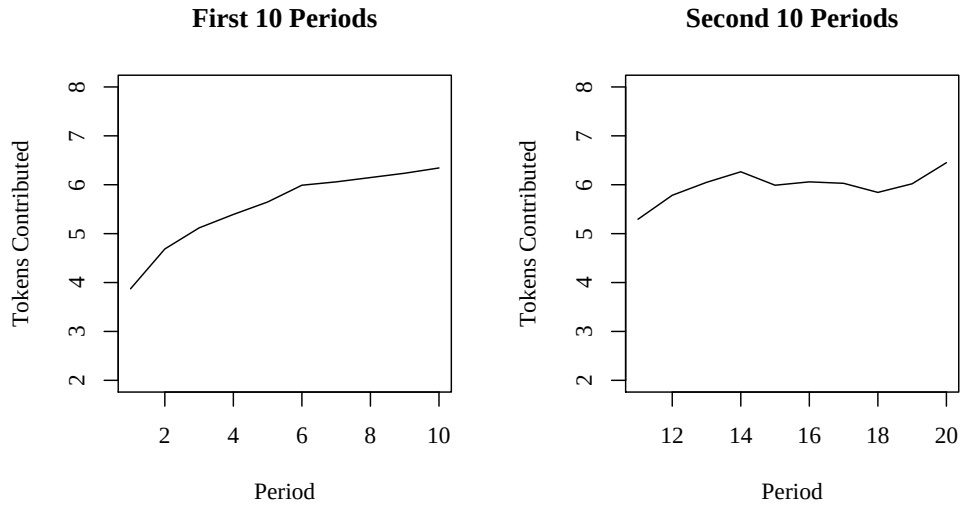


Figure 4.2: Contributions to the Target Group by Period

not surprising, as there is no financial incentive to contribute to the non-target group. Given these facts, analysis will be focused on contributions to the target group.

Figure 4.2 displays a plot of the overall average contributions to the target group over time in the 20 periods of the experiment. It seems clear that subjects at first tend to contribute at levels slightly below any of the benchmark equilibrium predictions, but then contribute progressively more over the first six periods until reaching an average level a full token above even the maximum predicted benchmark contribution. In the second ten periods of the experiment, there seems to be a modest restart effect with slightly lowered contributions, but the average contribution level quickly returns to an overall average contribution of about six tokens, again one token more than the maximum benchmark. Other than the restart effect, there seems to be little difference in the first and second sets of periods, at least on average.

Recall that, in section 4.3.2, I found reason to suspect that there may be systematic differences in the contribution levels of like and unlike pairings of subjects, particularly in the ideological treatment. An examination of overall results separated by like and unlike pairs empirically confirms this suspicions. Subjects in unlike pairs who were members in the target group contributed an average of 6.89 tokens to the target group, compared to only 5.95 tokens for members of the target group in like pairs; this difference is statistically significant ($p < .001$)[7] according to a Mann-Whitney test. Similarly, subjects in unlike pairs who were *not* members of the target group contributed an average of 6.14 tokens to the target group, compared to only 4.83 tokens for non-members of the target group in like pairs; this difference is also highly statistically significant ($p < .001$).

Due to this systematic difference among like and unlike groups, I examine the results of

[7]Unless otherwise noted, all reported p-values are for a two-tailed test.

Table 4.1: Contributions to the Target Group, Like Pairings

treatment and group membership	average	s.e.	p-value	Δbenchmark	p-value	NT
Non-ideological, in target group	5.96	0.163	-	+1.10	>0.001	320
Ideological, in target group	5.94	0.154	-	+1.08	>0.001	360
difference	0.021	-	0.881			
Non-ideological, not in target group	5.35	0.163		+1.02	>0.001	320
Ideological, not in target group	4.36	0.154		+0.03	0.448	360
difference	0.986	-	>0.001			

The statistical significance of the difference between ideological and non-ideological treatments was determined using a non-parametric Mann-Whitney rank-sum test. The statistical significance of the difference between overall outcomes and predicted benchmarks was determined using the non-parametric Wilcoxon signed-rank test. Reported p-values are substantively similar to those derived from standard ANOVA analysis.

these pairings separately. Table 4.1 shows averaged contributions to the target group among like pairings (that is, pairings where both subjects are members of the same group). The results are arranged separately according to whether subjects are members of the target group, as different theoretical expectations exist for these two cases as described above.

As shown, there is little substantive difference between the ideological and non-ideological treatments when subjects are members of the target group: in both cases, the average contribution is about 1 token greater than the benchmark prediction of 4.86. In both cases, the outcome is statistically distinguishable from the benchmark.

When subjects are not members of the target group, there is a detectable difference between the ideological and non-ideological treatments. Subjects in the non-ideological treatment have contribution levels that are very similar to those of target group members. Subjects in the ideological treatment, however, contribute 4.36 tokens on average, a result not statistically distinguishable from the benchmark prediction of 4.33 tokens. This contribution level is nearly a full token lower than that of the non-ideological treatment, a statistically significant difference.

Table 4.2 displays averaged contributions to the target group among unlike pairings wherein subjects are members of different groups. In the case where subjects were members of the target group, those in the ideological treatment contributed 1.38 more tokens than their counterparts in the non-ideological treatment, a highly statistically significant difference. In both cases, contributions are well over the benchmark theoretical prediction, with contributions in the non-ideological treatment similar to contribution levels to target group members in like pairs.

Results for subjects who were not members of the target group seem to tell essentially the same story. Again, subjects in the ideological treatment contribute more tokens than those in the non-ideological treatment, and the difference is statistically significant though substantively smaller than among target group members (0.587 tokens instead of 1.38 tokens). Also as before, contribution levels are significantly higher than their theoretical benchmark, a token and a half in the case of subjects in the non-ideological treatment and

Table 4.2: Contributions to the Target Group, Unlike Pairings

	average	s.e.	p-value	Δbenchmark	p-value	NT
Non-ideological, target group member	6.16	0.231	-	+1.31	>0.001	160
Ideological, target group member	7.54	0.393	-	+2.69	>0.001	180
difference	-1.38	-	>0.001			
Non-ideological, not in target group	5.83	0.231	-	+1.51	>0.001	160
Ideological, not in target group	6.41	0.218	-	+2.09	>0.001	180
difference	-0.587	-	0.0886			

The statistical significance of the difference between ideological and non-ideological treatments was determined using a non-parametric Mann-Whitney rank-sum test. The statistical significance of the difference between overall outcomes and predicted benchmarks was determined using the non-parametric Wilcoxon signed-rank test. Reported p-values are substantively similar to those derived from standard ANOVA analysis.

over two tokens in the case of subjects in the ideological treatment.

One other aspect of the data deserves mention: as has been observed in previous tournament experiments, there is a great deal of variance in the data at the individual level. The distribution of tokens contributed to the target group is shown in Figure 4.3. As has been observed in previous experiments (Bull, Schotter and Weigelt, 1987; Schotter and Weigelt, 1992), a great deal of heterogeneity in play exists among individual participants despite convergence to a stable average. Nor does this heterogeneity disappear over time, as Figure 4.4 shows: even in the second-to-last period of the experiment, there is a substantial degree of difference in the individual play choices of subjects.

4.4.2 Panel Analysis

Experimental methodology can eliminate many of the threats to inference that typically exist in a panel data set. In particular, random assignment of the key regressors ensures that omitted variable bias is not an issue. As bias from unobserved unit heterogeneity or omitted dynamic terms is a form of omitted variable bias, we need not worry about these threats to the validity of estimated coefficients. However, experimental methods cannot eliminate several potential problems with the data.

Threats to Inference

First, by design, the range of possible contributions is limited: subjects cannot destroy contributions to the target group by contributing negative tokens (though they can contribute to a different, opposed group), nor can they risk losing money by contributing more than 10 tokens. Furthermore, there is evidence of crowding at these barriers at both ends of the spectrum in Figure 4.3. This censoring process biases coefficients estimated by OLS regression, as it introduces correlation between the error and regressors (namely, those regressors that predict token contributions outside the boundary). In addition, the standard

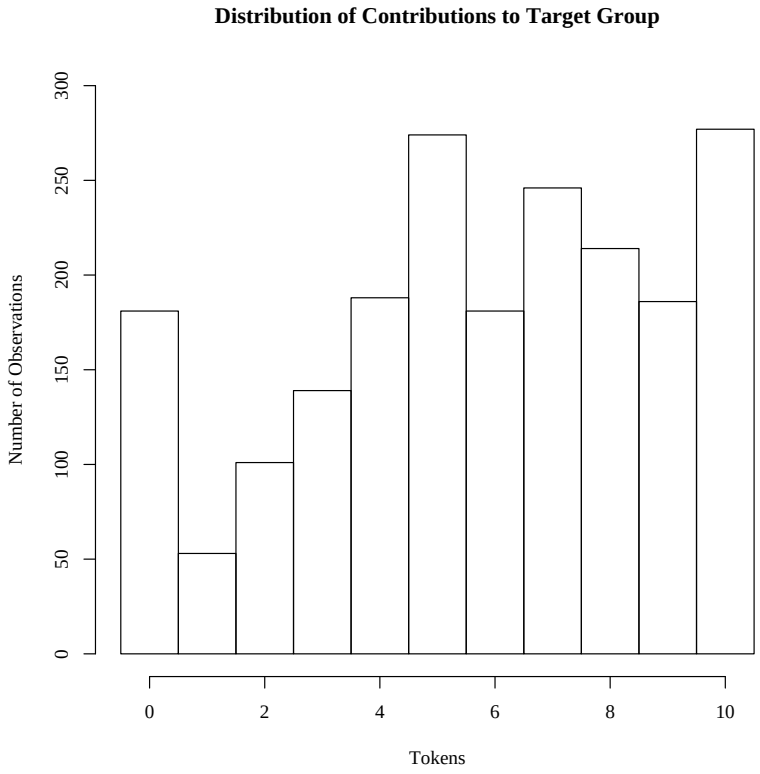


Figure 4.3: Distributions of Contributions to the Target Group

errors of the estimate can be deflated due to the crowding of observations at the barriers; only the bottom half of the error distribution is observed for cases with predicted contributions close to 10, and only the top half of the error distribution is observed for cases with predicted contributions close to 0.

Second, despite the fact that unobserved unit-level heterogeneity cannot be correlated with any of the independent variables due to random assignment, OLS regression may still yield incorrect estimates of the standard error. Hence, though magnitudes of estimated effects will not be influenced, it will be necessary to correct for unit heterogeneity to ensure that I can accurately determine whether these effects are statistically distinguishable from zero.

Finally, as indicated by Figure 4.2, there is certainly a trend in the data set over time. As the plot shows an apparently curvilinear growth over the first 10 periods, with a similar but less dramatic pattern in the second 10 periods. The plot also implies that the initial contribution level of the second set of periods starts higher than that of the first.

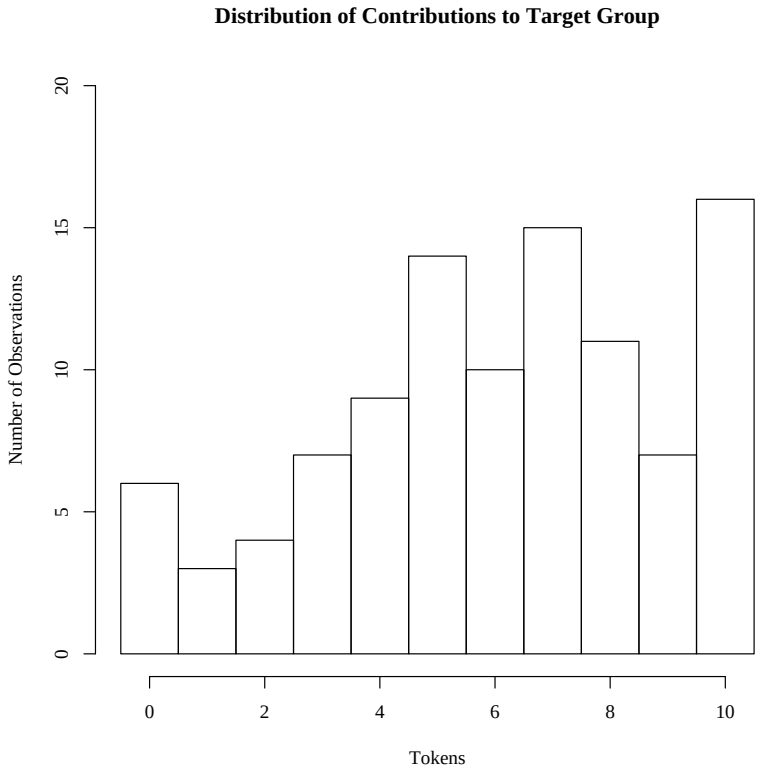


Figure 4.4: Distributions of Contributions to the Target Group, Period 19

Model Selection

Statistical approaches exist to deal with these problems. Censored observations can be explicitly modeled using Tobit regression, which separately models the process of (i) being censored and (ii) the observed value of the dependent variable conditional on not being censored. The problem of unit heterogeneity can be handled using a random effects approach, especially appropriate for this setting as zero correlation between the regressors and the unit effect is guaranteed by random assignment. Correcting for both of these problems simultaneously allows for two related choices: a random-effects Tobit model, or a censored negative binomial regression (Hilbe, 2005, 2007) that recognizes the count nature of the data, with standard errors clustered by subject to correct for panel heterogeneity. The random-effects Tobit model was preferred by the Akaike Information Criterion for every model I estimated; hence, these results are presented.

Controls are included to handle the trend and ordering effects present in the data. A logged period variable can effectively model the trend present in the data. An intercept term for the first set of sessions can handle the difference in the first and second sessions.

Table 4.3: RE Tobit, Pooled Observations

variable	coefficient	p-value
target is own group	0.666	>0.001
ideological groups	-0.317	0.447
own target * ideological groups	1.16	>0.001
Controls		
first sequence	-1.94	>0.001
ln periods 1-10	1.31	>0.001
ln periods 11-20	0.428	0.001
constant	5.01	>0.001

Results derived from a random-effects Tobit model estimated in Stata 9.2. N=102, T=20, NT=2040. 181 observations were left-censored, 277 were right-censored.

Results

Table 4.3 presents results for all observations, pooling subjects that were in like and unlike pairings. The results indicate that ideological affinity does not, on average, add significantly to the rebellion of subjects against contributions to a group that is not their own: while subjects in the ideological treatment contribute $\frac{1}{3}$ of a token less when the target group is not their own compared to subjects in the non-ideological treatment, this difference is not statistically significant. However, ideological affinity does tend to increase positive contributions to one's own group: compared with the non-ideological treatment, subjects in the ideological treatment contribute an average of 0.84 more tokens to the target group when they are members of the target group. This increase is statistically significant ($p = 0.043$).

Theoretical predictions and examination of raw descriptive statistics from the experiment show that examining the behavior of subjects in like and unlike pairings is appropriate. Tables 4.1 and 4.2 therefore conduct separate analysis for like and unlike pairings, respectively. In the case of unlike pairs, the result above holds: subjects in the ideological treatment contribute 1.81 more tokens ($p = 0.006$)[8] than those in the non-ideological treatment when they are members of the target group, while ideological and non-ideological groups are no different when not members of the target group. However, like pairs exhibit different behavior: subjects in the ideological treatment contribute 1.29 *fewer* tokens when not members of the target group, but behave no differently than the non-ideological group when members of the target group.

The results of the random-effects tobit models therefore confirm the initial conclusions drawn from the descriptive statistics. Contributions to the target group are lower in the

[8]This result is obtained using a combined significance test on $\left(\frac{\partial contribution}{\partial ideological}|target = 1\right) = \beta_{ideological} + \beta_{own\ target \cdot ideology} \cdot own\ target$; all such results reported in the text are based on combined significance tests where appropriate.

Table 4.4: RE Tobit, Subject Pair in Same Group

variable	coefficient	p-value
target is own group	0.814	0.026
ideological groups	-1.29	>0.001
own target * ideological groups	1.17	>0.001
Controls		
first sequence	-1.89	>0.001
ln periods 1-10	1.46	>0.001
ln periods 11-20	0.568	>0.001
constant	4.71	>0.001

Results derived from a random-effects Tobit model estimated in Stata 9.2. N=68, T=20, NT=1360. 146 observations were left-censored, 139 were right-censored.

Table 4.5: RE Tobit, Subject Pair in Different Groups

variable	coefficient	p-value
target is own group	0.338	0.337
ideological groups	0.690	0.298
own target * ideological groups	1.12	0.022
Controls		
first sequence	-2.07	>0.001
ln periods 1-10	1.03	>0.001
ln periods 11-20	0.151	0.547
constant	6.30	>0.001

Results derived from a random-effects Tobit model estimated in Stata 9.2. N=34, T=20, NT=680. 35 observations were left-censored, 138 were right-censored.

ideological treatment when both subjects are not members of the target group, supporting the conclusion that ideological valence tends to suppress contributions to an ideologically opposed group, even when competitive pressures to contribute exist. However, contributions are higher in the ideological treatment when a target group member competes with a non-target group member, indicating that ideological valence tends to boost contributions to one's own group.

4.5 Discussion

The results of the experiment generally support the predictions of Chapter 3, though with some important qualifications. Subjects in the experiment did converge to a positive contribution level, confirming the qualitative prediction that the tournament-style strategic budgeting incentives are capable of motivating positive output in an environment with competing pressures for group loyalty. In addition, the ideological treatment did create systematic differences in contribution levels that are consistent with the predicted change. However, several predictions of the theory were not confirmed, and merit additional discussion.

4.5.1 Overcontribution

While qualitatively accurate, the theory's quantitative prediction systematically underestimated the observed level of contribution in all but one case. This systematic underprediction extends even to those cases where groups were not ideological, and hence the underprediction cannot be solely attributed to this factor. Nor do previous tournament experiments observe the overcontribution observed in this experiment. While past experiments observe considerable heterogeneity at the individual level, average contribution levels are generally at or below predicted levels.

Aside from the aspect of ideological affinity, the present experiment differs in an important way from past experiments: subjects are organized into different groups, with contributions to one or the other group creating benefits for that group's members. The fact that this contribution creates a collective good for other people might spur higher contributions in all treatments, regardless of whether these groups are ideological in nature, due to a valuation that individual subjects assign to the welfare of others. Note that the value created for oneself via contribution to the public good is already included in the predicted contribution level; this is the constant 2 term in equation 4.2.

When past experiments have employed a group-based design (e.g., Nalbanian and Schotter, 1997), the groups contributed to a collective effort that determined whether everyone on the team won or lost a competition with another group. This design creates a free-rider problem within the teams, creating an incentive to shirk that could easily overwhelm any group affinity effect. By contrast, an individual subject's victory competition in my experiment does not depend on other subjects' contributions, eliminating a downward pressure on contributions and allowing a group affinity effect to be observed.

4.5.2 Ideology and Pairing Structure

The theory's predictions about ideological valence in like pairings were partially confirmed: compared to the non-ideological treatment, ideological subjects in like pairings contributed fewer tokens when they were not members of the target group. However, contributions are not higher for ideological subjects who are members of the target group participating in like pairs, as would be expected from the theory. In unlike pairings, this pattern of confirmed and unconfirmed predictions is reversed. Subjects in the ideological treatment contributed more tokens than subjects in the non-ideological treatment when

Table 4.6: RE Tobit, Subject in Target Group

variable	coefficient	p-value
subject in like pair	-0.619	0.404
ideological groups	1.52	0.050
like pair * ideological groups	-1.61	0.091
Controls		
first sequence	-1.35	0.013
ln periods 1-10	1.53	>0.001
ln periods 11-20	0.289	0.050
constant	6.07	>0.001

Results derived from a random-effects Tobit model estimated in Stata 9.2. N=102, T=20, NT=1020. 43 observations were left-censored, 157 were right-censored.

they were members of the target group, but contributions are not lower for subjects in the ideological treatment who are not members of the target group.

The effect of ideology, at least in the context of this experiment, appears to be contextual. Subjects only feel comfortable shirking contributions on an ideological basis when their competitor is in the same group, and presumably likely to share a motivation to shirk. The presence of a competitor from the target group group deters this form of shirking, perhaps because the subject knows that the competitor will not shirk for ideological reasons and will therefore gain an advantage in the competition if the subject shirks him or herself. In other words, ideological shirking might be compared to a form of risky cooperation, akin to a Prisoner's Dilemma. If both subjects cooperate by shirking, both enjoy an ideological benefit and maintain the same probability of receiving the penalty—but mutual shirking can only occur when both subjects will benefit, when neither is a member of the target group.

But the interaction between ideological affinity and mutual shirking goes beyond this narrow case. Table 4.6 shows the results of a regression performed on subjects who were in the target group; the main regressors of interest are (i) whether the subject was in a pair with a member of the same group, (ii) whether the groups were ideologically determined, and (iii) an interaction between the previous two regressors. In the non-ideological treatment, being in a like pair has no statistically discernible impact on the number of tokens a subject contributes to the target group. However, in the ideological treatment, being in a like pair results in 2.24 fewer tokens contributed to the target group ($p < .001$). This decline exists in spite of the fact that both subjects are members of the target group, and hence should derive financial and ideological benefits from contributions to that group. The same finding holds for subjects who are not in the target group, as shown in Table 4.7: being in a like pair results in a decline of 3.03 tokens contributed to the target group, but only in the ideological treatment.

It therefore appears that being in a common ideological group can serve to facilitate trust between subjects beyond whatever trust is created by a randomly assigned group, and this

Table 4.7: RE Tobit, Subject not in Target Group

variable	coefficient	p-value
subject in like pair	-0.616	0.789
ideological groups	0.813	0.383
like pair * ideological groups	-2.42	0.033
Controls		
first sequence	-2.57	>0.001
ln periods 1-10	1.06	>0.001
ln periods 11-20	0.590	0.002
constant	6.17	>0.001

Results derived from a random-effects Tobit model estimated in Stata 9.2. N=102, T=20, NT=1020. 138 observations were left-censored, 120 were right-censored.

trust facilitates collusion between subjects that undermines the incentives of the tournament. Similar problems are not unknown to the literature on bureaucratic duplication (e.g., Ting, 2003), but usually involve deriving benefit from another agency's work product rather than the "mutual disarmament" to avoid punishment by a superior observed here. The threat of such collusion is known to tournament theorists (see Bolton and Dewatripont, 2005, pp. 26-38), though the finding that shared ideology facilitates collusion is novel.

In addition to this collusion effect, ideological identification of groups can increase a subject's effort level by increasing his/her affinity with the task. The results in Tables 4.6 and 4.5 both indicate a direct, positive impact of ideology on contributions to the target group when subjects were in an unlike pairing. However, the collusion effect of ideology cancels out the effect of increased affinity for the task in cases where both subjects are in the target group. Hence, there is no net change in the number of tokens contributed to the target group. When both subjects are *not* in the target group, the negative effect of ideological affinity pushes subjects in the same direction as the power of collusion, and hence subjects contribute fewer tokens compared to subjects in the non-ideological group.

4.5.3 Strategic Budgeting Beyond the Laboratory

What do the results of the experiment mean for strategic budgeting in agencies beyond the laboratory? First, the experiment gives us a reason to expect that tournament-style incentives will effective in government bureaucracies. Under all conditions, subjects contributed as many tokens as predicted in equilibrium, and typically more. Hence, we have ample reason to believe that tournament-style incentives exert the expected causal effect on behavior, including in ideologically-charged environments. If we can observe a relationship between correlates of strategic budgeting and correlates of bureaucratic effort, as I do in Chapter 5, it is reasonable to conclude that this relationship is caused by strategic budgeting.

However, there are several qualifications to this conclusion. First, strategic budgeting could be least effective on ideologically homogeneous bureaucracies opposed to the incentivized goal. While the tournament incentives of the experiment did extract substantial positive contributions from subjects to a target group to which they were ideologically opposed, these contributions were lower than contributions from subjects in a comparable, but non-ideological environment. Hence, while strategic budgeting techniques may help an executive bring a hostile bureaucracy to heel, we may expect this bureaucracy to be less-efficient than one populated with friendly bureaucrats. Outside the laboratory, where exerting effort is both more consequential and more powerfully politically charged, we may expect resistance to strategic budgeting techniques to be correspondingly greater.

Second, strategic budgeting may be effective on ideologically homogeneous bureaucracies that support the incentivized goal, but collusion inside or between competing agencies may undermine this effectiveness. Although their agreement with the executive's goals might be expected to increase their compliance with executive directives, they also have an incentive (per Niskanen) to extract maximum slack from minimum effort. A homogeneous agency may find it easier to collude against the interests of the executive (and, presumably, the public) on this front.

Finally, strategic budgeting incentives are expected to have the greatest impact on ideologically heterogeneous agencies. This finding is good news for democracy: patronage (and its attendant corruption) is not required for a bureaucracy that is responsive to elected officials and to the public. Indeed, an agency that hires on the basis of merit and without regard to ideological conviction should, on the basis of this experiment, be easier to manage using strategic budgeting techniques.

CHAPTER 5

EVIDENCE FOR STRATEGIC BUDGETING IN THE AMERICAN STATES

It is helpful to know that there exists a theoretical rationale by which executives can control a bureaucracy, even an ideologically hostile bureaucracy, using strategic budgeting. Combined with evidence from the laboratory that confirms the causal power of budget incentives in an ideological environment, this theoretical argument provides strong reason to suspect that strategically manipulating agency budgets to create competition could provide the executive with a power lever of control, one not typically contemplated in the academic literature on inter-branch policy-making.

But none of this evidence indicates whether, in the causally chaotic environment of the external political world, strategic budgeting is actually used or would be an effective technique of bureaucratic control. The purpose of this chapter is to make the case that strategic budgeting can be effective when applied to bureaucratic agencies and may already be employed in state agencies in the United States. To this end, I present original quantitative evidence that suggests that strategic budgeting incentives may already be employed at the level of the state government.

5.1 Research Design

Testing for the presence of (efficacious) strategic budgeting incentives is not straightforward. First, the existence (or absence) of budget incentives for bureaucrats is not necessarily observable. The use of strategic budgeting incentives does not require a great deal of coordination or infrastructure, as compared with the direct review of agency actions or regulations. Indeed, not even the explicit articulation of a strategic management policy would be required. Assuming that the executive does not him/herself write the budget, all that is required is the existence of a group of individuals who share the executive's preferences and are informed enough about the behavior of the bureaucracy to compare them and appropriately reward or punish agencies according to their compliance. However, although we know that this requirement is met by the Office of Management and Budget at the federal level, the OMB's powers are not necessarily used to strategically budget, nor does OMB's scrutiny translate into easy quantification.

Second, the theory predicts a lack of dynamism in bureaucratic effort so long as the ideology and professionalism of the bureaucrats, as well as the structure of budget incentives

offered, do not change. I do not expect the composition of the bureaucracy nor the institutional structure of most American governments (state or federal) to change frequently or dramatically, meaning that most data sets are unlikely to show changes in behavior that would be attributable to the comparative statics of strategic budgeting theory. The theory *does* predict a change in bureaucratic effort if the preferences of the executive change, as this change in preference would cause a change in budget incentives offered, but it would be difficult to rule out alternative causes for this change in effort. For example, pure loyalty to elected officials would cause this same change, as would the existence of an alternative command-and-control system. Hence, short of actually varying the institutional context of the government, it may be hard to detect the power of budgetary incentives and would certainly be hard to separate the impact of budget incentives from rival causes.

Finally, and perhaps most importantly, changes in an agency's budget may cause changes in that agency's policy for reasons having nothing to do with the incentives provided by a strategic budgeting regime. Budget expansions provide additional resources that presumably enable a bureaucracy to increase its output. Likewise, budget contractions could force an agency to restrict its output in order to accommodate the decline in resources.[1] It will be important to separate this impact of budget changes from the incentive-based impact of budget changes.

I propose to test for evidence of the impact of strategic budgeting power using a technique that avoids these pitfalls. According to the strategic budgeting theory I proposed in Chapter 3, I expect policy effort to be more consistent with the executive's preferences in those governments where s/he has the power to offer strategic budgeting incentives—if those incentives are being employed. Hence, if policy effort exerted by bureaucracies systematically varies between cases where an executive has the power to strategically budget and those where s/he does not, controlling for the actual size of the bureaucracy's budget during that time, then we have evidence that strategic budget techniques are indeed being employed to control the bureaucracy. My technique requires several conditions to be present in the data:

1. Variation in the degree of control the executive has over the budget. Because institutional change of this nature is infrequent, this condition implies that comparative panel data that includes variation on this institutional dimension will be advantageous to deploy.

2. Variation in the preferences of the executive. In the comparative panel data context, the preferences of the executive need to vary *within* panels in order to demonstrate that the bureaucracy is more responsive to these changes in systems that allow for strategic budgeting.

3. Government budget data. Information on the size of the government's overall budget and the budget of the agency under study will be necessary to establish that the power of strategic budget incentives is distinct from resource constraints imposed by budget limitations.

[1] Of course, some budget changes may not compromise a bureaucracy's core outputs, but simply increase or decrease the levels of slack available to the bureaucracy. While affecting the perks available to bureaucrats, such fluctuations might not result in a change in policy output. I presume, however, that at least sufficiently large decreases should eventually force a bureaucracy to produce less of some policy outputs.

5.2 Hypotheses to Test

As noted above, I expect effort levels to be more consistent with an executive's preferences in systems where more powerful strategic budget incentives are being offered. This hypothesis follows from the results of Chapter 3 showing that (i) executives will prefer to offer positive budget incentives to induce effort, and (ii) productivity along the desired effort dimension rises as the budget incentives offered to a bureaucrat rise accordingly. That is, suppose that policy effort e exerted by a bureaucracy is some function $g()$ of the executive's preferences p and budget incentives b, and a set of other control variables S:

$$e = g(p, b, S)$$

With this notation in hand, I can concisely write the central conceptual hypothesis to be tested:

$$sign\left(\frac{\partial g}{\partial b}\right) = sign(p) \tag{5.1}$$

Equation 5.1 says that the relationship between effort and budget incentives should have the same sign as the executive's preference for effort (assuming that $p > 0$ indicates a preference for more effort and $p < 0$ indicated a preference for less effort). That is, executive preference is a contextual variable that determines the influence of budget incentives on bureaucratic behavior.

As I previously noted, I cannot directly observe the existence of magnitude of budget incentives. Rather, I will attempt to observe whether it is possible for budget incentives to be offered by measuring the degree of control the executive has over the budget. That is, I presume:

$$budget\ incentives = b(c)$$

The variable c measures executive control over the budget. I assume that:

$$\frac{db}{dc} > 0$$

Therefore, we now write effort as:

$$e = g(p, b(c), S)$$

The hypothesis to be tested is therefore:

$$sign\left(\frac{\partial g}{\partial b(c)}\frac{db}{dc}\right) = sign(p) \tag{5.2}$$

The comparative static depends on the value of $\frac{db}{dc}$: if there is a positive relationship between control over the budget and the offering of budgetary incentives, then using control over budgets as a proxy for budgetary incentives is a valid test of the original conceptual hypotheses in equation 5.1. Conceptually, control is a necessary condition for budgetary incentives to be offered; is is therefore safe to assume that $\frac{db}{dc} \geq 0$. Hence, if I find no relationship between executive control over the budget and bureaucratic effort, it is possible

that either (i) no relationship between budget incentives and bureaucratic effort exists, or (ii) control is necessary but not sufficient for strategic budgeting to be employed, and hence $\frac{db}{dc} = 0$.

Of course, equation 5.2 is still too abstract to test directly. Concepts of *budget control* and *executive preferences* must be measured, and a parametric form for $g()$ must be assumed,[2] before a test can proceed. These tasks are tackled in the next two sections.

5.3 Data

Recall that my theory of strategic budgeting assumes that a single executive has control over the distribution of an agency's budget, with multiple agencies (or agency sub-units) vying for control over larger budgets. In constructing an empirical test of the theory, I will operationally assume that the role of the executive is played by a state governor and that the agency in question is the state agency responsible for enforcing air pollution regulations. Furthermore, I operationalize policy effort exerted by the agency using the number of inspections performed by that agency.

American state environmental agencies provide an ideal forum for testing a theory of strategic budgeting for two reasons. First, there is a great deal of institutional variation among American states when it comes to the gubernatorial role in the budget process: the governor's control over the budget is considerably stronger in some states than in others. Strong gubernatorial control over the budget is strongly associated with that governor's ability to exert control over the bureaucracy via strategic budgeting. This variation allows us to compare policy effort in states with close gubernatorial control over the budget to states without such close control.

Secondly, policy effort by environmental protection agencies has a reasonably clear partisan valence that allows a researcher to infer the governor's preference for this particular policy. Depending on the precise structure of the model to be employed, I make one of two assumptions: either (i) on average, and net of a state's overall environmental tendencies, Democratic governors will prefer more environmental inspections than their Republican counterparts, or (ii) on average, a transition from a Republican governorship to a Democratic governorship in any state will result in an increase in preference for environmental inspections in a given state (and vice versa).

5.3.1 Dependent Variable: State Inspections

The dependent variable in my analysis is the number of *state inspections* performed on fixed-source air pollution emitters by 49 states[3] between the years 2000 and 2006, as retrieved from the federal EPA's Enforcement and Compliance History Online website, using data from

[2]Strictly speaking, this is not true: I could employ a non-parametric model. This course is inadvisable given the high dimensionality of the problem (9 independent variables, plus a lag) and the small number of data points.

[3]Nebraska was excluded due to its nonpartisan legislature; legislative partisanship is an important control variable in the analysis.

the Air Facility System (AFS).[4] The AFS codes a variety of inspection and enforcement actions taken by both state and federal agencies, with data at the individual (plant) level. I aggregated eight types of actions, ranging from owner/operator source tests observed by a state agent up through full compliance evaluations, taken by state environmental agencies into a count for each state for each of the seven years I observe.[5]

Figure 5.1 shows a kernel density plot of the distribution of state inspections in the data set. The plot demonstrates that state inspections have a mean far from zero ($\mu = 1020.857$, $\sigma = 194.991$), which can possibly justify ignoring the count aspect of the data. However, it is also clear from this plot that certain count aspects of the data are present, including a skewed, Poisson-like distribution and observations at or around the lower boundary (1 observation at zero, 3 below 10, 13 below 50, and 33 below 100).

Figure 5.2 provides a view of state inspections have changed in the 49 states during the period of the data set. In particular, the figure makes it easy to visualize three features of the data that are important to keep in mind during analysis. First, there is a great deal of heterogeneity in the overall level of state inspections performed between states: some seem to have a persistently large number of inspections, while other states have persistently fewer inspections (all states are on the same vertical scale in the figure). There are countless reasons for persistent differences among the states, including but not limited to differences in political culture, industry concentration, population size, and the balance of urban and rural population. While I will attempt to explicitly model several of these effects with measured control variables, it is unlikely that I will be able to completely eliminate this heterogeneity. Hence, a statistical model will probably need to recognize unit heterogeneity and eliminate any attendant biases it might create in regression estimates.

Second, Figure 5.2 strongly suggests that the number of state inspections is strongly related to the number of state inspections that took place before. A wide variety of empirical tests[6] reject the possibility of a unit root or panel cointegration in the data, implying that methods assuming covariance stationarity can be employed. Empirical analysis performed on this data set will therefore need to account for this dynamic effect to avoid omitted variable bias.

Third, though there is no evidence of a common trend over time—indeed, the Nyblom-Harvey statistic rejects the possibility of a common trend or any panel cointegration[7]—it

[4] Although the AFS records data from before the year 2000, "the data quality before November 2000 has not been assessed and should be considered unknown" (United States Environmental Protection Agency, 2008*a*), and therefore data prior to the year 2000 was excluded.

[5] The actions counted as state inspections were coded in the AFS database as follows: 3A (state-observed owner-operator source test), 5C (state inspection level 2 or greater), 6C (state source test conducted), FF (state off-site full compliance evaluation), FS (state on-site full compliance evaluation), PS (state on-site partial compliance evaluation), PX (state off-site partial compliance evaluation), and SI (state investigation initiated). More details are available in the U.S. EPA's manual about AFS action types (United States Environmental Protection Agency, 2003).

[6] Running the Levin-Lin-Chu (2002) and Im-Pesaran-Shin (2003) tests for unit roots in a panel setting both decisively reject the null of a unit root, with $p < .001$ for all tests. See Baltagi (2005, Chapter 12) for an overview of these tests. Stata 9.2 and the levinlin (Bornhorst and Baum, 2007) and ipshin (Bornhorst and Baum, 2006) packages were used to run the test. Results for panel cointegration are obtained using the Nyblom-Harvey statistic; see footnote 7 below.

[7] The Nyblom-Harvey statistic is described in Nyblom and Harvey (2000). It is a Lagrange multiplier test

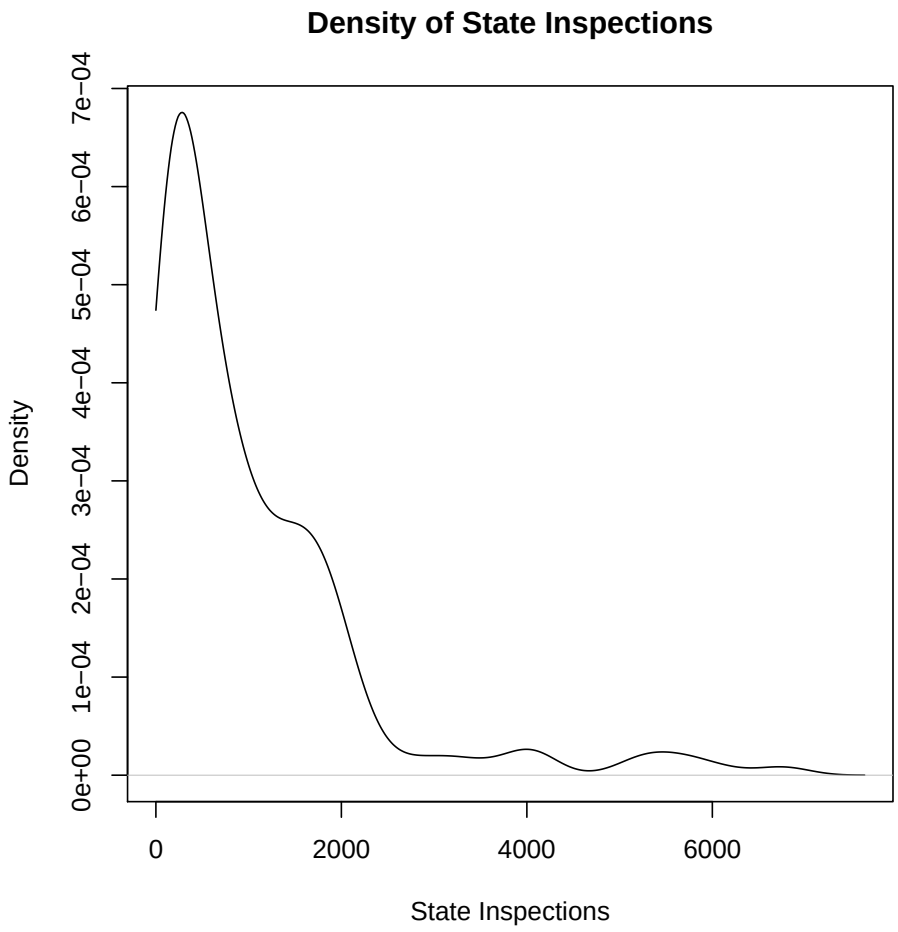


Figure 5.1: Density of State Inspections Recorded in the Air Facility System, 2000-2006

is possible that individual years are associated with common increases or decreases in state inspections. For example, 2002 seems to be associated with a rise in inspections in a large number of panels (indeed, the mean change is 204 inspections between 2001 and 2002). Including some form of control for year-specific effects therefore seems prudent.

---

of the existence of common trends in a panel data set, specifically testing whether the rank of the variance-covariance matrix of panel-specific lags has zero rank (i.e., there are no common trends in the data, and no panels are cointegrated). The nharvey package (Baum and Bornhorst, 2007) in Stata 9.2 yields N-H values of 1.71 and 2.38 with nonparametric adjustment for long-run variance using 2 and 4 lags, respectively; each of these is well below the level required for rejection of the null hypothesis of no cointegration.

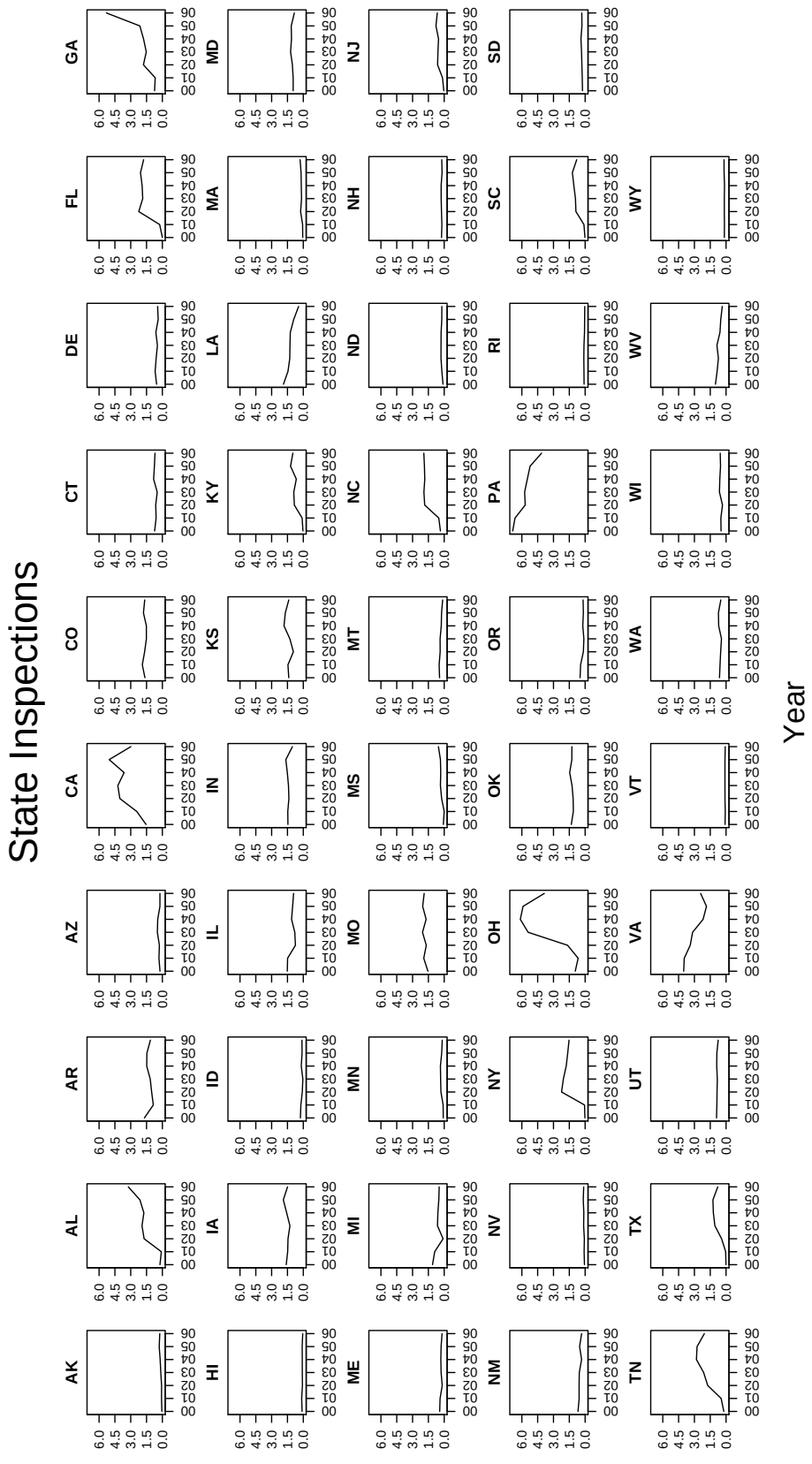


Figure 5.2: State Inspections in the Air Facility System by State, 2000-2006

Alternative Dependent Variable: Serious State Inspections

Of course, it is possible that some forms of state inspection are not particularly serious and undertaken primarily as a way of *signaling* concern for the environment to the public without imposing serious costs on business and industry, likely to be a key constituency. Hence, I develop a second measure, *serious state inspections*, that includes only full compliance evaluations (on- or off-site) undertaken by the state. These evaluations, which are thoroughgoing evaluations of a site's compliance with all air pollution standards[8], are likely to be disruptive and costly to the business.

5.3.2 Independent Variables: Governor's Partisanship and Gubernatorial Budget Power

There are two independent variables of primary interest. The first, *Republican Governor*, records whether the governor was a Democrat or Republican in the year of interest.[9] This variable was retrieved from the *Book of the States* for the appropriate year (*Book of the States*, 2000-2006). The second, *Budget Responsibility*, is a binary indicator of whether the governor is primarily responsible for preparation of the state budget as indicated in the *Book of the States*.[10] There is sufficient variation in this indicator for a test, primarily between panels but also within: while 40 of 49 states are coded as assigning budget responsibility to the governor in 2000, by 2006 only 34 of 49 states are thusly coded. To ensure the robustness of my results, I also employ an alternative measure of gubernatorial control over the state budget, based on the additive gubernatorial budget power index constructed by Barrilleaux and Berkman (2003), which combines four indicators of gubernatorial budget control into a single additive index.[11] This index, which ranges between 0 and 4, shows more variation than the binary measure, though it too shows initial concentration toward gubernatorial power with a tendency toward lesser gubernatorial power over time.

Recall that the hypothesis to be tested (equation 5.2) indicates that executive control over the budget should make bureaucratic effort more consistent with executive preferences. I have assumed that, on average, Democrats will tend to prefer more inspections than their Republican counterparts. Hence, 5.2 predicts an interactive effect between gubernatorial partisanship and gubernatorial budget responsibility. As Republican governors gain more control over the budget, I expect the number of state inspections to fall; as Democratic governors gain more budget control, I expect the number of state inspections to rise.

Given this implicit two by two quasi-experimental design, with budget responsibility and gubernatorial partisanship as the variables of interest, it is most important that there be sufficient information in every cell of this design. Table 5.1 indicates that sufficient variation exists for the binary budget responsibility indicator, although (as previously indicated) the

[8]See `http://www.epa.gov/compliance/monitoring/inspections/index.html#evaluation`.

[9]1=Republican governor, 0=Democratic governor. Governor Jesse Ventura of Minnesota, a member of the Reform Party, was coded as a Democrat based on information available about his stands on environmental policy (*OnTheIssues.org: Jesse Ventura on Environment*, N.d.). Note that no articles turn up on Lexis/Nexis for a search on "Jesse Ventura" and "environment."

[10]Note specific table location here.

[11]The elements of the index are noted in the appendix to this chapter.

Table 5.1: Distribution of Data, Binary Budget Responsibility

	Republican Governor	Democratic Governor
Gov. has Sole Responsibility	145	117
Gov. Shares Responsibility	43	38

Table 5.2: Distribution of Data, Budget Power Index

	Republican Governor	Democratic Governor
0	1	3
0.5	0	0
1	5	17
1.5	1	3
2	35	26
2.5	4	13
3	39	43
3.5	28	31
4	42	52

distribution is skewed toward governors with sole responsibility for budget preparation. Even greater variation exists for the budget power index, as shown in Table 5.2.

5.3.3 Control Variables

I also include six additional independent variables as controls for alternative causal processes. The most important of these are *ln Total Revenue* and *ln Health Spending*, which attempt to control for the resource availability effects of an expanding or contracting budget. The first, *ln Total Revenue*, is the natural log of the total general revenue of a state; *ln Health Spending* measures annual money spent by a state on air quality inspection activities and all other non-hospital health-related expenditures. This measure is quite imperfect, as it includes other health-related government spending (like rabies control programs)[12], but should be correlated with the government's level of support for environmental protection inspections. Both of these variables were retrieved from the U.S. Census Bureau's database on state government finance (N.d.).

Two of the additional variables are meant to control for the extent of the environmental

[12]According to the Census Bureau's *Government Finance and Employment Manual*, the health spending measure specifically includes "expenditures for general health activities, categorical health activities and programs, health-related inspections, community health care programs, regulation of air and water quality, rabies and animal control, and ambulance and emergency medical services.... Also includes state or local expenditure financed by Federal Government 'Superfund' for cleanup of hazardous waste sites" (2006, p. 5-37).

problem in the state, as larger environmental dangers might prompt more inspections while a larger industry presence might bring political pressure to decrease inspections. To measure the size of polluting industries in the state, I include *ln Production Employment*, the natural log of the number of persons employed in production industries, as recorded by the U.S. Bureau of Labor Statistics Occupational Employment Statistics program[13] (*Bureau of Labor Statistics Occupational and Employment Statistics Program*, N.d.). To measure the extent of the pollution problem in the state, I include *ln Population at Risk*, a count of the number of people living in areas designated in non-attainment of Clean Air Act standards[14] by the U.S. Environmental Protection Agency's *Green Book* (United States Environmental Protection Agency, 2008*b*). Note that *ln Population at Risk* counts people more than once if their area is in non-attainment for multiple pollutants,[15] so that the population at risk can be greater than the state's population; this counting procedure is used to allow a more continuous measurement of problem severity, rather than forcing all areas into binary categories of attainment or non-attainment.

Finally, to control for the external political environment of the state, I include an indicator of whether the legislature was under the unified control of the Republican party, the Democratic party, or whether control was divided as recorded in the *Book of the States* (*Book of the States*, 2000-2006). This indicator is broken into two dummy variables, one that indicates unified Republican control and another than indicates unified Democratic control; when both are zero, control is divided.

5.4 Statistical Analysis

I will employ the generalized linear model (GLM) framework to model my data, with extensions and modifications as discussed below. Generally speaking, this framework means assuming that the data generating process takes the following form:

$$\vec{e} = g\left(X\vec{\beta} + \vec{\varepsilon}\right)$$

That is, the column vector of bureaucratic effort $\vec{e}$ is a function $g()$ of the matrix of column vectors $X = \{1, x_1, x_2, \ldots, x_k\}$ mapped to an additive index by a column vector of coefficients $\vec{\beta} = \{\beta_0, \beta_1, \beta_2, \ldots, \beta_k\}$, plus a vector of error terms $\vec{\varepsilon}$, where $g()$ maps the additive index into the dependent variable space. Stated another way, consider an single observation i: that observation's effort level is assumed to be a function of the additive index of independent variables measured for that observation:

$$e_i = g\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + \varepsilon_i\right)$$

[13]The Bureau of Labor Statistics SOC code number for this category is 51. For example, dry cleaning workers, electrical power generation workers, chemical plant workers, and other manufacturing-related employment are all classified in this category. A complete description of all occupations in the category is available at the Bureau of Labor Statistics website (`http://stat.bls.gov/oes/current/oes_stru.htm`).

[14]These standards set maximum exposure levels for 8 pollutants: ozone (1-hour and 8-hour exposure), carbon monoxide, nitrogen dioxide, sulfur dioxide, particulate matter (PM-10 and PM-2.5), and lead.

[15]For example, if a person lives in an area that is in non-attainment for three pollutants, that person is counted three times.

We know from equation 5.2 that we expect the relationship between state inspections and gubernatorial budget responsibility to have the same sign as the executive's preference for effort. I have assumed that, on average, Democratic governors prefer more state inspections than Republican governors. Hence, in the context of a generalized linear model of state inspections, I assume the following relationship:

$$\begin{aligned} state\ inspections &= g(I) \\ I &= \beta_0 + \beta_1 \cdot republican + \beta_2 \cdot budget\ control + \\ &\quad \beta_3 \cdot republican \cdot budget\ control + \sum_{i \in S} \beta_i x_i + \varepsilon_i \end{aligned} \tag{5.3}$$

Equation 5.3 sets up four categories corresponding to the four cells of Table 5.1. The base category, captured in the constant term β_0, tells us the number of state inspections to expect under a Democratic governor without budget control. It is natural to expect that $\beta_0 > 0$, simply because some positive number of inspections will occur under any regime. To know whether preferences exert an impact on inspections in the absence of budget powers, we look to the coefficient on β_1, which captures the difference in state inspections between Republican and Democratic governorships when the governor exercises (comparatively) the least control over the budget. This point is seen by examining the derivative of equation 5.3 with respect to gubernatorial partisanship:

$$\frac{\partial\, state\ inspections}{\partial\, republican} = \frac{\partial g}{\partial X\beta}\left(\beta_1 + \beta_3 \cdot budget\ control\right) \tag{5.4}$$

Link functions are typically constructed such that they are monotonically positive with respect to the index $X\beta$. Hence, the sign of equation 5.4 depends on β_1 when the governor exercises the least control over the budget. If treated as a "general equilibrium" model—that is, one that completely captures the relationship between governor and bureaucracy—strategic budgeting theory predicts that the governor's preferences should make no difference when he/she cannot establish and control a competition over budgets. However, there are likely to be other mechanisms by which the governor can exert control over the bureaucracy, and even a governor with comparatively less power over a budget may still have sufficient power to strategically budget to some degree. Therefore, while finding that $\beta_1 = 0$ can be construed as a confirmation of a theoretical prediction, confirming this prediction is not necessarily strong evidence in support of strategic budgeting theory. Neither is rejecting this prediction strong evidence for rejecting the theory.

The prediction of equation 5.2 is a much more important test of the theory. In terms of the operationalizations proposed above, equation 5.2 predicts:

$$\begin{aligned} \left(\frac{\partial\, state\ inspections}{\partial\, budget\ responsibility} \mid republican = 0\right) &= \frac{\partial g}{\partial X\beta}\left(\beta_2\right) > 0 \\ \left(\frac{\partial\, state\ inspections}{\partial\, budget\ responsibility} \mid republican = 1\right) &= \frac{\partial g}{\partial X\beta}\left(\beta_2 + \beta_3 \cdot republican\right) < 0 \end{aligned}$$

That is, when a democratic governor is in office, the relationship between state inspections and budget responsibility should be positive. When a republican governor is in office,

this relationship should be negative. Because $\frac{\partial g}{\partial X\beta}$ is strictly positive, $\beta_2 > 0$ and $(\beta_2 + \beta_3 republican) > 0$ will constitute confirmation of these hypotheses.

5.4.1 Potential Threats to Inference

As discussed in Section 5.3, the panel data set I am employing is ideal for testing the hypothesis of strategic budgeting: the combination of cross-sectional and temporal variance allows for the comparison of bureaucratic behavior under different budgeting regimes when the executive's preferences change. However, relatively short panel data sets present multiple problems, including the following.

- **The dependent variable is a count.** Hence, some of the assumptions necessary to sustain OLS regression are violated. In particular, the dependent variable is discrete, bounded at zero, and heteroskedastic (Cameron and Trivedi, 1998, pp. 88-89). Without appropriate corrections, OLS yields biased estimates of coefficients and variance.

- **Unit heterogeneity is likely.** States systematically differ in ways that are difficult to observe. The number of inspections performed in a given state is influenced by political and economic circumstances unique to that state that are difficult to exhaustively measure but that change slowly over time. Ignoring these effects biases coefficient estimates if these unit effects are correlated with independent variables.

- **The dependent variable is dynamic.** The nature of bureaucratic governance (particularly year-to-year stickiness in budgets and operating procedures), make it likely that the number of inspections performed by a state in a given year is closely related to the number of inspections performed in past years. Ignoring dynamics can bias the results of a regression due to omitted variable bias.

- **Regressors are likely to be endogenous or weakly exogenous.** When making budget allocations for the future, such as deciding how much to spend on inspections related to air quality, policy makers will probably consider current air quality and the current number of inspections being performed. This introduces the possibility of weak exogeneity of the regressors, or correlation between regressors at time t and regression errors at time $t-1$ and before. If a lagged dependent variable is included to account for the dynamic aspect of the panel, this variable will certainly be endogenous, or correlated with the regression error at time t and before. These problems violate regression assumptions and cause potential bias in differenced estimators designed to correct for unit heterogeneity (Chamberlain, 1992; Wooldridge, 1997).

Statistical techniques exist to deal with each of these problems, though only recently have techniques emerged that claim to simultaneously handle them all. Generic count data can be captured using OLS with appropriate transformations of the dependent variable and the calculation of heteroskedastic-consistent standard errors, or (more commonly) generalized linear models involving a Poisson or negative binomial link function (Cameron and Trivedi, 1998, Chapter 3). Lags of the independent variable (or related techniques involving the calculation of autocorrelation coefficients) are often employed to account for dynamics,

including in the count data context (Brandt et al., 2000; Williams and Brandt, 2001). Fixed and random-effects models, typically involving the first-differencing of the dependent variable but sometimes directly employing unit dummy variables (Allison and Waterman, 2002), can correct for unit heterogeneity (Baltagi, 2005, Chapters 1 and 2), including in the count context (Hausman, Hall and Griliches, 1984).

However, employing both fixed or random effects *and* a lagged dependent variable is known to introduce bias by introducing an endogenous regressor (the lag) that is correlated with the error of the differenced equation. Although recent studies (Judson and Owen, 1999, e.g.,) indicate that this bias is typically small for coefficients other than that on the lag (Wilson and Butler, 2007), it is still prudent to pursue alternative estimators that avoid this problem altogether.

5.4.2 Model Choice

Two estimators have been proposed that specifically deals with the problem of unit heterogeneity in a dynamic panel model. One, the estimator of Arellano and Bond (1991), is based on the generalized method of moments (GMM) and uses instrumental variables for the lagged dependent variable (and other endogenous or predetermined regressors, if present) to eliminate correlation with the error in a model differenced to eliminate unit heterogeneity (Anderson and Hsiao, 1982, see also). The Arellano-Bond estimator is not designed for count data, but transforming the count using a square root makes the count's distribution closer to that of a continuous and homoskedastically-distributed variable and may make this model more appropriate to employ (Cameron and Trivedi, 1998, pp. 88-89).

A second, related GMM estimator is specifically designed for unit heterogeneity in dynamic panel models (Blundell, Griffith and Windmeijer, 2002), and will be the primary model I rely upon for inference. The Blundell et al. estimator corrects for unit heterogeneity in the context of a Poisson model using quasi-differencing techniques developed by Chamberlain (1992) and Wooldridge (1997) while allowing for the presence of a lagged dependent variable. ExpEND, a software package to estimate these models in GAUSS, has been released by Windmeijer (2002). Blundell et al. assume[16] the following specification:

$$E[y_{it}] \quad = \quad \gamma y_{it-1} + \exp\left(x_{it}^{'}\beta + \eta\right) = \gamma y_{it-1} + \mu_{it}\nu_i$$

For any observation,

$$y_{it} = \gamma y_{it-1} + \mu_{it}\nu_i + u_{it}$$

Here, $\mu_{it} = \exp\left(x_{it}^{'}\beta\right)$ represents the influence of regressors on the expected value of the dependent variable y_{it}, while $\nu = \exp(\eta)$ represents the effect of unmodelled unit heterogeneity. Recognizing the lower boundary of 0 on count data, the lag coefficient $\gamma > 0$, as are μ_{it} and ν_i by virtue of exponentiation. For any individual observation, an IID error term u_{it} is added. Quasi-differencing is used to eliminate the influence of ν_i.

[16]Notation and results in the following description are drawn from Blundell et al. (2002, pp. 120-121); see also (Windmeijer, 2002).

In this analysis, I employ the quasi-differencing technique of Wooldridge,[17] which uses the following transformation of the dependent variable when the linear feedback model—that is, the model assuming that a lag of y enters into the regressors—is employed:

$$q_{it} = (y_{it} - \gamma y_{it-1}) \frac{\mu_{it-1}}{\mu_{it}} - (y_{it-1} - \gamma y_{it-2}) = u_{it} \frac{\mu_{it-1}}{\mu_{it}} - u_{it-1}$$

Note that this transformation eliminates unit heterogeneity. Using this transformed dependent variable, and assuming that the independent variables x are predetermined, the following moment conditions for GMM estimation are established:

$$E\left(q_{it} | y_{it-2}, x_{it-1}\right) = 0$$

Hence, letting $Z_i = (y_{it-2}, x_{it-1})$, a matrix of instrumental variables, and W_N an optimal weight matrix ((see Blundell, Griffith and Windmeijer, 2002, p. 118)) the estimation algorithm will minimize:

$$\left(\frac{1}{N}\sum_{i=1}^{N} s_i' Z_i\right) W_N \left(\frac{1}{N}\sum_{i=1}^{N} Z_i' s_i\right)$$

Because of its close fit with the structure of the data, the Blundell et al. model is probably the best choice for this data set. However, to examine the robustness of my results under different specifications, I will also estimate the Arellano-Bond estimator with a square root transformation on the dependent variable. Two other models, (i) a differenced fixed effects negative binomial estimate with a lag, and (ii) a negative binomial model with dummy fixed effects and a lag, will also be used on some data sets as an additional robustness check.

5.4.3 Results

Because my data set includes two different measures of the dependent variable and two different measures of budget responsibility, there are a total of four sets of results to examine.

State Inspections and Budget Responsibility

The Arellano-Bond and Blundell et al. models give slightly different results when applied to the data set of state inspections with a binary indicator of budget responsibility. As Table 5.3 shows, the Arellano-Bond model indicates a negative relationship between budget responsibility and state inspections that is indistinguishable from zero for Democratic governors. However, as before, a negative coefficient on the interaction between Republican governorship and budget responsibility indicates that budget responsibility may exert a negative influence on state inspections under a Republican governor. Calculation[18] of the total coefficient on a change in budget responsibility under a Republican governorship yields

[17] Note that this transformation requires standardizing all independent variables to be deviations from their overall mean (Windmeijer, 2000), a transformation automatically performed in ExpEND.

[18] Using Stata 9.2's nlcom routine.

Table 5.3: Dynamic Panel Models of State Clean Air Act Inspections

	Arellano-Bond		Blundell et al.	
variable	coefficient	p-value	coefficient	p-value
ΔRepublican Governor	2.11	0.525	1.18	0.0693
ΔBudget Responsibility	-2.23	0.497	0.831	0.0605
ΔRep. Gov. * Budget Rep.	-3.85	0.200	-0.883	0.115
	Controls			
ΔLagged State Inspections	0.565	0.000	0.558	0.0077
ΔRepublican Legislature	-0.197	0.933	-0.614	0.339
ΔDemocratic Legislature	1.36	0.490	-0.355	0.453
Δln Production Employment	-2.14	0.866	0.818	0.0421
Δln Population at Risk	-0.113	0.433	-0.0024	0.934
Δln Total Revenue	-8.35	0.373	0.962	0.241
Δln Health Spending	2.66	0.494	-0.158	0.727
two-step AR(1)		0.0007		0.0860
two-step AR(2)		0.177		0.889

dep. var.:state CAA inspections, 2001-2006. Arellano-Bond uses square root transform on dv. $N = 49$, $NT = 245$. All models include year dummies. Arellano-Bond two-step Sargan χ^2: $p = 0.8433$. Blundell et al. two-step Sargan χ^2: $p = 0.7511$. Arellano-Bond estimated using xtabond in Stata 9.2. Blundell et al. estimated using ExpEND v1.02 by Windmeijer in Gauss Light 8.0 using MAXLIK 5.0. Reported p-values are two-tailed. DV is instrumented using GMM-type lags (max 2), all others 1 lag only.

a coefficient of -6.08 with a p-value of 0.014,[19] indicating that increases in gubernatorial budget responsibility under a Republican governorship tend to be associated with negative changes in the number of inspections performed.

The Blundell et al. model, however, tells a different story. The positive and significant coefficient on budget responsibility indicates that, under a Democratic governorship, increases in budget responsibility will tend to increase the number of state clean air inspections performed. In addition, there is a negative and significant coefficient on the interaction between republican governorship and budget responsibility, indicating that the direction of budget responsibility's impact on state inspections reverses under a republican governorship. However, the total coefficient on budget responsibility under a Republican regime is -0.0528 with a standard error of 0.245,[20] a result statistically indistinguishable from zero.

Hence, these two models tell related, but different stories. The Arellano-Bond model seems to indicate that budget responsibility only matters under Republican governorships,

[19]This quantity is equal to $(\beta_2 + \beta_3 republican)$, where β_2 is the coefficient on budget responsibility and β_3 is the coefficient on the interaction term.

[20]Results manually calculated using VCV matrix.

Table 5.4: Negative Binomial Models of State Clean Air Act Inspections

	FE Neg. Binomial		GLM w/ dummies	
variable	coefficient	p-value	coefficient	p-value
Republican Governor	0.308	0.003	0.326	0.024
Budget Responsibility	0.0685	0.469	0.0247	0.847
Rep. Gov. * Budget Rep.	-0.241	0.040	-0.263	0.099
	Controls			
Lagged State Inspections	0.000307	0.000	0.000281	0.000
Republican Legislature	0.0912	0.306	0.170	0.147
Democratic Legislature	0.00998	0.918	0.0359	0.778
ln Production Employment	-0.111	0.524	0.904	0.225
ln Population at Risk	-0.00842	0.149	-0.00323	0.690
ln Total Revenue	-0.147	0.508	0.130	0.763
ln Health Spending	-0.214	0.047	-0.171	0.245

dep. var.: state CAA inspections, 2000-2006. $N = 49$, $NT = 294$. All models include year dummies. Both models estimated using Stata 9.2. Reported p-values are two-tailed.

and its effect is negative. The Blundell et al. model, by contrast, says that budget responsibility matters only under Democratic governorships, and its effect is positive. Both stories are partially consistent with the theory, but not with each other.

The standard fixed effects negative binomial models with a lagged dependent variable do not unambiguously resolve this conflict. Table 5.4 shows results from fixed-effects negative binomial models with fixed effects and a lagged dependent variable. The first column shows the results using the Hausman, Hall, and Griliches (1984) procedure, while the second column shows results using the Allison and Waterman (2002) dummy variables approach.

In both cases, there is a positive but statistically insignificant relationship between budget responsibility under a democratic governorship. However, the negative and statistically significant coefficient on the interaction between Republican governorships and budget responsibility suggests that Republican governors may be able to use budget power to decrease the number of clean air inspections performed by the state environmental bureaucracy. Calculation[21] of the total coefficient on budget responsibility under a republican governorship yields a coefficient of $-.173$ with p-value of 0.045 for the Hausman, Hall, and Griliches model[22], confirming that gubernatorial budget responsibility is associated with a decrease in state inspections performed under a Republican governorship.

In some respects, the two negative binomial models tend to support the Arellano-Bond results, particularly in showing a statistically significant and negative effect of budget

[21] Using Stata 9.2's nlcom routine.

[22] For the dummy variable model, the coefficient is $-.239$ with a p-value of 0.051.

responsibility on state inspections under a Republican governorship. However, these models also do not repeat the Arellano-Bond model's finding of a negative coefficient on budget responsibility under a Democratic governorship, an odd finding if Democrats tend to prefer more environmental inspections than Republicans. It is possible that an out-of-control bureaucracy produces more inspections than either Democratic or Republican governors would prefer, and hence both use budget power to restrict inspections though Democrats restrict them less.

Although not of primary interest, it is interesting to note that there is an unexpected *positive* coefficient on Republican governor in both models. This suggests that, under a Republican governor with comparatively little budget control, the number of state inspections performed is greater than under a Democratic governor. While not a direct disconfirmation of the model (see Section 5.3), it is suggestive of several possible modifications required to the theory, which will be discussed further in Section 5.5.1.

State Inspections and Budget Power Index

Repeating the analysis using an alternative measure of gubernatorial budget power does not resolve this conflict between the Arellano-Bond and Blundell et al. models. Table 5.5 shows results of the Arellano-Bond and Blundell et al. models using the budget power index described earlier. Again, the Arellano-Bond model shows a statistically insignificant coefficient on budget power under a Democratic governor, but a negative and significant coefficient ($= -3.45$, $p < .001$) under a Republican governor. The Blundell et al. model, again, shows a positive and (if using a one-tailed test) statistically significant coefficient on budget power for Democratic governors, but a coefficient statistically indistinguishable from zero for budget power under Republican governors (-0.0777, s.e. $= .208$). Note also that the positive and significant coefficient on Republican governorship persists in these models.

Serious State Inspections and Budget Responsibility

Neither does employing the alternative dependent variable, serious state inspections, resolve this conflict. Table 5.6 shows the results from running the Arellano-Bond and Blundell et al. models on the data set using *serious state inspections* as the dependent variable. As in the previous models, the Arellano-Bond model reveals no relationship between budget responsibility and serious state inspections under a Democratic governorship, but does reveal a negative such relationship under a Republican regime (coefficient$= -4.70$, $p = 0.039$). As before, the Blundell et al., model, tells a different story than the Arellano-Bond model, though this story is more interesting than before.

As Table 5.6 displays, the Blundell et al. dynamic panel count model indicates a marginally statistically significant (one-tailed) positive relationship between budget responsibility and serious state inspections under a Democratic governorship, Meanwhile, under a Republican governorship, a marginally statistically significant (using a one-tailed test) and *negative* relationship exists between budget responsibility and serious state inspections (coefficient$= -.362$, s.e. $= .262$, $p = .168$). These results match the predictions of the theory of strategic budgeting. The results also indicate higher inspections under a Republican governor with low budget control, which is not predicted by the theory.

Table 5.5: Dynamic Panel Models of State Clean Air Act Inspections, Alternative Budget Measure

	Arellano-Bond		Blundell et al.	
variable	coefficient	p-value	coefficient	p-value
ΔRepublican Governor	7.03	0.056	1.37	0.0662
ΔBudget Power Index	-0.654	0.650	0.273	0.121
ΔRep. Gov. * Budget Power	-2.80	0.013	-0.350	0.0955
	Controls			
ΔLagged State Inspections	0.595	0.000	0.516	0.0069
ΔRepublican Legislature	-2.44	0.366	-0.521	0.238
ΔDemocratic Legislature	0.367	0.907	-0.409	0.3831
Δln Production Employment	-10.8	0.499	0.949	0.0077
Δln Population at Risk	-0.180	0.289	-0.0080	0.7600
Δln Total Revenue	-10.5	0.331	0.575	0.474
Δln Health Spending	1.77	0.645	0.236	0.739
two-step AR(1)		0.0010		0.0309
two-step AR(2)		0.533		0.476

dep. var.: state CAA inspections, 2001-2006. Arellano-Bond uses square root transform on dv. $N = 49$, $NT = 245$. All models include year dummies. Arellano-Bond two-step Sargan χ^2: $p = 0.9082$. Blundell et al. two-step Sargan χ^2: $p = 0.6376$. Arellano-Bond estimated using xtabond in Stata 9.2. Blundell et al. estimated using ExpEND v1.02 by Windmeijer in Gauss Light 8.0 using MAXLIK 5.0. Reported p-values are two-tailed. DV is instrumented using GMM-type lags (max 2), all others 1 lag.

Serious State Inspections and Budget Power Index

Finally, running the Blundell et al. model on serious state inspections using the alternative index measure of gubernatorial budget power generally continues the story told in all the other models, as shown in Table 5.7. Again, the budget power index under a Democratic governor is positively associated with serious state inspections, and (borderline) statistically distinguishable from zero. Under a Republican governor, however, the budget power index has a negative but statistically insignificant relationship with serious state inspections (coefficient= -0.201, s.e. $= 0.193$, p $= 0.298$).

The Arellano-Bond model, however, indicates that budget power is *negatively* and significantly associated with the budget power index under a Democratic governorship, and slightly more negatively associated with this index under a Republican governorship. While the negative relationship under a Republican governor is consistent with theoretical expectations, the negative relationship under a Democratic governor is more difficult to explain. Even the previous explanation I offered—that both governors prefer fewer inspections, but Republicans prefer fewer than Democrats—could not survive this model.

Table 5.6: Dynamic Panel Models of Serious State Inspections

	Arellano-Bond		Blundell et al.	
variable	coefficient	p-value	coefficient	p-value
ΔRepublican Governor	3.08	0.417	1.51	0.0079
ΔBudget Responsibility	-1.35	0.635	0.650	0.124
ΔRep. Gov. * Budget Rep.	-3.35	0.355	-1.01	0.0163
	Controls			
ΔLagged State Inspections	.600	0.000	0.475	0.0000
ΔRepublican Legislature	-1.13	0.569	-0.0463	0.915
ΔDemocratic Legislature	2.22	0.115	0.402	0.381
Δln Production Employment	-14.8	0.217	1.46	0.0000
Δln Population at Risk	-.0813	0.385	-0.0267	0.197
Δln Total Revenue	.997	0.923	0.316	0.744
Δln Health Spending	3.31	0.350	-0.647	0.422
two-step AR(1)		0.0001		0.395
two-step AR(2)		0.780		0.354

dep. var.: state CAA inspections, 2001-2006. Arellano-Bond uses square root transform on dv. $N = 49$, $NT = 245$. All models include year dummies. Arellano-Bond two-step Sargan χ^2: $p = 0.8514$. Blundell et al. two-step Sargan χ^2: $p = 0.1137$. Arellano-Bond estimated using xtabond in Stata 9.2. Blundell et al. estimated using ExpEND v1.02 by Windmeijer in Gauss Light 8.0 using MAXLIK 5.0. Reported p-values are two-tailed. DV is instrumented using GMM-type lags (max 2), all others 1 lag.

This result might indicate that all governors share a preference for fewer serious air inspections related to the Clean Air Act, possibly as an attempt to satisfy industry interests. More likely, the result is simply anomalous and a product of the relatively poor specification of the model.

5.5 Discussion

The results of Section 5.4 confirm many predictions of strategic budgeting theory. In all models but one, gubernatorial budget power is associated with state inspections pursuant to the Clean Air Act in the expected way under Republican or Democratic governors. Some models indicate that Republican governors use budget power to lower inspections, while other models indicate that Democratic governors use budget power to raise inspections (and one model indicates both, though statistical significance is borderline).

But these results are hardly a clear-cut, unambiguous confirmation of the theory: the results also provide several puzzles to resolve. Why is there a positive coefficient on Republican governor consistent throughout most models? Why do the Arellano-Bond and

Table 5.7: Dynamic Panel Models of Serious State Inspections, Alternative Budget Measure

	Arellano-Bond		Blundell et al.	
variable	coefficient	p-value	coefficient	p-value
ΔRepublican Governor	0.548	0.901	1.95	0.0153
ΔBudget Power Index	-2.25	0.083	0.209	0.113
ΔRep. Gov. * Budget Power	-0.325	0.807	-0.410	0.0587
	Controls			
ΔLagged State Inspections	0.614	0.000	0.388	0.0028
ΔRepublican Legislature	-1.66	0.430	-0.0393	0.907
ΔDemocratic Legislature	1.52	0.300	0.489	0.353
Δln Production Employment	-14.2	0.381	1.436	>0.001
Δln Population at Risk	-0.124	0.329	-0.0332	0.104
Δln Total Revenue	-1.49	0.901	0.452	0.633
Δln Health Spending	2.36	0.498	-0.795	0.347
two-step AR(1)		0.0002		0.195
two-step AR(2)		0.765		0.275

dep. var.: state CAA inspections, 2001-2006. Arellano-Bond uses square root transform on dv. $N = 49$, $NT = 245$. All models include year dummies. Arellano-Bond two-step Sargan $\chi^2 = 0.7747$. Arellano-Bond estimated using xtabond in Stata 9.2. Blundell et al. two-step Sargan χ^2: $p = 0.2766$. Blundell et al. estimated using ExpEND v1.02 by Windmeijer in Gauss Light 8.0 using MAXLIK 5.0. Reported p-values are two-tailed. DV is instrumented using GMM-type lags (max 2), all other variables with 1 lag.

Blundell et al. models provide related, but inconsistent results? If the Arellano-Bond results are to be believed, why isn't there a positive relationship between gubernatorial budget power and state inspections under a Democratic regime? And why does one Arellano-Bond model provide the anomalous result that both Democratic and Republican governors use budget powers to decrease the number of state inspections?

5.5.1 The Effect of a Republican Governor

One effect detected in most of the statistical models is the often positive and significant coefficient on Republican governorship. This somewhat odd result is certainly not predicted by the original incarnation of strategic budgeting theory, though it does not strongly reject that theory (see Section 5.4). Indeed, it is possible that this effect could be explained by a reasonable revision to that theory.

Strategic budgeting theory, as explained in Chapter 3, assumes that bureaucrats pay attention only to (i) their own preferences, (ii) strategic budgeting incentives offered by their peers, and (iii) the executive's preferences. In particular, (iii) assumes that, as bureaucrats

become more professionalized, they tend to exert effort more in line with the executive's preferences. It is, however, possible that professionalism does not induce bureaucrats to act in *accordance* with an executive's preferences, but rather tends to *diminish the incentive to rebel* against an executive's preferences.

There are several reasons why bureaucrats may choose to pursue their own preferences more strongly when the executive is opposed, particularly in areas (like Clean Air Act implementation) where their actions have the imprimatur of established law and cannot be completely eliminated by executive action. For example, if an executive's overall environmental policy tends to favor economic growth and neglect environmental protection, then environmentalist bureaucrats may feel a need to compensate by *increasing* the level of effort they exert in their own domain. The executive has power to effect changes in a wide variety of areas beyond bureaucratic budgeting, including the vetoing of new environmental legislation, the reorganization of agencies or the establishment of new agencies, and the elimination or weakening of programs not firmly established in law. Hence, as the gap between the executive's actions (and preferences) gets larger, bureaucrats could feel a greater pressure to exert more effort *against* the governor's preferences.

In this instance, rather than increasing compliance, professionalism may act to decrease the level of rebellion against executive preferences. This formulation suggests that the utility of the bureaucrat in the theoretical model may be better written as:

$$u_i\left(e_i, e_{-i}\right) = w + \pi\left(e_i, e_{-i}\right) B + \alpha_i e_i + \left(1 - r_i\right)\left(\alpha_i - \alpha_s\right) e_i - c_i\left(e_i\right) \tag{5.5}$$

Here, utility is a function of the gap between the supervisor's preferences and one's own preferences $(\alpha_i - \alpha_s)$, with greater values of professionalism r_i making this gap less important. More plainly, when r_i is larger, rebellion should decrease.

5.5.2 Model Inconsistency

Perhaps more troubling is the lack of consistency in results for the models presented here. Momentarily neglecting the significance of coefficients, all four Arellano-Bond models presented (using two different budget power indicators and two different state inspection measures) indicate that increased budget power is associated with fewer state inspections under both Democratic and Republican governors. All four Blundell et al. models, along with the two fixed-effects negative binomial models, indicate that budget power is associated with fewer state inspections under a Republican governor but more state inspections under a Democratic governor. These results provide reasonably firm evidence that Republican governors use budget power to reduce the number of inspections performed, a confirmation of one prediction from the strategic budgeting model. However, mixed signals are provided on how Democratic governors use budget power. How can this result be explained?

We might chalk up the Arellano-Bond results to mere misspecification. The Blundell et al. estimator is designed specifically for dynamic panel count data, whereas the Arellano-Bond model assumes a continuous dependent variable model. Though built on similar principles using the Generalized Method of Moments technique, the Blundell et al. estimator differs from the Arellano-Bond estimator in several ways, notably in the differencing technique it employs to remove fixed effects and the way that lags of the dependent variable enter the model. In both cases, these differences exist because the assumptions of the Blundell et al.

technique are better matches for count data; for example, the lagged dependent variable is structured in such a way that negative counts cannot be predicted (Blundell, Griffith and Windmeijer, 2002, pp. 120-121). This misspecification story is supported by the fact that both fixed-effects negative binomial models with lagged dependent variables give results that are consistent with the Blundell et al. technique in signs (if not in significance).

5.5.3 The Power of Strategic Budgeting Under Different Regimes

If we assign disproportionate credence to the Blundell et al. model, we can conclude that Democratic governors use budget power to increase the number of Clean Air Act inspections performed by state environmental agencies. However, these same models provide much weaker evidence of Republican governors decreasing inspections: though always in the correct direction, this effect is only statistically significant in one of four regressions (and then, only marginally). The sample size may simply be too small to statistically distinguish the size of this effect from zero, in which case the gathering of more data over time may bolster the findings and allow for a stronger confirmation of strategic budgeting theory. Hence, the collection of a longer panel must be a primary goal of future research, and one that can only be accomplished with time in this particular data given the limitations of pre-2000 data.

However, the collection of more data—even if it establishes the statistical significance of Republican governors' use of budget powers—cannot explain why strategic budgeting should be less efficacious under a Republican regime. One possible explanation comes from Chapter 4, where results from a laboratory experiment showed that strategic budgeting incentives are less effective at inspiring productive effort from those who are ideologically opposed to the goal in mind compared to those who are ideologically sympathetic with the goal. Extending this logic, a Democratic governor could be more effective at increasing effort from an environmental bureaucracy that largely agrees with that governor's goals than a Republican governor would be at forcing a bureaucracy to do policy work with which it disagrees (and displace the inspection activities that it agrees with).

If this finding is true, it further reinforces the "rebellion" term added to the bureaucrat's utility function in equation 5.5, which suggests that greater disagreement between executive and bureaucrat will prompt greater exertion of effort counter to the executive's preferences, implying that equally-sized budget incentives will have a much smaller effect on bureaucrat behavior when bureaucrats disagree with the executive compared to the case where executives and bureaucrats agree.

5.6 Appendix to Chapter 5: Elements of the Budget Powers Index

The following descriptions of the elements of the budget powers index are paraphrased from Barrilleaux and Berkman (2003, pp. 415-416). Each element of the index is coded as indicated in parentheses; the components are then added to create the budget powers index.

- Budget Responsibility: Binary variable from the *Book of the States* indicating whether the governor has primary responsibility for the budget (= 1), or shares responsibility

with the legislature (= 0). This is the same dichotomous variable used as the indicator of gubernatorial budget control in the analysis.

- Line Item Veto: Binary variable from the *Book of the States* indicating whether the governor has a line-item veto power (= 1) or not (= 0).
- Veto Override: Binary variable from the *Book of the States* indicating that the Legislature can override gubernatorial vetoes with a simple majority (= 0) or more than a majority is required (= 1).
- Document Review: Three-category variable from the *Book of the States* indicating whether the governor's budget is used as the working copy for the budget bill on the floor (= 1), whether the legislature has an opportunity to introduce budget bills of its own after committee review (= 0), or if either the legislature or executive can introduce another budget document later (= 0.5).

CHAPTER 6

CONCLUSION

This dissertation began with two questions: Are bureaucrats independent policy makers, or are they controlled by political masters in the elected branches of government? How is political control accomplished when bureaucrats enjoy extensive career protections against political retaliation? I argued that executives are well-placed to establish competition for budgets within and between bureaucratic agencies, and that this strategy is more effective at securing political compliance from a hostile or indifferent bureaucracy than better-known alternative management strategies. Bureaucrats' desire for larger budgets and the attendant slack resources they provide, originally advanced as a criticism of bureaucratic inefficiency, ironically makes this form of control possible.

I deduced this conclusion from a formal theory of the bureaucracy that incorporates a realistic model of human behavior: bounded rationality, ideological preferences, professional ethics, and self-interest are all incorporated into the model. The model also allows for arbitrarily large agencies, while most past work focused on two-agent tournament competition. Quantal response equilibrium is especially suited to the task, and I employed it as the basis for my formal model. Due to the complexity of the model, I relied on a strategy of deriving "partial equilibrium" analytic comparative static predictions and "general equilibrium" computational numeric predictions. Both of these techniques produced the same basic predictions: budget competition should motivate compliant behavior, be less-effective in ideologically-opposed agencies, and be unnecessary in ideologically-friendly or highly professionalized agencies.

Testing this theory is not straightforward. Reliable measures of bureaucratic ideology separate from executive ideology are difficult to obtain. The existence (and especially the degree) of strategic budget competitions is hard to observe. Strategic budgeting theory does not predict a great deal of dynamism over time in bureaucratic behavior, and plausible alternative avenues of influence on the bureaucracy through the budget exist. As a result, I employed a two-pronged testing strategy.

My two-pronged testing strategy was designed to combine the strong internal causal validity of laboratory experiments with the strong external validity of field data tests. In the laboratory, I found that competitive incentives were effective in an ideologically-charged environment, and that ideological agreement with the task had an unexpected intervening influence: it increased contribution in competitions with ideologically-opposed subjects, but *lowered* contributions in competitions with ideologically-alike subjects. This effect was ascribed to the tendency of ideological sameness to facilitate mutually beneficial

cooperation by agents against the principal. In environmental inspection data from the American states, I found that strong gubernatorial budget power amplified the association between gubernatorial partisanship and environmental agency inspection behavior. This finding is consistent with the prediction that budget incentives will increase the compliance of the bureaucracy with executive policy goals.

While I believe that I have offered a convincing account in this dissertation of how strategic budgeting can be an effective managerial technique for the bureaucracy, much remains to be done. For example, interviews of bureaucrats and elected officials could add a great deal to the story told in this dissertation. First, this information could help clarify whether strategic budgeting is explicitly employed by executives (and their employees), or more implicitly used. Second, finer-grained information about the nature of competition in the bureaucracy could be obtained. Topics of interest should include: cases when competition is most and least likely to succeed, factors that interfere with or enhance the effectiveness of competition, and the most appropriate level at which competition should be organized (at the individual level, between projects inside an agency subunit, between agency subunit, between entire agencies, or elsewhere).

Second, and relatedly, the specific details of implementation remain to be determined for policies to which strategic budgeting techniques may be helpfully applied. For example, as I noted in Chapter 2, President Bush's No Child Left Behind education initiative has encountered significant problems related to its use of a cardinal, benchmark-style standard for schools. The use of a relative, competition-style standard of performance like the one described in this dissertation may be able to improve the results obtained from the program. However, the details of implementation in this case would not be trivial nor straightforward. For example, as I also mentioned in Chapter 2, schools cannot be compared willy-nilly: comparing a poor, inner-city school to a rich suburban school in a privileged area would foreordain the winner (regardless of teacher effectiveness) and merely reinforce inequalities of race, class, and culture. This and other, similar details must be explored in depth for each policy target area in order to ensure the success of a competition-based plan for bureaucratic management. As I hope this dissertation has demonstrated, this exploration is worth performing, as the potential for benefit is great.

APPENDIX
HUMAN SUBJECTS APPROVAL FORMS

Florida State
UNIVERSITY

Office of the Vice President For Research
Human Subjects Committee
Tallahassee, Florida 32306-2742
(850) 644-8673 · FAX (850) 644-4392

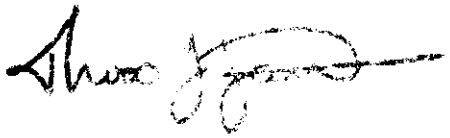
APPROVAL MEMORANDUM

Date: 1/30/2007

To:
Willam Berry
MC: 2230

Dept.: **POLITICAL SCIENCE**

From: **Thomas L. Jacobson, Chair**



Re: **Use of Human Subjects in Research**
A Study in Divided Control of Political Organizations

The forms that you submitted to this office in regard to the use of human subjects in the proposal referenced above have been reviewed by the Secretary, the Chair, and two members of the Human Subjects Committee. Your project is determined to be Expedited per 45 CFR § 46.110(b) 7 and has been approved by an accelerated review process.

The Human Subjects Committee has not evaluated your proposal for scientific merit, except to weigh the risk to the human participants and the aspects of the proposal related to potential risk and benefit. This approval does not replace any departmental or other approvals, which may be required.

If the project has not been completed by **1/29/2008** you must request renewed approval for continuation of the project.

You are advised that any change in protocol in this project must be approved by resubmission of the project to the Committee for approval. Also, the principal investigator must promptly report, in writing, any unexpected problems causing risks to research subjects or others.

By copy of this memorandum, the chairman of your department and/or your major professor is reminded that he/she is responsible for being informed concerning research projects involving human subjects in the department, and should review protocols of such investigations as often as needed to insure that the project is being conducted in compliance with our institution and with DHHS regulations.

This institution has an Assurance on file with the Office for Protection from Research Risks. The Assurance Number is IRB00000446.

Cc:
HSC# 2007.054

Florida State
UNIVERSITY

Office of the Vice President For Research
Human Subjects Committee
Tallahassee, Florida 32306-2742
(850) 644-8673 · FAX (850) 644-4392

APPROVAL MEMORANDUM (for change in research protocol)

Date: 2/20/2008

To:
Willam Berry
MC: 2230

Dept: POLITICAL SCIENCE

From: Thomas L. Jacobson, Chair

Re: Use of Human subjects in Research
Project entitled: A Study in Divided Control of Political Organizations

The memorandum that you submitted to this office in regard to the requested change in your research protocol for the above-referenced project have been reviewed and approved. Thank you for informing the Committee of this change.

A reminder that if the project has not been completed by 1/16/2009, you must request renewed approval for continuation of the project.

By copy of this memorandum, the chairman of your department and/or your major professor is reminded that he/she is responsible for being informed concerning research projects involving human subjects in the department, and should review protocols of such investigations as often as needed to insure that the project is being conducted in compliance with our institution and with DHHS regulations.

This institution has an Assurance on file with the Office for Protection from Research Risks. The Assurance Number is IRB00000446..

cc:
APPLICATION NO. 2008.0044-R

SUBJECT'S CONSENT FORM - *SAMPLE*

PURPOSE
I am being invited to participate voluntarily in this research experiment to study the economics of decision-making.

SELECTION CRITERIA
I am a randomly recruited student at Florida State University. Certain criteria (such as class at the university) may have played a role in how the set of subjects was narrowed down. Only persons 18 years of age or older may participate, and I affirm that I am 18 years of age or older.

PROCEDURE
This experiment will last up to 2 hours. I will be assigned to a computer terminal by chance, "like the flip of a coin" or "random arrival." I will be asked to make decisions at the computer terminal.

PARTICIPATION COSTS AND SUBJECT COMPENSATION
In addition to the $10 for showing up on time and participating, I have the opportunity to earn additional compensation, which will be based upon my decisions, the decisions of others who are in the experiment, and the rules within which those decisions are made. I am free to ask any questions about the rules as to how my compensation will be determined. Any compensation I receive as a result of my participation in this experiment may be reported for taxation purposes to appropriate federal and state agencies, but the results of the study will remain confidential and will not be forwarded to tax authorities. I am free to withdraw from the experiment without additional compensation and without incurring the ill will of the experimenters at any time. If I do so, I may keep my $10.00 show-up fee.

RISKS AND BENEFITS
There are no known health risks or health benefits for this experiment beyond those from any other typical activity in a Florida State University classroom or computer lab.

CONFIDENTIALITY
The confidentiality of any personal information will be protected to the extent allowed by law. To the extent allowed by law, our rule is that only the researcher and any research assistants conducting this experiment may know what my earnings are (subject to tax reporting requirements above) and only researchers affiliated with the experimental economics research group at Florida State University may have access to the data with my name. My name will not be reported with any results related to this research.

CONTACTS
I can obtain further information from Prof. William Berry at 850-644-7321. If I have questions concerning my rights as a research subject, I should call the Human Subjects Committee office at 850-644-8836.

BEFORE GIVING MY CONSENT, THE METHODS, INCONVENIENCES, RISKS, AND BENEFITS HAVE BEEN EXPLAINED TO ME AND MY QUESTIONS HAVE BEEN ANSWERED. I MAY ASK QUESTIONS AT ANY TIME AND I AM FREE TO WITHDRAW FROM THE PROJECT AT ANY TIME WITHOUT CAUSING BAD FEELINGS. MY PARTICIPATION IN THIS PROJECT MAY BE ENDED BY THE INVESTIGATOR OR BY THE SPONSOR FOR REASONS THAT WOULD BE EXPLAINED, BUT WHICH WILL CARRY NO BAD EFFECTS BEYOND THIS EXPERIMENT. SHOULD CURRENTLY UNKNOWN INFORMATION DEVELOP DURING THE COURSE OF THIS STUDY THAT MAY AFFECT MY WILLINGNESS TO CONTINUE IN THIS RESEARCH PROJECT, IT WILL BE GIVEN TO ME AS SOON AS IT BECOMES AVAILABLE. THIS CONSENT FORM WILL BE FILED IN A LOCKING FILE CABINET IN THE RESEARCHERS OFFICE WITH ACCESS RESTRICTED TO AN AUTHORIZED REPRESENTATIVE OF THE FLORIDA STATE UNIVERSITY ECONOMICS DEPARTMENT. I DO NOT GIVE UP ANY OF MY LEGAL RIGHTS BY MY CONSENT. A COPY OF THIS CONSENT FORM WILL BE GIVEN TO ME UPON REQUEST.

_____ Signature and date

_____ Print Your Name

REFERENCES

Alchian, Armen A. and Harold Demsetz. 1972. "Production, Information Costs, and Economic Organization." *The American Economic Review* 62(5):777–795.

Allison, Paul and Richard Waterman. 2002. "Fixed-Effects Negative Binomial Regression Models." *Sociological Methodology* 32:247–265.

Anderson, Simon P., Jacob K. Goeree and Charles A. Holt. 1999. "Stochastic Game Theory: Adjustment to Equilibrium Under Noisy Directional Learning.". URL: `http://ideas.repec.org/p/vir/virpap/327.html`.

Anderson, Simon P., Jacob K. Goeree and Charles A. Holt. 2002. "The Logit Equilibrium: A Perspective on Intuitive Behavioral Anomalies." *Southern Economic Journal* 69:21–47.

Anderson, T. W. and Cheng Hsiao. 1982. "Formulation and estimation of dynamic models using panel data." *Journal of Econometrics* 18(1):47–82.

Aragones, Enriqueta and Thomas R. Palfrey. 2004. "The Effect of Candidate Quality on Electoral Equilibrium." *American Political Science Review* 98:77–90.

Arellano, Manuel and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies* 58:277–297.

Baltagi, Badi. 2005. *Econometric Analysis of Panel Data*. 3rd ed. Wiley.

Barrilleaux, Charles and Michael Berkman. 2003. "Do Governors Matter? Budgeting Rules and the Politics of State Policymaking." *Political Research Quarterly* 56:409–417.

Baum, Christopher F and Fabian Bornhorst. 2007. "NHARVEY: Stata module to perform Nyblom-Harvey panel test of common stochastic trends." Software component provided by Boston College Department of Economics in its series Statistical Software Components with number S419703. URL: `http://ideas.repec.org/c/boc/bocode/s419703.html` accessed May 11, 2008.

Bendor, Jonathan. 1985. *Parallel Systems: Redundancy in Government*. University of California Press.

Bendor, Jonathan, A. Glaser and T. Hammond. 2001. "Theories of Delegation." *Annual Review of Political Science* 4:235–269.

Bernheim, B. Douglas and Michael D. Whinston. 1986. "Common Agency." *Econometrica* 54(4):923–942.

Blundell, Richard, Rachel Griffith and Frank Windmeijer. 2002. "Individual effects and dynamics in count data models." *Journal of Econometrics* 108:113–131.

Bolton, Patrick and Mathias Dewatripont. 2005. *Contract Theory*. The MIT Press.

Book of the States. 2000-2006. Vol. 38-32 Council of State Governments.

Bornhorst, Fabian and Christopher F. Baum. 2006. "LEVINLIN: Stata module to perform Levin-Lin-Chu panel unit root test." Software component provided by Boston College Department of Economics in its series Statistical Software Components with number S419702. URL: `http://ideas.repec.org/c/boc/bocode/s419704.html` accessed May 11, 2008.

Bornhorst, Fabian and Christopher F. Baum. 2007. "IPSHIN: Stata module to perform Im-Pesaran-Shin panel unit root test." Software component provided by Boston College Department of Economics in its series Statistical Software Components with number S419704. URL: `http://ideas.repec.org/c/boc/bocode/s419702.html` accessed May 11, 2008.

Brandt, Patrick, John Williams, Benjamin Fordham and Brian Pollins. 2000. "Dynamic Modeling for Persistent Event-Count Time Series." *American Journal of Political Science* 44(4):823–843.

Brehm, John O. and Scott Gates. 1999. *Working, Shirking, and Sabotage: Bureaucratic Response to a Democratic Public*. Ann Arbor, MI: University of Michigan Press.

Bull, Clive, Andrew Schotter and Keith Weigelt. 1987. "Tournaments and Piece Rates: An Experimental Study." *Journal of Political Economy* 95:1–33.

Bureau of Labor Statistics Occupational and Employment Statistics Program. N.d. URL: `http://stat.bls.gov/oes/home.htm` accessed May 6, 2008.

Burnett, William B. 1987. "Competition in the Weapons Acquisition Process: The Case of U.S. Warplanes." *Journal of Policy Analysis and Management* 7(1):17–39.

Cameron, Colin and Pravin Trivedi. 1998. *Regression Analysis of Count Data*. Cambridge University Press.

Chamberlain, Gary. 1992. "Comment: sequential moment restrictions in panel data." *Journal of Business and Economic Statistics* 10:20–26.

Chubb, John E. 1985. "The Political Economy of Federalism." *American Political Science Review* 79(4):994–1015.

Crawford, Vincent and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50:1431–1451.

Darling-Hammond, Linda. 2007*a*. "Evaluating 'No Child Left Behind'." *The Nation* . URL: `http://www.thenation.com/doc/20070521/darling-hammond` accessed June 6, 2008.

Darling-Hammond, Linda. 2007*b*. "Race, inequality and educational accountability: the irony of 'No Child Left Behind'." *Race Ethnicity and Education* 10(3):245–260.

Dixit, Avinash K. 1997. "Power of Incentives in Private versus Public Organizations." *American Economic Review* 87(2):378–382.

Dixit, Avinash K., Gene M. Grossman and Elhanan Helpman. 1997. "Common Agency and Coordination: General Theory and Application to Government Policy Making." *Journal of Political Economy* 105:752–769.

Downs, Anthony. 1967. *Inside Bureaucracy*. Boston, MA: Little, Brown, and Co.

Drago, Robert and John Heywood. 1989. "Tournaments, Piece Rates, and the Shape of the Payoff Function." *Journal of Political Economy* 97(4):992–998.

Eisner, Marc Allen and Kenneth J. Meier. 1990. "Presidential Control versus Bureaucratic Power: Explaining the Reagan Revolution in Antitrust." *American Journal of Political Science* 34(1):269–287.

Ferejohn, John and Charles R. Shipan. 1990. "Congressional Influence on Bureaucracy." *Journal of Law, Economics, and Organization* 6:1–20.

Fischbacher, Urs. 2007. "z-Tree: Zurich Toolbox for Ready-made Economic Experiments." *Experimental Economics* 10(2):171–178.

Gerhardt, Michael J. 2003. *The Federal Appointments Process*. Duke University Press.

Goeree, Jacob and Charles Holt. 2005. "An Explanation of Anomalous Behavior in Models of Political Participation." *American Political Science Review* 99:201–213.

Golden, Marissa Martino. 2000. *What Motivates Bureaucrats?* Columbia University Press.

Govindan, S. and R. Wilson. 2003. "A global Newton method to compute Nash equilibria." *Journal of Economic Theory* 110:65–86.

Green, Jerry R. and Nancy L. Stokey. 1983. "A Comparison of Tournaments and Contracts." *The Journal of Political Economy* 91:349–364.

Grossman, Sanford J. and Oliver D. Hart. 1983. "An Analysis of the Principal-Agent Problem." *Econometrica* 51(1):7–46.

Guarnaschelli, Serena, Richard D McKelvey and Thomas R Palfrey. 2000. "An Experimental Study of Jury Decision Rules." pp. 407–423.

Hammond, Thomas H. and Jack H. Knott. 1996. "Who Controls the Bureaucracy?: Presidential Power, Congressional Dominance, Legal Constraints, and Bureaucratic Autonomy in a Model of Multi-Institutional Policy-Making." *Journal of Law, Economics, and Organization* 12(1):119–166.

Hausman, Jerry, Bronwyn Hall and Zvi Griliches. 1984. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship." *Econometrica* 52(4):909–938.

Hilbe, Joseph. 2005. "CENSORNB: Stata module to estimate censored negative binomial regression as survival model.". URL: `http://ideas.repec.org/c/boc/bocode/s456508.html` accessed May 22, 2008.

Hilbe, Joseph. 2007. *Negative Binomial Regression*. Cambridge University Press.

Holland, Paul. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81(396):945–960.

Holmstrom, Bengt. 1982. "Moral Hazard in Teams." *The Bell Journal of Economics* 13(2):324–340.

Hursh, David. 2007. "Exacerbating inequality: the failed promise of the No Child Left Behind Act." *Race Ethnicity and Education* 10(3):295Ű308.

Im, Kyung So, M. Hashem Pesaran and Yongcheol Shin. 2003. "Testing for unit roots in heterogeneous panels." *Journal of Econometrics* 115(1):53–74.

Johnson, Ronald N. and Gary D. Libecap. 1989*a*. "Agency Growth, Salaries, and the Protected Bureaucrat." *Economic Inquiry* 27(3):431–451.

Johnson, Ronald N. and Gary D. Libecap. 1989*b*. "Bureaucratic Rules, Supervisor Beahvior, and the Effect on Salaries in the Federal Government." *Journal of Law, Economics, and Organization* 5(1):53–82.

Johnson, Ronald N. and Gary D. Libecap. 1994. *The Federal Civil Service System and the Problem of Bureaucracy: The Economics and Politics of Institutional Change*. University Of Chicago Press.

Judd, Kenneth L. 1998. *Numerical Methods in Economics*. The MIT Press.

Judson, Ruth and Ann Owen. 1999. "Estimating dynamic panel data models: a guide for macroeconomists." *Economics Letters* 65:9–15.

Laffont, Jean-Jacques and David Martimort. 2001. *The Theory of Incentives: The Principal-Agent Model*. Princeton University Press.

Landau, Martin. 1969. "Redundancy, Rationality, and the Problem of Duplication and Overlap." *Public Administration Review* 29(4):346–358.

Lazear, Edward P. and Sherwin Rosen. 1981. "Rank-Order Tournaments as Optimum Labor Contracts." *The Journal of Political Economy* 89:841–864.

Levin, Andrew, Chien-Fu Lin and Chia-Shang James Chu. 2002. "Unit root tests in panel data: asymptotic and finite-sample properties." *Journal of Econometrics* 108:1–24.

Lewin, Tamar. 2007. "States Found to Vary Widely on Education." *The New York Times* . URL: http://www.nytimes.com/2007/06/08/education/08scores.html accessed June 6, 2008.

Lewis, David E. 2008. *The Politics of Presidential Appointments: Political Control and Bureaucratic Performance*. Princeton University Press.

Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage Publications.

Loomis, Burdett. 2001. The Senate: An 'Obstacle Course' for Executive Appointments? In *Innocent Until Nominated: The Breakdown of the Federal Appointments Process*, ed. G. Calvin Mackenzie. Brookings Institution Press chapter 5, pp. 160–172.

Mackenzie, G. Calvin. 1981. *The Politics of Presidential Appointments*. The Free Press.

Maddala, G. S. 1986. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press.

Madison, James. 1999. Federalist No. 51: The Strucure of the Government Must Furnish the Propert Checks and Balances Between the Different Departments. In *The Federalist Papers*, ed. Clinton Rossiter. New American Library pp. 288–292.

Martin, Douglas. 2004. "Anne Gorsuch Burford, 62, Reagan E.P.A. Chief, Dies." *New York Times* . URL: http://query.nytimes.com/gst/fullpage.html?res=9F0CE6D81E3AF931A15754C0A9629C8B63 accessed July 6, 2008.

McCaffery, Jerry and L. R. Jones. 2001. *Budgeting and Financial Management in the Federal Government*. Greenwich, CT: Information Age Publishing.

McCubbins, Mathew D., Roger G. Noll and Barry R. Weingast. 1987. "Administrative Procedures as Instruments of Political Control." *Journal of Law, Economics, and Organization* 3(2):243–277.

McCubbins, Mathew D. and Thomas Schwartz. 1984. "Congressional Oversight Overlooked: Police Patrols versus Fire Alarms." *American Journal of Political Science* 28(1):165–179.

McKelvey, Richard D., Andrew M. McLennan and Theodore L. Turocy. N.d. "Gambit: Software Tools for Game Theory, Version 0.2007.12.04.". URL: http://gambit.sourceforge.net/.

McKelvey, Richard D. and Thomas R. Palfrey. 1996. "A Statistical Theory of Equilibrium in Games." *Japanese Economic Review* 47:186–209.

McKelvey, Richard and Thomas R. Palfrey. 1995. "Quantal Response Equilibria for Normal Form Games." *Games and Economic Behavior* 10:6–38.

McKelvey, Richard and Thomas R. Palfrey. 1998. "Quantal Response Equilibria for Extensive Form Games." *Experimental Economics* 1:9–41.

McNeil, Linda McSpadden, Eileen Coppola, Judy Radigan and Julian Vasquez Heilig. 2008. "Avoidable Losses: High-Stakes Accountability and the Dropout Crisis." *Education Policy Analysis Archives* 16(3). URL: `http://epaa.asu.edu/epaa/v16n3/` accessed June 6, 2008.

Meyer, Annette. 2002. *Evolution of United States Budgeting, Revised and Expanded Edition*. Praeger.

Meyers, Roy T. 1994. *Strategic Budgeting*. University of Michigan Press.

Miranda, Rowan and Allan Lerner. 1995. "Bureaucracy, Organizational Redundancy, and the Privatization of Public Services." *Public Administration Review* 55(2):193–200.

Mirrlees, James A. 1976. "The Optimal Structure of Incentives and Authority within an Organization." *The Bell Journal of Economics* 7(1):105–131.

Moe, Terry. 1987. "An Assessment of the Positive Theory of Congressional Dominance." *Legislative Studies Quarterly* 12(4):475–520.

Moe, Terry M. 1985. "Control and Feedback in Economic Regulation: The Case of the NLRB." *American Political Science Review* 79(4):1094–1116.

Nalbanian, Haig and Andrew Schotter. 1997. "Productivity under Group Incentives: An Experimental Study." *American Economic Review* 87(3):314–341.

National Council of State Legislatures. 1997. "Gubernatorial Veto Authority with Respect to Major Budget Bills." Online. URL: `http://www.ncsl.org/programs/fiscal/lbptabls/lbpc6t3.htm` accessed June 5, 2008.

Niskanen, William A. 1971. *Bureaucracy and Representative Government*. Chicago: Aldine, Atherton.

Nyarko, Y. 2001. "Learning, Quantal Response Equilibrium and Equilibrium in Beliefs.". URL: `http://ideas.repec.org/p/cvs/starer/01-03.html`.

Nyblom, Jukka and Andrew Harvey. 2000. "Tests of Common Stochastic Trends." *Econometric Theory* 16:176–199.

OnTheIssues.org: Jesse Ventura on Environment. N.d. URL: `http://www.ontheissues.org/Governor/Jesse_Ventura_Environment.htm` accessed May 3, 2008.

Ortiz, Brandon. 2008. "Ex-principal 'driven by test scores;' Report accuses her of violations." *Lexington Herald-Leader* . URL: `http://www.kentucky.com/779/story/394252.html` accessed on June 6, 2008.

Reagan, Ronald. 1982. "State of the Union Address." Online. URL: `http://www.c-span.org/executive/transcript.asp?cat=current_event&code=bush_admin&year=1982` accessed June 5, 2008.

Rosen, Bernard. 1989. *Holding Government Bureaucracies Accountable*. 2nd ed ed. New York: Praeger.

Russell, Gregory D., James A. Conser and Rebecca Paynich. 2005. *Law Enforcement in the United States*. Jones & Bartlett.

Schotter, Andrew and Keith Weigelt. 1992. "Asymmetric Tournaments, Equal Opportunity Laws, and Affirmative Action: Some Experimental Results." *Quarterly Journal of Economics* 107:511–539.

Shipan, Charles R. 2004. "Regulatory Regimes, Agency Actions, and the Conditional Nature of Congressional Influence." *American Political Science Review* 98(3):467–480.

Signorino, Curtis. 1999. "Strategic Interaction and the Statistical Analysis of International Conflict." *American Political Science Review* 93:279–297.

Simon, Herbert A. 1997. *Administrative Behavior, 4th Edition*. 4 sub ed. Free Press.

Spence, Michael. 1973. "Job Market Signaling." *Quarterly Journal of Economics* 87(3):355–374.

Taubman, Philip. 2008. "Lesson on How Not to Build a Navy Ship." *The New York Times* . URL: `http://www.nytimes.com/2008/04/25/us/25ship.html` accessed June 5, 2008.

Ting, Michael. 2003. "A Strategic Theory of Bureaucratic Redundancy." *American Journal of Political Science* 47(2):274–292.

Turocy, Theodore L. 2005. "A dynamic homotopy interpretation of the logistic quantal response equilibrium correspondence." *Games and Economic Behavior* 51:243–263.

United States Bureau of the Census. 2006. *Government Finance and Employment Classification Manual*.

United States Bureau of the Census. N.d. "State Government Finances." Online. URL: `http://www.census.gov/govs/www/state.html` accessed May 6, 2008.

United States Environmental Protection Agency. 2003. "AFS National Action Types." Online. URL: `http://www.epa.gov/compliance/resources/publications/data/systems/air/actions.pdf` accessed May 2, 2008.

United States Environmental Protection Agency. 2008*a*. "Enforcement & Compliance History Online." Online. URL: `http://www.epa-echo.gov/echo/idea_download.html` accessed May 1, 2008.

United States Environmental Protection Agency. 2008*b*. "Green Book." Online. URL: `http://epa.gov/oar/oaqps/greenbk/index.html` accessed May 6, 2008.

U.S. House of Representatives Committee on Government Reform. 2004. *United States Government Policy and Supporting Positions*. U.S.Government Printing Office. URL: `http://www.gpoaccess.gov/plumbook/index.html` accessed July 6. 2008.

Verstegen, Deborah. 1990. "Education Fiscal Policy in the Reagan Administration." *Educational Evaluation and Policy Analysis* 12(4):355–373.

Warshaw, Shirley Anne. 1995. "White House Control of Domestic Policy Making: The Reagan Years." *Public Administration Review* 55(3):247–253.

Weingast, Barry R and Moran. 1983. "Bureaucratic Discretion or Congressional Control? Regulatory Policymaking by the Federal Trade Commission." *Journal of Political Economy* 91(5):765–800.

Whitford, Andrew B. 2006. "Incentives and Tournaments in Public Organizations." Online. URL: `http://ssrn.com/paper=873488`.

Williams, John and Patrick Brandt. 2001. "A Linear Poisson Autoregressive Model: The Poisson AR(p) Model." *Political Analysis* 9(2):164–184.

Williamson, Oliver E. 1998. *The Economic Institutions of Capitalism*. New York: Free Press.

Wilson, James Q. 1991. *Bureaucracy: What Government Agencies Do and Why They Do It*. Reprint ed. Basic Books.

Wilson, Sven and Daniel Butler. 2007. "A Lot More to Do: The Sensitivity of Time-Series Cross-Section Analyses to Simple Alternative Specifications." *Political Analysis* 15:101–123.

Windmeijer, Frank. 2000. "Moment conditions for fixed effects count data models with endogenous regressors." *Economics Letters* 68:21–24.

Windmeijer, Frank. 2002. "ExpEND, A Gauss Programme for Non-linear GMM Estimation of EXPonential Models with ENDogenous Regressors for Cross-Section and Panel Data." *Centre for Microdata Methods and Practice Working Paper* CWP14/02.

Wood, B. Dan. 1988. "Principals, Bureaucrats, and Responsiveness in Clean Air Enforcements." *American Political Science Review* 82(1):213–234.

Wood, B. Dan and James E. Anderson. 1993. "The Politics of U.S. Antitrust Regulation." *American Journal of Political Science* 37(1):1–39.

Wood, B. Dan and Richard Waterman. 1991. "The Dynamics of Political Control of the Bureaucracy." *American Political Science Review* 85(3):801–828.

Wood, B. Dan and Richard Waterman. 1993. "The Dynamics of Political-Bureaucratic Adaptation." *American Journal of Political Science* 37(2):497–528.

Wooldridge, Jeffrey M. 1997. "Multiplicative panel data models without the strict exogeneity assumption." *Econometric Theory* 13:667–678.

Yi, Kang-Oh. 2005. "Quantal-response equilibrium models of the ultimatum bargaining game." *Games and Economic Behavior* 51:324–348.

BIOGRAPHICAL SKETCH

Justin Esarey is Assistant Professor of Political Science at Emory University in Atlanta, GA. He graduated in 2008 with a Ph.D. in Political Science from Florida State University in Tallahassee, FL. He also holds a BS in Economics and a BA in Political Science from Bowling Green State University in Bowling Green, OH.

Professor Esarey's research interests are in statistical methodology, game theory, and American public policy. He has published scholarly papers on these topics in *Political Analysis* and the *Journal of Theoretical Politics*. He is the recipient (with Charles Barrilleaux and Tim Salmon) of a 2007 National Science Foundation grant to study income redistribution policy using laboratory experiments.