The London School of Economics and Political Science

The Organizational Basis of Government in Developing Countries: Management and Policy Implementation in Ghana's Public Sector

Martin J. Williams

A thesis submitted to the Government Department of the London School of Economics and Political Science for the degree of Doctor of Philosophy

London, January 2016

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 68,747 words.

Abstract

Organizations are the literal bricks-and-mortar of government, the means by which political choices are turned into delivered goods and services, yet they are largely absent from prominent theories of state capacity, institutions, and political economy in developing countries. In three papers, I show that this omission is theoretically and empirically unjustified: not only is there a great deal of heterogeneity among organizations within the same government, but the complexities of organizations can interact with institutional and political economy factors in a manner that alters these literatures' conclusions about public service delivery, bureaucracy, and reform. The first paper, *"From Institutions to Organizations in the Study of State Capacity"*, draws on interviews with senior managers from 40 organizations in Ghana's central government to document the vast range of variation in management quality among them, explore its roots in theories of relational contracts, and connect theories of organizational performance to the institution-centric literature on state capacity in developing countries. The second paper, *"Policy Implementation, Distributive Politics, and Fiscal Institutions"*, analyzes an original database of 14,000 small infrastructure projects in Ghana's local governments, and shows that the fiscal institutions used to fund projects are associated with large differences in completion rates, even after controlling for project characteristics and district, community, and contractor fixed effects. I develop a theory of policy implementation as intertemporal bargaining among political actors who face commitment problems with respect to project distribution, and show that fiscal institutions can mitigate the negative effects of these distributive pressures. The final paper, *"One Size Does Not Fit All"*, shows that the quality of budget execution and compliance with budget processes varies dramatically across ministries in Ghana's government, and that the drivers of budget performance are heterogeneous and often idiosyncratic.

Acknowledgements

The idea for this dissertation came while I was working in Ghana's Ministry of Trade and Industry (MOTI). I owe a great debt to my former colleagues in MOTI and elsewhere in Ghana's government for their friendship and humor, and for helping me understand – first as a colleague, then as a researcher – why the system works the way it does. My research was motivated by the many public servants who pour themselves into their work with the utmost professionalism and dedication, in spite of all the obstacles and often with little recognition.

I am grateful also to the numerous public servants, far too numerous to give full recognition to individually, who made my research possible. In particular, my thanks go to two Heads of Ghana's Civil Service, Nana-Agyekum Dwamena and before him Woeli Kemevor, for their admirable support for research and evidence-based policymaking, and their belief that I might produce something useful. I hope their faith has been justified. However, no financial support was received from the Government of Ghana during the study, and the opinions expressed herein do not necessarily represent the views of these institutions and individuals and should not be attributed to them.

I have been fortunate to have two dedicated and thoughtful supervisors, Patrick Dunleavy and Joachim Wehner, who have encouraged me to let my mind wander and explore while also reminding me to occasionally bring it back. The London School of Economics and Political Science and the Tyng Scholarship at Williams College provided generous financial support for my doctoral studies, for which I am very thankful, and the African Center for Economic Transformation graciously hosted me during fieldwork in Accra.

For the first paper, I thank Ghana's Office of the Head of Civil Service for facilitating this research, and to the numerous public servants in Ghana who generously gave their time to be interviewed and otherwise support this project. My thanks are also due to Nana Agyekum-Dwamena, Patrick Dunleavy, Ben Eghan, Jean-Paul Faguet, Doug Gollin, Arvind Nair, Dan Rogger, Christian Schuster, Daniela Scur, John Van

Reenen, Joachim Wehner, and participants at the American Political Science Association Annual Conference 2014 and a workshop at Ghana's Civil Service Training Centre for helpful discussions and comments.

The second paper has benefited from advice, data, and discussions with individuals of numerous institutions, including the National Development Planning Commission, Ministry of Local Government and Rural Development, Local Government Service Secretariat, Ministry of Finance Fiscal Decentralization Unit, Office of the Head of Civil Service, District Assemblies Common Fund Administrator, Ghana Audit Service, the District Development Facility Technical Committee, Canadian International Development Agency, World Bank, Ministry of Education, GETFund Secretariat, and Ghana Road Fund Secretariat. Without their time and generous assistance this project would not have been possible. The analysis of this data has been conducted independently and the opinions expressed herein should not be attributed to any of these institutions. I am grateful for comments from Nana Agyekum-Dwamena, Daniel Berger, Stefano Caria, Darin Christensen, Douglas Gollin, Nahomi Ichino, Ryan Jablonski, Peter John, Pia Raffler, Imran Rasul, Dan Rogger, Anisha Sharma, Joachim Wehner, Erik Wibbels, and workshop participants at the Civil Service Training Centre (Ghana), London School of Economics, the Overseas Development Institute, the Department for International Development (UK), EPSA 2015 Annual Conference, and APSA 2015 Annual Conference, and to Nahomi Ichino and Erik Wibbels for sharing data. This research was funded by the International Growth Centre grant 1-VRG-VGHA-VXXXX-89105. Allan Kasapa and Joseph Napen provided excellent research assistance.

For the third paper, I thank Ghana's Ministry of Finance for facilitating this research, in particular Eva Mends and numerous officers of the Budget Division. My thanks are also due to Patrick Dunleavy, Christopher Humphrey, Philipp Krause, Arvind Nair, Dídac Queralt, Christian Schuster, Joachim Wehner, three anonymous referees, and workshop participants at the Overseas Development Institute and the European Political Science Association 2014 conference for helpful discussions and comments.

More personally, I have been extremely fortunate to have had the support of many friends in London, Accra, and elsewhere, who have somehow managed to keep me sane. They are too numerous to name individually, but they know who they are and how grateful I am for their friendship and patience. That said, a special thanks is due to Zoe Marks – way above the game.

Most of all, thanks to my mum, the bravest person I know, and to Kuukuwa, for believing in me and making me laugh.

Any remaining mistakes are my own.

Table of Contents

List of figures, tables, and photographs

Note: For simplicity, tables and figures are numbered within each chapter rather than cumulatively throughout the dissertation.

Chapter One

INTRODUCTION:

THE ORGANIZATIONS OF THE STATE

Chapter Abstract

This chapter introduces the dissertation's motivation and the key distinction between viewing the administrative apparatus of the state as a set of organizations and viewing it as an institution. I survey how this distinction applies to prominent literatures on bureaucracy and public service delivery, develop an overarching theoretical framework that ties together the three substantive papers, and introduce each paper. I then provide some brief background on Ghana's state bureaucracy in comparative perspective, and conclude by delimiting the scope of the dissertation.

Introduction:
The Organizations of the State

The starting point of this dissertation is the observation that the administrative apparatus of the state is composed of a set of organizations. While “the state” exists in the abstract, it takes tangible form in these ministries, agencies, departments, and commissions, each with their own office buildings, authorizing laws, and human resource divisions. Organizations are the literal bricks-and-mortar of the state: they are the street-level “front office” of the state that interacts with citizens as well as the administrative “back office” through which organizations are funded, staffed, monitored, and audited by still other organizations. While this observation may seem so obvious as to be trivial, it has far-reaching consequences for both theoretical and empirical approaches to studying government performance and the development of state capacity.

First, thinking about the state bureaucracy as a unitary institution rather than a set of linked organizations influences the choice of the appropriate unit of analysis – in particular, it lends itself to an analytical focus on variation in performance *across* governments rather than *within* governments. This focus on measuring and explaining variation across geographically defined units – societies, nations, sub-national regions – has dominated theoretical and empirical approaches to studying state capacity, governance, and public service delivery. It is undoubtedly true that variation in performance across governments is significant, but a growing number of studies from diverse theoretical and methodological perspectives have documented that the variation within governments can also be significant, and perhaps even greater than the variation across governments (e.g. Ingraham, Joyce, and Donahue 2003; Leonard 2010; Gingerich 2013; Rasul and Rogger 2014). From both a theoretical and empirical perspective, then, the assumption that within-government variation is second-order to across-government variation is just that: an assumption.

Second, studying the state as a unitary institution has implications for the types of determinants and mechanisms that are examined in the effort to explain performance. If performance is assumed to vary primarily at the societal level, then inevitably the variables that are called on to explain this variation must also vary at the same level.

Thus, while the determinants explored by the empirical literature on long-run economic and institutional development are bewilderingly diverse, what they share is that they do not vary *within* the societal unit of analysis. This focus on across-government variation biases the choice of mechanisms to be examined towards those that are long-term, abstract, and out of the control of any individual actor or group – since, almost by definition, only these types of determinants and mechanisms can operate at the level of an entire society. While there are advantages to this approach, it means that the existing body of literature on institutions and state capacity has little of practical relevance to say about the key questions of whether and why some organizations within a given government perform better than others, and what individual actors or groups can do to improve government performance.

Finally, neglecting the organizational basis of states can lead scholars of developing country governments also to overlook the rich bodies of work from organization theory, public administration, American politics, and organizational economics that address precisely the questions that bedevil the institutionalist approach to states – such as whether and how political principals can actually control public bureaucracies, the extent to which organizational functioning is determined by the organization's external political and institutional context, and the internal dynamics of organizational performance. Yet while these literatures have thoroughly interrogated these dynamics in the context of developed countries, there is much less work that examines these same issues in developing countries, where many of the presumptions of this literature – for example that laws and administrative regulations will be (or even can be) enforced – may not hold.

The three papers of this dissertation represent three attempts to grapple with the organizational basis of government in developing countries. Each makes a novel theoretical argument that starts from this organizational perspective, and each supports this argument with empirical evidence drawn from original fieldwork. While each paper attempts to provide some guidance for efforts at reform, my broader hope is that the dissertation demonstrates the need to start asking different *questions* about governments in developing countries. In particular, rather than asking why one country's government performs better than another country's, scholars and policymakers alike can often learn more by asking why one organization in a country's government performs better than another. This dissertation does not claim to provide a complete answer to this genre of question – and indeed I will later suggest that the complexities of organizations may make a truly complete answer impossible – but it does aim to provide theoretical, methodological, and empirical steps towards a more insightful and relevant approach to the problem of government performance in developing countries.

Before proceeding, a definitional note: throughout this section and indeed this dissertation, I use terms like state capacity and organizational performance that are multi-dimensional and highly contested (Rothstein and Teorell 2008, Talbot 2010, Fukuyama 2013, Rotberg 2014). I regard these as essentially contested concepts (Gallie 1956; Collier, Hidalgo, and Maciuceanu 2006) in the sense that definitional debates about them are productive but also unresolvable. Since my primary focus is on highlighting and understanding within-government variation in performance, I do not seek to adjudicate or add to these debates, although each paper makes clear the specific sense in which I am using performance and defends my methodology for measuring it. Likewise, for convenience and to illustrate my argument I often discuss them as unidimensional constructs (e.g. "better" and "worse" organizations). While a more nuanced treatment may be a fruitful area for future research, this simplification facilitates the more basic task of this dissertation: to point out the salient fact of variation in performance within governments and to consider its causes, mechanisms, and consequences.

The remainder of this introduction proceeds as follows. The next section provides a very brief and selective review of the main bodies of literature that bear on these questions, focusing on how they conceive the relationships among states, institutions, and organizations. Section 1.2 then integrates these diverse approaches into a common conceptual framework for studying the determinants of organizational performance in government. Section 1.3 situates the three papers of this dissertation within this framework, and Section 1.4 discusses the relevance of the Ghanaian case in comparative perspective. Section 1.5 concludes by delineating the scope of the dissertation.

1.1 STATE AS INSTITUTION OR STATE AS ORGANIZATIONS?

The literature on the long-run development of state capacity and governance quality focuses heavily on differences across countries, arguing that these are driven by national-level variation in various geographical, historical, institutional, political economy, cultural, and even genetic factors (North 1990; Putnam 1993; Gallup, Sachs, and Mellinger 1999; Acemoglu, Johnson, Robinson 2001; Besley and Persson 2009; Acemoglu, Garcia-Jimeno, and Robinson 2012; Ashraf and Galor 2013). Another strand of research examines bureaucratic quality and formal political institutions more directly, by using cross-national governance indicators (Evans and Rauch 1999; Kaufmann, Kraay, and Mastruzzi 2010; Teorell, Dahlström, and Dahlberg 2011). While these studies differ in their hypotheses and conclusions, they share in common an inability to explain significant organization-level differences within the same government, since these organizations are typically located within a few miles of each other in the capital city, are staffed from the same population pool, and are subject to the same formal institutions and national-level political context. In their focus on cross-country variation, these literatures implicitly assume that variation among organizations is negligible, either because such variation is not substantial or is solely a matter of exceptional cases, and that organizational performance is constrained fairly tightly by these national-level variables.

This perspective has come to dominate thinking about governance and reform in recent decades. As I will argue more fully later in this dissertation, much of this follows from North's (1990) distinction between institutions as the formal and informal "rules of the game" and organizations as the "players" of this game; his conscious decision to focus on institutions rather than organizations; and his accompanying assumption that organizations are rational actors that maximize some objective function within the constraints defined by their institutional context. While this institutionalist approach has influenced a large literature that focuses explicitly on questions of how institutional development stems from various political economy, historical, and other factors (Acemoglu, Johnson, and Robinson 2005; North, Wallis, and Weingast 2009), its basic assumption – that the state is best seen as an institution – has also become pervasive in the more applied practitioner literature on governance (e.g. World Bank 2002). Even critics of the one-size-fits-all approach to governance such as Grindle (2011) and Andrews (2013) frame their critiques and recommendations in terms of *institutional* reform, providing little basis for understanding why there might be variation among organizations within the same government. Likewise, the mainly case study-based literature on “islands” or “pockets” of effectiveness in developing country governments (Grindle 1997, Tendler 1997, Leonard 2010, Roll 2013) emphasizes that such variation is “exceptional”, thus reinforcing the notion that organizations are tightly constrained within the institutional and political context within which they operate.

In stark contrast, a large literature in American politics and public administration explicitly examines whether and how the operation of government organizations can be controlled by political principals and legal frameworks (McCubbins and Schwartz 1984; McCubbins, Noll, and Weingast 1987; Moe 1990; Macey 1992; Epstein and O’Halloran 1999; de Figueiredo 2002; Huber and Shipan 2002). These authors point out that political control of the bureaucracy may be imperfect due to the costs of monitoring and the incomplete nature of laws and regulations, giving rise to a potentially significant amount of agency discretion. The primary analytical focus of this literature is on the control mechanisms that politicians select to induce the bureaucracy to do what they want – in particular, the tradeoff between rules and discretion – and the strategic opportunities and considerations involved in this. However, this literature has limited applicability to developing country contexts, because it assumes that administrative laws are enforceable (even if incomplete) and that bureaucrats are competent. While these may be appropriate default assumptions in countries with relatively strong states, these conditions are often absent in weak states. Indeed, Huber and McCarty (2004) show that integrating low bureaucratic capacity into a standard delegation model weakens politicians’ control of the bureaucracy, reverses many of the standard conclusions of the literature, and diminishes politicians’ incentives for reform. This literature thus provides considerable grounds for scepticism that bureaucracies are tightly constrained by their institutional and political contexts.

The large literatures on differences in organizational performance in both the public and private sectors (primarily of developed countries) provide a further reason to doubt the appropriateness of North’s assumption that organizations are tightly constrained by their institutional environments. Studies of government performance in OECD countries have long taken organizations as their unit of analysis and documented substantial differences in performance among them (Wilson 1989; Lynn, Heinrich, and Hill 2000; Ingraham, Joyce, and Donahue 2003; Meier, O'Toole, and Nicholson-Crotty 2004; Talbot 2010). There is even stronger evidence in the private sector, where numerous studies have documented large and persistent differences in productivity and management quality among organizations even within the same narrowly defined field (Gibbons and Henderson 2013), with evidence most heavily concentrated in the manufacturing sector (Bloom and Van Reenen 2007, Syverson 2011) but extending also to other fields such as hospitals (Carrera and Dunleavy 2010; Bloom, Propper, Seiler, and Van Reenen 2015). While average levels of management and productivity do vary from country to country and are influenced by common factors, such as contract enforcement and competition levels, the importance of heterogeneity in performance among organizations has been widely recognized and become a key element of influential models even outside of organizational economics, such as international trade (e.g. Melitz 2003). If these "persistent performance differences" (Gibbons and Henderson 2013) can exist even in ruthlessly competitive for-profit sectors where organizations' outputs and outcomes are easily measured, then presumably the scope for differences in performance among *ex ante* identical organizations is even larger in the public sector, where outputs and outcomes are typically difficult to measure, incentives are weak and diffuse, competition is scarce or non-existent (Wilson 1989, Prendergast 2003), and there is no built-in mechanism to allow inefficient organizations to exit.[1]

This cross-sectional evidence that organizational performance may not be tightly constrained is supported by a growing body of experimental and quasi-experimental studies that demonstrate that specific interventions and policy design features can improve policy outcomes in both the public (Reinikka and Svensson 2004; Olken 2007; Zamboni and Litschig 2013; Banerjee *et al* 2014; Yanez-Pagans and Machicado-Salas 2014) and private sectors (Bloom *et al* 2013; Atkin *et al* 2015; Blader *et* al 2015; Martinez *et al* 2015). These findings are convincing in the context of particular programs over relatively short time frames, but a shortcoming of this

[1] I am grateful to Patrick Dunleavy for this last point.

approach to studying organizations is that it is unclear whether these interventions would still be effective outside of narrow and carefully controlled implementation settings, in bureaucratic and political equilibrium, or once scaled up to national level (Rodrik 2009, Deaton 2010, Olken and Pande 2012). Because this literature is geared towards understanding discrete policy changes, it has largely neglected to address the organizational context in which these policies are implemented. Yet complex issues like corruption, management, and governance are rarely "solved" with a single policy change; rather, addressing them requires sustained and iterative effort over a long period of time, both through internal organizational processes and external oversight and control mechanisms. This is the conceptual starting point for this dissertation's analysis of organizational performance in the developing country governments.

1.2 A CONCEPTUAL FRAMEWORK FOR ORGANIZATIONAL PERFORMANCE IN THE PUBLIC SECTOR

This section presents a conceptual framework that differentiates the state as an institution from the state as a set of organizations, and arrays the factors that influence organizational performance accordingly. The framework is not intended to be comprehensive, to privilege certain hypotheses, or to propose specific mechanisms, but rather to reconcile the conflicting perspectives and assumptions discussed in the previous section and show how they relate to each other. Likewise, the distinctions made by the framework are not always black-and-white, and many theoretical perspectives concern the connections between different levels of explanations; nevertheless, the framework provides a useful way of organizing and reconciling an extremely diverse set of theories and literatures.

The framework rests on two distinctions. First, some determinants of organizational performance apply to *all* organizations in a given government, while others apply to only a *subset* – or just a single organization. Obvious examples of the former would be any of the variables put forward by the cross-national literature on institutions, governance, and state capacity. For instance, a country's wealth, average level of education, judicial system effectiveness, generalized social trust in organizations, or quality of political leader all influence the performance of government organizations, but presumably do so relatively consistently across all organizations.[2] Less obvious examples are what Ménard (2014) refers to as "meso-institutions": the cross-cutting institutional and organizational arrangements that instantiate abstract institutions and provide the means through which particular organizations are embedded in their institutional context. In the public sector context this could include central management agencies such as the finance ministry, state audit organization, and civil service personnel management organization, as well as more general societal institutions such as judicial courts and administrative tribunals. When these function more effectively, this presumably enhances the operation of all the operational ministries and agencies that they regulate or provide services to.

Second, among the determinants of organizational performance that vary within a government, I distinguish between determinants that are *external* to an organization and those that are *internal* to it. The external factors that influence an organization's performance are those related to its specific context and its structural characteristics, such as: the task characteristics of the organization, for example the degree to which its outputs or outcomes can be measured (Wilson 1989) or the degree of discretion required by frontline bureaucrats (Mintzberg 1983); its degree of legal, financial, policy, and managerial autonomy to go about these tasks (Carpenter 2001); its budget structure (Dunleavy 1991); and organization-specific political economy factors such as interest group configurations, policy networks, and the alignment of the organization's critical tasks with these stakeholders' interests (Wilson 1989). These factors are external to the organization in the sense that they are not under the organization's direct control, at least in the short- to medium-term, although in the long term organizations can of course cultivate relationships with interest groups, seek changes to their authorizing laws, develop alternative sources of revenue to give them more autonomy, and so on.

On the other hand, factors internal to the organization are those over which the organization *does* exercise direct control, even if actually controlling them is not

[2] This may not always be the case, for example if trust or education is more important to the activities of some organizations than to others'. This would be an example of an interaction between general societal factors and organization-specific structural characteristics. Exploring such interactions would be a fruitful area for further study.

trivial. The most obvious example, and the one on which this dissertation dwells the most, is management: broadly understood as the structures, processes, and practices that are used to get work done on an everyday basis. This also includes self-reflective processes related to the organization's efforts to measure its own performance, diagnose problems, and identify and implement improvements. Importantly, management consists both of the formal, codifiable, *de jure* way in which the organization's staff are intended to behave, but also on the informal, difficult to communicate, *de facto* implementation or non-implementation of these formal features. While actually improving management in an organization can be difficult – see for example Gibbons and Henderson's (2013) useful typology of problems of perception, inspiration, motivation, and implementation in improving management practices – organizations are rarely directly constrained from doing so by external forces. This distinguishes internal factors like management from external aspects of the organization's context or structural characteristics over which it has considerably less control.

Combining these two distinctions yields a stylized three-level framework for classifying the potential determinants of organizational performance in the public sector: 1) *government-wide* factors, that affect all organizations in a government and are external to any one organization; 2) aspects of *organizational context and structural characteristics*, that affect only a subset of organizations but are external to these organizations; and 3) *management*, which varies at the organizational level and is internal to the organization. Table 1 illustrates this framework.

Although these distinctions are conceptually useful, they are not meant to be absolute, as some hypothesized determinants span multiple categories, interact across categories, or could be classified in different categories in different situations. For example, a factor like personnel quality could influence organizational performance across all three levels: the overall level of education in a polity, the quality of human resources in a particular sector (for example if a country has an especially strong tradition or educational system in a particular field), and the effectiveness of a particular organization's hiring practices at recruiting talented individuals into the organization. Likewise, the effect of organizational leadership on performance may operate largely through mechanisms internal to the organization, but especially in the public sector is often determined externally through appointments. Similarly, in the public sector some internal management practices are in fact regulated or mandated by externally defined laws and administrative regulations, either government-wide or sector-specific (in an organization's authorizing law, for example). While these considerations certainly nuance the application of the framework to thinking about the determinants of government performance, the distinctions on which the framework rests are nonetheless important ones.

Table 1: Conceptual framework for the determinants of organizational performance

Type of determinant	External/ internal to organization	Scope within government
Government-wide *e.g. National-level economic variables, political institutions, social norms, judicial system effectiveness, quality of central management agencies*	External	All organizations
Organizational context and structural characteristics *e.g. Task characteristics, autonomy, budget structure, policy networks, interest group orientation*	External	Subset of organizations
Management *e.g. Organizational processes, organizational slack, "X-inefficiency", relational contracts, leadership, organizational culture*	Internal	Subset of organizations

Note also that the possibility of multiple equilibria in internal organizational factors means that these factors may not collectively be fully determinative of organizational performance. In particular, factors that are external to the organization may operate as *constraints* on the performance of organizations, much as institutions function as constraints on the actors embedded in them.[3] If organizations operate efficiently within these constraints, then these constraints directly determine organizational performance: relaxing them (for example, if a government's finance ministry improves its management of the budget) leads to an improvement in organizational performance, while tightening them worsens performance. However, recognizing that many determinants of organizational performance are internal to the organization implies that many organizations may be operating inefficiently, *within* rather than *on*

[3] In this dissertation's first paper, I discuss the roots of this view of institutions as constraints in North's (1990) influential work.

their notional production frontier. If this is the case, then relaxing or tightening the external constraints to organizational performance may have no effect – for a badly managed organization, the effectiveness of the ministry of finance may not be a binding constraint.

The idea that organizations often operate inefficiently has a long history in organization theory and management studies (Cyert and March 1963, Leibeinstein 1966) and the more recent economics literature on relational contracts has shown how this inefficiency can develop within an organization through a dynamic process due to the challenge of developing shared understandings, expectations, and norms.[4] The dynamic way in which relational contracts develop inside organizations suggests that there could be multiple performance equilibria possible for any given organization, and thus that *ex ante* identical public bureaucracies – as defined by the combination of government-wide factors and organizational context and structural characteristics – could perform at significantly different levels. This multiple-equilibria characteristic of organizational performance creates a role for the type of path dependence that is usually associated (in analysis of the state) with the literature on historical institutionalism (Mahoney and Thelen 2010; Hacker, Pierson, and Thelen 2015). While I explore this idea in more depth later in this dissertation, I note it now to point out that the conceptual framework put forth in this section is broad enough to encompass not only different sets of determinants of organizational performance, but also different logics about the mechanisms through which these determinants map onto organizational performance.

1.3 THREE PAPERS ON WITHIN-GOVERNMENT VARIATION

The three papers of this dissertation each ask different research questions, use different methods, and draw on different sources of data. The papers also fit into the conceptual framework presented above in different and complementary – although by no means comprehensive – ways. What they share is a common motivation: understanding within-government variation in performance. Accordingly, each paper asks two questions:

[4] Gibbons and Henderson (2013) provide a useful review.

1. What is the extent of within-government variation in performance?
2. What mechanisms generate this variation?

The first, primarily descriptive, question is of interest because each paper demonstrates the existence of substantial variation in contexts where existing empirical evidence of variation is very limited, and where existing theoretical approaches do not provide adequate explanation for this variation. The second, more analytical, question then uses this variation to motivate and examine a novel theoretical argument. Although each paper also makes other, more topic-specific contributions, these two central questions run through all three papers.

The first paper, "From Institutions to Organizations: Management and Informality in Ghana's Public Bureaucracies", addresses within-government variation in organizational performance head-on by gathering data on management quality in 40 central government organizations and analyzing the resulting distribution. Using management quality as a proxy for organizational performance, it finds that not only is there a broad range of variation in management quality across organizations – whether measured by quantitative benchmarks or qualitative description of everyday management practices – but that this variation is substantial and systematic, and is not limited to a handful of exceptional "islands" of excellence. Instead, the distribution of organizational performance in government is smooth and continuous, ranging from extremely dysfunctional organizations to exceptionally well managed ones. In between these two extremes, organizations differ from one another in marginal yet collectively significant ways.

This variation is poorly explained by the institution-centric theories of state capacity in developing countries, which focus exclusively on government-wide determinants of organizational performance, so I develop an alternative but complementary approach grounded in theories of relational contracts that can explain variation in management quality, organizational culture, and organizational performance even among *ex ante* identical organizations. I find qualitative evidence consistent with this explanation, and link it to theories of informal institutions.

In terms of the conceptual framework presented above, the paper aims to focus in on variation in internal management practices while holding constant the government-wide factors on which much previous research has focused. As I discuss in the paper, isolating the effects of internal factors from organizational characteristics is extremely difficult because the vast number of potential organizational characteristics to consider makes any attempt to identify them empirically vastly overdetermined: although my sample size of 40 organizations is larger than most qualitative, case study-based approaches by an order of magnitude, the number of organizational characteristics that could conceivably affect organizational performance is an order magnitude larger still, even without considering interactions among them. I deal with this in a limited way by demonstrating that quantitative variation also exists within sub-groups defined according to key organizational characteristics, as well as by drawing on qualitative evidence from interviews to establish that the patterns of variation are qualitatively consistent with relational contracts theory. Future research with larger sample sizes could help address this challenge, but the finite number of government organizations in most small- and medium-sized governments – about 200, in Ghana's central government – nevertheless poses a challenge. This is therefore an area in which mixed-method approaches are likely to be particularly important, and this paper contributes one methodology for doing this in an integrated manner.

The second paper, "Policy Implementation, Distributive Politics, and Fiscal Institutions: Evidence from Unfinished Infrastructure in Ghana", takes a very different approach. Drawing on an original database of over 14,000 small infrastructure projects undertaken by Ghana's local governments, it examines how the fiscal institutions through which projects are funded affect project completion. These fiscal institutions are similar to Ménard's meso-institutions, in that they provide a tangible connection between the organizations that actually implement policy and the more abstract national-level budget institutions and resource management by Ghana's state. I am able to examine the effect of these fiscal institutions on project completion by taking advantage of a unique feature of the dataset: the same types of projects are executed in the same districts and communities by the same pool of contractors through different funding sources. This lets me control for all the observable and unobservable characteristics of the politicians, bureaucrats, and local authorities where the projects are being implemented – corresponding to the second and third levels of the conceptual framework – and thus isolate the operational effects of the fiscal institutions. (I use coefficient stability tests to examine issues of endogenous project selection into funding mechanisms, and argue that such a selection mechanism is unlikely to account for more than a fraction of the observed variation.) I also demonstrate one mechanism through which fiscal institutions affect project completion: by mitigating the consequences of intertemporal bargaining failures over project distribution among local elites facing commitment problems. I present evidence consistent with this view, most notably that the effect of fiscal institutions on project completion is strongly conditioned by the ruling party's vote share in the district, and show evidence against alternative explanations.

This paper complements the first in several ways. Whereas the first paper relies on a process-based measure of performance – management quality – gathered through intensive qualitative interviews that are then post-coded with a numerical benchmarking scheme, the second paper examines an output-based measure of performance – whether infrastructure projects are completed or left abandoned mid-construction – measured using secondary data I digitized from government administrative records. And while the first paper investigates technical aspects of management that are internal to the organization, the second paper uses district fixed effects to partial out differences in internal management or personnel quality across districts and instead seeks to understand how political factors – namely distributive pressures and bargaining among local political actors – affect policy implementation.

The third paper, "One Size Does Not Fit All: Budget Institutions and Performance in Ghana's Spending Ministries", combines aspects of both of these papers. It uses fiscal data and structured interviews with 23 budget schedule officers in Ghana's Ministry of Finance to investigate differences across Ghana's spending ministries in the quality of budget management and in various measures of budget performance. These ministries are all subject to the same formal budget institutions and national-level political and institutional context, so variation must be driven by organization- or sector-specific factors. As with the first paper, the existing comparative literature on budget institutions and performance focuses almost entirely on differences across governments rather than within them, thus giving little explanation for any variation within countries. I find that not only is there substantial variation across ministries in expenditure outturns and volatility, compliance with formal budget processes, and other outcomes such as extra-budgetary spending, but that there is no correlation between procedural compliance and budget outcomes. That is, performing the on-paper aspects of budget management has little relationship with budget performance. Instead, I find that the drivers of ministries' budget performance are heterogeneous and often idiosyncratic to the particularities of individual sectors and organizations.

Like the first paper, "One Size Does Not Fit All" focuses on within-government variation in order to cancel out the government-wide factors affecting budget performance. However, it spends more time considering the relationship between the external context and structural characteristics of organizations and their internal management in determining performance. It does so using both quantitative and qualitative approaches, although it is again constrained in this by the inherently limited sample size of ministries within a given country. At the very least, though, its findings suggest that variation in performance is only weakly explained by the types of structural characteristics of organizations that might be thought to be important determinants. While a more sweeping or generalizable conclusion about the determinants of budget performance would make a more easily digestible "takeaway" message, the idea that the mechanisms that generate organizational performance could be heterogeneous and highly idiosyncratic – both in terms of external and internal determinants – perhaps provides a deeper challenge to scholars and reformers alike.

1.4 GHANA'S STATE IN COMPARATIVE PERSPECTIVE

Ghana is a useful site in which to study these issues because it is something of a "best case" for governance and institutional quality in Africa. Ghana is a stable and institutionalized democracy with two dominant parties that have alternated in power twice since 1992. The country has little recent history of large-scale violent conflict, and appears to have set its more recent history of *coups d'état* behind it. Ghana's economy is relatively prosperous and its population relatively well educated, and the country has a free and active press and a vibrant civil society. Although some ethnic tensions do exist and both major parties draw to some extent on ethno-regional bases of support, these ethnic divides are less severe and politically mobilized than in many other countries. Together, these features mean that Ghana is often viewed internationally as a bright spot for governance in Africa.

Despite these apparently favorable institutional and political conditions, Ghana's public bureaucracies have been persistently perceived as inefficient, poorly managed, and corrupt, resulting in a long string of reform programs since the 1980s that have each aimed to improve management and productivity in government: the Public Administration Restructuring and Decentralization Implementation Committee (PARDIC), Civil Service Reform Programme (CSRP), National Institutional Renewal Programme (NIRP), Civil Service Performance Improvement Programme (CSPIP), Public Financial Management Reform Programme (PUFMARP), Public Sector Re-Invention and Modernization Strategy (PUSERMOS), Public Sector Incomes Administration Reform, Subvented Agency Reform, and Single Spine Pay Policy (SSPP) (Ayee 2001; Owusu 2006; Antwi, Analoui, and Nana-Agyekum 2008; Eghan 2008; Wereko 2009; Annan-Prah and Ohemeng 2015). The history of these reforms and fluctuations broadly tracks international trends in public management and donor programming: Africanization and central planning in the 1950s and 1960s; tensions between civilian and military rule in the 1970s; retrenchment, decentralization, cash budgeting, and agencification in the 1980s and 1990s; medium-term expenditure frameworks, poverty reduction plans, anti-corruption initiatives, integrated financial management systems, e-government, and performance management in the 2000s and 2010s. This long history of both government- and donor-led reforms – which is shared by many other African countries – suggests that if Ghana's public bureaucracies are inefficient, it is not for lack of reform effort and international "best practice".

In terms of political institutions, Ghana has a hybrid regime with a president elected by national popular vote and a unicameral parliament composed of representatives elected from single-member constituencies. Although the presidency and parliament could, in principle, be controlled by different parties, this has not occurred since the return to democracy in 1992. Parliament's role in policymaking and implementation is limited to debating and voting on bills proposed by the executive branch, and conducting *ex post* scrutiny and hearings on audit reports and occasionally other issues on an *ad hoc* basis. I describe the country's administrative and bureaucratic structures in more detail in each of the three papers, but broadly speaking the central government bureaucracy retains many elements of its colonial-era British heritage, and since the 1980s the country has undergone a partial but progressive decentralization of some government functions – particularly those related to the delivery of small-scale, basic public goods and services – to district governments. This combination of executive dominance, colonial bureaucratic origins, and partial decentralization is shared by many countries in sub-Saharan Africa, so the study of bureaucracy and within-government variation in performance in Ghana is likely to have wider relevance.

5. SCOPE OF THE DISSERTATION

Having discussed what this dissertation seeks to accomplish, it is worth noting some things that it does *not* seek to do. Although I study within-government variation in performance among organizations, I do not seek to provide a general answer to the question "What causes some organizations to perform better than others?" This questions seeks a simple answer of the form "variable *X* causes organizational performance", and thinks about organizational performance in the same way that cross-country regression studies approach the determinants of economic growth or institutional quality.

Although it is an intuitive question to ask, I argue that it is in many ways the *wrong* question. I have discussed one practical reason for this: governments are composed of a finite number of organizations that each have dozens or even hundreds of potentially relevant characteristics, so organizational performance is empirically overdetermined. Even absent practical considerations, however, the conceptual framework presented in Section 1.2 suggests a deeper reason related to the multiple levels of determinants of organizational performance: while "*X* causes organizational performance" may be a logically coherent way of thinking about external factors that influence an organization's performance, it is more problematic when *X* is a factor that is internal to the organization, such as management or organizational culture. Since internal determinants of organizational performance are (by definition) endogenous to the organization, their causal effects cannot be identified using observational data. While they can be identified through experimental methods, this raises serious concerns about external validity, since any experimental intervention that is plausibly exogenous must (also by definition) be initiated by external forces and thus can no longer be truly internal. Finally, the multiple equilibria character of management and the important distinction between *de jure* and *de facto* practices limit the generalizability of any such intervention, since its effects are strongly contingent on organizational context and other non-verifiable aspects of its implementation.

Instead of asking what causes organizational performance, this dissertation instead focuses on understanding the mechanisms through which some organizations perform better than others. The first paper measures variation in management quality and explores relational contracts as a mechanism that generates this variation, but does not attempt to explain why some organizations have "good" relational contracts. The second paper measures variation in project completion rates across different fiscal institutions (project funding sources) and identifies a political economy mechanism that generates these differences, but does not attempt to explain the origin of these fiscal institutions or why they are designed differently. The third paper comes closer to investigating the determinants of organizational performance, but nevertheless finds that both observable organizational characteristics and compliance with budget procedures have little correlation with budget performance, instead concluding that the drivers of performance tend to be heterogeneous across and idiosyncratic to particular organizations.

To return to the cross-country growth regression analogy, in emphasizing mechanisms and heterogeneity my approach is much more similar to the "growth diagnostics" approach that, instead of asking "What causes economic growth on average?", asks "What is the binding constraint to economic growth in this particular country at this time?" (Hausmann, Rodrik, and Velasco 2005; Hausmann, Klinger, and Wagner 2008). My approach helps provide a more robust theoretical foundation to the idea propounded by the applied governance and development literature that "good enough governance" (Grindle 2011) or "problem-driven iterative adaptation" (Andrews 2013) are preferable to one-size-fits-all approaches to governance reform. However, this dissertation is focused on understanding the roots of variation in performance rather than contributing directly to the design of reforms – diagnosis, rather than prescription. While I note some specific implications for reform in each of the three articles, I instead reserve discussion of reform for the dissertation's conclusion.

REFERENCES

Acemoglu, Daron, Simon Johnson, and James Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91(5): 1369-1401.

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2005. "Institutions as a Fundamental Cause of Long-Run Growth." In Aghion, Philippe, and Steven N. Durlauf (eds), *Handbook of Economic Growth, Volume 1A*, 385-472.

Acemoglu, Daron, Camilo García-Jimeno, and James A. Robinson. 2012. "Finding Eldorado: Slavery and Long-Run Development in Colombia." *Journal of Comparative Economics* 40 (4): 534–64.

Andrews, Matt. 2013. *The Limits of Institutional Reform in Development: Changing Rules for Realistic Solutions*. Cambridge: Cambridge University Press.

Annan-Prah, Elizabeth, and Frank Ohemeng. 2015. "Improving Productivity Through Performance Management in Public Sector Organizations in Ghana: Is Change Management the Answer?" Mimeo, 30 March.

Antwi, K.B., F. Analoui, and D. Nana-Agyekum. 2008. "Public Sector Reform in Sub-Saharan Africa: What Can Be Learnt From the Civil Service Performance Improvement Programme in Ghana?" *Public Administration and Development* 28: 254-264.

Ashraf, Quamrul, and Oded Galor. 2013. "The 'Out of Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development." *American Economic Review* 103(1): 1-46.

Atkin, David, Azam Chaudhry, Shamyla Chaudhry, Amit Khandelwal, and Eric Verhoogen. 2015. "Organizational Barriers to Technology Adoption: Evidence from Soccer-Ball Producers in Pakistan." NBER Working Paper 21417, July.

Ayee, Joseph. 2001. "Civil Service Reform in Ghana: A Case Study of Contemporary Reform Problems in Africa." *African Journal of Political Science* 6(1): 1-41.

Banerjee, Abhijit, Raghabendra Chattopadhyay, Esther Duflo, Daniel Keniston, and Nina Singh. 2014. "Can Institutions be Reformed from Within? Evidence from a Randomized Experiment with the Rajasthan Police." Mimeo.

Besley, Timothy, and Torsten Persson. 2009. "State Capacity, Conflict, and Development." *Econometrica* 78(1): 1-34.

Blader, Steven, Claudine Gartenberg, Rebecca Henderson, and Andrea Prat. 2015. "The Real Effects of Relational Contracts." *American Economic Review Papers and Proceedings* 105(5): 452-456.

Bloom, Nicholas, and John Van Reenen. 2007. Measuring and Explaining Management Practices Across Firms and Countries. *Quarterly Journal of Economics,* 122**,** 1351-1408.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. "Does Management Matter? Evidence from India." *Quarterly Journal of Economics* 128(1): 1-51.

Bloom, Nicholas, Carol Propper, Stephan Seiler, and John Van Reenen. 2015. "The Impact of Competition on Management Quality: Evidence from Public Hospitals." *Review of Economic Studies* 0: 1-133.

Carpenter, Daniel P. 2001. The Forging of Bureaucratic Autonomy: Reputations, Networks, and Policy Innovation in Executive Agencies, 1862-1928. Princeton: Princeton University Press.

Carrera, Leandro, and Patrick Dunleavy. 2010. "Why does productivity vary across NHS hospital trusts in England? Untangling how management competence and the use of ICTs shape hospitals' performance" LSE Public Policy Group Mimeo, March.

Collier, David, Fernando Daniel Hidalgo, and Andra Olivia Maciuceanu. 2006. "Essentially Contested Concepts: Debates and Applications." *Journal of Political Ideologies* 11(3): 211-246.

Cyert, Richard, and James March. 1963. *Behavioral Theory of the Firm*. Englewood Cliffs: Prentice-Hall.

de Figueiredo, Rui. 2002. "Electoral Competition, Political Uncertainty, and Policy Insulation." *American Political Science Review* 96(2): 321-333.

Deaton, Angus. 2010. "Instruments, Randomization, and Learning About Development." *Journal of Economic Literature* 48: 424–455.

Dunleavy, Patrick. 1991. *Democracy, Bureaucracy, and Public Choice: Economic Explanations in Political Science.* London: Prentice-Hall.

Eghan, Ben. 2008. "Enhancing the Performance of the Public Service in a Developmental State: Ghana Case Study." African Association for Public Administration and Management Annual Roundtable, 6-10 October, Accra.

Epstein, David, and Sharyn O'Halloran. 1999. *Delegating Powers: A Transaction Cost Politics Approach to Policy Making Under Separate Powers*. Cambridge: Cambridge University Press.

Evans, Peter, and James Rauch. 1999. "Bureaucracy and Growth: A Cross-National Analysis of the Effects of 'Weberian' State Structures on Economic Growth." *American Sociological Review 64 (5)*: 748-765.

Fukuyama, Francis. 2013. "What Is Governance?" *Governance* 26(3): 347–68.

Gallie, W.B. 1956. “Essentially Contested Concepts.” *Proceedings of the Aristotelian Society* 56: 167-98.

Gallup, John Luke, Jeffrey Sachs, and Andrew Mellinger. 1999. “Geography and Economic Development.” *International Regional Science Review* 22(2): 179-232.

Gibbons, Robert, and Rebecca Henderson. 2013. “What Do Managers Do?” In *The Handbook of Organizational Economics*, ed. Robert Gibbons and John Roberts. Princeton: Princeton University Press.

Gingerich, Daniel. 2013. “Governance Indicators and the Level of Analysis Problem: Empirical Findings from South America.” *British Journal of Political Science* 43 (3): 505-540.

Grindle, Merilee. 1997. “Divergent cultures? When public organizations perform well in developing countries.” *World Development* 25: 481-95.

Grindle, Merilee. 2011. “Governance Reform: The New Analytics of Next Steps.” *Governance* 24(3): 415-418.

Hacker, Jacob, Paul Pierson, and Kathleen Thelen. 2015. “Drift and Conversion: Hidden Faces of Institutional Change.” In Mahoney, James, and Kathleen Thelen (eds), *Advances in Comparative-Historical Analysis*, Cambridge: Cambridge University Press, pp. 180-210.

Hausmann, Ricardo, Bailey Klinger, and Rodrigo Wagner. 2008. “Doing Growth Diagnostics in Practice: A ‘Mindbook’.” CID Working Paper No. 177, September.

Hausmann, Ricardo, Dani Rodrik, and Andrés Velasco. 2005. “Growth Diagnostics.” Mimeo, March.

Huber, John, and Nolan McCarty. 2004. “Bureaucratic Capacity, Delegation, and Political Reform.” *American Political Science Review* 98(3): 481-494.

Huber, John, and Charles Shipan. 2002. *Deliberate Discretion? Institutional Foundations of Bureaucratic Autonomy*. New York: Cambridge University Press.

Ingraham, Patricia W., Philip G. Joyce, and Amy Kneedler Donahue (2003). *Government Performance: Why Management Matters*. London: The Johns Hopkins University Press.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2010. "The Worldwide Governance Indicators Methodology and Analytical Issues." World Bank Policy Research Working Paper 5430. www.govindicators.org

Leibenstein, Harvey. 1966. "Allocative Efficiency vs "X-Efficiency." *American Economic Review* 56: 392-415.

Leonard, David. 2010. "'Pockets' of Effective Agencies in Weak Governance States: Where Are They Likely and Why Does It Matter?" *Public Administration and Development* (30): 91-101.

Lynn, Laurence, Carolyn Heinrich, and Carolyn Hill. 2000. "Studying Governance and Public Management: Challenges and Prospects." *Journal of Public Administration Research and Theory* 10(2): 233-261.

Macey, Jonathan. 1992. "Organizational Design and Political Control of Administrative Agencies." *Journal of Law, Economics, and Organization* 8(1): 93-110.

Mahoney, James, and Kathleen Thelen (eds). 2010. *Explaining Institutional Change: Ambiguity, Agency, and Power*. Cambridge: Cambridge University Press.

Martinez, Elizabeth, Nancy Beaulieu, Robert Gibbons, Peter Provonost, and Thomas Wang. 2015. "Organizational Culture and Performance." *American Economic Review Papers and Proceedings* 105(5): 331-335.

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

McCubbins, Matthew 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.

Meier, Kenneth, Laurence O’Toole, and Sean Nicholson-Crotty. 2004. “Multilevel Governance and Organizational Performance: Investigating the Political-Bureaucratic Labyrinth.” *Journal of Policy Analysis and Management* 23(1): 31-47.

Melitz, Marc. 2003. “The Impact of Trade on Intra-industry Reallocations and Aggregate Productivity Growth.” *Econometrica* 71(6): 1695–1725.

Ménard, Claude. 2014. “Embedding Organizational Arrangements: Towards a General Model.” *Journal of Institutional Economics* 10(4): 567-589.

Mintzberg, Henry. 1983. *Structure in Fives*. Englewood Cliffs, NJ: Prentice Hall.

Moe, Terry. 1990. “The Politics of Structural Choice: Towards a Theory of Public Bureaucracy.” In Williamson, Oliver (ed), *Organization Theory: From Chester Barnard to the Present and Beyond*, Oxford: Oxford University Press, 116-153.

North, Douglass C. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.

North, Douglass, John Joseph Wallis, and Barry Weingast. 2009. *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History*. Cambridge: Cambridge University Press.

Olken, Benjamin. 2007. “Monitoring Corruption: Evidence from a Field Experiment in Indonesia.” *Journal of Political Economy* 115(2): 200-249.

Olken, Benjamin, and Rohini Pande. 2012. "Corruption in Developing Countries." *Annual Review of Economics* 2012(4): 479-509.

Owusu, Francis. 2006. "Differences in the Performance of Public Organisations in Ghana: Implications for Public-Sector Reform Policy." *Development Policy Review* 24(6): 693-705.

Prendergast, Canice. 2003. "The Limits of Bureaucratic Efficiency." *Journal of Political Economy* 111(5): 929-958.

Putnam, Robert. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton: Princeton University Press.

Rasul, Imran, and Daniel Rogger. 2014. "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service." University College, London, June.

Reinikka, Ritva, and Jakob Svensson. 2004. "Local Capture: Evidence from a Central Government Transfer Program in Uganda." *The Quarterly Journal of Economics* 119(2): 679-705.

Rodrik, Dani. 2009. "The New Development Economics: We Shall Experiment, but How Shall We Learn?" in Cohen, Jessica, and William Easterly (eds), *What Works in Development? Thinking Big and Thinking Small*, Washington DC: Brookings Institution, 24-47.

Roll, Michael (ed). 2013. *The Politics of Public Sector Performance: Pockets of Effectiveness in Developing Countries*. London: Routledge.

Rotberg, Robert. 2014. "Good Governance Means Performance and Results." *Governance* 27 (3): 511-518.

Rothstein, Bo, and Jan Teorell. 2008. "What Is Quality of Government? A Theory of Impartial Government Institutions." *Governance* 21 (2): 165-190.

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature* 49: 326–365.

Talbot, Colin. 2010. *Theories of performance : organizational and service improvement in the public domain,* Oxford, Oxford University Press.

Tendler, Judith. 1997. *Good government in the tropics*. Baltimore: Johns Hopkins University Press.

Teorell, Jan, Carl Dahlström, and Stefan Dahlberg. 2011. "The QoG Expert Survey Dataset." University of Gothenburg: The Quality of Government Institute. http://www.qog.pol.gu.se

Wereko, T.B. 2009. *Public Sector Reforms in Ghana: In Search of Effective Service Delivery*. Institute for Democratic Governance and The African Capacity Building Foundation: Ghana Speaks Lecture/Seminar Series No. 1.

Wilson, James. 1989. *Bureaucracy*. New York: Basic Books.

World Bank. 2002. *World Development Report 2002: Building Institutions for Markets*. Oxford: Oxford University Press for the World Bank.

Yanez-Pagans, Monica, and Gustavo Machicado-Salas. 2014. "Bureaucratic Delay, Local-Level Monitoring, and Delivery of Small Infrastructure Projects: Evidence from a Field Experiment in Bolivia." *World Development* 59: 394-407.

Zamboni, Yves, and Stephan Litschig. 2013. "Audit Risk and Rent Extraction: Evidence from a Randomized Evaluation in Brazil." Mimeo, June.

Chapter Two

FROM INSTITUTIONS TO ORGANIZATIONS:

MANAGEMENT AND INFORMALITY IN GHANA'S PUBLIC BUREAUCRACIES

Chapter Abstract

Studies of state capacity in political science and economics have largely sought to explain variation in bureaucratic quality over space or time through the lens of national-level factors, notably the quality of institutions. However, these institutional approaches are unable to explain variation among organizations within the same government. I develop a complementary approach grounded in organizational economics, in particular relational contracts theory, that is consistent with organizational-level variation in bureaucratic quality within governments and generates predictions about the changing qualitative relationship between formal and informal practices across this spectrum. I find empirical support using original interview-based data on management quality from 40 organizations in Ghana's central government. The range of variation in management quality across organizations is substantial and systematic, is not limited to a handful of exceptional "islands" of excellence, and is qualitatively more consistent with theories of relational contracts and organizational culture than prevailing theories of formal and informal institutions. I suggest a number of ways in which institutional and organizational approaches can usefully complement each other in the study of government bureaucracies and their reform.

From Institutions to Organizations:
Management and Informality in Ghana's Public Bureaucracies

A large and influential literature has sought to explain variation in societies' economic and political performance through the framework of institutions – the formal and informal "rules of the game" that define the action sets within which actors operate (North 1990). Since state capacity plays an important role in economic and political development, this institutional lens has also come to dominate studies of state capacity and the quality of government bureaucracies, especially in developing countries. For example, the World Bank's widely used World Governance Indicators list "government effectiveness" as one dimension of governance quality alongside other abstract elements of institutional quality such as "rule of law" and "protection of property rights".[5] Across a wide range of theoretical and methodological perspectives, from rational choice to historical institutionalism to the literature on good governance, the language of institutions has become the default vocabulary for the study of state bureaucracies and their reform.

Yet for all its theoretical and empirical successes, the institutions framework has run up against important limitations in understanding the functioning of government bureaucracies. First, the institutions literature has focused heavily on explaining variation in performance at the national or societal level across space or time, and thus cannot explain within-government variation among ministries and agencies that face identical institutional, historical, geographic, and socio-cultural contexts. While the implicit assumption of the institutions literature is that such organization-level variation is non-existent (or at least is second-order to society-level variation), an increasing body of case studies (Grindle 1997, Tendler 1997, Leonard 2010, Roll 2013) and large-N studies (Gingerich 2013, Rasul and Rogger 2014) shows that within-government variation cannot be assumed away even as a first approximation. This empirical evidence also presents a deep theoretical challenge to the institutions literature: if we observe a wide range of outcomes across organizations operating

[5] This is defined as "perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies." (Kaufmann, Kraay, and Mastruzzi 2010, p.4)

within the same institutional context, how much do these institutions actually determine or constrain performance? Understanding this cross-sectional variation can therefore shed new light on the difficult question of how institutional equilibria change over time that has challenged both rational choice and historical scholars of institutions (North 1990; Greif and Laitin 2004; Mahoney and Thelen 2010; Hacker, Pierson, and Thelen 2015).

Second, while institutional perspectives recognize that both formal and informal aspects of institutions are important and potentially interactive, they offer limited analytical tools for understanding the interactions of formal and informal factors in actual bureaucratic practice (North 1990, Ostrom 1990, La Porta et al 1999, Helmke and Levitsky 2004, Stein and Tommasi 2008, Andrews 2013). This is because the analytical focus on informal *institutions* is simultaneously very broad and very restrictive, and because it overlooks the role that organizational hierarchies play in consciously and continuously trying to shape the informal norms and practices of their members. Informal institutions is a broad category in that it potentially encompasses everything from subconscious socio-cultural cognitive frames to highly regularized and actively enforced norms, and factors that are general to an entire society as well as those that are propagated within specific institutional sub-fields or organizations. Yet it excludes or elides a wide range of informal processes, actions, and behavioral regularities that are extremely important for the functioning of bureaucracies, but are not institutionalized. Furthermore, while much of the literature on informal institutions examines how they can emerge organically from strategic interaction among actors, informal norms and practices within organizations are the product not just of organic interactions but also the deliberate interventions of leadership and managers to cultivate useful norms (i.e. “organizational culture”) and govern day-to-day practices that require the exercise of discretion by its members. Even formal structures and processes, such as incentive schemes and performance monitoring, may be consciously designed to affect informal norms and practices. While institutionalist approaches thus correctly emphasize that informality is an important feature of government bureaucracies, an approach tailored to the specificities of how informality functions within organizations could provide richer insights.

The central argument of this paper is that, for many purposes, there are theoretical and empirical advantages to studying government bureaucracies from an *organizational* rather than an *institutional* perspective. Organizations are the literal bricks-and-mortar of governments; political choices over laws, policies, and resource allocations are implemented not by an abstract notion of "government effectiveness" but by particular ministries, departments, and agencies through complex, painstaking, and sometimes ineffective bureaucratic processes. Though organizations have long been a key unit of analysis for scholars of public administration and management, particularly in wealthy democracies (Wilson 1989; Ingraham, Joyce, and Donahue 2003; Talbot 2010), studies of the state as institution and state as organizations have increasingly diverged. This seemingly semantic distinction has significant consequences for research, because institutional and organizational approaches are associated with different research questions, different bodies of theory, and different sets of empirical methodologies. This divergence is even more evident in the assessing the impact of the New Institutional Economics (NIE) on studies of state capacity: while the Northian "institutions" branch of the NIE has thoroughly transformed the way political scientists and economists think about the development and functioning of government bureaucracies, the influence of the Williamsonian "organizations" branch of the NIE has been largely restricted to particular niches of the public administration sub-discipline.

More specifically, I argue that the theoretical and methodological tools of organizational economics and organization theory provide insights into the two shortcomings of institutional approaches identified above: 1) explaining within-government variation in organizational performance, and 2) analyzing the interaction of formal and informal processes in these bureaucracies. A large body of empirical research shows large and persistent differences in productivity and management quality among organizations even within the same narrowly defined field (Gibbons and Henderson 2013; Bloom and Van Reenen 2007; Syverson 2011; Dunleavy and Carrera 2013; Bloom, Propper, Seiler, and Van Reenen 2015). This literature also explores the micro-foundations of variation in performance among *ex ante* identical organizations, outlining theories based on: the necessarily incomplete nature of formal contracts, since appropriate actions cannot always be codified *ex ante* or verified *ex post*; and the repeated nature of principal-agent interactions within organizations.

Together, these create the potential for multiple equilibria in organizational performance, and make informal methods of management and control necessary complements to formal ones.

These theories of organizational performance as determined by *relational contracts* provide an integrated view of formal processes and informal norms and practices within organizations.[6] In this view, the organization functions as a *nexus of institutions*, both formal and informal. It is also an important agent of *endogenous institutional change*, in that it acts calculatively and purposively to shape the incentives, norms, and behavior of its members. The informal factors on which it seeks to act include both general, organization-wide norms (e.g. organizational culture), as well as the task-specific exercise of discretion by agents. This organizational approach not only provides a more appropriate basis for the analysis of the day-to-day functioning of organizations but also sheds light on the workings of institutional mechanisms by revealing the ways in which they are – and are not – constraining on the actors that operate within them.

I empirically illustrate the contributions of an organizational approach by drawing on original interview-based data on management practices from 40 organizations in Ghana's central government, using an adaptation of Bloom and Van Reenen's (2007) influential methodology for benchmarking management quality. These organizations are all located near each other in the capital city, sometimes even sharing the same building, and are thus subject to identical sets of institutional, historical, socio-cultural, and national-level political factors; in many cases formal uniformity of structures and even specific practices is dictated by a common set of legal and

[6] In contrast, historical institutionalism makes an explicit distinction between institutional enforcement (Levitsky and Murillo 2009) and the analysis of informal institutions (Helmke and Levitsky 2004). In showing how formal structures can co-exist with and reinforce informal practices, relational contracts bears some similarity to Greif and Laitin's (2004) theory of endogenous institutional change that rests on the concept of quasi-parameters that are exogenous in the short run but endogenous in the long run. But whereas Greif and Laitin's quasi-parameters typically occur as an unintended byproduct of formal institutional structures, the development of positive relational contracts in organizations – organizational culture – is typically the result of deliberate efforts of managers and is directly and often consciously linked to the design of formal structures, processes, and incentives within the organization.

administrative regulations. Yet the range of variation in management quality across organizations is substantial and systematic, and is not limited to a handful of exceptional “islands” of excellence. Across this spectrum, a qualitative pattern emerges: in the worst-managed organizations, formal management processes are either undefined or unexecuted, and substitutive informal norms are absent or rare; somewhat better-managed organizations execute most formal processes but without the support of complementary informal norms and practices, thus following the letter of the process but not its spirit – perfunctory performance, in the language of relational contracts (Williamson 1975, Hart and Moore 2008); and the best-managed organizations not only define and execute formal processes, but also cultivate and encourage complementary informal processes to ensure that management processes are executed according to their spirit – consummate performance). Both of these empirical results are consistent with the predictions of an organizational approach based in relational contract theory, but are not well explained by prevailing institutionalist approaches in their rational choice, historical, or governance and development varieties.

The remainder of this paper proceeds as follows. Section 2.1 discusses the relationship between organizational and institutional theories of government bureaucracy, and draws implications for empirical analysis. Section 2.2 details the methodology, sample, and context for the analysis of management practices in Ghanaian government organizations. Section 2.3 shows that the changing qualitative relationship of formal and informal processes across the spectrum of management quality is consistent with relational contract theories of management, and Section 2.4 demonstrates that aggregating these practice-level differences reveals the existence of substantial and systematic variation in management quality at the organization-level. Section 2.5 concludes by discussing the relationship of an organizational perspective on government bureaucracies to approaches rooted in institutions and political economy, and considers other implications for research on states, institutions, and development.

2.1 INSTITUTIONS, ORGANIZATIONS, AND RELATIONAL CONTRACTS

This section highlights two shortcomings of institutionalist approaches to studying government bureaucracy: their inability to explain organization-level variation within governments, and their weak predictions about the interaction of formal and informal factors. It argues that alternative theoretical approaches from organization economics, in particular relational contracts theory, can address both of these shortcomings. Finally, the section sets out the quantitative and qualitative empirical implications of this theoretical framework, and discusses how the empirical distribution and patterns of bureaucratic performance can provide guidance on whether (and when) it is more appropriate to analyze government bureaucracy as *an institution* or as *a set of organizations*.[7]

Institutions and organizational performance

The explosion of literature on institutions, governance, and development in the past two decades has been fundamentally shaped by North's distinction between institutions as the formal and informal "rules of the game" and organizations as the "players" of this game (1990, 1-2).[8] Like North, this literature has largely focused – empirically and theoretically – on institutions, to the neglect of organizations. In large part, this neglect stems from North's assumption that organizations operate efficiently within the constraints of a society's institutions. North is explicit about this: after referring to Williamson's (1975, 1985) transaction cost theories of organization, he writes "Whatever the merits of these alternatives…the focus in this study is on

---

[7] The purpose of this discussion is not to hypothesize a specific model or determinant of bureaucratic performance; indeed, organizational approaches to bureaucracy and relational contracts theory are both broad enough to encompass numerous potential models and determinants. Nonetheless, the distribution of a variable is informative about the process through which it is generated (Jones et al 2009), and it is this broader question of the applicability of institutional and organizational frameworks with which this paper is concerned.

[8] For example, Acemoglu, Johnson, and Robinson's (2005, p.388) influential review of the literature on institutions and development cites North's definitions. North's definition of institutions is a broad one, encompassing everything from formal political institutions to diffuse factors such as culture, social capital, and trust. While these are all important theoretical constructs in their own right, in this paper I follow North's distinction and group them together for brevity in order to more clearly juxtapose institutional and organizational theories of bureaucracy.

organizations as purposive entities designed by their creators to maximize wealth, income, or other objectives defined by the opportunities afforded by the institutional structure of the society." (1990, 73) With this assumption, a theory of institutions as *constraints* on organizations becomes a theory of institutions as *determinants* of organizational performance – that is, organizations are assumed to be operating at their notional production frontier.[9] Since institutions are defined as society-wide, they provide no theoretical explanation for organization-level variation in performance within governments. While institutionalists would likely not deny that any organization-level variation is possible within institutions, the assumption is nevertheless that such variation is either negligible or second-order.

Likewise, the empirical and theoretical literature has overwhelmingly focused on explaining the causes and consequences of variation in institutional quality across countries or other geographic units (Tilly 1990, Evans and Rauch 1999; Acemoglu, Johnson, and Robinson 2001; Gennaioli and Rainer 2007; Acemoglu, Garcia-Jimeno, and Robinson 2012; Acemoglu, Ticchi, and Vindigni 2011; Nunn and Wantchekon 2011). While more recent models of state capacity have sought to distinguish between different aspects of state capacity, such as fiscal and legal capacity (Besley and Persson 2009), even these more sophisticated models do not provide a general explanation for variation among organizations in the same government. Even the qualitative, case study-based literature that seeks to identify "islands" or "pockets" of effectiveness in developing country governments (Grindle 1997, Tendler 1997, Leonard 2010, Roll 2013) inadvertently reinforces this homogenizing assumption by emphasizing that such variation is "exceptional".[10]

[9] This is a consequence of how North defines the scope of his work: "the emphasis in this study is on the institutions that are the underlying rules of the game and the focus on organizations (and their entrepreneurs) is primarily on their role as agents of institutional change; therefore the emphasis is on the interaction between institutions and organizations." (1990, 3) While North's view of organizations as agents of institutional change does recognize that causality may also run from organizations to institutions in the long term, institutions are nevertheless determinative of organizational performance at any given point in time.

[10] For instance: "It is well established that even in countries that have poor governance and weak public sectors, exceptional well-functioning government and government-supported agencies do exist." (Leonard 2010, 91)

What larger-N research does exist at the organizational level, however, suggests that variation within government may be the norm rather than the exception. In their study of 63 government organizations in Nigeria, Rasul and Rogger (2014) document extensive variation both in management quality and output delivery. Gingerich (2013) draws on a survey of civil servants in 30 organizations across three countries in Latin America to argue that within-country variation across organizations is actually greater than cross-country variation, for a variety of survey items related to politicization, capacity, and corruption – this calls into question the institutionalist assumption that organization-level variation is second-order compared to institution-level variation. Studies of government performance in OECD countries have long viewed organizations as a fundamental unit of analysis (Wilson 1989, Ingraham *et al* 2003, Talbot 2010), and have documented substantial differences in performance among them.[11]

Outside of the public sector, the potential for organizations to operate inefficiently has long been a key theme in the study of organizations. Early theorists questioned economic models of firms as perfectly rational maximizers with concepts such as organizational slack (Cyert and March 1963), X-inefficiency (Leibenstein 1966), and organizational culture (Schein 1985).[12] A growing body of empirical research shows large and persistent differences in productivity and management quality among organizations even within the same narrowly defined field (Gibbons and Henderson 2013), with evidence most heavily concentrated in the manufacturing sector (Bloom and Van Reenen 2007, Syverson 2011) but extending also to other fields such as hospitals (Bloom et al 2015, Carrera and Dunleavy 2010) and schools (Bloom et al 2014). These "persistent performance differences" among *ex ante* identical organizations appear to be the norm, not the exception, within organizational fields (Gibbons and Henderson 2013). Both theoretically and empirically, the institutionalist treatment of organizations as purposive rational maximizers within institutional

[11] Some of this literature also discusses inherent differences across organizations in the nature of management, for example due to different types of tasks or budget structures (Mintzberg 1983, Wilson 1989, Dunleavy 1991). While these can be important, for brevity my discussion focuses on theoretical and empirical literature on true inefficiency in the sense of organizations operating within their notional production frontier, given external constraints.

[12] Other perspectives view cultural dispositions as semi-inherent to organizations (e.g. Douglas 1992).

constraints is at odds with the vast scholarship on management and organizational performance.

Relational contracts and management practices

Relational contract theories of dispersion in organizational performance among *ex ante* identical organizations rest on the idea that contracts are necessarily incomplete. [13] This is particularly true of the employment contract and the management practices that specify the actions of managers and employees within this contract. Contracts may be incomplete in two senses: future states of the world and the associated actions for each party to take may be impossible to fully anticipate and specify *ex ante* (Simon 1951, Kreps 1996); and/or some value-creating aspects of parties' actions may be unverifiable by third parties *ex post*, and thus unenforceable (Hart and Moore 2008). This incompleteness implies that it is generally efficient for both parties to retain some level of discretion – indeed, as a practical matter it is almost impossible to completely eliminate discretion. [14] In a repeated game framework, this creates short-term incentives for both employees and managers to play Defect strategies (by shirking on effort or reneging on promises of rewards, for example), even if the cooperative equilibrium is both socially and privately efficient in the long term (Kreps 1996).

Management in organizations is therefore both technical – in the specification of tasks, contingencies, and the design of incentives – and relational, in that it requires building shared expectations, understandings, and norms over time.[15] This creates the potential for substantial long-term divergences in performance among organizations, along with path dependence arising from minor stochastic shocks (Chassang 2010).

[13] I focus on relational contracts that are within rather than between organizations. My discussion of relational contracts here is highly stylized and oriented towards building intuition about why performance might differ among organizations embedded in identical institutional, historical, socio-cultural, political, and task contexts. See Gibbons and Henderson (2012) for a concise formal exposition of these models.

[14] Agent discretion has long been an important element of theories of management and public administration (Mintzberg 1983, Wilson 1989), and much of the recent work on relational management practices in organizational economics formalizes these arguments.

[15] Nelson and Winter's (1982) work on routines in organizations places a similar emphasis on the accretion of shared understandings and processes over time.

The relational aspects of management are likely to be even more important in public sector organizations than private sector ones: outputs of public sector organizations are often non-priced and/or difficult to measure (Wilson 1989, Prendergast 2003)[16]; and public sector organizations often have multiple principals, thus weakening the potential effectiveness of incentive schemes (Dixit 1996).[17]

The idea that shared norms can be used to sustain cooperation without the intervention of formal enforcement mechanisms is, of course, a familiar one for scholars of institutions (Ostrom 1990, Greif 1993, Putnam 1993, Nunn 2008, Nunn and Wantchekon 2011), and studies such as La Porta et al (1997) have demonstrated that variations in generalized trust in large organizations across countries are positively correlated with levels of economic and institutional development. The contribution of an organizational perspective is to emphasize that socio-cultural norms, expectations, and cognitive frames not only vary across organizations within a given society from their societal "mean", but also that they are explicitly *organizational* phenomena in that organizations invest a great deal of effort and resources into shaping and maintaining them. Indeed, hierarchy in organizations exists explicitly in order to facilitate cooperation among and coordination amongst members of the organization, and the cultivation of shared understandings and beliefs amongst individuals constitutes much of what managers actually do in organizations (Gibbons and Henderson 2013). The determinants and consequences of "organizational culture" are central questions for scholars of management (Schein 1985) and increasingly for empirical organizational economics (Martinez et al 2015, Gibbons and Kaplan 2015).

[16] As Prendergast (2003) notes, public sector tasks are in the public sector precisely because they cannot be efficiently delivered through market mechanisms, and thus observed (non-commercial) public bureaucracies are necessarily inefficient.

[17] Dixit's (1996) multiple principal model is one of many non-relational contract theoretical models that can be used to explain why the political principals of public organizations may not be able to realize their preferred outcomes. But whereas inefficiency arises in the multiple principal model arises due to competing incentives being offered to the agent by the multiple principals, the *lack* of strong incentives is a more common scenario in developing country bureaucracies.

In this sense, the organization functions as a *nexus of institutions*.[18] While they are certainly influenced by more general societal institutions, whether formal (e.g. contract enforcement) or informal (e.g. trust), organizations design and promulgate among their members a mutually reinforcing bundle of formal rules and informal norms. Whereas changing a formal institution in society typically requires negotiation and coordination among multiple actors, organizations centralize the design and enforcement of many formal rules in a calculative hierarchy that exists largely for this purpose. Likewise, informal institutions in society arise either organically through long-term repeated interactions or (occasionally) through deliberate coordination amongst elites. In contrast, organizational hierarchies attempt to directly shape informal norms amongst their members – albeit with great difficulty and uncertain success – and link this effort to the design of formal structures and processes, such as incentive schemes. Organizations are therefore far from being passive "price-takers" of institutions, at least with respect to their internal functioning.

In addition to general, organization-wide cultures and norms, a key function of management is to develop shared informal norms about the execution of specific tasks that require agents to exercise discretion. Since the optimal exercise of discretion cannot always be specified *ex ante* or verified *ex post*, the performance of any tasks involving discretion cannot be fully governed by formal management tools like organizational manuals, standard operating procedures, or administrative regulations. Examples of tasks that require agents' discretion include those where the principal cannot perfectly observe either the state of the world or the agent's action, or those that involve an important non-measurable quality dimension. In these cases, the utility to the principal of an agent performing only the formal (fully specifiable and verifiable) aspects of a task may be substantially less than if the agent performs both the formal and the informal aspects of the task. Hart and Moore (2008) model these as perfunctory and consummate performance – actions consistent with the letter of the contract versus those consistent with the spirit of the contract, respectively.[19]

---

[18] This is analogous to Alchian and Demsetz's (1972) theory of the firm as a nexus of property rights.

[19] Hart and Moore (2008) assume that the agent is indifferent between perfunctory and consummate performance; more generally, deadweight loss is created (and multiple equilibria are possible) if and only if the agent's cost of performing the informal aspect of the task is less than the utility it creates for

Likewise, discretion for agents also implies discretion for principals (managers), in deciding whether the task was performed perfunctorily or consummately and thus in how the reward or punish the agent. As Gibbons and Henderson (2012, 2013) argue, building these relational understandings often requires employees and managers not just to solve a problem of *credibility* of mutual commitments over the use of discretion (sustaining cooperative equilibria) but also a simultaneous problem of *clarity* (building shared understanding of what actions constitute cooperation and defection in which states of the world). Informality is thus an important feature of bureaucracies even when formal contracts and management practices can be perfectly enforced.

This view of the relationship between formal processes and informal norms and practices in organizations improves over institutionalist approaches in several ways. For example, Helmke and Levitsky (2004) categorize informal institutions as potentially complementary, substitutive, accommodating, or competing according to their convergence with formal institutions and the strength of formal institutions. While useful, this approach takes as given the design and strength of both formal and informal institutions, whereas a relational contracts approach emphasizes that within organizations these are jointly and consciously shaped. Likewise, by defining informal institutions as those that are shared "*created, communicated, and enforced outside of officially sanctioned channels*" (Helmke and Levitsky 2004, p. 727; emphasis in original), it explicitly excludes that much of what managers in bureaucratic hierarchies do is create, communicate, and enforce management practices that involve both formal and informal behaviors.[20] Finally, this literature limits its focus purely to *institutionalized* informal behaviors rather than those driven simply by incentives or other "behavioral regularities" (Helmke and Levitsky 2004, p. 727), whereas management in organizations is typically concerned with all of these. Corruption, for example, can be thoroughly institutionalized or a simple result of agency problems and imperfect monitoring. Andrew's (2013) discussion of

the principal. The language of perfunctory and consummate performance is originally from Williamson (1975, p. 69).

[20] To take an extreme example, Gibbons and Henderson (2013) refer to the employee handbook of the retailer Nordstrom which, for many years, consisted solely of the sentence: "Use good judgment in all situations."

institutional reform comes closer in recognizing that public sector reform requires changes not just to formal structures but also to informal socio-cultural norms and cognition. But whereas this approach regards informal factors as "content" to be "supplied" by reforms alongside formal content (Andrews 2013, p. 66), relational contract theory emphasizes that, by its very nature, informal "content" cannot be specified *ex ante* in the same way as formal reforms can be. This qualitative understanding of the relationship between formality and informality in organizations is thus crucial to theories of bureaucratic performance and reform.

Implications for empirical analysis

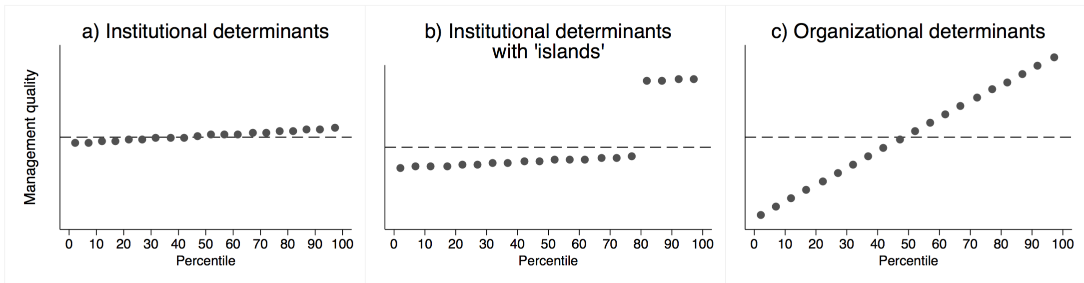
This paper's aim is to explain why we observe variation in performance among organizations within a common institutional setting and to explore the mechanisms that generate this variation, not to hypothesize a specific set of determinants of organizational performance. Indeed, relational contracts theory implies that asking "what are the exogenous determinants of organizational performance?" is in some ways the wrong question, since organizations that share the same context and external features may perform very differently for purely internal, endogenous reasons. Instead, this paper's empirical analysis seeks to examine the extent to which quantitative and qualitative patterns of variation in management quality among organizations are consistent with and well explained by organizational (as opposed to institutionalist) theoretical approaches. These empirical predictions are not about the determinants of organizational performance, but rather about its *distribution* – the extent and distribution of dispersion in performance among organizations within a government – and its qualitative *mechanisms* – *how* organizations are well or poorly managed, rather than *why*. The distribution and mechanisms of management in turn shed light onto the processes through which it is generated, and thus can inform hypotheses about how these processes might be influenced.

The distribution of performance among organizations operating in identical institutional contexts is an indirect indicator of how constraining institutions actually are on organizations.[21] Institutionalist perspectives on state bureaucracies explicitly or

[21] The definitions of government performance and good governance are heavily contested (Rothstein and Teorell 2008, Talbot 2010, Fukuyama 2013, Rotberg 2014). While I focus on one specific measure

implicitly assume that organizations are purposive, rational actors that maximize some objective function within overall institutional constraints (North 1990), and so there is little reason to expect substantial variation in outcomes among them (other than perhaps some random noise). This corresponds to the stylized distribution illustrated in Figure 1(a), where deviations from the overall mean are uniformly low. If this distribution pertains in reality, then the assumption that institutions tightly constrain and thus determine bureaucratic quality is a sound one, and an organizational perspective would contribute little. The literature on "islands" of effectiveness in developing country governments takes this distribution and theoretical framework as a baseline, then argues that "exceptional" well functioning organizations can exist (Grindle 1997, Tendler 1997, Leonard 2010, Roll 2013). But because this literature relies exclusively on small-N case studies, it cannot identify whether such organizations are indeed exceptional – as illustrated in Figure 1(b) – or whether they are simply the upper end of a continuous distribution that includes not just very strong and very weak organizations, but also a full spectrum in between.

Figure 1: Three stylized distributions of management quality



Note: Dotted line indicates mean of simulated distribution.

This is the perspective implicitly adopted by medium- or large-N organization-focused empirical work (Ingraham *et al* 2003, Talbot 2010, Gingerich 2013, Rasul and Rogger 2014). Figure 1(c) illustrates this as a uniform distribution with large deviations from the mean. For simplicity a uniform distribution is shown to illustrate

of performance – management quality – the broader point about within-government variation is generally applicable.

the framework's distributional prediction, but it is equally compatible with other alternatives (such as a normal distribution) that are both continuous – to distinguish them from 1(b) – and have a substantial degree of variation – to distinguish them from 1(a). If organizational factors are important and systemic determinants of performance above and beyond institutional factors, the overall distribution of performance across organizations will resemble the distribution in Figure 1(c). The following section discusses methodological details of how this paper defines and measures organizational performance, and compares its distribution in Ghana's government to these hypothetical distributions.[22]

On the question of mechanisms, the relational contracts view of management predicts that the way in which formal and informal factors interact will change qualitatively when moving along the spectrum of overall management quality. That is, although formal and informal factors always interact, the difference between good and bad organizations is not simply "better" interactions, but qualitatively different ones. Table 1 illustrates three qualitative typologies of management. Perfunctory performance of a task involves carrying out the formal aspects of the task that are fully codifiable and verifiable, but not the associated informal aspects that also create value. Consummate performance, however, involves carrying out both the formal *and* the informal aspects of the task. In better-performing organizations, then, qualitative examination of day-to-day tasks would show managers and employees collaborating not just to design and enforce formal rules, but also to cultivate the complementary informal practices that enhance the efficiency of the formal practice. In less well-managed organizations, however, agents might follow formal rules, but only to the

[22] The definition of "substantial" in this context is, of course, contestable, since organizational performance has no natural scale. One option is to compare it to cross-country variation, as Gingerich (2013) does in arguing that within-country variation in governance quality across organizations actually exceeds variation across-countries, using self-reported survey data. However such comparisons raise methodological issues about cultural and contextual differences affecting item interpretation, as well as other possible sources of survey bias. This paper's approach is to use an objectively defined methodology for benchmarking management that covers a range of variation that is "substantial" quantitatively in that it has been widely used in other contexts and shown to be predictive of significant variations in performance, and qualitatively in that the benchmarks cover a range of variation in practices comparable to what public sector reform efforts seek to realize. Section 2.2 discusses measurement in more detail.

extent they are codifiable and verifiable. Finally, in the worst organizations, agents may not even execute these formal processes, and managers may not attempt to or be able to enforce them. In this state of non-performance, positive informal practices (now substitutive rather than complementary of formal practices) only exist on an *ad hoc* basis – due to exceptional individuals, for example – rather than being systematically fostered and institutionalized, and thus should be empirically rare. Note that this framework is agnostic about the *design* of these formal and informal practices; rather they are shaped on an ongoing basis through a relational process, and the assumption is that regardless of whether a given management practice is theoretically optimal, executing it is better than not executing it.[23]

Table 1: Formality, informality, and performance in organizations

	Non-performance	Perfunctory performance	Consummate performance
Formal processes defined and executed	No	Yes	Yes
Supporting informal norms and practices cultivated	No	No	Yes

Qualitatively, then, the empirical analysis examines the extent to which this spectrum of performance – from non-performance to perfunctory performance to consummate performance – is an accurate characterization of day-to-day bureaucratic functioning and management in government organizations. Unlike the distributional predictions, these qualitative predictions of relational contract theory are not necessarily in tension with institutionalist approaches to formal and informal institutions; rather, they are considerably more specific. If they are accurate, then, their specificity provides a stronger – or at least complementary – analytical basis for understanding the functioning of government bureaucracy than institutionalist approaches. They also

[23] This may not be true in cases where the organization's political principles heavily incentivize outcomes but also impose rigid and poorly designed formal rules that hamper the achievement of these outcomes; Helmke and Levitsky (2004) cite the Soviet *blat* system as an example of an "accommodating" informal institution that improved efficiency by working around a formal institution. These seem less likely to be important in the contexts of developing country bureaucracies, where outcomes are typically not strongly incentivized and formal rules are often weakly enforced.

predict that, to the extent they result from relational processes in organizations rather than generalized socio-cultural institutions, informal norms and expectations will vary substantially across organizations, and that better-managed organizations will try to actively foster positive informal norms through official channels.

I address these quantitative and qualitative predictions in reverse order: Section 2.3 examines qualitative variation at the level of two specific management practices, and Section 2.4 then aggregates these practice-level variations up to the organizational level in order to draw a picture of the overall distribution of variation in management quality in Ghana's central government. Before this, however, the next section discusses details of measurement and methodology.

2.2 MEASURING MANAGEMENT

To examine the extent to which performance varies across organizations within the same government, I conduct semi-structured interviews on management practices with senior managers in 40 organizations in Ghana's central government. Broadly defined as the cumulative operation of formal structures and processes and informal norms and practices within an organization, management is a close proxy for organizational performance: better-managed organizations perform better than worse managed ones, all else equal. To measure management quality, I adapt the methodology of Bloom and Van Reenen (2007; BVR henceforth) from organizational economics, which uses semi-structured interviews with senior managers to benchmark the effectiveness of a variety of management practices in the organization.[24] This method produces an indicator of performance that is more comparable across organizations than the outcome-based measures used by most cross-national studies, and is also largely attributable to and within the control of the organization itself.[25]

[24] This method has been shown to be highly correlated with productivity in organizations as diverse as manufacturing firms, schools, hospitals, and non-governmental organizations, in developed as well as developing countries (see Bloom et al 2014 for a survey).

[25] In contrast, organizational outputs depend on the supply of resource inputs and outcomes may depend on the interaction of these outputs with societal factors or the actions of other government organizations. While certain aspects of management are not entirely within an organization's control (for example due to labor laws or budgeting processes), in Ghana these largely apply government-wide

Likewise, it is also more comparable across organizations than qualitative small-N methods, making it possible to incorporate a large enough sample to be representative of the overall population of organizations in a government.[26]

In contrast to most survey-based approaches to studying management, the BVR method uses interviews with managers to investigate the *actual* state of practice in their organization, rather than asking what the management practices are *supposed to be* on paper or what interviewees' *perceptions* of management quality are. This focus is particularly important in the context of governments in developing countries, since it is widely recognized that *de facto* practices frequently diverge from *de jure* ones in this context (Andrews 2013) and much of the comparative governance literature has so far relied on perception-based measures of experts or the subjects themselves (Kaufmann et al 2010; Teorell et al 2011, Gingerich 2013), both of which may be subject to various biases (Schwarz 1999, Bertrand and Mullainathan 2001, Olken 2009). To get at *de facto* practice, the conduct of the interviews included a number of measures to elicit honest answers and avoid social desirability bias, such as starting with open questions (e.g. "How do individual staff know what tasks and targets they are responsible for?") rather than the types of closed questions ("Do individual staff have tasks and targets?") that are common to most surveys; following up with more specific probing questions ("So how do officers know what to do when they get to the office on Monday morning?") and requests for examples; and asking whether "on paper" management practices were actually carried out and meaningful in reality ("Do these targets actually influence what individuals do on a day-to-day basis, or are they

---

and thus cannot explain variation. In fact, examining the implementation of these common *de jure* processes provide a useful test to alleviate concerns that variation is being driven by organizations choosing to implement management practices that are different but not necessarily better or worse.

[26] The BVR approach has also been adapted to African public sectors by Rasul and Rogger (2014), who use it with government organizations in Nigeria. Although there are many similarities between Rasul and Rogger and this article, the two projects were independently conceived and implemented. I first became aware of the Rasul and Rogger study in May 2013, mid-way through fieldwork for this article. However, the details of this article's methodology differ in important ways from both BVR and Rasul and Rogger's adapation of it – these distinctions are discussed below.

just on paper?"). Finally, all interviews were conducted anonymously at both the individual and organization level, so respondents had no incentive to misreport.[27]

The interviews focused on 15 aspects of management across four sub-groups: operations management, target setting, performance monitoring, and human resources management. Following BVR, these practices are evaluated against a pre-defined scoring grid on a scale of 1-5, with scores defined according to objective benchmarks rather than subjective relative perceptions (such as Likert scales).[28] Appendix 2.1 provides full details of each practice, sample questions associated with it, criteria for benchmarking, and examples. The benchmarking criteria are designed to be neutral with respect to the *style* of management and instead focus on whether for each practice the organization has a process or routine that is consciously designed and followed in reality, and which it continuously seeks to improve. The framework is thus largely agnostic with respect to the many prominent debates in the academic literature on public management that focus more on defining optimal practices (distinguishing between the 90th and 99th percentiles) than on explaining organizations' failures to adopt anything resembling an optimal practice (distinguishing between the 10th and 90th percentiles). To aggregate the scores, each practice was converted into a normalized z-score and then averaged across the organization, with missing practices excluded, and the resulting average was then re-converted to a z-score for use in analysis.[29]

---

[27] It is of course still possible that respondents could lie about actual management practices, but the combination of open-ended questions and probing follow-ups means that to do so, a respondent must both: 1) want to lie despite having no incentive to do so; and 2) be aware of what a better management practice would be.

[28] While this approach is more effective at eliciting information about sensitive and/or informal practices than typical closed-ended survey questions, it also requires the interviewer to use more discretion in posing and coding questions. While this could engender concerns about possible interviewer bias, this method is at least more transparent about how discretion and judgment are used by the interviewer than the strictly qualitative case study approaches that have thus far been used to study informal practices and institutions.

[29] To check whether reported practices depended heavily on the individual being interviewed, I conducted two interviews with individuals from different divisions in each of the 20 ministries. The correlation of overall z-scores between interviewees in the same organization is very strong (0.84) and highly statistically significant. Figure A1 in Appendix 2.2 shows this relationship graphically. The

Table 2: Organizations in sample

Ministries (20)	Agencies and Departments (20)
Administration sector	
• Ministry of Chieftaincy and Traditional Affairs • Ministry of Finance and Economic Planning • Ministry of Information • Ministry of Justice and Attorney-General's Department • Ministry of Local Government and Rural Development	• Department of Community Development • Ghana Revenue Authority • Information Services Department • Registrar-General's Department
Economic sector	
• Ministry of Environment, Science, Technology, and Innovation • Ministry of Food and Agriculture • Ministry of Lands and Natural Resources • Ministry of Tourism, Culture, and Creative Arts • Ministry of Trade and Industry	• Environmental Protection Agency • Fisheries Commission • Ghana Irrigation Development Authority • Ghana Standards Authority • Ghana Tourism Authority • Minerals Commission
Infrastructure sector	
• Ministry of Communications • Ministry of Energy and Petroleum • Ministry of Roads and Highways • Ministry of Transport • Ministry of Water Resources, Works, and Housing	• Department of Rural Housing • Ghana Highway Authority • Ghana Investment Fund for Electronic Communications • Ghana Shippers' Authority • National Petroleum Authority
Social sector	
• Ministry of Education • Ministry of Employment and Labour Relations • Ministry of Gender, Children and Social Protection • Ministry of Health • Ministry of Youth and Sports	• Department of Children • Department of Cooperatives • Ghana Library Authority • National Health Insurance Authority • National Youth Authority

Note: Organization names are reported as at mid-2013, when fieldwork for this article was conducted. Some organizations have since been renamed, merged, or split. Sectoral classifications based on Government of Ghana budget classification, with the exception of the Ministry of Energy, which is coded as Infrastructure rather than Economic sector due to its expenditure profile.

difference is due to some combination of measurement error and genuine differences in management practices between divisions in the same organization. This correlation is higher than that found in re-interviews by BVR. This may be explained in part by methodological differences – BVR's resurveys were of a manager in the same organization but a different plant, whereas I focused on two individuals within the same building. The BVR resurveys were also conducted by different interviewers, which likely increased the variance of scores. For the purpose of analysis, in the organizations where two interviews were completed, I first averaged each practice score between the two interviews before converting them to z-scores and aggregating. Although averaging the two interviews mechanically reduces the variance of scores among ministries somewhat, alternative aggregation schemes, such as using only the first interview in each organization, do not change the conclusions of the analysis.

I conducted these interviews in person with management-level staff in 40 organizations in Ghana's central government, comprising all twenty of Ghana's ministries (excluding security sectors such as the Ministry of Defense) and one randomly selected agency of each ministry, thus covering a representative sample of organizations across nearly the whole of government.[30] Respondents in specific positions within each organization were targeted for interviews, in order to minimize the potential for selection bias, and although timing and availability necessarily played some role in selection, there is no evidence that responses or management scores differed systematically by interview or interviewee characteristics (see Appendix 2.2 for more details).

In addition to these numerical measures, I also treat the interviews as rich sources of qualitative data. This allows me not only to capture the nuanced interactions of formal and informal factors in day-to-day management practices, but also to make comparisons across a much larger sample than the small-N methods usually associated with research into informal institutions would allow. The qualitative data underlying these codings continues to serve to validate these codings and provide an important check for consistency in the coding process, which helps to alleviate the concerns about interviewer bias often associated with survey research as well as qualitative fieldwork methods.[31]

---

[30] The three security sector ministries – Defense, Interior, and Foreign Affairs and Regional Cooperation – were excluded from the study. At the time of the study the Ministry of Fisheries and Aquaculture Development was newly created and not yet operational, but was being created out of the extant Fisheries Commission, which was included in the sample. To select agencies, I first excluded all organizations that are: primarily commercial (state-owned enterprises); very small (below 30 staff); the two largest agencies (Ghana Education Service and Ghana Health Service), which are each an order of magnitude larger in terms of staff size than any other organization in Ghana's government; and an assorted handful of organizations whose specific functions or structure made them non-comparable to other organizations. A random selection of one eligible agency per ministry was then made. The Ministry of Chieftaincy and Traditional Affairs had no eligible agencies.

[31] This qualitative data was referred to throughout the interview process in order to ensure that codings were consistent. All individual practice scores were finalized *before* aggregation and analysis, however, in order to minimize the potential for data mining and other forms of bias. Drawing on qualitative data

2.3 FORMALITY AND INFORMALITY IN GOVERNMENT ORGANIZATIONS

This section examines qualitative variation within two of the fifteen management practices covered by the interviews: defining the tasks and responsibilities of employees, and assessing and rewarding performance. There is a wide range of variation in these across organizations. However, this variation is characterized not by a gaping chasm between a handful of exceptional organizations and a mass of poor ones, but by an accumulation of marginal yet collectively significant differences. Qualitatively, the observed qualitative relationship between formal processes and informal norms and practices corresponds to that hypothesized by relational contracts theory, ranging from non-performance to perfunctory performance to consummate performance of the practice.

The definition of individual tasks and responsibilities and the assessment of individual performance are both critical aspects of management in any organization. Both are also subject to some degree of standardization across public service organizations in Ghana. For the assessment of individual performance this level of intended uniformity is very high, as the process is standardized even down to the specific form used for every government employee's annual appraisal.[32] This makes it a particularly strong test of this paper's theoretical arguments: the formal "rules" (and any society-wide informal norms) are constant across organizations, so variation in their execution must be due to organization-level factors; if this variation comprises not just the execution of formal processes but also the fostering of complementary informal norms and practices, then these organization-level factors have a strongly relational element.

---

to ensure the consistency of coding is a partial compensation for the inability to statistically control for interviewer bias.

[32] Ghana's governmental structure makes a distinction between the Civil Service and the Public Service, each of which have their own staff appraisal system. The staff of ministries and departments, which comprise the Civil Service, therefore undergo a slightly different appraisal process than staff of the agencies, who are part of the Public Service but not the Civil Service. For the purposes of this study, however, the differences between these two appraisal schemes are negligible, and each scheme is formally identical across organizations within each service.

Defining individual tasks and responsibilities

The interview item Defining Tasks and Responsibilities deals whether individuals know what they are responsible for and do it, as opposed to waiting around to be told what to do. In other words, when individuals come to work on Monday morning, how do they know what they should be doing? In the best organizations, individual objectives and tasks are clearly defined through a combination of formal target setting, periodic review meetings, to-do lists, and informal check-ins. This means that staff can handle routine tasks with a high degree of autonomy, and understand how these connect to the work of their colleagues and the division and organization as a whole. In these cases autonomy complements rather than replaces hierarchical supervision, as individuals still consult regularly with their directors in one-on-one and team meetings and are directly assigned tasks that arise in addition to their defined duties.

In organizations on the other end of the spectrum, in contrast, direct hierarchical supervision is the sole method for assigning tasks, with the result that officers spend much of the time idle and supervisors feel overloaded because they are responsible for managing not only their own daily workload, but also that of all their subordinates. Service-wide regulations mandate that individual tasks and targets be defined annually, as part of the appraisal process, but in poorly managed organizations this is either not carried out, or is carried out in a formalistic way using vague or recycled objectives that are forgotten as soon as they are written and do not actually guide individuals' work during the year.

As the histogram in Table 3 shows, both of these extremes are represented, but the majority of organizations find themselves somewhere in between. Often this means partial execution of these processes: targets are defined for individuals, but not regularly; job descriptions provide some guidance about general responsibilities; and some officers can carry out certain routine tasks autonomously, but periods of idleness while waiting for direct orders from a superior are still common.

Table 3: Defining individual tasks and responsibilities

Examines how well the organization allocates tasks and defines individual responsibilities

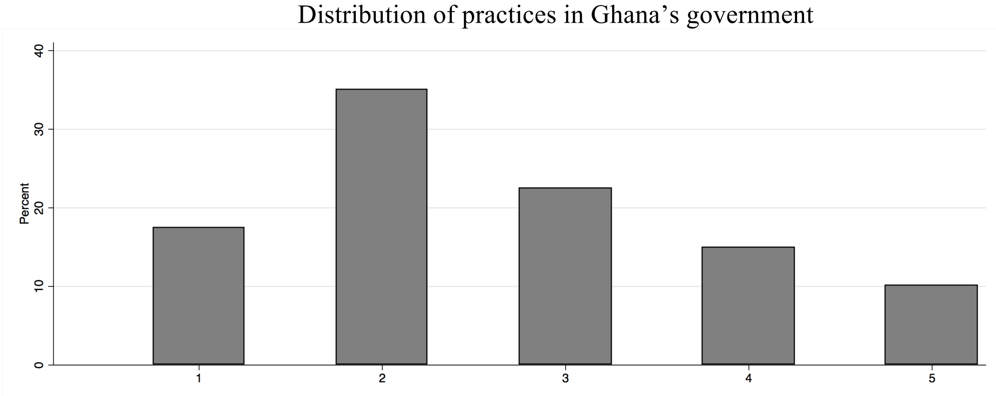
Sample questions

• How are objectives set for individual staff members? • Could everyone in the organization say what they are responsible for and how that will be assessed?	• When a staff member comes to work on Monday morning, how do they know what to do? • How do you know which tasks are best suited to different staff?

Benchmarking criteria

Score 1	***Score 3***	***Score 5***
Objectives are undefined or very vague. Staff are expected simply to "do what they're told". Tasks are allocated to individuals on an *ad hoc* basis, with little or no consideration of individual competences or workloads. Staff are often unsure what they should be doing.	Objectives may be formally defined but are vague or inconsistently applied. Staff have a general sense what they are responsible for, but completion of most tasks relies mainly on the giving/ following of direct orders. There is some consideration of which tasks should be allocated to certain individuals, but this is partial, inconsistent, or based only on an officer's formal position.	Individuals understand clearly what they are responsible for and what they should be doing at all times. Objectives are meaningful and are defined through a two-way process. Tasks are allocated based on identified competences.

Distribution of practices in Ghana's government



Selected examples

Below average (score 1)	***Average (score 2.5)***	***Above average (score 4.25)***
• Officers in a social-sector ministry have targets defined through the appraisal process, but in practice they wait for work to be minuted to them by a superior. • Targets are seen as a formality that don't actually influence work. "People set targets, but the fact is that in most situations in our parts of the world these things are only done when there is something critical that occurs, like the officer wants a promotion." • Staff are often idle because they don't have work. "It's not like you come to work every time and there is something special that you do." • One consequence of this is that management have to micro-manage and are often overworked.	• In one economic-sector ministry, directors apportion their division's objectives to officers at the start of the year. This is supposed to be annual but is not always done. • Job descriptions provide some guidance about individual responsibilities, but not enough that officers are always aware of what specific tasks they should be doing. • "It's not everyone who knows what they're responsible for and can get on with it without waiting for director's minutes...[but] some of us who have routine work know."	• In an infrastructure ministry, each officer sits with their boss at the start of every year and reaches mutual agreement on a set of specific tasks and targets (e.g. submitting a report within 48 hours of a workshop). • Individual tasks are derived directly from the division's targets and areas of responsibility. • "People are pretty clear on what they need to be doing." • Each division meets regularly to discuss the status of each officer's work, the connections between their roles, and the organization's targets.

Assessing and rewarding individual performance

A similar range of practices exists with respect to assessing and rewarding individual performance. Personnel decisions and payscales are subject to a high degree of central control and regulation, and on paper, the process is completely standardized across organizations.[33] Yet in actual practice, there is a significant amount of positive and negative deviance from the intended standard.

On the negative side, in many organizations individual performance is so unimportant that the annual appraisal process is not even carried out annually, as intended; rather, individuals often complete several years' worth of appraisals retrospectively as they near the date for promotion interviews. In these organizations the appraisal itself is essentially a self-assessment that the supervisor merely signs off on, except in cases of extreme misconduct, and there is no link between formal or informal assessments of performance and promotion or other rewards.

On the positive side, some organizations not only carry out the formal appraisal process as designed, but also institute complementary informal measures to make the formal process more meaningful. Supervisors hold the standard annual and mid-year performance review meetings with each of their staff, but also hold informal discussions on performance in weekly divisional meetings so that individuals are not only kept aware of their own performance, but also how they compare to others in the division. Because there are strong centrally-imposed constraints on pay and promotions, organizations find other ways to recognize and reward good performance: awards nights for top performers as voted by their colleagues, occasional written letters of recognition for exemplary work, opportunities for training, and so on. While accelerated formal promotions are rare, it is common for these organizations to reward star performers with "informal promotions": appointment to committees, roles as a focal person, and other opportunities that further career development and may also entail some financial benefit. In between the

[33] The Office of the Head of Civil Service and Public Services Commission are heavily involved in almost all human resources decisions pertaining to pay, promotion, hiring, and firing, so for all intents and purposes these are formally identical within the Civil Service and Public Service. In general, however, the Public Service organizations have a slightly higher degree of autonomy in some of these respects.

Table 4: Assessing and rewarding performance

Examines whether individual performance is assessed transparently and rewarded proportionately

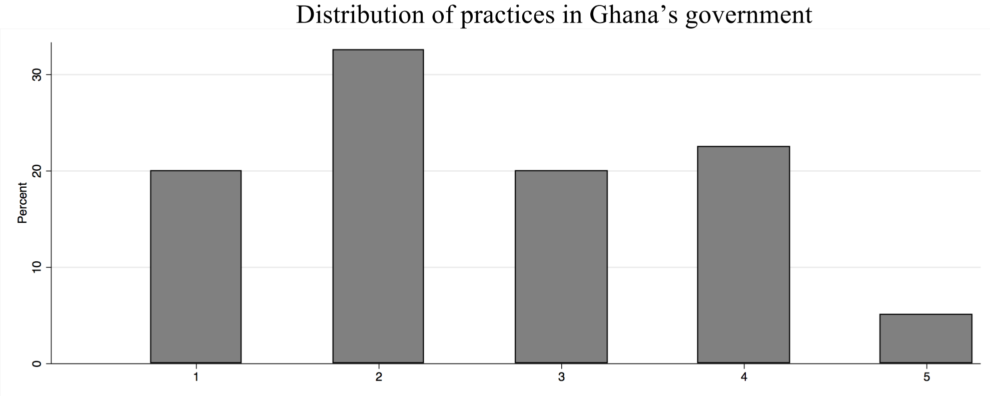
Sample questions

- How does your appraisal system work? Are the appraisals actually completed regularly?
- Are there any financial or non-financial (e.g. awards, training, travel opportunities) rewards related to this review?
- Are better performers likely to be promoted faster or are promotions given solely on the basis of tenure / seniority?

Benchmarking criteria

Score 1	***Score 3***	***Score 5***
People are rewarded in the same way irrespective of their performance. Promotions are solely based on tenure.	There is an evaluation system which in principle awards good performance (financially or non-financially), but awards are not based on clear criteria/ processes. Performance can sometimes influence career progression.	There is an evaluation system which rewards individuals (financially or non-financially) based on performance. Rewards are given as a consequence of well-defined and monitored individual achievements. Performance is a key criterion for promotion decisions.

Distribution of practices in Ghana's government



Selected examples

Below average (score 1)	***Average (score 2)***	***Above average (score 3.5)***
• In an economic-sector ministry, staff appraisals are only completed once every three years, when it is time for the officer to be considered for promotion. • The appraisal is a self-assessment; actual discussions about officers' performance rarely happen. • "The appraisals don't improve the system. They're a formality." • There is no link between performance and promotion. Recommendations for promotion are based solely on seniority. • In principle performance can be incentivized with training opportunities, but the allocation of these is viewed as subjective. "One director even says he'll send you 'if your face looks nice'."	• Appraisals in one social-sector ministry are always done annually, but there are no rewards and the process is viewed mainly as a formality. • Better performing officers are sometimes more likely to be recommended as a focal person for an activity and receive additional responsibilities, but this is viewed as subjective and dependent on one's director and the chief director. • In principle non-performing officers can be not recommended for promotion, but in practice this does not happen and recommendation for promotion is based entirely on years served. • The ministry gives officers a bag of rice at Christmas as a bonus, but this is given to all staff regardless of performance.	• In one infrastructure ministry, appraisals are always done annually, and individuals' performance is also discussed weekly in divisional meetings, so officers are also aware of their peers' performance. • Each year the ministry holds an awards night. Within each division staff vote for a top performer, and then these individuals vote for an overall top performer. • Accelerated promotions are rare, but non-performing individuals tend to be recommended more slowly, so better performers rise through the ranks faster. • The HR division makes an effort to recognize officers for exceptional work, for example by writing letters of commendation for outstanding performance on a particular task.

extremes of good and bad practice, similar types of informal recognition or rewards sometimes exist but are often less transparent in their selection criteria and linked only indirectly, or not at all, to performance assessments – bags of rice or small bonuses for all staff at Christmas, for example, or highly subjective allocation of training opportunities by supervisors.[34]

Qualitative variation and relational contracts

For both of these practices, the variation among organizations is clearly qualitatively significant, in the sense that improving the practice would require dramatically different actions in each organization. Some organizations fail to execute even simple required procedures, while others not only execute them consistently but also take initiative to go above and beyond these minimum requirements and find ways to make them more than mere formalities. In between these extremes, organizations' management practices vary in ways that are more marginal, yet still non-trivial. The distribution of practice across the other 13 management practices is similar, albeit with some variation across practices (see Appendix 1 for qualitative examples, and Appendix 2 for the quantitative distributions).

The qualitative nature of this variation is also consistent with relational contracts theory, as outlined in Table 1, but is not well explained by theories of informal institutions. While parts of this variation can be characterized using institutional theory – the enforcement of formal institutions (Levitky and Murillo 2009), or the complementary nature of informal norms in well-managed organizations (Helmke and Levitsky 2004) – other qualitatively important aspects of management are not. For example, the same complementary informal norms in well-managed organizations do not exist as substitutive informal norms in poorly managed organizations, as Helmke and Levitsky's (2004) classification of the relationship between formal and informal institutions might predict. Relational contracts theory is clear about why this is,

[34] One of the key results of Rasul and Rogger (2014) is that incentive-linked management practices are negatively related to project completion in Nigeria's civil service; they explain this as a distortion of effort and cooperation as a result of incentives that are stronger than optimal. However, I found no evidence of any incentive schemes in Ghana's government that could potentially undermine organizational performance through distortion of effort. Whether this is a difference between methodologies or between public management in Ghana and Nigeria is unclear.

whereas an institutional framework has little explanation for why institutions *do not* exist in a particular setting. Likewise, actual management practice clearly is shaped by and seeks to shape numerous behavioral incentives, regularities, and actions that fall outside the realm of “institutions”. These can only be understood by viewing informal norms and practices in management as arising due to the need to guide and coordinate the use of discretion by managers and agents.

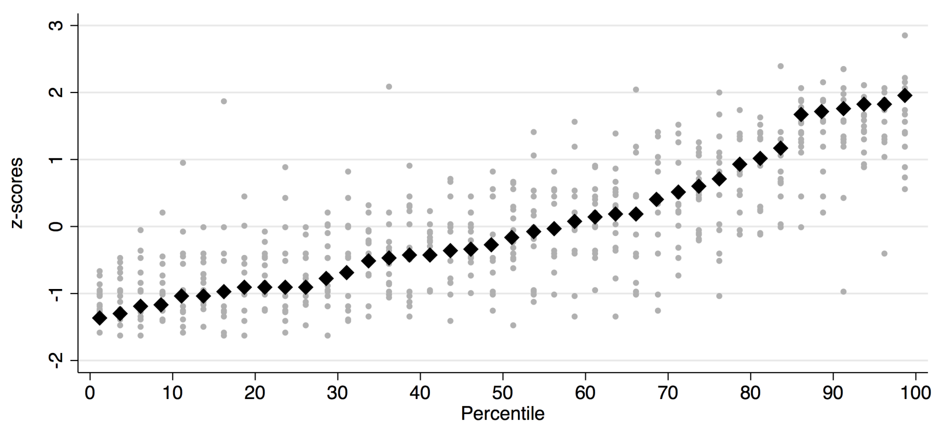
2.4 THE DISTRIBUTION OF MANAGEMENT QUALITY ACROSS ORGANIZATIONS

The previous section demonstrated that there is considerable micro-level variation in management practices among organizations within Ghana’s government. This section aggregates these micro-level differences to the level of the organization, and examines the resulting distribution against the three stylized distributions discussed in Section 2.2. If practice-level variation is just random noise – idiosyncratic decisions by individual managers, perhaps, or simple stochastic error accumulation over time – it will average out across organizations and the overall distribution of management quality will be relatively flat. If the quality of management practices is correlated within organizations, however, then these micro-level differences will aggregate into significant differences across organizations. This corresponds to the organization-focused view of performance and management, which views management as both a technical and relational problem that has to be deliberately and systematically shaped over time by organizations.

Variation across organizations

Figure 2 plots the normalized practice z-scores along with the overall organization score (the renormalized mean of the individual practice z-scores). The distribution is slightly right skewed – the mean is below the median – but there is continuous incremental variation in overall scores across the entire spectrum. There is a strong within-organization correlation of practice scores – the correlation coefficients between each practice score and the overall organization score all fall between 0.61 and 0.88, and are all highly statistically significant. At the extremes, the within-organization distribution of the best organizations dominates that of the worst organizations. That is, the most lagging management practice in the top few organizations is still better than the best practice of the bottom organizations. Despite this, the variation within organizations is still meaningful. An improvement of a quartile, from the 10^{th} percentile to the 35^{th}, would require an overall improvement of just 0.65 standard deviations, equivalent to an average increase of just 0.58 marks in each practice. If each organization simply improved all its practices to the level of the 75^{th}-percentile practice *in that organization*, this alone would increase the overall organization score by 0.60 standard deviations.

Figure 2: Variation in management quality across Ghana's government



Note: Small dots represent practice z-scores normalized within each practice, large dots represent means of practice z-scores re-normalized at organization-level.

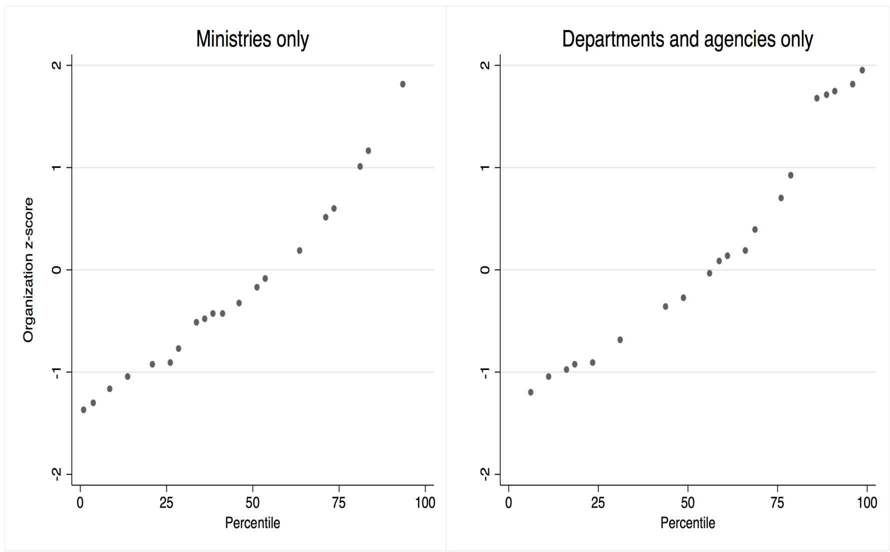
This distribution is inconsistent with the view that national- or societal-level institutions tightly constrain or determine organizational performance. Although management quality has no natural scale, the numerical codings of individual management practices are based on qualitative descriptions benchmarked to objective scales (as opposed to subjective, perception-based, Likert-type scale); this makes it possible to link the overall organizational management score to the variation at the practice-level that is qualitatively meaningful, and thus show that the differences between the worst, the average, and the best organizations are indeed meaningful ones. Accumulated across the numerous aspects of management, these minor differences add up to organizations that function in fundamentally different ways. This is consistent with the findings of the case study literature on "islands" of excellence in developing country governments, which also draws on primarily qualitative evidence. However, the empirical distribution in Figure 2 is not consistent with this literature's view of variation in management as a feature solely of "exceptional" organizations. Rather, the distribution is smooth and continuous, consisting of minor and marginal differences that collectively amount to significant heterogeneity. The largest discontinuity in this distribution is a minor gap at the 85th percentile, but this is equivalent to just under half a standard deviation and is far from systematic enough to be considered evidence of a bimodal pattern of "islands" of excellence.

Variation among similar organizations

While this is strong evidence that management quality varies substantially at the organizational level within Ghana's government, it does not show whether this variation arises from factors internal or external to the organization. For example, organizations might have different task characteristics that are inherently more or less difficult to carry out, and thus better management could be a reflection of these contextual differences rather than (or in addition to) an independent causal factor in its own right. If, hypothetically, variation was due entirely to external factors, then organizational performance would still be tightly constrained by contextual variables, as institutionalist perspectives presume; however, these constraints would not be national-level or purely institutional in nature, but specific to particular organizations. Investigating the extent to which management quality is driven by internal and external factors is beyond the scope of this paper – indeed, it is an entire research agenda in its own right (see for example Gibbons and Henderson 2013, Martinez *et al* 2015). In principle, this could be tested by conditioning on a full set of observable and unobservable characteristics of organizations, but this is obviously unfeasible for measurement and the potential number of variables (and their interactions) would be vastly larger than the universe of government organizations in any country.

Instead, this paper's more humble claims are that: a great deal of variation seems to be driven by organizational factors (whether internal or external) rather than institutional ones; this variation is qualitatively consistent with relational management processes playing an important role; and since relational processes imply that management quality can vary substantially among *ex ante* identical organizations (i.e. those with identical external contexts), external factors (whether institutional or organizational) are unlikely to fully determine management quality. While conditioning on the full set of observable and unobservable characteristics of organizations is impossible, it is nonetheless possible to examine as a simple robustness check two ways of restricting the analysis to organizations that are more likely to be comparable to each other: ministries only, and organizations within the same sector.

Figure 3: Range of variation by organization type

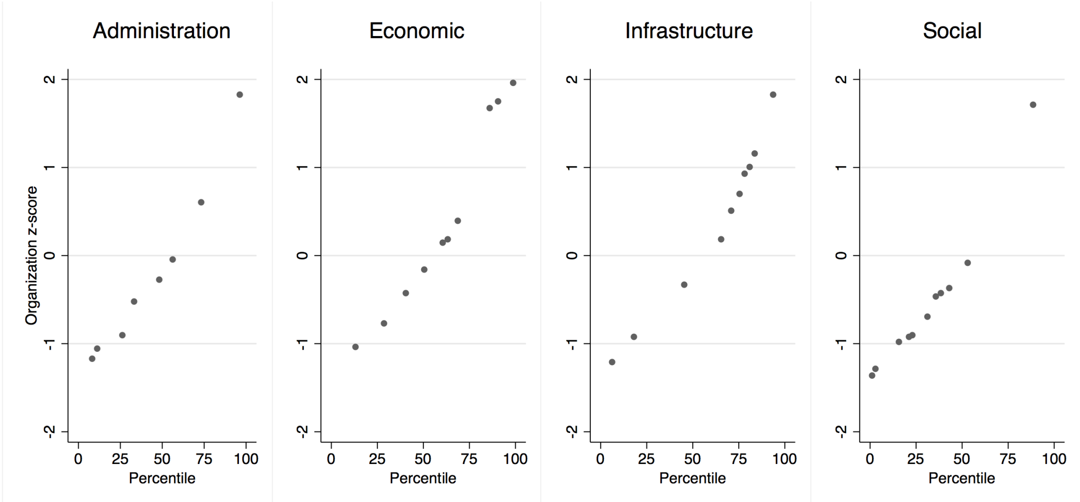


Note: Organization-level z-scores shown. See Table 2 for list of ministries, departments, and agencies.

Management within the subset of 20 ministries is even more comparable than among the sample as a whole, because Ghana's structure of government gives all ministries identical functions – policymaking, and oversight of subsidiary departments and agencies – while all implementation activities are undertaken by departments and agencies. Thus the core tasks of the agencies and departments in the sample vary somewhat, but those of ministries do not. Figure 3 demonstrates that the range and distribution of variation among ministries is comparable to that in the full sample.

Next, Figure 4 shows the distribution of organization-level management scores across the four broad sectors of Ghana's government: Administration, Economic, Infrastructure, and Social (Table 2 lists these by organization). The only significant difference across sectors is the lower average performance of social sector organizations, although the range of scores in this sector is comparable to that of other sectors.[35]

Figure 4: Range of variation within sectors



Note: Sectoral classifications based on Government of Ghana budget classification, with Ministry of Energy coded as Infrastructure rather than Economic sector due to expenditure profile.

While these sub-sample comparisons illustrate that substantial organization-level variation exists even after conditioning on some particularly salient observable external characteristics of organizations, they are only suggestive (but not conclusive) evidence that these external characteristics do not fully determine organizational performance. Nonetheless, this study has shown empirical evidence that is both quantitatively and qualitatively consistent with internal management factors – based in relational contracts theory – playing a large role in the day-to-day functioning of organizations. At a minimum, this theory and evidence suggests that, rather than presuming that organizational performance is fully determined by institutional or external organizational-level factors, the new default assumption for studies of bureaucratic quality should be that these factors are not fully constraining. The

[35] This difference is statistically significant at the five percent level, albeit with an extremely small sample size.

appropriate question for research on state capacity is not whether management matters – it is how much it matters, and how it can be changed.

2.5 DISCUSSION AND IMPLICATIONS

This paper has argued that institutionalist approaches to studying state capacity should be complemented by organizational approaches, for both empirical and theoretical reasons. Empirically, taking organizations rather than institutions as the key unit of analysis makes it possible to understand and begin to explain within-country variation in bureaucratic quality. This variation across organizations is empirically important, and may in some cases even exceed cross-country variation (Gingerich 2013). Studying variation among organizations in the same government also sheds light on the operation of institutions, by revealing the ways in which institutions are – and are not – constraining on the actors that operate within them.

Theoretically, an analytical focus on organizations enables scholars to draw on a rich body of theory from organizational economics, organization theory, and management. In particular, relational contracts theory provides both an explanation for overall differences in management quality among organizations, and an account of its micro-foundations in the governance of discretion within organizations. It also offers qualitative predictions about the relationship between formal structures and processes and informal norms and practices. While organizational approaches are not necessarily incompatible with institutionalist frameworks for analyzing the relationship between formal and informal institutions, they nonetheless bring to the fore the ways in which organizations act as a nexus for institutions, by deliberately and calculatively using both formal and informal means to shape the incentives and behavior of individuals. Relational contracts theory also permits the analysis of behaviors and control mechanisms that are not institutionalized but are nonetheless important, whereas these are beyond the scope of institutionalist approaches.

This is not to say that institutionalist approaches to studying government bureaucracy and state capacity are never useful. Rather, the claim is that bureaucratic outcomes are the joint products of both institutional *and* organizational factors. To draw on a statistical metaphor, if institutional factors represent national averages, then organizational factors represent deviations up or down from that average. This requires researchers to choose their framework according to their question, in ways that require going beyond a simple distinction between cross-country versus within-country. Focusing on institutional differences may be appropriate for cross-country studies of mean levels of overall bureaucratic quality, but not for cross-country comparisons of particular sectors or agencies – as Gingerich (2013) points out, national-level averages may be uninformative about particular organizations if organization-level variance within countries is high. On the other hand, an organizational framework may be appropriate for studies of the *relative* performance of different organizations within a country (since they share the same national-level institutions), but a longitudinal study of *changes* in the performance of a particular organization would need to consider changes in both national-level institutional factors as well as changes in organization-specific factors, both external and internal to the organization. While these distinctions may seem basic, they are nonetheless crucial – yet absent from much scholarship on bureaucracy and its reform.

Deeper engagement with the role of organizations and management also raises a difficult set of questions about identifying causality. Empirical studies of the determinants of institutional quality often identify causal effects by focusing on factors that are deeply historical; this solves the endogeneity problem, but also excludes any determinants of institutional quality that present-day actors could conceivably affect. Likewise, empirical studies of the determinants of organizational performance can estimate the effects of external, exogenously defined characteristics of organizations, such as task characteristics or organization-specific political economy factors; while these are important and largely under-researched, they also exclude by design factors that are internal to the organization, such as management. At the level of particular management practices, experimental methods can establish the impact of interventions on both formal and informal aspects of management (Bloom *et al* 2013; Martinez *et al* 2015), yet the external validity of any intervention that targets informal norms and practices is necessarily limited because these practices are as much relational as technical, so the specifiable details of any intervention are incomplete. For both quantitative and qualitative studies, then, there is a major role to play for descriptive research on mechanisms and processes, as well as research that focuses on cleanly identifying causal impacts.

Finally, these same factors that make organizational performance and management difficult topics for scholars also make improving them a non-trivial challenge for their political principals. While institutionalist approaches to state capacity have largely elided this reality, focusing instead on understanding the conditions under which political actors rationally want to invest in state capacity (e.g. Besley and Persson 2009), an organizational approach would bring this "investment" technology to the foreground. Even when political actors have long time horizons and an interest in a stronger state, are they always able to achieve this, and through what mechanisms?[36] Conversely, when and how can state capacity improve if these preconditions are absent, as is perhaps the most prevalent case in developing countries? What is the role of specific organizations in the development of broader societal institutions? Questions such as these provide a rich agenda for a new wave of scholarship on state capacity. Answering them requires a new perspective on the quality of government bureaucracy as not just as an abstract institution, but as the outcome of the day-to-day workings, inefficiencies, and complexities of the organizations that comprise it.

2.6 CONTRIBUTIONS

This paper's primary objective is to consider how theories of the internal dynamics of organizations might be relevant for studies of state capacity. In so doing, it makes theoretical and methodological contributions to several literatures.

From the perspective of organizational economics, the paper's main theoretical contribution is to introduce serious consideration of internal organizational dynamics into the New Institutional Economics literature on the functioning and development of institutions (North 1990, Greif and Laitin 2004), in particular the branch of this literature that focuses on the development of states' administrative capacity (Greif 2007, Besley and Persson 2009). By emphasizing the ways in which the organization functions as a nexus of formal and informal institutions and the role of management

[36] Indeed, a large literature beginning with McCubbins, Noll, and Weingast (1987) treats political control of the bureaucracy as an open question rather than a given. While this literature focuses almost exclusively on advanced democracies, Huber and McCarty (2004) show that low bureaucratic capacity weakens political control of the bureaucracy as well as politicians' incentives to reform.

in shaping the formal incentives, norms, and cognitive frames of the organization's members, the paper suggests the theoretical possibility that organizations might function as agents of endogenous institutional change. A related theoretical contribution is the paper's novel suggestion that, in the context of management within public bureaucracies, the institutionalist literature's theories of formal and informal institutions (Helmke and Levitsky 2004, Stein and Tommasi 2008, Mahoney and Thelen 2010, Andrews 2013) can be usefully linked to and recast in terms of theories of relational contracts (Hart and Moore 2008, Gibbons and Henderson 2013) from the organizational economics literature.

Secondarily, the paper extends the growing literatures on differences in management quality and organizational performance in organizational economics (Bloom and Van Reenen 2007; Syverson 2011; Bloom, Propper, Seiler, and Van Reenen 2015) and public management (Ingraham, Joyce, and Donahue 2003; Talbot 2010; Dunleavy and Carrera 2013) to the empirical setting of developing country states. It also contributes methodologically by introducing a novel interview-based, mixed-method approach to studying management practices that is based on Bloom and Van Reenen's (2007) influential method and is particularly well suited to contexts where informality and inconsistent execution of formal processes are salient aspects of management. Integrating the quantitative and qualitative aspects of data collection and analysis in this way serves as a useful alternative to other mixed-method approaches that combine both quantitative and qualitative methods of analysis but base them on separate data sources (Lieberman 2005).

Viewed from comparative politics, this paper's theoretical and empirical argument about the importance of within-government, cross-organization variation in management quality and performance is not in itself novel. Indeed, this empirical variation is consistent with studies that have emphasized the role of particular government organizations in individual countries' historical development (Fisher and Lundgren 1975, Johnson 1982, Amsden 1989, Wade 1990), mostly focused on Europe or East Asia, as well as the literature on business-state relations (Bräutigam, Rakner, and Taylor 2002; Opoku 2010) which emphasizes the potential role of organization-specific constellations of interests in determining bureaucratic functioning. This paper is distinct from this literature in that it focuses on theoretical mechanisms driven by factors internal to the organization – relational contracts that evolve endogenously – rather than on the role of external contextual factors (especially political economy) in determining organizational outcomes. This paper thus complements and extends this existing comparative politics literature on state capacity. As the conceptual framework in this dissertation's introduction makes clear, organizational performance is jointly determined by factors that are both external *and* internal to the organization, and rich case studies such as Carpenter (2001) also point to the ways in which they can interact over the medium- to long-term. The relative balance and interaction between these two sets of determinants, and developing a fuller theory of their interaction, is an important area for further research.

Finally, the more applied development and governance literatures on "islands" or "pockets" of effectiveness in developing country governments (Grindle 1997, Tendler 1997, Leonard 2010, Roll 2013) also empirically document and seek to explain – through a combination of external and internal factors – the existence of variation in management and performance across organizations. This paper contributes empirically to this literature by demonstrating that this variation is not due solely to "exceptional" organizations (Leonard 2010, 91) but rather to a more continuous range of variation across government. While this point about the distribution of management quality is an empirical one, the empirical distribution of a variable can generate insights about the theoretical mechanism that generates it (as Jones *et al* [2009] point out in a different context). The larger sample of this paper thus complements this literature's overwhelming focus on small-N, case study-based approaches to study organization-level variation in performance in government.

REFERENCES

Acemoglu, Daron, Simon Johnson, and James Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91(5): 1369-1401.

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2005. "Institutions as a Fundamental Cause of Long-Run Growth." In Aghion, Philippe, and Steven N. Durlauf (eds), *Handbook of Economic Growth, Volume 1A*, 385-472.

Acemoglu, Daron, Davide Ticchi, and Andrea Vindigni. 2011. "Emergence and Persistence of Inefficient States." *Journal of the European Economic Association* 92(2): 177-208.

Acemoglu, Daron, Camilo García-Jimeno, and James A. Robinson. 2012. "Finding Eldorado: Slavery and Long-Run Development in Colombia." *Journal of Comparative Economics* 40 (4): 534–64.

Alchian, Armen, and Harold Demsetz. 1972. "Production, Information Costs, and Economic Organization." *American Economic Review* 62: 777-795.

Andrews, Matt. 2013. *The Limits of Institutional Reform in Development: Changing Rules for Realistic Solutions*. Cambridge: Cambridge University Press.

Bertrand, Marianne, and Sendhil Mullainathan. 2001. "Do People Mean What They Say? Implications for Subjective Survey Data," *American Economic Review Papers and Proceedings* 91 (2): 67–72.

Besley, Timothy, and Torsten Persson. 2009. "State Capacity, Conflict, and Development." *Econometrica* 78(1): 1-34.

Bloom, Nicholas, and John Van Reenen. 2007. Measuring and Explaining Management Practices Across Firms and Countries. *Quarterly Journal of Economics,* 122**,** 1351-1408.

Bloom, Nicholas, Carol Propper, Stephan Seiler, and John van Reenen. 2010a. "The Impact of Competition on Management Quality: Evidence from Public Hospitals." Centre for Economic Performance Discussion Paper 983.

Bloom, Nicholas, and John Van Reenen. 2010b. "New Approaches to Surveying Organizations." *American Economic Review Papers and Proceedings* 100 (May): 105-109.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. "Does Management Matter? Evidence from India." *Quarterly Journal of Economics* 128(1): 1-51.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, Daniela Scur, and John Van Reenenn. 2014. "The New Empirical Economics of Management." *NBER Working Paper 20102*, May.

Bloom, Nicholas, Carol Propper, Stephan Seiler, and John Van Reenen. 2015. "The Impact of Competition on Management Quality: Evidence from Public Hospitals." *Review of Economic Studies* 0: 1-133.

Carrera, Leandro, and Patrick Dunleavy. 2010. "Why does productivity vary across NHS hospital trusts in England? Untangling how management competence and the use of ICTs shape hospitals' performance" LSE Public Policy Group Mimeo, March.

Chassang, Sylvain. 2010. "Building Routines: Learning, Cooperation, and the Dynamics of Incomplete Relational Contracts." *American Economic Review* 100(1): 448-465.

Cyert, Richard, and James March. 1963. *Behavioral Theory of the Firm*. Englewood Cliffs: Prentice-Hall.

DiMaggio, Paul, and Walter Powell. 1983. "The Iron Cage Revised: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review* 48(2) :147-160.

Dixit, Avinash. 1996. *The Making of Economic Policy: A Transaction-Cost Politics Perspective*. London: The MIT Press.

Douglas, Mary. (1992). *Risk and Blame: Essays in Cultural Theory.* London: Routledge.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan. 2014. The Value of Discretion in the Enforcement of Regulation: Experimental Evidence and Structural Estimates from Environmental Inspections in India. Mimeo, February.

Dunleavy, Patrick. 1991. *Democracy, Bureaucracy, and Public Choice.* Hemel Hempstead: Harvester Wheatsheaf.

Dunleavy, Patrick, and Leandro Carrera. 2013. *Growing the Productivity of Government Services*. Cheltenham: Edward Elgar.

Evans, Peter, and James Rauch. 1999. "Bureaucracy and Growth: A Cross-National Analysis of the Effects of 'Weberian' State Structures on Economic Growth." *American Sociological Review 64 (5)*: 748-765.

Fukuyama, Francis. 2013. "What Is Governance?" *Governance* 26 (3): 347–368.

Gallup, John Luke, Jeffrey Sachs, and Andrew Mellinger. 1999. "Geography and Economic Development." *International Regional Science Review* 22(2): 179-232.

Gennaioli, Nicola, and Ilia Rainer. 2007. "The Modern Impact of Precolonial Centralization in Africa." *Journal of Economic Growth* 12 (3): 185–234.

Gibbons, Robert, and Rebecca Henderson. 2012. "Relational Contracts and Organizational Capabilities." *Organization Science* 23(5): 1350-1364.

Gibbons, Robert, and Rebecca Henderson. 2013. "What Do Managers Do?" In *The Handbook of Organizational Economics*, ed. Robert Gibbons and John Roberts. Princeton: Princeton University Press.

Gibbons, Robert, and Robert Kaplan. 2015. "Formal Measures in Informal Management: Can a Balanced Scorecard Change a Culture?" *American Economic Review Papers and Proceedings* 105(5): 447-451.

Gingerich, Daniel. 2013. "Governance Indicators and the Level of Analysis Problem: Empirical Findings from South America." *British Journal of Political Science* 43 (3): 505-540.

Greif, Avner. 1993. "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition." *American Economic Review* 83(3): 525-548.

Greif, Avner, and David Laitin. 2004. "A Theory of Endogenous Institutional Change." *American Political Science Review* 98(4): 633-652.

Grindle, Merilee. 1997. "Divergent cultures? When public organizations perform well in developing countries." *World Development* 25: 481-95.

Grindle, Merilee. 2004. "Good Enough Governance: Poverty Reduction and Reform in Developing Countries." *Governance* 17 (4): 525-548.

Hacker, Jacob, Paul Pierson, and Kathleen Thelen. 2015. "Drift and Conversion: Hidden Faces of Institutional Change." In Mahoney, James, and Kathleen Thelen (eds), *Advances in Comparative-Historical Analysis*, Cambridge: Cambridge University Press, pp. 180-210.

Hart, Oliver, and John Moore. 2008. "Contracts as Reference Points." *Quarterly Journal of Economics* 123(1): 1-48.

Hausmann, Ricardo. 2008. "The Other Hand: High Bandwidth Development Policy." CID Working Paper No. 179, September.

Helmke, Gretchen, and Steven Levitsky. 2004. "Informal Institutions and Comparative Politics: A Research Agenda." *Perspectives on Politics* 2(4): 725-740.

Huber, John, and Nolan McCarty. 2004. "Bureaucratic Capacity, Delegation, and Political Reform." *American Political Science Review* 98(3): 481-494.

Ingraham, Patricia W., Philip G. Joyce, and Amy Kneedler Donahue (2003). *Government Performance: Why Management Matters*. London: The Johns Hopkins University Press.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2010. “The Worldwide Governance Indicators Methodology and Analytical Issues.” World Bank Policy Research Working Paper 5430. www.govindicators.org

Kreps, David. 1996. “Markets and Hierarchies and (Mathematical) Economic Theory.” *Industrial and Corporate Change* 5(2): 561-595.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer and Robert Vishny. 1997. “Trust in Large Organizations.” *American Economic Review Papers and Proceedings* 87(2): 333-338.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer and Robert Vishny. 1999. “The Quality of Government.” *Journal of Law, Economics, and Organization* 15(1): 222-279.

Leibenstein, Harvey. 1966. “Allocative Efficiency vs “X-Efficiency.” *American Economic Review* 56: 392-415.

Leonard, David. 2010. “‘Pockets’ of Effective Agencies in Weak Governance States: Where Are They Likely and Why Does It Matter?” *Public Administration and Development* (30): 91-101.

Levitsky, Steven, and Maria Victoria Murrillo. 2009. “Variation in Institutional Strength.” *Annual Review of Political Science* 12: 115-133.

Mahoney, James, and Kathleen Thelen (eds). 2010. *Explaining Institutional Change: Ambiguity, Agency, and Power*. Cambridge: Cambridge University Press.

Martinez, Elizabeth, Nancy Beaulieu, Robert Gibbons, Peter Pronovost, and Thomas Wang. 2015. "Organizationl Culture and Performance." *American Economic Review Papers and Proceedings* 105(5): 331-335.

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

Mintzberg, Henry. 1983. *Structure in Fives*. Englewood Cliffs, NJ: Prentice Hall.

Nelson, Richard, and Sidney Winter. 1982. *An Evolutionary Theory of Economic Change*. London: Belknap Press of Harvard University Press.

North, Douglass C. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.

Nunn, Nathan. 2008. "The Long-Term Effects of Africa's Slave Trade." *Quarterly Journal of Economics* 123(1): 139-176.

Nunn, Nathan and Leonard Wantchekon. 2011. "The Slave Trade and the Origins of Mistrust in Africa." *American Economic Review* 101(7): 3221-3252.

Olken, Benjamin. 2009. "Corruption Perceptions vs. Corruption Reality." *Journal of Public Economics* 93: 950-964.

Ostrom, Eleanor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. New York: Cambridge University Press.

Prendergast, Canice. 2003. "The Limits of Bureaucratic Efficiency." *Journal of Political Economy* 111(5): 929-958.

Putnam, Robert. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton: Princeton University Press.

Rasul, Imran, and Daniel Rogger. 2014. "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service." University College, London, June.

Roll, Michael (ed). 2013. *The Politics of Public Sector Performance: Pockets of Effectiveness in Developing Countries*. London: Routledge.

Rotberg, Robert. 2014. "Good Governance Means Performance and Results." *Governance* 27 (3): 511-518.

Rothstein, Bo, and Jan Teorell. 2008. "What Is Quality of Government? A Theory of Impartial Government Institutions." *Governance* 21 (2): 165-190.

Schein, Edgar H. 1985. *Organizational Culture and Leadership*. San Francisco: Jossey-Bass.

Schwarz, Norbert. 1999. "Self-Reports: How the Questions Shape the Answers." *American Psychologist* 54 (2): 93-105.

Scott, Richard. 2008. *Institutions and Organizations: Ideas and Interests*. 3rd ed. (1st ed. 1995). London: Sage Publications.

Simon, Herbert. (1951), "A Formal Model of the Employment Relationship." *Econometrica* 19: 293-305.

Stein, Ernesto, and Mariano Tommasi (eds). 2008. *Policymaking in Latin America: How Politics Shapes Politics.* Washington, DC: Inter-American Development Bank.

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature* 49: 326–365.

Talbot, Colin. 2010. *Theories of performance : organizational and service improvement in the public domain,* Oxford, Oxford University Press.

Teorell, Jan, Carl Dahlström, and Stefan Dahlberg. 2011. *The QoG Expert Survey Dataset.* University of Gothenburg: The Quality of Government Institute. http://www.qog.pol.gu.se

Tendler, Judith. 1997. *Good government in the tropics*. Baltimore: Johns Hopkins University Press.

Tilly, Charles. 1990. *Coercion, capital, and European states, AD 990-1990*. Cambridge: Blackwell.

Williamson, Oliver. 1975. *Markets and Hierarchies.* New York: Free Press.

Wilson, James. 1989. *Bureaucracy*. New York: Basic Books.

APPENDIX 2.1: INTERVIEW GUIDE AND PRACTICE EXAMPLES

(1) Operations (Routine)

Examines how well procedures for routine work are structured (based on most common routine task identified by respondent)

- Can you briefly talk me through your procedure for carrying out these day-to-day functions?
- How structured/ standardized is the process?
- How often do challenges or difficulties arise with the current way of doing it?
- How have you tried to address these challenges?

	Score 1	**Score 3**	**Score 5**
Key criteria	Procedures for undertaking routine operations are undefined or are rarely followed in practice. The organization relies almost entirely on top-down assignment to get work done, so execution depends on the individuals involved. Main rationale for current practice is "that's the way it's always been done".	Procedures for routine operations are defined and generally followed, but may be inefficient or arbitrary. Where responsibility for work is assigned to individuals, follow-up and checks sometimes occur but are inconsistent. There is some consideration of how best to undertake day-to-day operations but this is mainly on an *ad hoc* basis.	Routine operations procedures have been consciously structured to improve quality and efficiency, and staff consistently implement these procedures.
Examples	A social sector ministry frequently has severe problems compiling its quarterly reports. To solicit inputs they send a circular requesting information, but these often arrive late or not at all. "The office environment doesn't help…my Director is two floors away, the divisions aren't together at all." Little effort is made to address these recurring problems.	An economic sector ministry has a semi-standardized process for undertaking field monitoring visits. Monitoring teams have a standard composition, but the process for initiating and reviewing the visits can vary on an *ad hoc* basis. Disagreements among divisions and "turf" wars in writing the reports are common, but the ministry has not taken any action to resolve these.	To monitor the performance of its decentralized offices, monitoring staff of an infrastructure ministry visit each office for one week, with arrival on Sunday night and an inception meeting on Monday. Days are used to visit sites, while files from the local offices are reviewed at the hotel in the evening. On Friday an exit meeting is used to review the findings with the local office staff and discuss disagreements. The report goes to Ministry management the next week, after which the office is given one month to respond. These issues are checked up in the next year's monitoring visit. Monitoring officers noticed that the monitoring reports were sometimes not being read fully, so they tweak the report format every year: "If you read the same thing over and over you'll stop paying attention, but this way it stays fresh."

(2) Operations (Occasional)

Examines how well procedures for one-off work are structured (based on most common occasional task identified by respondent)

- Can you briefly talk me through your procedure for carrying out these occasional functions?
- How structured/ standardized is the process?
- How often do challenges or difficulties arise with the current way of doing it?
- How have you tried to address these challenges?

	Score 1	Score 3	Score 5
Key criteria	There is no structure or procedure for undertaking or managing one-off operations; all one-off work is done entirely *ad hoc*. The organization relies almost entirely on top-down assignment to get work done, so execution depends on the individuals involved. Main rationale for current practice is "that's the way it's always been done".	The organization makes use of some structures and processes to help undertake and manage one-off operations, but these may be inefficient or arbitrary. Where responsibility for work is assigned to individuals, follow-up and checks sometimes occur but are inconsistent. There is some consideration of how best to undertake one-off operations but only on an *ad hoc* basis.	Appropriate structures and processes have been put in place to improve quality and efficiency of one-off operations, and staff consistently implement these procedures.
Examples	An economic sector ministry contracts out policy development to consultants, but aside from an inception report there is little engagement until the final report is submitted, at which time it is often found that the consultants have made unsuitable recommendations and an in-house committee has to be convened to fix the report. Despite the recurrent nature of this problem, no changes have been made to the process.	An economic sector ministry has a standard procedure for initiating and developing new policy documents, but staff receive little training in executing it and so the process sometimes yields "amateurish results."	To develop new policies, an infrastructure ministry goes through a systematized process. An individual will develop a project proposal and initial scoping study, based on which a formal project team will be constituted to develop a detailed proposal and budget to be submitted to the minister for approval. This process is followed for all projects, albeit with some flexibility for unexpected or urgent projects.

(3) Process Documentation and Continuous Improvement
Examines processes for and attitudes towards continuous improvement

- When problems occur, how do they typically get exposed and fixed?
- How often do you discuss and challenge existing processes?
- Do officers ever suggest process improvements? Can you think of any examples?

	Score 1	**Score 3**	**Score 5**
Key criteria	Problems are only detected after they have affected work. No process improvements are usually made when problems occur. Frontline staff do not contribute to improving processes.	Problems are sometimes detected before they have negative impacts, but there is no structured, proactive process to do so. Improvements are made on an *ad hoc* basis, but depend entirely on individuals taking initiative. Resolution of problems involves some, but not all, of the appropriate individuals and staff groups, or includes non-relevant individuals or groups.	Exposing problems in a structured way is integral to individuals' responsibilities and resolution occurs as a part of normal business procedures rather than by extraordinary individual effort or *ad hoc* teams. Resolution of problems involves all appropriate individuals and staff groups.
Examples	A social sector ministry becomes aware of problems only once they have affected work. Management occasionally meets but this is not regular: "We hardly meet as a team." Management relies entirely on officers to take the initiative to raise problems, but even in these cases the focus is on logistical problems rather than operational performance. "It's the way things are done here."	A social sector ministry has quarterly sector working group meetings, as well as monthly meetings for a particular programme. Challenges and new ideas could be discussed at these, although they are viewed mainly as routine briefings. However, most problems are still not identified until they have negative impacts: "You would see if there was a problem because it affects your work."	An infrastructure ministry holds biweekly directors meetings and each division holds weekly meetings. The divisional meetings are very open, and operational staff see it as their job to make their directors aware of issues so they can be raised at a higher level. Feedback about the division from other divisions is discussed – for example whether the division is meeting deadlines or submitting receipts to the Finance division properly. Almost all officers keep electronic diaries so they know what is due at what time and when they should expect responses. They also use the register of outgoing letters kept by the division's secretary to track outstanding issues. All monitoring reports are shared with all directors on a routine basis.

(4) Adopting Practice Improvements

Examines whether practice improvements are effectively disseminated throughout the organization

- How does the organization encourage identifying and incorporating new practices and more efficient ways of doing things?
- How are these new practices shared around the organization?
- How does the organization ensure that officers are utilizing these new practices? How often does this happen?

	Score 1	Score 3	Score 5
Key criteria	There is no effort to incorporate new ideas or practices. When practice improvements do happen, there is no effort to disseminate them through the organization.	New ideas or practices are sometimes adopted but in an *ad hoc* way. These are sometimes shared informally or in a limited way, but the organization does not actively encourage this or monitor their adoption.	Seeking out and adopting practice improvements is an integral part of the organization's regular work. Improvements are systematically disseminated throughout the organization and monitored for adoption.
Examples	An infrastructure sector agency has occasional (but not regular) meetings of all staff, at which new ideas could in principle be raised, but in practice discussion revolves around administrative and logistical issues. "We hardly discuss actual work issues and new ideas."	In an administrative sector ministry, discussing new ideas is seen as a matter of individual initiative between an individual and his/her supervisor. However this is encouraged to an extent, and in some cases new ideas raised this way have been integrated into routine practice and shared with the rest of the organization.	An economic sector agency has a Core Transformation Team of nine or ten officers who meet every quarter and come up with a list of suggestions to improve the organization. These are put into a memo and sent to management, and fourteen days later management will meet with the Team to discuss. The members of the team include some division heads but also many younger officers, and were selected because they are bold and outspoken. The agency is also trying to create an Idea Forum as an additional way of encouraging staff to suggest innovations, and sharing new ideas within the division is seen as a routine part of work.

(5) Target Balance (Organization- and Division-Level)

Examines whether the organization defines an appropriate range of meaningful objectives and/ or targets to operationalize its goals

- What types of objectives and targets are set for the organization and its divisions? *(Note: Operational objectives and targets, not overarching policy direction)*
- Are objectives/ targets organized into different categories? Do they cover processes, outputs, and outcomes?
- How much are these objectives/ targets determined by statute/ external parties, as opposed to the organization itself?
- What about goals that are not externally set (i.e. are chosen by the organization itself)?

	Score 1	**Score 3**	**Score 5**
Key criteria	Objectives and targets are vague, only externally imposed, or non-existent. The only targets are lists of activities drawn up as part of the budget process.	Some relevant objectives and targets are defined, but these mainly consist of lists of activities and related outputs. These may be loosely linked to outcome targets.	A balanced range of objectives and targets is defined for the organization, including process targets and other performance-related targets.
Examples	An administration-sector ministry's only targets are activities or outputs. Outcomes and processes are neither targeted nor monitored.	A social-sector ministry has targets for inputs, outputs, and a handful of interim outcomes, although most key outcomes are not measured at all. "The [National Development Planning Commission] requires that [we] have an M&E plan with indicators on four levels: input, output, process, and outcome. But we focus only on input and output." One division has slightly more thorough monitoring of outcomes, collected mainly in conjunction with a donor program in that sub-sector. The efficiency of internal processes is not targeted or monitored.	In an economic-sector agency, the agency as a whole is focused mostly on outcome targets, while each division has output targets. These output targets contribute to the organization's overall outcome targets. Each division has key objectives (long-term); strategies for achieving them (medium-term); and specific targets (annual outputs). The organization also has some process targets, for example dealing with permit applications in a certain time period. These are defined in the agency's authorizing act.

(6) Defining Individual Tasks and Responsibilities

Examines how well the organization allocates tasks and defines individual responsibilities

- How are objectives set for individual staff members?
- Could everyone in the organization say what they are responsible for and how that will be assessed?
- When a staff member comes to work on Monday morning, how do they know what to do?
- How do you know which tasks are best suited to different staff?

	Score 1	**Score 3**	**Score 5**
Key criteria	Objectives are undefined or very vague. Staff are expected simply to "do what they're told". Tasks are allocated to individuals on an *ad hoc* basis, with little or no consideration of individual competences or workloads. Staff are often unsure what they should be doing.	Objectives may be formally defined but are vague or inconsistently applied. Staff have a general sense what they are responsible for, but completion of most tasks relies mainly on the giving/ following of direct orders. There is some consideration of which tasks should be allocated to certain individuals, but this is partial, inconsistent, or based only on an officer's formal position.	Individuals understand clearly what they are responsible for and what they should be doing at all times. Objectives are meaningful and are defined through a two-way process. Tasks are allocated based on identified competences.
Examples	Officers in a social-sector ministry have targets defined through the appraisal process, but in practice they wait for work to be minuted to them by a superior. Targets are seen as a formality that don't actually influence work: "People set targets, but the fact is that in most situations in our parts of the world these things are only done when there is something critical that occurs, like the officer wants a promotion." Officers are often idle because they don't have work. "It's not like you come to work every time and there is something special that you do." One consequence of this is that management are often overworked.	In one economic-sector ministry, directors apportion their division's objectives to officers at the start of the year. This is supposed to be annual but is not always done. The Scheme of Service provides some guidance about an individual's responsibilities, but not enough that officers are always aware of what specific tasks they should be doing: "It's not everyone who knows what they're responsible for and can get on with it without waiting for director's minutes...[but] some of us who have routine work know." Most officers below Deputy Director-level have only a general sense of how their work connects to organizational performance.	In an infrastructure ministry, each officer sits with their boss at the start of every year and reaches mutual agreement on a set of specific tasks and targets (e.g. submitting a report within 48 hours of a workshop). Individual tasks are derived directly from the division's targets and areas of responsibility. For example one technical division has an officer responsible for each sub-sector within that division; these officers take charge of handling letters and issues within that sub-sector, in consultation with their director: "People are pretty clear on what they need to be doing." Each division meets regularly to discuss the status of each officer's work, the connections between their roles, and the organization's targets: "All our work connects."

(7) Target Interconnection and Clarity

Examines how easily understandable objectives/ targets are and how well they cascade down the organization

- How do organizational objectives/ targets connect with divisional and individual activities and objectives?
- Does anyone complain that the targets are too complex, ambiguous, or unclear?
- If I asked one of your (mid-level) staff directly about these targets, what would he or she tell me?

	Score 1	**Score 3**	**Score 5**
Key criteria	Goals do not cascade down the organization, are unclear, and are not clearly understood. Beyond just following orders, individuals do not know how their work is related to overall performance.	Goals do cascade, but only to some staff and/ or division heads. Individuals have only a general sense of how their work contributes to performance and meeting overall objectives.	Goals increase in specificity as they cascade, ultimately defining individual expectations for all staff groups. Individuals understand clearly how their work contributes to overall performance.
Examples	In an administration sector ministry staff just have a general sense that their work has to do with the mission of the organization, but do not see specific connections.	Divisions in an economic sector ministry take their activities from the overall workplan, but whether directors brief their staff on these targets depends on the director.	An infrastructure ministry distributes organizational targets to its divisions, which then meet and agree a distribution of tasks to produce these outputs, as well as specific individual-level targets.

(8) Target Time Horizon

Examines whether the organization links short-term targets and long-term goals

- What kind of time scale are you looking at with your objectives/ targets?
- Are the long-term and short-term goals set independently? Could you meet all your short-run goals but miss your long-run goals?
- Which goals receive the most emphasis?

	Score 1	**Score 3**	**Score 5**
Key criteria	The organization focuses only on a single time horizon; either short-term targets or long-term goals are absent, or do not meaningfully influence the organization's actual functioning.	The organization has both short and long-term goals, but in practice the annual budget cycle dominates. Short- and long-term goals are set independently so they are not necessarily linked to each other.	Long-term goals are translated into specific short-term targets so that short-term targets become a 'staircase' to reach long-term goals. Meeting goals for each time horizon is taken seriously.
Examples	A social sector department only has annual goals, defined in each year's budget.	An administration sector ministry has annual targets that are broken down quarterly. These are not explicitly linked to the organization's overall mission.	An infrastructure ministry has qualitative long-term goals that are broken down into more specific medium-term targets, which are subdivided into annual and quarterly targets. Officers take short-term targets seriously because they view them as flowing into the long-term goals.

(9) Target Stretch

Examines whether objectives/ targets are appropriately demanding

- How tough are your objectives/ targets to achieve? Do you feel pushed by them? Are they realistic?
- How are the level of your objectives/ targets set?
- On average, how often would you say that you meet your objectives/ targets?

	Score 1	**Score 3**	**Score 5**
Key criteria	Objectives/ targets are either too easy or impossible to achieve, and may be little more than a 'laundry list' of desired outcomes. Target levels are chosen with no rationale or simply based on the previous years' with no re-evaluation.	Some objectives/ targets are challenging but these are often externally imposed. Target levels are based on a rough guess of what is possible given previous years' performance and available resources. Some parts of the organization are not held to the same standards ("sacred cows").	Objectives/ targets are genuinely demanding for all parts of the organisation and are developed through a negotiated process. Target levels are set based on analysis of relevant technical information.
Examples	An infrastructure agency, individuals set targets for themselves and these are never checked. "Speaking as a professional, my targets should be co-determined by myself, my department, and the organization as a whole, and should be based on the organization's focus for the year. But each year we set our targets and don't get any confirmation or discussion from the system, the organization. We don't sit and challenge each other about them. If I do it or don't do it, it's nobody's business."	In a social sector ministry, targets are set at the previous year's level, with adjustments for perceived changes in the resource levels and the organization's delivery capacity.	In an economic sector agency, operational divisions first propose targets, then the Planning division challenges them on their justification and ambition. The divisions then negotiate, with the head of the organization serving as mediator.

(10) Performance Tracking

Examines whether organizational performance is tracked using meaningful metrics and with appropriate regularity

- How do you monitor how well the organization is achieving its goals?
- Do you use any indicators to track performance? What kind (quantitative/ qualitative)?
- How frequently are these indicators measured?
- How are these indicators related to the organization's objectives/ targets? Are they directly under the organization's control?

	Score 1	**Score 3**	**Score 5**
Key criteria	Measures tracked are not appropriate or do not indicate directly if overall objectives are being met. Tracking is an *ad hoc* process and most processes aren't tracked at all.	Performance indicators have been specified but may not be relevant to the organization's objectives. Some important performance indicators are tracked formally.	Performance is continuously tracked, both formally and informally, using appropriate indicators.
Examples	A social sector department has no functioning monitoring and evaluation unit, and no indicators are defined or tracked. There is an annual report, but "It's a make-believe kind of approach to assessing ourselves, it's not scientific."	An administration sector agency writes monthly and quarterly reports, but indicators are limited to activities. They are aware of some outcomes but these are not tracked formally.	A social sector ministry tracks a number of outcome, input, and financial performance indicators through a decentralized online system. Some of this indicators are for particular agencies or districts while others are negotiated on a sector-wide basis with agencies and development partners. These are all measured annually, and some can be updated more frequently.

(11) Performance Review and Dialogue

Examines whether organizational performance is meaningfully reviewed with appropriate frequency and communicated to staff

- How do you review performance, formally or informally? How often?
- Who is involved in these meetings? What type of feedback occurs during these meetings? Who gets to see the results of this review?
- If I were to ask an officer who is involved in implementation, could he/she tell me how the organization is doing against these indicators?
- For a given problem, how would you identify the root cause?

	Score 1	**Score 3**	**Score 5**
Key criteria	Performance is reviewed infrequently or in an unmeaningful way (only success or failure is noted). The right information for a constructive discussion is often not present. Conversations are 'one-way', focus overly on data that is not meaningful, and exclude most staff. The purpose of the review is not explicitly stated. Next steps are not clearly defined.	Performance is reviewed periodically with appropriate data, and both successes and failures identified. Objectives of meetings are clear to all participating, but results are communicated only to senior staff and conversations often do not drive to the root causes of problems. Next steps are mentioned but not well defined.	Performance is continually reviewed, based on the indicators tracked, with a focus on problem solving and addressing root causes. Purpose and follow-up steps are clear to all. Meetings are an opportunity for constructive feedback and coaching. Results are communicated to all staff.
Examples	A social-sector ministry holds monthly management meetings in principle, but these have not been regular for the past year. Management meetings are "more for human management and the condition of the office and that kind of thing than results management. There's very little discussion of what are we supposed to do and where are we." The ministry has quarterly reviews for some of its programs, but only the donor-led ones are regular. Quarterly and annual reports are seen as a formality undertaken to meet external requirements, not an opportunity to review performance.	Another social sector-ministry has monthly management meetings, which are regular, plus an annual review. These are focused on performance. It also holds meetings of all staff (about three per year) to discuss administrative issues. Directors are supposed to brief their officers after management meetings, but this does not always happen. More senior officers would learn about the organization's performance through working with their directors. Agreed actions would be minuted to the officer responsible – "That's if the minutes can be done in good time, sometimes they can take long."	An economic-sector agency has several levels of performance review: weekly, quarterly, and annual. Each level involves different groups of staff, so everyone gets to know about the agency's performance. Management meetings are held every Monday morning. Each division's quarterly report feeds into the agency's quarterly report, which is viewed as an opportunity to review performance. Directors meet quarterly with their divisions (usually, but not 100% of the time), and general staff meetings are held occasionally. There is an annual retreat for senior management to assess performance and plan for the next year; the heads of departments brief their staff after returning from this. At the beginning of the year a review is held to discuss the previous year's performance.

(12) Consequence Management

Examines how well the organization follows up on and enacts plans to fix identified problems (department/ process-level)

- Let's say you've agreed to a follow-up plan at one of your meetings, how would you carry that plan forward? What would happen if the plan was not enacted?
- How long does it typically go between when a problem is identified to when it is solved? Can you think of a recent example?
- How do you deal with repeated failures in a specific department or area of process?

	Score 1	**Score 3**	**Score 5**
Key criteria	Failure to achieve agreed objectives does not carry any consequences. Responsibility for following up is poorly defined.	Responsibility for follow-up may be allocated, but there would be no other specific actions beyond this. Failure to achieve agreed results is tolerated for a period before action is taken and sanctions are rare. Difficult problems can persist indefinitely.	Responsibility for follow-up is clearly defined and accompanied by other complementary steps. Failure to achieve agreed objectives is quickly addressed and may lead to sanctions. Ongoing efforts are made to fix even difficult problems.
Examples	In an economic sector ministry, if an agreed action was not taken, "they would say it was 'due to unforeseen circumstances' and they would be given another week. I've never seen any punishment given."	An economic sector agency assigns problems and follow-up actions to directorates, which then have to report back at the next meeting. If the action was not taken, at the following meeting a deadline might be given, but nothing beyond this.	An infrastructure ministry assigns problems to individuals and notes this in the minutes, which are circulated to all directors within 48 hours of the meeting. If these actions were not taken without good reason a formal reprimand would be issued, although the ministry is shifting to a more team-based system of work and thus reducing the use of individual reprimands.

(13) Assessing and Rewarding Performance

Examines whether individual performance is assessed transparently and rewarded proportionately

- How does your appraisal system work?
- Are the appraisals actually completed regularly?
- Are there any financial or non-financial (e.g. awards, training, travel opportunities) rewards related to this review?
- Are better performers likely to be promoted faster or are promotions given solely on the basis of tenure / seniority?

	Score 1	**Score 3**	**Score 5**
Key criteria	People are rewarded in the same way irrespective of their performance. Promotions are solely based on tenure.	There is an evaluation system which in principle awards good performance (financially or non-financially), but awards are not based on clear criteria/ processes. Performance can sometimes influence career progression.	There is an evaluation system which rewards individuals (financially or non-financially) based on performance. Rewards are given as a consequence of well-defined and monitored individual achievements. Performance is a key criterion for promotion decisions.
Examples	In an economic-sector ministry, staff appraisals are only completed once every three years, when it is time for the officer to go for promotion. The appraisal is basically a self-assessment, actual discussions about officers' performance rarely happen: "The appraisals don't improve the system. They're a formality." There is no link between performance and promotion. Recommendations for promotion are based solely on seniority, and out-of-turn promotion never happens. In principle performance can be incentivized with training opportunities, but the allocation of these is viewed as subjective: "One director even says he'll send you 'if your face looks nice'."	Appraisals in one social-sector ministry are always done annually, but there are no rewards and the process is still viewed as a formality. Better performing officers are sometimes more likely to be recommended as a focal person for an activity and receive additional responsibilities, but this is viewed as subjective and dependent on one's director and the chief director. In principle non-performing officers can not be recommended for promotion, but in practice this does not happen and recommendation for promotion is based entirely on years served. The ministry gives all officers a bag of rice at Christmas as a bonus, but this is given to all staff regardless of performance.	In one infrastructure ministry, appraisals are always done annually, and individuals' performance is also discussed weekly in divisional meetings, so officers are also aware of their peers' performance. Each year the ministry holds an awards night. Within each division staff vote for a top performer, and these individuals vote for an overall top performer amongst themselves. Accelerated promotions are rare, but non-performing individuals tend to be recommended more slowly, e.g. after five years instead of three. The ministry also makes an effort to recognize officers for exceptional work: one young officer was asked to give an important presentation on behalf of his director who had traveled. Later he received a written letter of congratulations from the Director of Human Resources Management on behalf of the Chief Director. This officer was highly motivated by this gesture – he has kept the letter and even read the letter to his children.

(14) Identifying Skill Needs and Attracting Human Capital

Examines the emphasis put on attracting talented applicants and applicants with particular skills

- How does the organization identify skill needs and ensure it has enough personnel with particular skills?
- How does the organization seek out individuals with these particular skills?
- How do you decide when to send staff for training? How does the organization select which staff to send?

	Score 1	Score 3	Score 5
Key criteria	The organization does not go out of its way to recruit talented applicants or applicants with particular skills, so hiring is driven entirely by staff numbers and availability of funds. Training opportunities are not actively sought, and when available are not allocated based on capacity and/ or need.	The organization occasionally tries to recruit applicants with particular skills, but identification of needed skills is not systematic and recruitment is only occasionally based on identified skill needs. The organization rarely goes out of its way to recruit talented applicants. Training opportunities are occasionally sought, but driven mostly by availability.	The organization actively recruits talented applicants and applicants with needed skills, based on regular assessments of skill needs. Training opportunities are sought and allocated to the staff who can most benefit from them.
Examples	An administrative sector ministry has no needs assessment or training plan, and its only targeting of recruits is for those that have a university degree and are computer literate.	A social sector ministry recruits mainly by selecting National Service personnel who they would like to stay on, or by handpicking staff from its agencies and having them put on secondment, but this depends on the intervention of the Chief Director and there is no systematic process for it.	An infrastructure ministry has an HR Development Plan based on a needs assessment, and also initiates specific needs assessments for emerging issue areas. Although they used to rely heavily on long-term consultants funded by donor projects, they now have brought enough expertise in-house that they only hire consultants for specific short-term contracts. To attract and retain staff the ministry actively seeks out training opportunities for staff, and has set up a provident.

(15) Dealing with Poor Performers

Examines whether the organization is able to deal with underperformers

- If an officer was not performing his/her job well, what would happen? Can you give me a recent example?
- How long would under-performance be tolerated? How common is it for officers to be removed from their position (not necessarily fired) because of underperformance?

	Score 1	Score 3	Score 5
Key criteria	Poor performance is not addressed or is inconsistently addressed. Poor performers rarely suffer consequences or are removed from their positions.	Poor performance is addressed, but on an ad hoc basis. Use of intermediate interventions, such as training, is inconsistent. Poor performers are sometimes removed from their positions under conditions of repeated poor performance.	Repeated poor performance is systematically addressed, beginning with targeted intermediate interventions. Persistently poor performers are moved to less critical roles or out of the organization.
Examples	In an administration sector ministry, poor performance is never addressed. In principle the person could be spoken to by a supervisor, but interviewees were not able to recall an instance when this had happened.	An economic sector ministry takes tasks away from poor performers and in some cases transfers them. The main sanctions are social; formal sanctions only occur in extreme cases that border on criminal behavior.	An economic sector agency addresses poor performance by shifting problem individuals to different departments or regions to see if this improves their performance, and offering retraining in some cases. Though these steps are common, firings never happen. *[Note: this example was scored 4; no organization in the sample scored a 5 on this practice.]*

APPENDIX 2.2: ADDITIONAL DETAILS OF INTERVIEW METHODOLOGY

This appendix details additional details of the selection of respondents, conduct of the interviews, and data quality checks undertaken, noting important differences to the standard BVR methodology and Rasul and Rogger's (2014) adaptation.

Following BVR, the interviews targeted individuals who were senior enough to have an overall picture of the organization's functioning but also low enough to be involved in the day-to-day functioning of the organization. Most commonly this meant directors of divisions (the second-highest level of the organization), although many interviews were also held with deputy directors and a handful with other senior officers or the organization's chief executive. To set up the interviews, Ghana's Office of the Head of Civil Service wrote to each ministry to request their participation in the research project. I then made contact with each ministry and arranged for interviews with officers of an appropriate level and position.

In order to focus as closely as possible on the actual functioning of the organization, the interviews focused on officers in technical, operational divisions as opposed to purely administrative ones like finance. The most common interviewee was the organization's Director of Policy, Planning, Monitoring, and Evaluation (PPME); for consistency, one of the two interviews in each ministry was conducted with an officer from this division or its analogue. Targeting officers in specific positions for interviews in this way minimized the potential for selection bias, and although timing and availability necessarily played some role in selection, there is no evidence that responses or management scores differed systematically by interview or interviewee characteristics. For interviewee selection in the agencies, I either received a personal referral from the sector ministry to the individual in the agency responsible for policy and monitoring functions, equivalent to the organization's Director of PPME, or was directed to this individual through formal correspondence, and then proceeded from this introduction as in the ministries. In each case selection bias is not a concern, as interviewee selection was still driven by position.

The interviews lasted 75 minutes on average and were conducted in person by the author in the interviewee's office. Responses were transcribed by hand during the interview (but not recorded) and subsequently post-coded. To encourage honest answers, all interviewees were guaranteed that they personally would remain anonymous and that their answers would not be associated with their organization. Again, this guarantee was particularly important in the context of discussing sensitive information in hierarchical government organizations. This follows established practice in BVR and other surveys, in conformity with LSE's human subjects research ethics guidelines. In order to focus on actual objective practices rather than interviewees' own subjective perceptions of management quality, interviewees were informed that they were participating in a research project on variation in management practices in the Ghana's public service, but were not told in advance that their responses were being benchmarked according to a numerical metric. Again, this aspect of the research design follows BVR. As in their work, not informing interviewees in advance that their responses would be analyzed using numerical post-codings as well as qualitatively was not deemed a problem because the potential for deception was minimal, posed no harm due to the strict anonymity protections imposed by the study, and interviewees were sent a report of the study after completion of the research.

Conducting the interviews in person is a significant departure from BVR, who conducted all interviews by telephone. In-person interviews were preferred in the developing country public service setting due to the greater emphasis on interpersonal relationships and trust, and were made possible by the location of all organizations' headquarters in Accra. This meant that interviews were effectively single-blind rather than double-blind, as in BVR. While this could engender concerns about interviewer bias, any such effect is presumably counteracted by the greater consistency achieved by using a single interviewer and by the increased data quality arising from in-person interviews. This approach differs slightly from that of Rasul and Rogger, who also used in-person interviews but did so in groups, and explicitly asked respondents to score management practices rather than having the interviewer do so. The impact of these methodological subtleties on data quality is an important area for future empirical work.

Most of these practices are analogous to practices included in various versions of the BVR surveys, albeit with slight adaptations to suit the public sector context, while others were added after pilot interviews with civil servants in Ghana and the United Kingdom. Appendix 2.1 contains the full list of the practices and scoring grid. The most notable adaptation was to broaden the emphasis of the benchmarking criteria from a focus on quantitative performance measures and financial incentives to give equal treatment to qualitative outcomes and non-financial rewards for staff. This was done to reflect the relative difficulty of measuring many public sector outputs and the lack of consensus about the appropriateness of hard incentive schemes in public sector organizations. This treatment of incentives differs significantly from Rasul and Rogger's (2014) own adaptation of the BVR methodology, and may help to explain why incentive-related practices were negatively correlated with project completion in their study.

Figure A1: Within-organization correlation of management scores

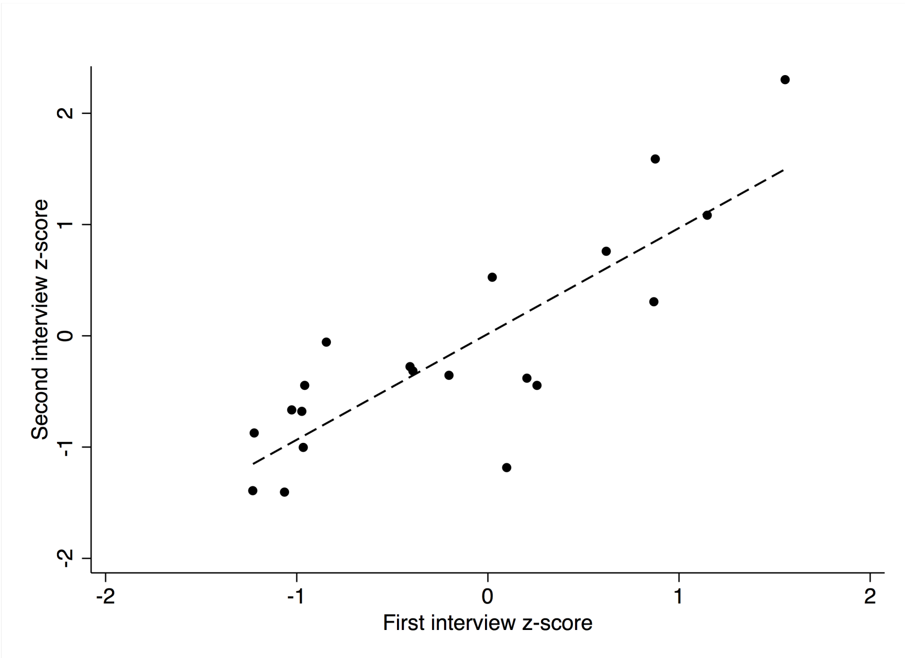
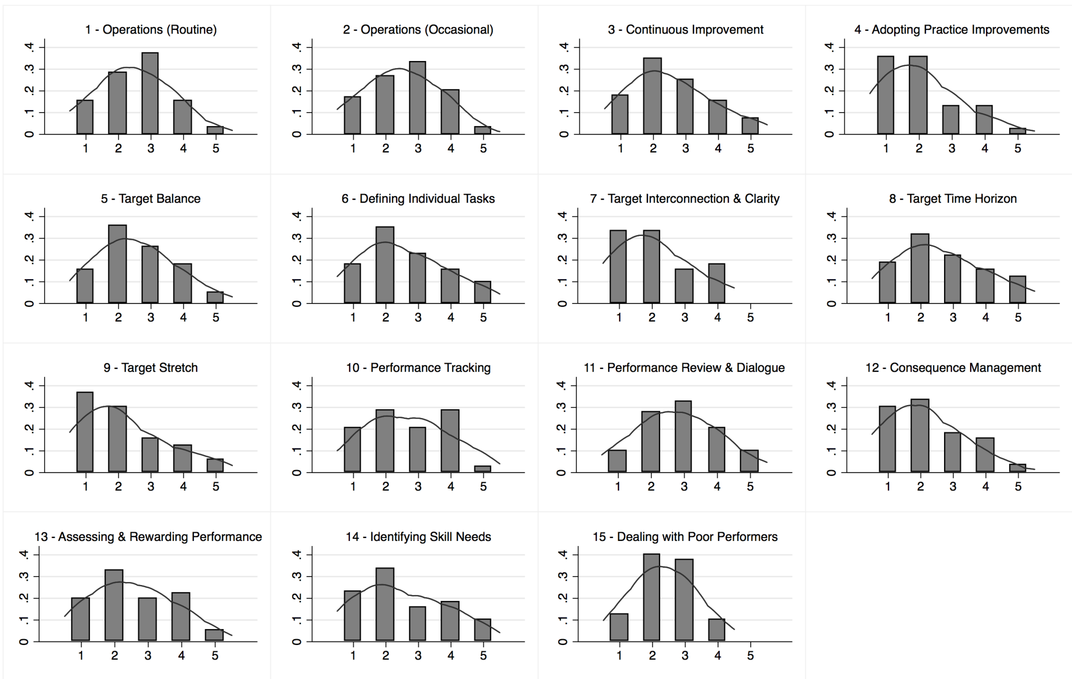


Figure A1 shows that the overall z-scores between two interviews undertaken in each organization are strongly correlated. In addition, there is a high degree of correspondence between the qualitative details given by interviewees within an organization. The most objective indication of this was given by Practice 13, Assessing and Rewarding Individual Performance. Since the same annual appraisal process is used by all ministries and an analogous annual process in used by agencies, all respondents reported this appraisal process, and in each case, I asked whether this annual process was *actually* always carried out every year. In many cases it was not, but importantly, the responses within an organization across the two interviewees were always consistent about whether the appraisals were or were not carried out annually in practice. This is reassuring evidence that reported variation across organizations is driven by real differences in practices, not individual idiosyncrasies or misreporting.

Figure A2: Distribution of scores within each practice



Note: Histogram of raw practice scores with Epanechnikov kernel density plot. Half values rounded up.

Finally, Figure A2 shows the distribution of scores for each of the 15 management practices that comprise the overall index. The distributions vary somewhat across the different practices; explaining why these distributions differ would be a useful task for future research.

Chapter Three

POLICY IMPLEMENTATION, DISTRIBUTIVE POLITICS, AND FISCAL INSTITUTIONS:

EVIDENCE FROM UNFINISHED INFRASTRUCTURE IN GHANA

Chapter Abstract

I study the relationship between policy implementation and distributive politics, and how fiscal institutions mediate this relationship. Using an original database of over 14,000 small local government infrastructure projects in Ghana, I show that approximately one-third of projects are started but never completed, consuming one-fifth of total capital expenditure, even though they are simple, popular, and visible. To explain this, I develop a theory of policy implementation as intertemporal bargaining among political actors who face commitment problems with respect to project distribution, and predict that the fiscal institutions used to fund projects can mitigate the negative effects of these distributive pressures. I find large differences in completion rates *within* districts across different project financing mechanisms, even when these projects are selected and implemented by the same politicians, bureaucrats, and contractors. Consistent with the theory, the relationship between project completion and districts' partisan alignment *varies* across these delivery mechanisms. I present evidence that this is due to bargaining failures among local political actors, rather than corruption, tactical redistribution, or other alternative explanations. I discuss implications for theories of public service delivery, distributive politics, and fiscal decentralization.

Policy Implementation, Distributive Politics, and Fiscal Institutions: Evidence from Unfinished Infrastructure in Ghana

> *"We all have demands on government but we have to be moderate in these demands in order that we do not disrupt the budget and government expenditure…As the head of your family, you cannot provide all that your family needs at once…Alhassan needs a shoe, Fuseini needs a bicycle, your wife needs a new cloth, and your mother needs a refrigerator. You cannot provide all these at once and so what you have to do is to prioritize…The fact that government has not yet provided a certain development that was requested does not mean that government will not provide it."*
>
> – John Mahama, President of Ghana (TV3 Network 2015)

* * *

Poor policy implementation is a salient feature of states with weak government bureaucracies, yet is not well understood. A particularly striking example is the phenomenon of unfinished infrastructure projects, which is anecdotally widespread in developing countries: schools, latrines, culverts, and bridges are started, partially constructed, but then abandoned and never finished. These types of local public goods are technically simple, relatively affordable, highly visible, politically popular, and desperately needed in contexts with massive infrastructure deficits. In short, they are exactly the type of goods that politicians seek to deliver in theories of pork-barreling, clientelism, and distributive politics. Why, then, would governments expend their scarce resources to half-build projects that have no value to constituents?

In this paper, I develop a novel theoretical explanation for partial policy implementation that revolves around the tension between the technical demands of project delivery and the political processes that determine the distribution of projects. Politicians typically face greater demand for public goods than they can supply at any given time, so they have to prioritize and sequence delivery. Sequencing delivery requires intertemporal exchange among political actors, but this is difficult since these actors cannot commit to future distributional decisions. When this intertemporal bargaining process fails, the path of policy implementation may be inefficiently erratic – in particular, politicians may devote resources to starting new projects to appease distributive pressures rather than to finishing existing ones. Since projects are implemented in multiple stages and efficient delivery requires continuity once a project is started, this political instability increases the overall cost of completing projects and leads to some projects being abandoned mid-construction.

A second question is how government institutions can be designed to minimize this problem. Delegating implementation to a neutral and capable bureaucracy would be one solution but such bureaucracies are the exception rather than the rule in developing country governments, and creating them may require reforms across dozens or even hundreds of sub-national authorities and central ministries and agencies. An appealing alternative for governments and donors is to focus their efforts on reforming fiscal institutions – the bureaucratic structures through which funds are delivered to implementing units – over which they have more direct control. Yet there is little evidence on the operational effectiveness impacts of alternative fiscal institutions, and attempts to make such comparisons across polities are usually confounded by concerns over unobserved heterogeneity and endogenous institutional choice.

To address these issues empirically, I have collected, digitized, and coded administrative records of over 14,000 local government infrastructure projects in Ghana from 2011 to 2013. These are predominantly small, highly visible, and technically simple projects such as classroom blocks and latrines. The median project had a budget roughly equivalent to USD 36,000 and was scheduled to last five months, and 88.8 percent of projects were scheduled to be complete within one year. However, after one year less than half were finished, and even after three years over a third of projects remained unfinished and are unlikely ever to be completed. These unfinished projects consumed nearly 20 percent of all capital spending, equivalent to USD 26.6 million annually, or enough to fully construct over 700 three-room schools *every year*. These descriptive findings represent the most comprehensive evidence available to date on the scope and shape of the problem of unfinished infrastructure projects in any developing country.

I examine whether fiscal institutions affect project completion by taking advantage of a unique feature of the dataset: the same types of projects are executed in the same districts and communities by the same pool of contractors through different funding sources. Even after controlling for all these factors, completion rates for observably identical projects (e.g. three-room schools) differ significantly across the three main funding sources. This within-district variation is of comparable magnitude to the variation in completion rates *across* districts – a project from the best fund source in the 25th-percentile district has the same completion probability as a project from the worst fund source in the 70th-percentile district. I examine issues of endogenous project selection into funding mechanisms using coefficient stability tests, and argue that it is unlikely to account for more than a fraction of the observed variation.

To understand the mechanism through which fiscal institutions affect project completion, I narrow the comparison to two lump sum transfer programs for which the same local-level politicians and bureaucrats have complete control over project selection and implementation: an unconditional grant funded by the central government, and a mainly donor-funded grant that requires districts to budget to complete existing projects before starting new ones. Consistent with the predictions of the intertemporal exchange theory of policy implementation, projects funded by the donor grant have a higher completion rate, but this effect is strongest in districts that voted heavily for the opposition party and disappears entirely in districts that voted heavily for the ruling party. I argue that this supports the idea that project non-completion is driven by bargaining failures among local-level political elites, and present evidence against alternative explanations.

This paper connects and contributes to several literatures. First, within the body of work on policy implementation in developing countries, Rasul and Rogger (2014) is the only other study (to my knowledge) to examine project completion in a large-N setting, while most studies focus instead on financial or material losses due to corruption (Reinikka and Svensson 2004, Olken 2007, various in Olken and Pande 2012). None of these studies link implementation to distributive politics. Second, a scattered literature on fiscal institutions and fiscal decentralization examines how budget institutions affect overall spending and borrowing (Ferejohn and Krehbiel 1987; von Hagen and Harden 1995; Alesina, Hausmann, Hommes, and Stein 1999; Rodden, Eskeland, and Litvack 2003; Rodden 2006) or how the source of funds affects their management and public service outcomes (Fisman and Gatti 2002, Gadenne 2015). This paper is unique in focusing on the operational impacts of financing mechanisms on project-level data. Third, I extend the recent literature in American politics that examines how administrative agencies mediate the workings of distributive politics (Bertelli and Grose 2009, Berry and Gersen 2010, Gordon 2011, Ting 2012) into a new context and suggest a new mechanism.

More broadly, this study's findings also have significant implications for the burgeoning empirical literature on distributive politics and clientelism, which tends to focus *either* on input-based measures of expenditure choice and allocation (Ward and John 1999, Keefer and Khemani 2009, Banful 2011, Jablonski 2014) *or* on output- or outcome-based measures of access or service quality (Besley et al 2004, Franck and Rainer 2012, Golden and Min 2013, Kramon and Posner 2013, Harding 2015), and assume that these are the outcomes of tactical targeting strategies by parties, politicians, or incumbents.[37] Yet this only holds if the determinants of project implementation are orthogonal to the determinants of project distribution – an extremely strong assumption that is contradicted by this paper. In contrast, I view both project distribution and implementation as collective choice processes among multiple principals that are more akin to logrolling in their potential for instability and inefficiency (Riker and Brams 1973, Weingast and Marshall 1988), so outputs and outcomes may differ systematically from inputs and intentions.

The remainder of this paper proceeds as follows. The next section develops a theoretical explanation for the occurrence of unfinished projects rooted in the technical attributes of project implementation and the commitment problems facing political actors, and discusses how fiscal institutions might mediate this relationship. Section 3.2 operationalizes this model in the context of infrastructure delivery in Ghana's local governments and lays out empirical hypotheses, and Section 3.3 presents the paper's original dataset. Section 3.4 examines the extent and dynamics of project non-completion, shows that there are substantial differences in completion rates across fund sources, and argues that these effects are largely causal rather than due to selection on unobservables. Section 3.5 shows that the efficiency impacts of

[37] Burgess *et al* (2015) use both a measure of expenditure and a measure of physical completion as dependent variables, but do not link the two. Implementation considerations are also absent from the foundational theoretical models in this area (Cox and McCubbins 1986, Lindbeck and Weibull 1987, Dixit and Londregan 1996).

fiscal institutions are conditional on districts' partisan alignment, and examines alternative explanations. Section 3.6 concludes with a brief discussion of implications for theory and policy.

3.1 PROJECT IMPLEMENTATION AS INTERTEMPORAL BARGAINING

This section presents a theoretical explanation of why and how the implementation of government projects may be undermined by distributive conflict over their allocation, and how this can be mediated by the fiscal institutions used to fund projects. This theory builds on two insights that are largely absent from the existing literature on the political economy of public service delivery in developing countries: 1) all but the smallest government projects must be implemented in multiple stages; and 2) in many cases project selection is the outcome of a collective choice process among many political actors rather than the revealed targeting strategy of a unitary politician, party, or incumbent engaging in tactical redistribution.

The basic theoretical mechanism is as follows. Governments typically face greater demands for projects than they can supply at one time but projects are discrete and expenditure on them is imperfectly fungible, so efficient implementation requires that they prioritize and sequence projects. This requires intertemporal exchange among political actors, but these actors face commitment problems and thus may not be able to strike and maintain such bargains, in which case the government's expenditure path will be erratic. This is inefficient from a technical standpoint, and leads some projects to be abandoned mid-implementation. These problems could be avoided by delegating implementation to a neutral and capable bureaucracy but this is typically not possible in developing countries; however, fiscal institutions may be able to mitigate the negative efficiency consequences of distributive pressures by increasing the cost of non-cooperation among political actors.

My exposition proceeds in three steps. First, I discuss the technical attributes of project delivery, in particular the multi-stage nature of their implementation and the imperfect fungibility of project expenditure across space and time. Second, I show how this "lumpiness" requires intertemporal exchange among political actors who face commitment problems, which if unsuccessful leads some projects to be started but never finished. Finally, I discuss the factors that determine whether political actors are able to strike intertemporal bargains over project distribution, and show how fiscal institutions provide a second-best solution when this bargaining process fails. For concreteness I situate my theoretical discussion in the context of the delivery of small infrastructure projects by decentralized governments dependent on inter-governmental transfers or donor funds – a common situation in developing countries (Gadenne and Singhal 2014). However, the underlying mechanism may also be applicable to project implementation in other settings.

Implementing projects

The implementation of all but the smallest development projects happen across multiple stages, rather than instantaneously, yet the consequences of this have been overlooked in the recent wave of literature on the political economy of public service delivery. Numerous studies (mainly focused on corruption) have documented that large gaps can arise between inputs and outputs over the course of implementation (Reinikka and Svensson 2004; Olken 2007; Beekman, Bulte, and Nillesen 2014), but view this implicitly as a static loss rather than examining the dynamic consequences.[38] This contrasts with the mostly qualitative, case study-based implementation literature in public administration, in which changing political and technical contexts and intertemporal decision-making processes during implementation play a key role (Pressman and Wildavsky 1973, Grindle 1980, Cornell 2014). Likewise, a growing number of studies on infrastructure project delivery document that delays, cost overruns, and other implementation problems are the norm rather than the exception (Flyvbjerg, Holm, and Buhl 2004; Flyvbjerg 2007, 2013; Budzier and Flyvbjerg 2013).

Figure 1 illustrates the temporal dimension of project implementation with a stylized timeline of the construction of a typical small public infrastructure project.[39] After

[38] Macro-level studies have also found similar gaps between inputs and outputs: for example, IMF (2015) calculates measures of investment efficiency using measures of public capital investment stock and infrastructure access.

[39] I focus on simple build-transfer projects, in which a private contractor is procured to build a government-designed project and the contractor's involvement ends upon physical completion when it hands the project over to the government. The large literature on public-private partnerships (PPPs), in

projects are selected, planned, and budgeted, the government goes through a procurement process and awards the contract to a private contractor. The contractor begins work on the project, and after completing the first stage requests partial payment from the government, which then inspects the work completed and, if satisfactory, makes payment.[40] This cycle then repeats itself: the contractor completes the project and requests full payment, and once the government has inspected the work and made full payment the project is handed over to the government and can be put into use. The multi-stage nature of this process means that, even after making the initial policy choice to start the project, the government must periodically renew its commitment to the project by releasing funds to the contractor, as stipulated in the contract. Failure to do so stops physical progress on the project, and unless payment is subsequently made and work resumed, the project will remain unfinished. Non-completion of projects entails major efficiency losses because projects typically generate few benefits for users until they are entirely complete – a school with walls but no roof is of little use.[41]

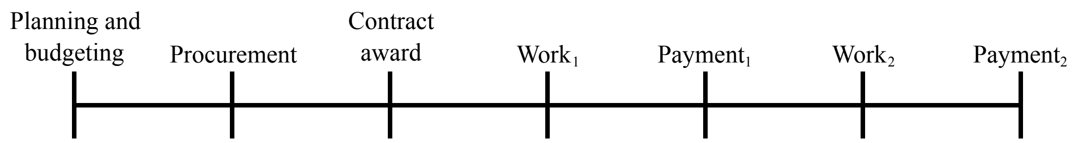
---

contrast, focuses mainly on projects in which the private sector party also maintains a post-completion operational role. While I focus on physical infrastructure projects, these features pertain to many types of discrete government undertakings. For example, it was reported to the Public Accounts Committee of Ghana's Parliament that the preparation of a Land Bill and Land Use and Planning Bill had overrun its planned expenditure by 317 percent and, years later, still had not been submitted to Parliament (Daily Guide 2015). The discrete nature of development *projects* (e.g. building a school) differentiates them from more routine services (e.g. teaching in the school); this distinction between projects and routine operations is a common one in public administration and management, but has not been explored by the literature on the political economy of service delivery.

[40] In some cases the government may make an up-front mobilization payment to the contractor prior to the start of work.

[41] This is a slight exaggeration, since projects nearing completion can sometimes be used without full functionality. For example, in site visits for this study the research team encountered instances of classes being held in a classroom block that had been roofed but construction had stalled prior to installation of windows, cladding of the building, and furnishing. Nonetheless, most infrastructure projects are characterized by a discontinuous relationship between implementation progress and use value. This non-linear relationship between expenditure on projects and their value to users distinguishes them from other types of policies, such as targeted grants or routine services, for which partial implementation can still generate relatively proportional benefits.

Figure 1: Stylized Timeline of an Infrastructure Project



This multi-stage character of project implementation interacts with the imperfect fungibility of project expenditure. Project expenditure is imperfectly fungible across time periods because mid-project interruptions or delays can increase the overall cost of completing the project, thus reducing the benefit-to-cost ratio of the project and making it less likely that it will ever be completed. This can be for engineering reasons – leaving half-constructed buildings exposed to the elements can result in physical decay within just a few months – as well as economic reasons such as contractors' financing and relocation costs, general price inflation, or recontracting costs.[42] Likewise, there are scale economies in the project implementation process itself arising from contractors' fixed costs of capital and managerial supervision. Efficient implementation thus requires continuity once a project has been started. This "lumpy" nature of projects also makes expenditure imperfectly fungible across potential projects: whereas fungible distributive benefits (e.g. subsidies, tax credits, budget allocations) can be divided and shared according to some targeting logic without losing value to beneficiaries, expenditure on discrete projects must cross a minimum threshold – the cost of completing the project – before it creates use value.[43]

---

[42] For a sample of OECD transportation projects, Flyvbjerg, Holm, and Buhl (2004) estimate that each year of delay is associated with a 4.6 percent cost increase; in developing countries with higher inflation and interest rates and weaker legal systems, this figure would presumably be even larger. The effect of these cost increases is compounded by nominal rigidities inherent in budgeting, policymaking, and procurement processes, which require renewed political and bureaucratic approval of nominal contract adjustments even if the real cost of the project does not increase. If nominal rigidities make it difficult to adjust contract amounts mid-project, then projects that are destined to overrun their initial cost could be abandoned mid-construction rather than completed at an increased cost.

[43] This property is also noted by Dixit in his multi-principal model of government agencies: "multiple principals may find it difficult to agree on the split of the total gain from cooperation, or be unable to make the internal transfers necessary to implement an agreed-upon split. This is particularly important in the political context, where the benefits of the principals are often nonmonetary and are measured in

Distributive politics, project sequencing, and commitment problems

A core premise of the vast literature on pork-barrel politics and clientelism is that politicians seek to deliver benefits to certain constituencies according to some type of tactical targeting logic in order to gain electoral support, and these benefits often take the form of locally targeted club goods, like small infrastructure projects, that benefit the recipient community but not other communities in the polity. From this tactical perspective, unfinished projects must have been left unfinished deliberately, either because the intended beneficiary was not the service users but the contractor (Samuels 2002), to give clients an incentive to return the politician to power to complete the project (similar to Robinson and Torvik 2005), to punish opponents, or simply as a low-cost signal of intent or promise.

However, political authority is rarely vested in a unitary actor in this way, even between elections. Politicians often have to share decision-making power with either a legislative assembly, are embedded in informal downward accountability relationships (such as interest group lobbying, personalistic ties, clientelist networks, or politically costly protests), or are vertically accountable to higher-level politicians. These are all especially prevalent in the context of local public good provision.[44] Politicians may therefore be constrained in enacting their preferred targeting strategy by the need to maintain the support of a minimum coalition of these political actors in any given period. Again this support could be formal, such as obtaining votes for a budget in a legislature, or informal, such as interest groups refraining from protest.[45]

---

noncomparable, non-transferable units, whereas monetary compensations are often illegal" (1996, p.98-9).

[44] On informal accountability ties, see for example Helmke and Levitsky (2004) and Tsai (2007). Leaver's (2009) model of "bureaucratic "minimal squawk behavior" provides an example of how interest groups can hold appointed officials accountable by appealing to their superiors.

[45] For generality, I refer in this section to one politician who must maintain some level of support from multiple political actors, without specifying the identity of the other political actors (who may also be elected politicians), the form of the accountability relationship (formal or informal, upward or downward), or the amount of support the politician must maintain (which could be more or less than a simple majority, and will also depend on the form of the accountability relationship). In the following section I will specify these actors and relationships in order to operationalize this model in the context

Yet at any given time, resource constraints make it impossible for a politician to simultaneously deliver projects to all (or even most) of her constituencies. Since project expenditure is imperfectly fungible, this means that the politician and the other political actors who have influence over project selection must make collective choices about the relative priorities of projects and their beneficiary constituencies, and select some projects to be implemented before others.[46] However, reaching and maintaining such an intertemporal bargain may be difficult because the non-contemporaneous benefit flows create a commitment problem: neither the politician nor the other political actors can commit to their actions in future periods because all have clear incentives to renege on their promises *ex post*.[47]

---

of decentralization in Ghana. However, the logic is also potentially applicable to other contexts, such as the relationship between a government minister and interest groups.

[46] The literature on issue attention (Jones and Baumgartner 2005) explores an alternative mechanism through which the necessity of prioritization can affect policy outcomes.

[47] Weingast and Marshall (1988) make this point about barriers to political exchange in legislative logrolling. The theoretical literature that emphasizes the credible commitment problems inherent in clientelism recognizes that commitment problems mean that observed policy outcomes may not stem directly from political actors' first-best targeting strategy (Keefer 2007, Keefer and Vlaicu 2008, Robinson and Verdier 2013). However, this perspective differs from mine in that the commitment problem in question is the mapping of electoral promises onto policy choice, rather than policy choice onto project implementation, and the inefficiency created is in the choice between private and public goods rather than in the implementation of the public goods. More closely related is Robinson and Torvik (2005), which suggests that politicians might deliberately support inefficient projects – "white elephants" – as a form of credible commitment to redistribution to their clients. However, this seems implausible as a general explanation for unfinished projects in the Ghana context, since: 1) most of these projects are social infrastructure rather than projects that generate fiscal returns, on which Robinson and Torvik's model focuses; and 2) there is a near-universal deficit of such social infrastructure in Ghana, so nearly all these projects potentially create positive social surplus (or at least a positive-return project of equivalent cost in the same community could have been identified). It is equally implausible that unfinished projects are simply used by politicians as a way to direct private transfers to communities (as opposed to contractors) in the form of labor income; not only is labor a relatively small share of most capital projects, but a politician could still achieve this aim by seeing a project through to completion and also reap the additional political benefits of having provided a club or public good.

What are the consequences of this commitment problem for project implementation? If politicians and political actors are able to reach and maintain intertemporal bargains, then there are no negative efficiency effects: projects get implemented sequentially according to the plan agreed *ex ante*. (I discuss below the conditions under which such bargains are more or less difficult to reach.) However if the politician and political actors cannot reach such a bargain, then implementation will be erratic, with some new projects started before ongoing ones are completed. Since mid-project delays and interruptions increase the overall cost of completing projects, this inconsistent implementation path leads to some interrupted projects never being finished.[48]

Why might politicians not simply form a stable minimum winning coalition with a subset of political actors? First, note that this may not be tactically optimal for a politician who wishes to win votes (or whose higher-level superiors wish to win votes) from all constituencies under her rather than just a simple majority.[49] Second, there may be uncertainty over the future distribution of political power. While the transactions cost literature on politicians' commitment and credibility problems focuses on political alternation as a mechanism of political uncertainty (Alesina 1988, Dixit, Grossman, and Gul 2000, de Figueiredo 2002), a large literature emphasizes the inherent instability of logrolling coalitions in legislatures; for example, Weingast and Marshall argue that "[p]roblems concerning the durability and enforceability of bargains are ubiquitous in legislative settings" (1988, p.142). This instability is likely to be heightened in the context of developing countries, where formal stabilizing institutions (such as committees) are often weak and informal politics are pervasive. Third, the salience of political actors' claims may be endogenous to project delivery:

[48] An alternative mechanism through which projects could be left incomplete is if politicians and political actors resolve their commitment problems by spreading expenditure thinly across many projects. However, unfinished projects can still emerge in two ways: spreading expenditure thinly will lengthen the completion time of the project beyond what is technically optimal, thus raising the cost and increasing the likelihood that the project will become unprofitable for the contractor; and even if starting many projects simultaneously avoids the need to make explicit sequencing choices in the first period, political actors still cannot commit to finishing all these projects in future periods.

[49] The extent to which this is true may vary across electoral systems, although a full discussion of this is beyond the scope of this paper.

if two communities are lobbying for a project to be built, starting a project in one community is likely to increase grievances and spur mobilization in the other community while decreasing mobilization in the first community.

On the other side of the coin, although theories of repeated games suggest many factors are likely to influence whether political actors can sustain intertemporal exchange, here I will focus on intra-party mechanisms that are proposed by the literatures on legislative bargaining and fiscal decentralization. Weingast and Marshall (1988) propose that parties may be able to enforce trades by offering rewards to individual representatives for cooperation or punishments for non-cooperation. Likewise, Rodden's (2006) analysis of soft budget constraints in fiscal federalism argues that partisan links between local and central governments provide the most likely avenue through which an inefficient over-borrowing equilibrium can be avoided. Partisan ties may also effectively increase the benefits to political actors of projects that do not directly benefit their constituency either due to perceived collective electoral benefits or affective ties. Polities where politicians and political actors are linked by partisan ties are thus more likely to be able to achieve a cooperative implementation outcome.

Fiscal institutions as a second-best solution to bargaining failure

The literature on political control of the bureaucracy suggests two potential solutions to the commitment problem described above: delegation to an independent bureaucrat, and the adoption of rigid rules to constrain future policy choices. However, these options are less applicable to developing countries whose governments tend to be characterized both by low bureaucratic capacity and the weak enforcement of laws and administrative procedures.[50] Project implementation also poses a significant challenge for public financial management systems, since it requires the mobilization of significant resource sums delivered in a timely fashion according to a pre-specified plan, budget, or contract. In contrast, budget systems in developing countries are frequently marked by the weak execution of budgets and medium-term plans, unpredictable and delayed fund releases, and a general gap

[50] For example, Huber and McCarty (2004) demonstrate that many standard results of the delegation literature no longer hold when bureaucratic capacity is low.

between *de jure* structures and processes and their *de facto* operation (Andrews 2010; de Renzio, Andrews, and Mills 2011; Wehner and de Renzio 2013). Contract enforcement by the judiciary in these settings is also likely to be slow, inconsistent, and expensive (especially relative to the cost of small infrastructure projects), thus removing another potential institutional remedy for instability. These features of many developing country polities allow for a significant degree of discretion in resource allocation even after plans, budgets, and contracts have been approved and signed.

In this context, the fiscal institutions used to fund projects provide an alternative potential bureaucratic remedy. Sub-national governments in developing countries are usually highly dependent on inter-governmental transfers (Gadenne and Singhal 2014), and these transfer schemes may impose conditions not only on what type of expenditures they can fund but also on their management and performance. One such condition that is of particular theoretical interest here is a requirement that recipient governments budget to finish existing projects before starting new ones (in the next section I show that this institution has an empirical counterpart in Ghana), with the threat of some sort of punishment (such as withholding future funds). This condition could induce cooperation among local political actors by in effect raising the cost of non-cooperation.

Indeed, the macro-historical literature on the development of state capacity emphasizes that efficiency, distributional considerations, and institutional arrangements have to be analyzed jointly, and that strong institutions facilitate welfare-improving inter-temporal bargains by mitigating credible commitment problems (Acemoglu, Johnson, and Robinson 2005; Besley and Persson 2009). But whereas this literature treats the emergence of institutions as endogenous to distributive preferences, in many polities certain aspects of project financing and delivery mechanisms are imposed in ways that are at least partly exogenous. This is most obvious in cases where foreign donors supply resources to recipient governments but impose conditions on their use (for example review of procurement processes or auditing requirements). However, such supervisory and control mechanisms can also be imposed by central government on lower levels of government in federal or decentralized systems.

To the extent that fiscal institutions can indeed induce cooperation among local political actors there are major implications for the ability of governments to improve public service delivery, since altering the operation of one fiscal institution may well be more feasible than improving the operation of dozens or even hundreds of local-level bureaucratic and political systems. The following section operationalizes this theoretical discussion in the context of infrastructure project provision in Ghana's local governments and develops specific hypotheses for empirical analysis.

3.2 INFRASTRUCTURE PROVISION IN GHANA'S LOCAL GOVERNMENTS

Ghana's local governments provide an ideal setting in which to examine these dynamics empirically. Ghana is a competitive democracy in which public good provision is highly salient and voters have been shown to respond to the delivery of visible public goods such as the ones I study (Briggs 2014, Weghorst and Lindberg 2013, Harding 2015). Yet citizens and voters alike frequently complain about projects being abandoned mid-construction.[51] National elections are closely contested and dominated by two stable parties, the National Democratic Congress (NDC, ruling since 2008) and the National Patriotic Party (NPP). The President is elected by national popular vote, so there is a strong incentive for the incumbent to try to deliver public goods in order to win votes and increase partisan turnout across all districts.

At the political level, districts are headed by a District Chief Executive (DCE) who is appointed by the President and presides over a directly elected non-partisan district assembly, in which 70 percent of members are elected from geographically defined single-member electoral areas and 30 percent are appointed by the President.[52] DCEs do not have a fixed term and so would expect to serve until their party loses power or

[51] For example, unfinished projects are widely reported on by the media (e.g. Ghana News Agency 2014), and action against them has been pledged both by Parliament (Citi FM 2014) and the President (Joy News 2014).

[52] For convenience and clarity I refer generically to DCEs, but in Municipal Assemblies this position is known as Municipal Chief Executive, and likewise Metropolitan Chief Executives in Metropolitan Assemblies.

they are removed. They are also downwardly accountable to assemblies, only weakly in a formal sense (through a requirement that their appointment be approved by a two-thirds majority of the assembly) but more strongly through informal means such as the threat of protest, which could (and occasionally does) lead to removal. Although district-level politics are formally non-partisan, district assemblies and DCEs are widely perceived in Ghana as politicized, partisan actors. For example, political parties often lend informal support to candidates during district assembly elections (GNA 2015), DCEs are often accused of unduly politicizing their roles (e.g. Today 2015), and a nationwide study by Ghana's National Commission on Civic Education found that 68.6 percent of respondents believed that district assemblies were partisan in their operation (NCCE 2015).

Presidential elections were held in 2012 and won by the incumbent party. Assembly members are elected to four-year terms in elections that occur simultaneously nationwide; these took place in 2010, and so the composition of assemblies was constant throughout this paper's study period. District bureaucracies are composed of professional public servants, including dedicated staff for planning, engineering, and finance. [53] Posting and transfer decisions are made centrally in Accra without formal input from the DCE or assembly, and staff are commonly rotated and posted to districts outside their home regions. Project selection decisions are made jointly by DCEs and assemblies in an annual budgeting process in which the DCE has a significant amount of formal and informal agenda-setting power [54], and while implementation is formally the responsibility of the DCE and the district bureaucracy, assembly members nevertheless have a significant influence on budget execution and project implementation through periodic assembly meetings during the year, as well as informal monitoring and lobbying.

Districts are primarily responsible for delivering a variety of public goods and services, including facilities for basic education, sanitation, construction and maintenance of non-trunk roads, primary care clinics, local economic infrastructure

[53] Districts can also use their internally generated funds (but not DACF or DDF transfers) to hire their own staff, but tend to do so for areas such as tax collection rather than infrastructure development.

[54] DCEs chair the important Executive Committee of the assembly and have a large degree of *de facto* discretion over district assembly affairs.

such as marketplaces, and the provision of housing and offices for local government staff. Provision of these facilities is highly politically salient, especially for members of the district assembly. While districts are restricted to relatively small projects that fall below a statutory size threshold, total local government expenditure on capital projects is nonetheless substantial: GH 317 million (equivalent to just over US $135 million) in 2013, or 42.5% of their total revenue.[55] However, these resources are still only enough to deliver projects to a fraction of the communities in any district at a given time.

For example, Kwabre East District Assembly spent GH 1.36 million on capital projects in 2013 (just above the national median of GH 1.24 million), and was comprised of 42 geographically distinct settlements represented by 31 elected assembly members. In the same year, the mean cost for a six-room classroom block (the modal project type) was GH 169,909, meaning that even if the district spent its infrastructure budget entirely on classroom blocks, it could still only deliver eight of them within the year – so roughly two-thirds of its assembly members and over four-fifths of its communities would not receive anything. A district could spread projects somewhat more widely by choosing smaller projects, but this nonetheless illustrates the necessity of prioritization and sequencing of project delivery by district governments.

This empirical context closely matches the theoretical situation described in the previous section: the delivery of locally targeted projects to single-member constituencies must be sequenced, but the representatives have no way to commit to stages of this sequence beyond the present year (or even within-year period). Since forming stable logrolling coalitions may be difficult, this could lead to the assembly adopting an erratic policy implementation path that leaves many projects half-constructed. While DCEs can use their agenda-setting power to stabilize these coalitions and enforce an agreed-upon project sequence, their ability to do this is conditioned by the underlying partisan sympathies of the assembly: DCEs are likely

---

[55] Schedule 3 of the Public Procurement Act (2003, Act 663) fixed this threshold at GHS 200,000, or approximately USD 85,000. Calculations based on an exchange rate of USD 1 = GHS 2.35, as at 31 December, 2013.

to be better able to reward cooperation or punish defection and commit to do so more effectively when dealing with their co-partisans. Likewise, the project selection preferences of the DCE are more likely to align with the collective choices of the assembly (either for instrumental or affective reasons) when the assembly is dominated by members sympathetic to the ruling party rather than to the opposition; to the (considerable) extent that DCEs have discretion in implementing decisions of the assembly, this preference alignment may also stabilize project implementation paths.

This context also provides an opportunity to understand how fiscal institutions influence project delivery and these distributive dynamics. Districts are heavily reliant on inter-governmental transfers funded by central government and donors. The two largest are the District Assemblies Common Fund (DACF) and District Development Facility (DDF), both of which are periodic, lump sum, formula-based transfers to districts (similar to block grants in a federal system).[56] For both transfer programs, project selection, procurement, and implementation are entirely conducted at the district level, by the same politicians and bureaucrats and through the same medium-term planning and annual budgeting process. Both are restricted to use on capital investments rather than recurrent expenditures or private transfers[57], both are financially audited by the same central government auditors as part of the routine

[56] The DACF allocation formula is approved every year by Parliament and changes frequently. While there is evidence that this formula has previously been manipulated to target funds to favored constituencies (Banful 2011), this should affect the number of projects completed or undertaken by districts, not the completion rate, and thus is not a concern for this study. The DDF allocation formula is based in part on an annual assessment of compliance with statutory regulations and administrative processes. Districts that perform better receive incrementally higher allocations, while districts that do not meet the minimum requirements do not receive funds for investment in that year's allocation. However, failure to meet minimum requirements has been rare after first year (prior to this sample) and the disbursements are made with a two-year lag from the assessment year, and as with the DACF any cross-district differences in quantity of funds received should affect the level rather than rate of project completion. For both fund sources, funds are disbursed to all districts simultaneously so these delays affect all districts equally, and there is no evidence of any differential delay in releases or manipulation thereof.

[57] Despite this restriction, districts occasionally spend DACF or DDF funds on non-investment activities, but these irregularities are reported in audit reports and are relatively infrequent.

annual district audit process, and both are administered by small central government secretariats to which annual budgets and reports are submitted but which play no direct role in implementation. But whereas the DACF is enshrined in the 1992 Constitution as a minimum percentage of central government revenue to be transferred to districts, the DDF was created in 2008-09 as a multi-donor funding pool (with the central government also providing minority co-financing).[58] As a result, the DACF imposes little operational control or oversight on districts' use of funds (other than spending them on capital projects), while the DDF requires that districts budget to complete existing DDF projects before starting new ones.[59]

I also examine a third major source of project funding, the Ghana Education Trust Fund (GETFund), which differs from the DACF and DDF in that districts propose a list of projects from which a central government committee selects some to fund. Unlike the DACF and DDF, GETFund resources are not transferred directly to districts but are used centrally to pay contractors directly for work done in the districts.[60] While I suggest that the central government also faces distributive pressures that could undermine implementation, I cannot test this directly because (unlike with the DACF and DDF) the different set of bureaucratic and fiscal procedures associated with GETFund projects make it impossible to isolate this effect. Nonetheless the completion of GETFund projects provides an indication of the priorities and preferences of the central government which will be useful later, so I include it in the analysis. These three major fund sources together comprise 48.4 percent of all projects.[61]

---

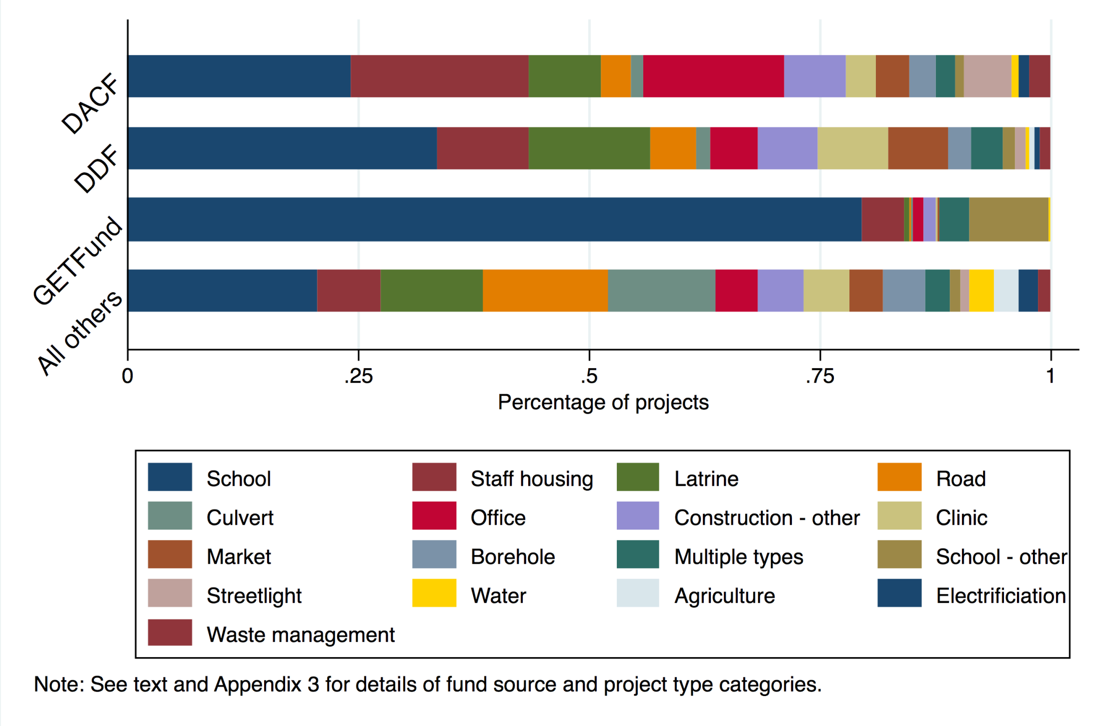
[58] The other significant difference between the DACF and DDF is that, due to aggregate fiscal constraints, the Ministry of Finance is often delayed in disbursing DACF funds by six months or more. For example, the third quarter 2013 funds were actually disbursed in the second quarter of 2014 (Kunateh 2014). In contrast, the DDF is disbursed to districts promptly and in full. However, I show evidence later that these aggregate budget release delays are not driving the differences in completion across these sources.

[59] This is enforced through a combination of *ex ante* budget scrutiny and *ex post* physical monitoring of a small number of projects, although discovery through both means is far from perfect.

[60] There is no formal rule specifying how GETFund projects or funds should be distributed across districts.

[61] The remainder is comprised of projects from districts' internally generated funds, projects being implemented in the district by the central government or donors (sometimes in partnership with the

Figure 2: Project Types by Fund Source



This overlapping set of project finance and implementation mechanisms gives rise to two hypotheses:

> *H1: The fund source through which a project is financed influences its probability of completion.*

> *H2: The completion rate of DDF projects is higher than the completion rate of DACF projects. The relative completion rate of GETFund projects is indeterminate.*

I argue in the following section that estimated differences across fund sources should be interpreted as mostly causal, rather than being primarily due to endogenous sorting on unobserved project characteristics.

---

district assembly), Ghana Road Fund projects, projects from the Member of Parliament's small share of the DACF allocation, and Urban Development Grant projects. Project fund source is not reported for approximately one-fifth of projects; since half of these are likely to be DACF, DDF, or GETFund projects, the estimated effects for these fund sources will be biased somewhat towards zero.

However, H1 and H2 focus on whether fiscal institutions impact project completion, not whether they do so through the theoretical mechanism I propose. Directly testing the prediction that district political alignment in positively related to completion rates is problematic, because estimating this relationship across districts introduces potential confounds due to observable or unobserved characteristics of districts that are correlated both with project completion rates and with district partisan alignment. For example, politicians or bureaucrats in districts that support the ruling party may be of different quality, have different affective preferences, or face different career concerns and private incentives. Instead, I examine this mechanism indirectly, by examining how district partisan alignment interacts with the effect of fiscal institutions on project completion.

> *H3: The difference in completion rates between DACF and DDF projects is decreasing in the ruling party's vote share in the district.*

Intuitively, project fund source should matter less in districts with a high ruling party vote share[62] because the DCE and assembly members are more likely to be able to sustain cooperation without this institutional commitment mechanism, so the requirement imposed by the DDF to finish existing projects before starting new ones may be redundant. In districts where a large part of the district assembly is likely to be sympathetic to the opposition, however, institutions that mitigate the impact of intertemporal bargaining failures among political actors may have an especially strong effect.

[62] The ruling party's vote share from the 2008 presidential elections was chosen because these elections preceded district level elections in 2010 as well as the entire sample period of 2011-13 and thus serve as an indicator of the underlying affiliation of the district, as opposed to the results of the December 2012 elections which could have been endogenous to public good delivery in 2011-12. Because Ghana has a strongly two-party dominant system, using ruling party vote share as a measure of political competitiveness is essentially identical to other measures, such as the winning margin or distance between the NDC and the main opposition party, the New Patriotic Party. The NDC and NPP were the top two parties in every constituency nationwide in the 2008 election and there were only a small number of constituencies where any third party obtained a significant number of votes, so these measures are all extremely highly correlated.

3.3 DATA

To measure the completion of infrastructure projects, I collected, digitized, and coded all available district Annual Progress Reports (APRs) for the years 2011-13. These reports are written annually by the district bureaucracy and submitted to the National Development Planning Commission.[63] Each report includes a table listing basic information about projects that were ongoing or active during the calendar year. Appendix 3.1 contains a sample page from one of these tables. However, APRs were not available for all district-years, and some reports had insufficient information on infrastructure projects; altogether it was possible to locate 479 out of a potential 602 APRs (79.6 percent), of which 407 (67.6 percent) had sufficient information to be entered into the database. While this level of coverage is perhaps surprisingly high for a previously unused source of administrative data, it nonetheless raises concerns that reporting completeness may be correlated with project completion. This does not appear to be the case, however, as project completion rates are balanced across districts' reporting completeness, as are a wide range of district characteristics (see Appendix 3.2). After cleaning and removing non-infrastructure projects, this yielded a database of 14,246 projects, which is the core dataset used in this study's analysis.

All reports were manually double-entered into the database and variables were coded algorithmically from the resulting text strings, with manual disambiguation for key variables (see Appendix 3.3). Projects were coded into 17 different basic "types", such as schools, clinics, and staff housing. For the parametric analysis in Sections 3.4 and 3.5, the category "schools" was broken down into five sub-categories according to the number of units in each school block, leaving a total of 22 project type categories. Project completion was coded as a binary variable by combining information from three raw variables. For instance, project completion was coded as 1 for values such as "COMPLETE", "100%", or "INSTALLED AND IN USE", and 0 for values such as "ONGOING", "90%", or "LINTEL LEVEL". This yielded a unique completion code for 91.6 percent of observations; the remainder were

[63] In practice, however, some districts do not prepare or submit APRs, while others submit them to the Regional Coordinating Council (RCC) but not the NDPC. To collect the APRs, I first worked with staff of the NDPC to locate all the soft and hard copy reports that were available centrally at the NDPC, then collected all extant APRs that were not available centrally from the ten regional capitals.

disambiguated by visual inspection if possible, and given a missing value if it was impossible to determine the project's status conclusively. Since many projects did not have commencement dates reported and few districts made use of unique project identification numbers, tracking the progress of projects across years was challenging and there is significant attrition. Appendix 3.4 discusses the three alternative methodologies that were used to track project progress across years and their likely directions of bias – the three methods yield qualitatively similar findings – and shows that project attrition is uncorrelated with fund source and thus poses little risk of bias to this paper's key empirical results.

The information in the project database is based on the self-reports of the local governments themselves and is thus subject to some concern about accuracy, either due to poor reporting quality or deliberate misrepresentation.[64] This study has therefore taken multiple approaches to probing the accuracy of the APR data. First, it should be noted that districts have had little incentive to deliberately misreport any data in the APR, because these reports were submitted in hard copy and subject to only cursory scrutiny by supervisory institutions, so careful and systematic analysis of them was not possible.[65] Data quality and reporting honesty are most likely to vary at the district level, but since this paper's analysis focuses mainly on within-district variation, misreporting bias would only be a problem if it were differential across fund sources. District officials would have no incentive to do so, however, since the APRs are not submitted directly to any of the funding institutions and were not being used by these institutions for monitoring purposes. Second, for one of the project fund sources the government maintains its own monitoring database which is compiled and

---

[64] Note however that if district APRs deliberately misreporting completion rates to exaggerate their achievements the completion rate would be biased upward, meaning that unfinished projects are even more prevalent than estimated here.

[65] Interviews with officers at these institutions have confirmed this, and have not yielded any anecdotal evidence of district-level officers being subject to any form of sanction or reproach due to information contained in the APRs. There were no concerns on the part of any of these officers that completion data on the APRs was being deliberately exaggerated, although some acknowledged that some of the finer details (such as the percentage of work completed) may be somewhat rough. Scrutiny of district operations is somewhat more intense in terms of financial management due to the Ghana Audit Service's annual audits, but these are not based on the APRs and if anything provide an incentive for district-level officers *not* to misreport the financial status of projects on APRs.

maintained independently of the APR reporting system. While this database differs somewhat in structure, format, and scope from the APR database, the estimates of project completion rates are similar in magnitude.[66]

Finally, I conducted physical site visits to a small sub-sample of projects that had been reported as complete in 2013, spread across four randomly selected districts in two regions. The physical evidence from this limited sample suggests that while construction quality and maintenance may be issues with projects reported as complete, there is little evidence that districts reports of completion status are systematically biased.[67]

3.4 FUND SOURCE MATTERS FOR PROJECT COMPLETION

This section first presents key descriptive statistics on the extent and dynamics of project non-completion in Ghana. Although anecdotes about unfinished infrastructure

[66] The Ministry of Education's internal monitoring database of 1,146 GETFund projects reports that of 6-unit classrooms and dormitories started between 2009 and 2013 nationwide, only 36.6 percent had been completed. (The GETFund is described in more detail in Section 4.1). Unfortunately it is not possible to disaggregate this by year of project commencement, and the date of reporting is not indicated (these figures are based on a database provided by the Ministry of Education in January 2015). In addition, the Ghana Audit Service reports that a June 2013 monitoring effort of 179 school projects in seven regions started in 2010 and 2011 found that 27 percent were complete, despite scheduled completion times of six to twelve months – a similar length to most GETFund projects in the APR database (Ghana Audit Service 2014, 290). While these estimates differ slightly in timespan and project coverage, they are in the same range as APR database estimates that GETFund projects have one-year completion rates of 24.0-25.4 percent and three-year completion rates of 32.0-56.1 percent. Likewise, the June 2013 monitoring report found that 61 percent of the contract value of these incomplete projects had been disbursed; the equivalent figure in the APR database is 48.4 percent.

[67] Resource constraints made it impossible to conduct physical site visits on a nationally representative scale. Seventeen of the twenty projects were fully complete, while the remainder were functionally complete but with minor areas of incompleteness (e.g. no windows, untiled floors, holes in roof, some roofing remaining to be done). Sixteen of the projects were in full use; of the remaining four, one was in partial use, one was out of use because of cracks and accessibility issues, and one had not been commissioned yet. The site visits were conducted in October 2014, ten months after the project had been reported as complete; in only one case did people present at the project site report that the project had actually been finished in 2014 rather than 2013.

projects in Ghana and other developing countries are widespread, empirical data is rare. For scholars of public good provision, measuring what percentage of government projects get finished is as fundamental a task as measuring poverty headcounts is for studies of poverty reduction, so this ignorance is worrying.[68] I estimate that approximately one-third of projects are never finished, and show that non-completion patterns are qualitatively consistent with this paper's theoretical mechanism that emphasizes how time consistency problems on the part of the government interact with the temporal dimension of project implementation.

I then take advantage of the overlap between fund sources and project types in Ghana's local governments to test H1 and H2, regarding the relationship between project financing and delivery mechanisms and completion rates. Since the same types of projects are financed and delivered through different fund sources within the same districts and narrowly defined geographic communities, and sometimes even by the same contractors, it is possible to control for a broad range of observable and unobservable factors that can influence project completion. While the causal effects of fund sources are not cleanly identified, the range of controls is so extensive that the estimated differences among fund sources are difficult to attribute to anything but the effect of the fund sources. I examine the potential for endogenous sorting of projects into fund sources on unobservable project or community characteristics using coefficient stability tests and other methods; while the possibility of sorting cannot be entirely ruled out, I argue it is unlikely to account for more than a small fraction of the observed differences among fund sources.

The dynamics of project completion

Less than half of projects (45.8 percent) finish within their first year, and even after three years of implementation 35.5 percent remain unfinished. The completion hazard rate decreases monotonically after the first year of a project, so projects that are not finished promptly become decreasingly likely to be completed.[69] These patterns hold

[68] What little data there is, however, confirms the severity of the problem: in the only other such large-N dataset available in a developing country (to my knowledge), Rasul and Rogger (2014) estimate that only 31 percent of small government infrastructure projects are fully completed within the budget year.

[69] Tracking project completion across years is complicated by attrition and missing data; I discuss my methodology for doing this in Appendix 3.4. I also present upper- and lower-bound estimates for these

across different types of project, as Figure 3 shows for the three most common project types (schools, latrines, and housing for local government staff).[70] Likewise, completion rates remain broadly consistent across years – annual project completion rates were 48.6 percent for 2011, 40.7 percent for 2012, and 44.7 percent for 2013 – indicating that the phenomenon of unfinished projects is not primarily explained by pre- or post-election dynamics (i.e. politicians starting or finishing projects *en masse* just before or after an election) relating to Ghana's December 2012 election.

These findings could emerge normally if many projects were scheduled to last longer than three years, but this is not the case. Most projects are intended to be brief: five months for the median, and 88.8 percent less than twelve months. Projects that get finished tend to be punctual: the median delay is just one month, 81.0 percent are finished in less than twelve months, and just 10.7 percent are delayed by a year or more. The median incomplete project, however, is twelve months past its expected completion date – a 200 percent delay – and there is a long tail of over a quarter of incomplete projects that are more than two years past their planned completion date.[71] While it is impossible to conclusively say that these projects will never be completed, just 5.3 percent of completed projects were finished two or more years past their due date, so it is empirically improbable. It is only a slight exaggeration to say that projects either finish promptly or not at all. This bifurcation of project outcomes is consistent with the mechanism from the theoretical model: due to the non-fungibility of project expenditure across time, mid-project delays or interruptions increase the cost of finishing the project for engineering, financial, and contractual reasons, and thus make it unlikely the government will return to complete the project.

---

completion rates under alternative sets of assumptions, which produce quantitatively different estimates but do not change the main patterns identified.

[70] It is noteworthy that project incompletion is an issue even for types of projects, such as staff housing and office buildings, for which the staff of the districts are themselves the direct beneficiaries.

[71] These figures likely *understate* the extent to which incomplete projects are behind schedule, since many incomplete projects attrite out of the dataset.

Figure 3: Completion Rate by Project Year

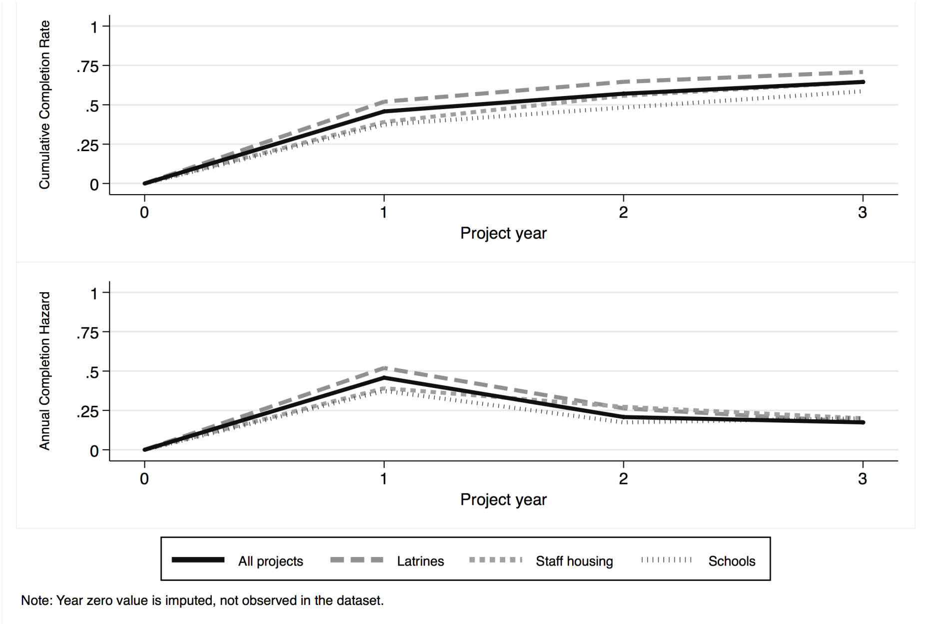


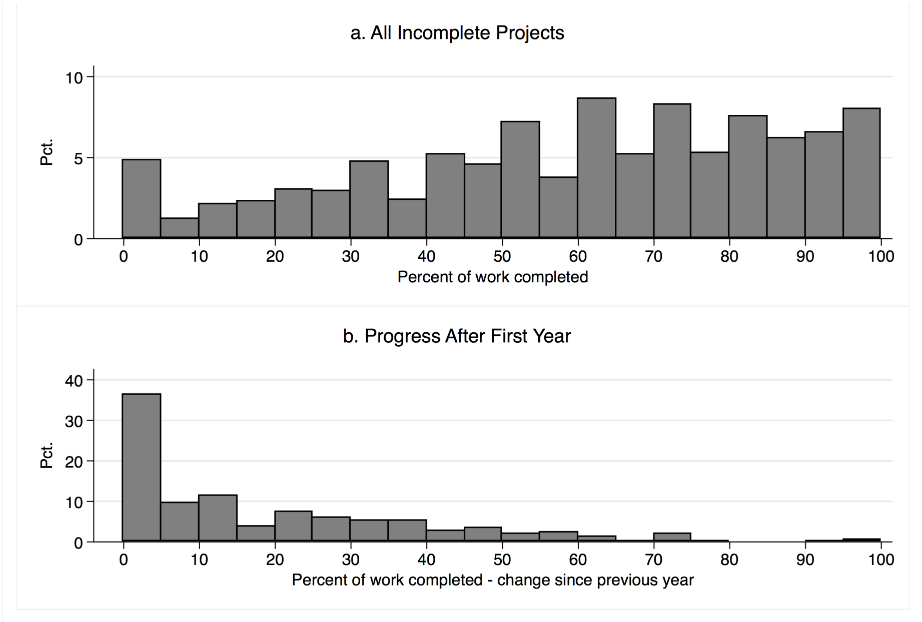
Figure 4(a) shows most incomplete projects have had a significant amount of work done on them – 60 percent, for the median. It is thus not the case that the phenomenon of unfinished projects is due merely to costless signaling of distributive intentions by politicians using projects that are never intended to be completed. Further consistent with this paper's theoretical mechanism, nearly half of projects are not touched after their first year (4b).[72] This is consistent with unfinished projects being due largely to intertemporal consistency problems combined with the multi-stage nature of the construction process, rather than a "smoother" underlying completion function which would be more suggestive of low capacity and resource constraints as driving factors.

Of course, these patterns are also consistent with another potential cause of project non-completion: corruption and campaign finance. If politicians' (or bureaucrats') objective is to embezzle funds, they may direct contracts to favored contractors (who may pay kickbacks) who are then placed under little pressure to complete the project. In addition, contractors themselves may be important clients of politicians and

[72] Reported progress is precisely zero for 32.7 percent of projects, but is 10 percent or less for 55.0 percent. Given that the percent of work complete variable may be reported somewhat imprecisely, reported progress in this range is more likely to be due to measurement error than genuine work completed.

rewarding them with contracts which they are not expected to complete may be a clientelist strategy. Finally, politicians and bureaucrats could embezzle the funds directly by reporting that they had been disbursed to contractors but instead taking them themselves.[73]

Figure 4: Work Done on Incomplete Projects



To examine this directly, I construct an analog of Olken's (2007) "missing expenditures" measure by subtracting the percentage of physical work complete on a project from the percentage of its budget that has been disbursed relative to the original contract value.[74] Unlike Olken's measure, however, this measure is two-

[73] Samuels (2002) makes this argument in the Brazilian case, and Luna (2015) presents evidence that similar relationships exist among politicians, bureaucrats, and contractors in Ghana's local governments. However,

[74] This measure is an analog to Olken's missing expenditures measure, not an exact replication. Whereas Olken's measure focuses on embezzled funds by measuring sub-standard material quality in completed roads, this paper's measure focuses exclusively on percentage physical completion (regardless of quality). These "quality" and "quantity" margins are not mutually exclusive. The extent to which these margins are interactive – if theft of resources on the quality margin increases the cost of finishing a project, thus leading to its abandonment – would be an interesting question for future research.

sided: contractors can be either overpaid or underpaid relative to the work they have done. For complete projects (5a), positive values thus represent cost overruns, while for incomplete projects (5b) they represent "missing expenditures".[75] In both cases, negative values represent delayed payment or non-payment for work done. If corruption were the cause of unfinished projects, we would expect to observe positive values: a 60 percent complete project with 100 percent of its budget disbursed, for example, might suggest that the remaining 40 percent had been embezzled by one party to the transaction. Instead, the opposite appears to be the case. Defining overpayment as projects with values greater than +10 percent and underpayment as values less than -10 percent, underpayment is almost five times more common (44.2 percent of projects) than overpayment (8.1 percent).[76] Overpayment ("missing expenditures") is slightly more common on incomplete projects than completed ones (15.8 percent, compared to 2.4 percent), but underpayment is still several times more frequent (50.9 percent of incomplete projects, 39.2 percent of complete projects).[77]

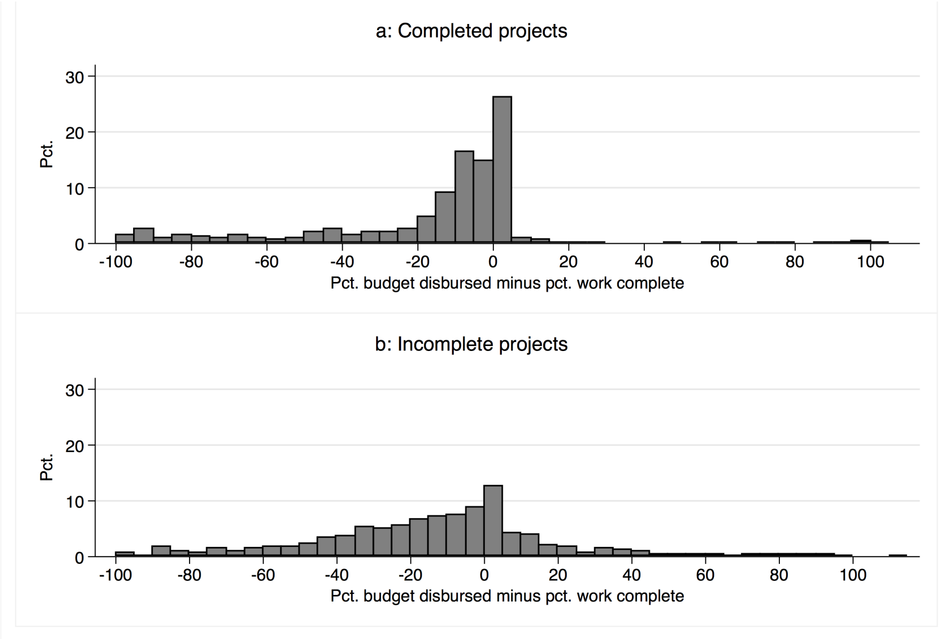
---

[75] As Olken emphasizes, "missing expenditures" can arise for reasons other than simple theft, even if they are highly suggestive of some form of either corruption or extremely poor contract management (2007, 203).

[76] The observed differences between over- and underpayment are far too great to be explained by normal time lags in processing payment. Governments usually have a contractual period of a few weeks or months after contractors have submitted payment requests in which to inspect work and make payment. However, there is no correlation between the over/underpayment measure and the month of project completion (not shown) – since both financial and physical status is reported in the APRs at the end of the year, this correlation would be positive if underpayment was primarily a short-term phenomenon. Thus, negative values of "missing expenditure" really do seem to be measuring severely delayed payment or non-payment rather than normal lags in processing payment.

[77] Contractors in Ghana frequently complain about delayed payment and non-payment for work by Government; there are reports of contractors going unpaid for periods of a year or more for work done (Abotsi 2013). This not only affects the completion probability of the project in question (especially if the delay increases the nominal or real cost of finishing the project), but also can reduce the completion probability of other projects by increasing the cost of tender bids as contractors build the interest costs of delayed payment into their estimates, and reducing contractors' financial capacity. Indeed, one seminar participant with extensive infrastructure experience in developing countries referred to the "contractor death spiral" that frequently occurs when relatively small contractors in countries with weak financial systems encounter payment delays.

Figure 5: Cost Overruns, Missing Expenditures, and Underpayment for Work



The total cost to society of unfinished projects is economically significant, whether measured in terms of the direct fiscal cost to government, the opportunity cost of foregone public goods, or the financial costs of non-payment to contractors. Since 35.5 percent of projects are unfinished after three years and expenditure on these averages 55.5 of the contract value, 19.6 percent of total local government capital expenditure – roughly one-fifth – is spent on projects which are almost certain never to be finished. Scaling this percentage by the total capital expenditure of Ghana's local governments in 2013 of USD 135 million implies that annual spending on projects abandoned mid-construction is approximately USD 26.6 million.[78] This wasted expenditure would be enough to fully construct 706 three-room school buildings *per year*. In addition to these costs borne by government and service users,

[78] Calculations based on an exchange rate of USD 1 = GHS 2.35, as at 31 December, 2013. These figures are indicative, since the expenditure on these projects would have spanned the period 2011-13 and thus could have been somewhat higher or lower depending on exchange rates fluctuations, the temporal incidence of expenditures, and changes in aggregate spending. Likewise, the set of project fund sources included in the overall local government capital expenditure figures is slightly different than the set of fund sources included in the APR database, but it is not possible to disaggregate them by fund source to arrive at a more precise estimate.

the underpayment of contractors implies an additional annual economic cost of USD 6.7 million borne by contractors.[79]

Project completion across fund sources

How are these completion patterns affected by the way in which projects are funded, and the different delivery mechanisms associated with these fund sources? This question bears directly on H1 and H2:

> *H1: Projects funded from different sources have different probabilities of completion.*
>
> *H2: The completion rate of DDF projects is higher than the completion rate of DACF projects. The relative completion rate of GETFund projects is indeterminate.*

Both hypotheses find support in simple comparisons of completion rates. Figure 6 shows the three-year completion rates of projects for the main three fund sources, both for all projects (6a) and for school buildings only (6b). The differences across fund sources are substantial: after three years 78.5 percent of DDF projects are completed, compared to just 64.0 percent of DACF projects and just 44.8 percent of GETFund projects. The same pattern emerges even when restricting the sample to school buildings only. The differences between DDF and DACF projects are particularly striking, since these projects are planned, procured, and executed by the exact same district-level bureaucrats and politicians, thus controlling for the quality, composition, and management of politicians and bureaucrats.

To estimate these differences parametrically while controlling for a broader range of other factors, I estimate a linear probability model of the following form:

$$(1) \qquad y_{i,j,k,t} = \boldsymbol{F_{i,j,k}}\boldsymbol{\beta} + \boldsymbol{P_{i,j,k}}\boldsymbol{\gamma} + \lambda_j + \delta_k + \nu_t + \epsilon_{i,j,k,t}$$

[79] The average physical completion percentage of projects that are unfinished after their third year is 64.9 percent, or 9.4 percentage points greater than the percent of the budget disbursed. Scaling this by total local government infrastructure expenditure implies an annual underpayment of USD 6.7 million

where $y_{i,j,k,t}$ represents the completion status of project *i* in district *j* of type *k* in year t,[80] $\boldsymbol{F_{i,j,k}}$ is a vector of fund source indicator variables, $\boldsymbol{P_{i,j,k}}$ is a vector of project characteristics including construction type (i.e. greenfield projects vs. maintenance/repair) and project year (first, second, or third)[81], λ_j is a vector of district fixed effects, δ_k is a vector of project type fixed effects[82], ν_t is a vector of year fixed effects, $\epsilon_{i,j,k,t}$ is an error term, and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are vectors of parameters to be estimated. The coefficients $\boldsymbol{\beta}$ - specifically, the differences between them – are the parameters of primary analytical interest.[83]

---

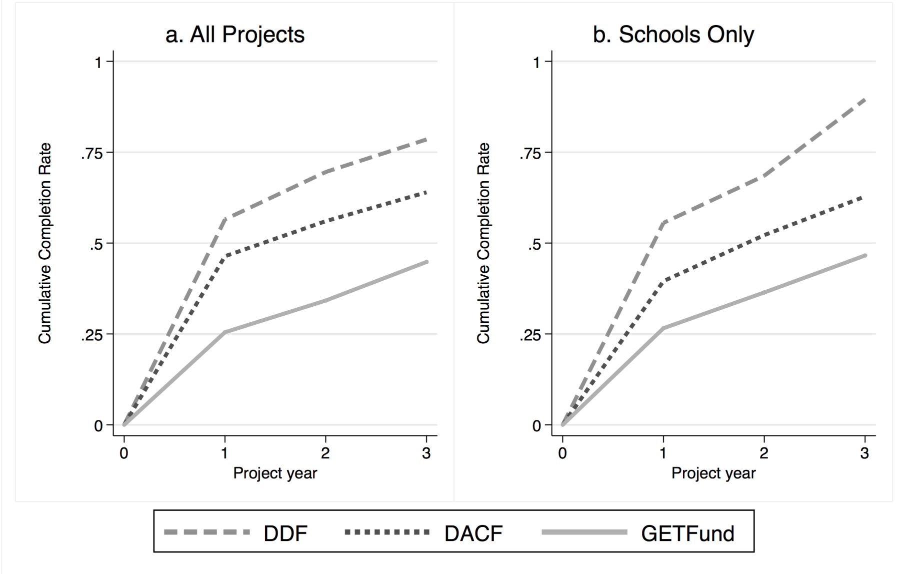
[80] Using the annual completion status of a project does not take into account whether the project is a first or subsequent year project (although I control for this using indicator variables in most specifications). The annual completion status has the disadvantage of differing from the eventual cumulative completion status of a project, and is potentially vulnerable to cross-district bias if attrition differs systematically across districts. However, using three-year completion rates for full sample estimation and cross-district comparisons is not feasible since these cannot be calculated for districts with missing report years or that do not report project commencement dates, or for projects that started in 2012 or 2013. Where it is possible to calculate both, annual completion rates are highly correlated with three-year completion rates across project types and fund sources, and there are no observed instances for any major sub-group (such as fund source or project type) of one sub-group having a higher annual completion rate but lower three-year completion rate than another sub-group. The annual completion rate is thus a robust measure to use both for within- and across-district comparisons.

[81] Project year is coded using the manual linking process; all projects that were not identified as continuations of a previous year's project are coded as year one. This coding scheme was chosen to make use of the entire sample while taking account, where possible, of whether a project was in its first or subsequent year. While there are theoretical concerns about this variable's endogeneity to project completion and potential bias from differential reporting quality or attrition across districts, in practice the inclusion of these variables has little effect on the coefficients of interest or the model as a whole, any cross-district bias would be cancelled out by the district fixed-effects, and attrition is independent of fund source.

[82] Project type is disaggregated into seventeen categories, with schools further disaggregated into 6 categories according to the number of rooms in the school block (with a residual category for schools with an unknown numbers of rooms), for a total of 22 categories. See Appendix 3 for more details.

[83] I use the linear probability model (LPM) due to the large number of fixed effects variables used (thousands, when adding location or contractor fixed effects) which make estimation by logit or probit computationally difficult and potentially biased. Where possible, however, estimating the model as a logit or probit produces similar results. Estimating a Cox proportional hazards model is problematic due to the short time frame and resulting truncation of the data, and the difficulties in linking project

Figure 6: Cumulative Project Completion by Fund Source



The results are presented in Table 1. Column 1 shows simple mean annual completion rates of the three major fund sources, with the constant term being the completion rate of projects from the minor or unknown fund sources. Column 2 adds district and year fixed effects, and Column 3 further adds project type fixed effects and characteristics. The bilateral differences between the three coefficients of interest are highly statistically significant in each case, and also economically significant: the DDF annual completion rate is in the range of 10-12 percentage points greater than that of the DACF, which is in turn 15-20 percentage points greater than that of the GETFund.[84]

---

observations across years discussed in the previous section. Nevertheless, using this estimator generates similar differences across fund sources.

[84] As noted above, these estimates are likely to be biased towards zero and thus understate the true differences among fund sources, since roughly two-fifths of the omitted fund source category is composed of projects with unknown fund source, and this presumably comprises a significant number of DDF, DACF, and GETFund projects.

Table 1: Project Fund Source and Completion

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All projects	All projects	All projects	All projects	Schools only	All projects	Offices & staff housing
DACF	-0.0186	-0.00364	-0.00101	-0.0220	-0.0243	-0.0888	-0.0632
	(0.0409)	(0.0247)	(0.0255)	(0.0429)	(0.0418)	(0.0525)	(0.0454)
DDF	0.0907	0.120**	0.122**	0.0949*	0.111*	-0.0143	0.0173
	(0.0613)	(0.0278)	(0.0287)	(0.0412)	(0.0499)	(0.0509)	(0.0591)
GETFund	-0.218**	-0.201**	-0.146**	-0.112*	-0.108*	-0.181**	-0.163*
	(0.0688)	(0.0365)	(0.0321)	(0.0439)	(0.0418)	(0.0600)	(0.0706)
Classroom block additional facilities							
"Ancillary facilities"					-0.0633*		
					(0.0261)		
Latrine					-0.0198		
					(0.0605)		
Office/ Store/ Library					0.0249		
					(0.0264)		
Constant	0.464**	0.515**	0.408**	0.611**	0.379**	0.823**	0.594**
	(0.0588)	(0.0212)	(0.0600)	(0.145)	(0.0540)	(0.195)	(0.0474)
Coeff. equality tests (prob > F)							
DDF = DACF	0.001	0.000	0.000	0.006	0.001	0.166	0.102
DACF = GETFund	0.000	0.000	0.000	0.056	0.050	0.192	0.109
DDF = GETFund	0.000	0.000	0.000	0.000	0.000	0.017	0.014
Observations	13,339	13,339	13,339	10,420	4,204	8,801	2,066
R-squared	0.031	0.217	0.244	0.699	0.281	0.765	0.306
Fixed effects							
District		Yes	Yes		Yes	Yes	Yes
Year		Yes	Yes	Yes	Yes	Yes	Yes
Project type			Yes	Yes	Yes	Yes	Yes
Location				Yes			
Contractor						Yes	
Project characteristics			Yes	Yes	Yes	Yes	Yes

** p<0.01, * p<0.05

Note: Dependent variable is project completion. Project characteristics are construction type (construction/ maintenance) and project year indicators (based on manual linking). Huber-White robust standard errors clustered at district level in Columns 1-3, 5, and 7, location level in Column 4, and contractor in Column 6.

The specification of Column 3 is already a demanding one, but Columns 4-7 show that the key findings are robust to even narrower comparisons. Column 4 includes fixed effects for the location of projects – the 5,248 unique villages and neighborhoods in which the projects are located – in order to address concerns that projects might be sorted into fund sources according to unobserved within-district heterogeneity across locations. Even after controlling for all the observed and unobserved features of these narrowly defined locations, there is still a gap of 11.7 percentage points between the completion probability of DDF and DACF projects, and another 9.0 percentage points between DACF and GETFund projects. To test whether differences are being driven by within-type variation in the size or complexity of the project, Column 5 restricts the sample to the construction of school buildings only, with number of units in the classroom block controlled for in the project type fixed effects and additional control variables indicating whether the classroom block is being constructed together with various types of additional facilities, such as latrines or offices.[85] This specification leaves the estimated differences between fund sources almost unchanged, and they remain highly significant.

Column 6 estimates the model with fixed effects for the 4,546 unique contractors that executed the projects – the fund source coefficients are thus effectively estimated within the 15.9 percent of contractors that executed projects from more than one fund source.[86] Even with this extremely demanding specification, the differences between the DDF-DACF and DACF-GETFund coefficients are still 7.4 and 9.2 percentage points, respectively, and both are statistically significant at the ten percent level. It is all the more remarkable that these results hold in this specification since the choice of contractor is endogenous to the procurement process, and the selection of qualified contractors (as opposed to less qualified but politically connected or bribe-paying firms) is itself one of the main channels that could affect implementation. Finally, Column 7 shows that the estimated effects remain substantial even when restricting the sample to district administration offices and housing for district staff, both types of projects for which bureaucrats themselves are the primary beneficiaries, although the DACF-DDF and DACF-GETFund coefficient differences are statistically significant only at the ten percent level. As discussed below, the continued significance of fund source differences even for these non-patronage project types suggests that the differences reflect true causal effects of fund source, not merely endogenous sorting of projects into fund sources based on considerations of political economy or clientelism. Appendix 6 shows that these results are further robust to a range of other sample restrictions and control variables.

---

[85] Appendix 3.3 contains details of the definition and coding of these additional facilities variables.

[86] A total of 5,113 unique contractors appear in the APR dataset, as detailed in Appendix 3.3, but the number used to estimate Column 10 of Table 3 is slightly smaller since some contractors implemented only non-infrastructure projects (which are excluded from this paper's analysis) and some projects are missing completion data.

The estimated differences in completion rates across fund sources are economically as well as statistically significant. In the baseline model in Column 3, the estimated gap between DDF projects and GETFund projects is 26.7 percentage points. By way of comparison, the standard deviation of annual completion rates across districts is 18.8 percentage points.[87] This means that a DDF project in the 25th-percentile district has approximately the same likelihood of completion as an equivalent GETFund project in the 70th-percentile district. Given that the cross-district differences include all the measurable and unobserved effects of leadership, district socioeconomic characteristics, political affiliation, and other variables, the magnitude of the estimated within-district differences across fund sources is remarkable.

Endogenous sorting and alternative explanations

The results presented in Table 1 show a significant and robust association between project fund source and project completion but, in the absence of experimental variation or convincing instrumental variables for fund source assignment, these associations are not cleanly identified as causal. In particular, the estimated differences between fund sources could be biased upwards if there is endogenous sorting of projects into fund sources based on unobservable project attributes. Although the scope for this is reduced by the unusually comprehensive set of controls for observables I am able to use, a selection effect could exist if there were some type of unobservable *within-location* variation in community attributes that are correlated with completion probability and according to which projects are endogenously sorted into fund sources. Local ethnic or partisan composition (Besley *et al* 2004; Ichino and Nathan 2013; Nathan 2015) could be one such source of unobserved heterogeneity: politicians might care more about delivering projects to residents in one part of the location and therefore allocate projects for that area into a fund source with a higher completion rate. Part of the estimated difference among fund sources would then be due to these omitted unobservable factors rather than any causal effect of the fund source itself.

[87] These district-level completion rates are calculated for the 59 districts for which three years of APR data are available.

To investigate the potential that such a selection effect could be driving the results, I build on the coefficient stability approach developed by Altonji, Elder, and Taber (2005), Bellows and Miguel (2009), Nunn and Wantchekon (2011), Gonzalez and Miguel (2015), and Oster (2015). The intuition behind this approach is that one way to gauge the potential impact of selection on unobservables is to examine how regression coefficients react to the inclusion of observable control variables; if the inclusion of powerful controls has only a small effect on the coefficient of interest, then selection on unobservables would have to be extremely strong in order to explain away the estimated coefficient. While such an effect cannot be ruled out, this approach can give a sense of how plausible the alternative selection explanation is and thus how robust the estimated coefficient is likely to be.

To examine the potential for project selection into fund sources based on unobservable within-location heterogeneity, I compare the estimated differences across fund sources across two different regression specifications: a "restricted" regression that uses fixed effects at district level, as well as controls for year, project type, and project characteristics (the baseline regression from Table 1, Column 3); and an "unrestricted" regression, which uses fixed effects at the location rather than district level. The decrease in estimated coefficient differences between these two specifications reveals the degree of project selection into fund sources *across* locations – that is, the extent to which projects from "better" fund sources are targeted to "better" locations in the district.[88] The results in Columns 1 and 2 of Table 2 show that moving from district to location controls does decrease the estimated differences among fund sources, but for the DDF-DACF comparison these movements are quite small relative to the effect size: the coefficient difference decreases from 13.2 percentage points to 11.7 in Panel (a) using a common sample, or from 12.3 to 11.7 in Panel (b) that makes use of all observations. The movements in the DACF-GETFund and DDF-GETFund differences are somewhat larger – 13.5 to 9.0 and 26.8 to 20.6 percentage points, respectively. These movements provide an estimate of the degree of selection on observables.

[88] However, a selection process of this nature would only bias the estimated differences upwards, not make them zero, since the rationale for the sorting of projects into fund sources is precisely that some fund sources have higher completion rates than others.

To see how strong selection on unobservable within-location heterogeneity would need to be in order to explain away the estimated effect, I calculate the ratio of the estimated effect (from the "unrestricted" specification with location controls) to the reduction in the effect due to selection on observables. The simple ratio – calculated following Altonji et al (2005), Bellows and Miguel (2009), and Nunn and Wantchekon (2011) – is presented in Column 3.[89] Oster (2015) refers to this ratio as the coefficient of proportionality, δ.[90] It implies that selection of projects into fund sources on within-location unobservable heterogeneity would have to be between 1.627 and 19.964 *times stronger* than the selection across locations in order to explain away the entire estimated differences among fund sources.

However, Oster (2015) and Gonzalez and Miguel (2015) observe that this simple ratio relies on strong implicit assumptions about the covariance properties of the sets of observable and unobservable covariates with the dependent variable, and show that changes in the coefficients between the restricted and unrestricted regressions should be scaled by movements in the R-squared. This correction requires making an assumption about R_{MAX}, the maximum possible percentage of variation in the dependent variable that could be explained by a notional regression with the full set of observable and unobservable explanatory variables; while 1 is an upper bound, R_{MAX} may well be below 1 due to measurement error or, as this paper does, using a linear estimator on a binary dependent variable.[91]

---

[89] The simple ratio in Column 3 is calculated as:

$$\delta^{DDF-DACF} = \frac{\hat{\beta}_U^{DDF} - \hat{\beta}_U^{DACF}}{(\hat{\beta}_R^{DDF} - \hat{\beta}_R^{DACF}) - (\hat{\beta}_U^{DDF} - \hat{\beta}_U^{DACF})}$$

where $\hat{\beta}_U^{DDF}$ is the estimated DDF coefficient from the unrestricted regression, $\hat{\beta}_R^{DDF}$ is the estimated DDF coefficient from the unrestricted regression, and so on.

[90] I follow Oster (2015) in notating this coefficient as δ.

[91] For Columns 4-7, the corrected proportionality coefficient is calculated as:

$$\delta^{DDF-DACF} = \frac{(\hat{\beta}_U^{DDF} - \hat{\beta}_U^{DACF}) * (R_U^2 - R_R^2)}{(\hat{\beta}_R^{DDF} - \hat{\beta}_R^{DACF}) * (R_{MAX} - R_U^2)}$$

where R_U^2 is the R-squared of the unrestricted regression and R_R^2 is the R-squared of the restricted regression.

Table 2: Coefficient stability and potential selection on unobservables

			Coefficient of proportionality δ calculated under different R_{MAX} assumptions				
			No R_{MAX} correction	Reliability ratio	Binary DV correction	Cutoff from randomized data	Most conservative case
	Restricted specification (District FE)	Unrestricted specification (Location FE)	*Altonji et al, Bellows & Miguel, Nunn & Wantch.*	*Gonzalez & Miguel (2015) using Baird et al (2008)*		*Oster (2015)*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel (a): Common sample							
DDF – DACF	0.132	0.117	7.668	3.704	2.122	1.777	1.240
	(0.026)	(0.043)					
DACF – GETFund	0.135	0.09	1.950	2.768	1.585	1.327	0.926
	(0.032)	(0.047)					
DDF - GETFund	0.268	0.206	3.374	3.230	1.850	1.549	1.081
	(0.035)	(0.048)					
R^2	0.278	0.699					
R_{MAX}			-	0.800	0.875	0.909	1.000
Observations	10,420	10,420					
Panel (b): All observations							
DDF – DACF	0.123	0.117	19.964	4.305	2.466	2.065	1.441
	(0.023)	(0.043)					
DACF – GETFund	0.145	0.09	1.627	2.800	1.604	1.343	0.937
	(0.03)	(0.047)					
DDF - GETFund	0.267	0.206	3.389	3.491	1.999	1.674	1.169
	(0.031)	(0.048)					
R^2	0.244	0.699					
R_{MAX}			-	0.800	0.875	0.909	1.000
Observations	13,339	10,420					

Note: Robust standard errors clustered at district (Column 1) or location (Column 2) level. Both the restricted and unrestricted regressions also include year and project type fixed effects and project characteristic controls (see Table 1 for details).

I therefore present four additional sets of proportionality coefficients calculated under different assumptions about R_{MAX}: a value of 0.8 proposed by Gonzalez and Miguel (2015) based on reliability ratios from resurveys in Baird et al (2008), which is a conservative estimate of survey measurement error (Column 4)[92]; a correction I develop for the mechanical error introduced by using a linear estimator on a binary dependent variable (Column 5)[93]; a cutoff proposed by Oster (2015) based on experimental studies where selection is random (Column 6)[94]; and a most conservative case where R_{MAX} = *1* (Column 7). The corrected proportionality coefficients only drop below 1 – the robustness threshold proposed by Oster (2015) – for one comparison, the most conservative case for the DACF–GETFund difference (δ = *0.926* in the common sample, δ = *0.937* in the full sample). For the DDF-DACF comparison that is of most interest, the corrected δ values range from 1.240 to 4.305.

For selection on unobservable within-location heterogeneity to be explaining the differences across fund sources, then, would require that this within-location selection process is from one and four times stronger than the selection process across locations. To further assess the plausibility of this, consider that the within-location

[92] Baird et al (2008) resurveyed the same individuals in a rural Kenyan sample at a three-month interval, and found that the auto-correlation of a simple, objective, time-invariant response (father's education) was in the region of 0.8. Gonzalez and Miguel (2005) therefore propose 0.8 as a best-case R_{MAX} for survey data in a rural African setting, while noting that other high quality data sources (e.g. US labor market surveys) also have reliability ratios in the range of 0.7 – 0.9, and that in Africa many datasets on economic outcomes exhibit autocorrelations as low as 0.2; see also McKenzie (2012).

[93] The linear probability model (LPM) generates an R-squared less than 1 regardless of the goodness-of-fit, since the dependent variable is binary but the predictions $\boldsymbol{X\beta}$ are continuous. The resulting R-squared thus depends on the strength of the model, as well as on unexplained variation introduced mechanically. Calculating an R_{MAX} requires separating out this mechanical component of the error, which depends on the distribution of the dependent variable and the predicted $\boldsymbol{X\beta}$. I do so empirically by using the unrestricted regression to generate binary predicted values that match the distribution of the dependent variable, then re-running the unrestricted regression with these binary predicted values as the dependent variable. The resulting R-squared is 0.875. This value is therefore a mechanical R_{MAX} for the use of the LPM on the unrestricted regression.

[94] Oster (2015) uses her method to recalculate coefficient bounds from a sample of experimental papers, where selection should be random and thus any coefficient movements due to the introduction of additional observables should be due to noise rather than an actual selection effect. She calculates that a value of $R_{MAX} = 1.3 * R_U^2$ allows her to retain the results of 90 percent of experimental studies.

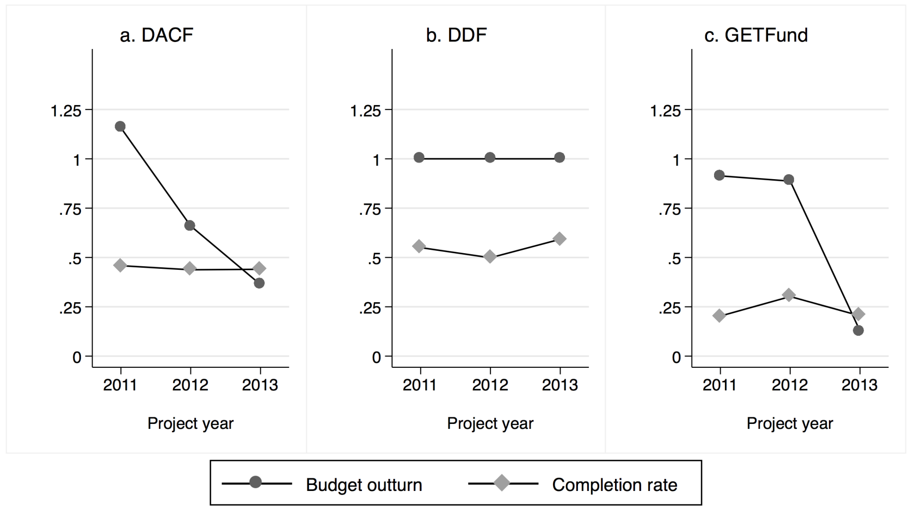
local partisan geography hypothesis discussed above relies on localities being sufficiently large that projects can be *club goods* accessible only to favored clients rather than *public goods* accessible to all residents of the location. Since the most common location definitions are village names (for rural areas) and neighborhood names (for urban areas), however, it is unlikely that many of them are larger than the catchment area of a school (for example), especially in contexts where existing service access and provision is low. While it is not presently possible to link project location to demographic information at the sub-district level, a very rough calculation suggests an upper bound on the mean population of these locations of 3,000 to 5,000.[95] While the possibility that infrastructure projects are locally excludable even within these relatively narrow geographic areas cannot be ruled out, it nonetheless seems implausible that sorting of projects into fund sources in order to engage in within-location targeting of projects could lead to differential completion rates across fund sources of the magnitude observed. Finally, large differences in completion rates are observed even for types of projects that are not normally associated with patronage because the district administration itself is the main beneficiary, as Column 7 of Table 1 shows. While this does not exclude the possibility that other project types are subject to highly localized targeting, it nonetheless suggests that this effect is not driving the results.

I also find little support for other alternative explanations for the observed differences among fund sources that – while not inconsistent with fund source having a causal effect on project completion – emphasize other mechanisms. One possibility is that the DDF's higher completion rates are simply due to prompt and full release of funds to districts, as opposed to the delays that often characterize the Ministry of Finance's releases to the DACF and GETFund secretariats. While district-level data on

[95] For the 56 districts that reported at least some project locations and for which three years of APR data are available, the average number of locations is 46.1, giving a mean population per location of 3,350 (based on the 2010 Census). The number of projects (mean population per project) for Metropolitan Assemblies (the most urbanized) in this sub-set is 160.0 (4,610), for Municipal Assemblies is 53.9 (3,031), and for District Assemblies (the most rural) is 38.4 (3,380). However since some locations presumably did not have projects in the period 2011-13 and thus do not appear in the dataset, the actual number of locations per district (and thus population per location) is likely to be even higher (lower) than this.

allocations and actual releases are not available, Figure 7 shows that there is little correlation between aggregate budget outturns for each fund source – which vary strongly across years – and annual completion rates for projects – which vary little across years. Since funds are released simultaneously to all districts, the non-release of funds cannot be driving the observed differences.

Figure 7: Fund Source Completion Rates and Budget Outturns



Note: Budget outturns and project completion are nationwide aggregates by fundsource-year. Outturns for DACF and GETFund are the annualized 9-month outturn reported in the next year's budget statement.

Another potential explanation of the results is that fund source-level completion rates are driven not by the institutional mechanisms through which they implement policy, but by politicians anticipating differences in voters' attribution of projects from different fund sources. For example, politicians might care more about completing projects for which voters think politicians are directly responsible, and therefore put more effort into completing those projects. If this were driving variation across fund sources, however, one would not expect the donor-supported DDF to have a higher completion rate than the DACF, which is entirely funded and implemented by the government. More generally, attribution and accountability considerations are important elements of institutional design for service delivery; to the extent that attribution effects account for variation in completion across fund sources, then, this is a point about the *mechanism* through which projects' fund source affects completion, not about *whether* the institutional design of this fund source affects completion.[96]

3.5 PARTISAN ALIGNMENT AND PROJECT COMPLETION

The previous section demonstrated that non-completion of projects is a common phenomenon in Ghana's local governments, that the fiscal institutions used to finance projects have a robust association with completion probability (and that there is little evidence that selection on unobservables explains more than a fraction of this effect), and that the patterns and dynamics of project non-completion are qualitatively consistent with this paper's theoretical mechanism that emphasizes how time consistency problems on the part of the government interact with the temporal dimension of project implementation. In this section, I test the theoretical argument more directly by examining how the estimated differences in completion rates across fund sources are related to the district's partisan alignment. The results are consistent with the model's predictions and are not consistent with alternative theoretical explanations that focus on tactical targeting strategies or the private incentives of politicians and bureaucrats.

Empirical strategy and results

Recall H3:

[96] However, even assuming that the estimated fund source effects are entirely causal, the marginal effect of shifting a project from (for example) DACF to DDF could be smaller than these estimates if it induced actors to change their behavior along other margins. For example, if corruption were more prevalent in DACF projects than DDF projects due to stricter monitoring, shifting a DACF project to the DDF might lead politicians, bureaucrats, or contractors to substitute corrupt behavior towards other DACF projects or non-infrastructure activities. Thus the marginal effect could be to increase the completion probability of the project in question but reduce that of other DACF projects, although the net marginal effect would still be positive (assuming that the demand curve for corruption is downward sloping). A related concern is that the effect of shifting projects to more effective fund sources may diminish over time, as Olken and Pande (2012) document often occurs in response to anti-corruption interventions. This concern may be slightly smaller in this case, since the programs under study are well established and thus the effects are already estimated in bureaucratic and political equilibrium.

H3: The difference in completion rates between DACF and DDF projects is decreasing in the ruling party's vote share in the district.

The intuition behind this hypothesis is that project fund source should matter less in districts with a high ruling party vote share because intra-party ties among the DCE and assembly members are more likely to facilitate intertemporal exchange and thus reduce the need for institutions that resolve this time consistency problem. I focus mainly on the DACF-DDF comparison in this section because these projects are selected and implemented simultaneously by the same politicians and bureaucrats, thus holding constant the identity, ability, preferences, and private incentives of politicians and bureaucrats and isolating the effect of the financing and control mechanisms associated with these fund sources.

To test H3, I estimate the equation:

$$(2) \qquad y_{i,j,k,t} = \boldsymbol{F_{i,j,k}\beta} + \boldsymbol{P_{i,j,k}\gamma} + \boldsymbol{F_{i,j,k}} * V_j * \boldsymbol{\tau} + \lambda_j + \delta_k + \nu_t + \epsilon_{i,j,k,t}$$

where all variables and parameters are as in equation (1) from the previous section, but with an interaction between fund source indicators $\boldsymbol{F_{i,j,k}}$ and the vote share of the current ruling party (the National Democratic Congress, NDC) from the first round of the 2008 presidential elections, V_j.[97] The coefficents $\boldsymbol{\tau}$ are the parameters of primary interest, in particular the difference between τ^{DDF} and τ^{DACF}.

Note that in this baseline specification $\boldsymbol{V_j}$ is not included in the model by itself, since it is time-invariant and thus captured by the district fixed effect λ_j. This lets me examine (one aspect of) the relationship between $\boldsymbol{V_j}$ and the outcome $y_{i,j,k,t}$ while still

[97] The ruling party's vote share from the 2008 presidential elections was chosen because these elections preceded the entire sample period of 2011-13 and thus serve as an indicator of the underlying affiliation of the district, as opposed to the results of the December 2012 elections which could have been endogenous to public good delivery in 2011-12. Ghana has a strongly two-party dominant system: the NDC and NPP were the top two parties in every constituency nationwide in the 2008 election and there were only a small number of constituencies where any third party obtained a significant number of votes, so using ruling party vote share is essentially identical to other measures, such as the winning margin or distance between the NDC and the main opposition party.

using district fixed effects λ_j to control for all observable and unobservable district characteristics that could otherwise confound the observed relationship. While numerous papers have studied how partisan composition is related to variations in governance outcomes *across* polities or geographic units, to my knowledge this is the first study to examine how polities' partisan characteristics affect within-polity variation in outcomes in this way.

Table 3 presents the results of this estimation. Column 1 shows that by itself, ruling party vote share is not significantly correlated with project completion; however, this could be due to omitted variable bias from other district characteristics that are correlated with both project completion and partisan vote share. More directly relevant to H3 is that the relationship between ruling party vote share and project completion varies markedly between DDF projects and DACF projects, as the interaction terms in Columns 2-5 show.

In Column 2 (without district fixed effects), DACF completion rates are strongly increasing in ruling party district vote share, while DDF completion rates are essentially unrelated to it. For the specifications in Column 3-6 it is the relative rather than absolute correlations that are of interest, and the gap between DDF and DACF completion probability is decreasing in the ruling party's vote share in the district across specifications, consistent with H3. The statistical significance of these differences varies across specifications, with p-values ranging from 0.018 to 0.137, but this is perhaps not surprising given the demanding specification and the likely presence of measurement error in both the dependent and explanatory variables, and the sign of the effect remains consistent.[98]

[98] The signs of these effects do not change when estimated separately across years rather than pooled for any of these specifications (estimates not shown), although the magnitude and significance does vary somewhat from year to year.

Table 3: Vote share and fund source interactions

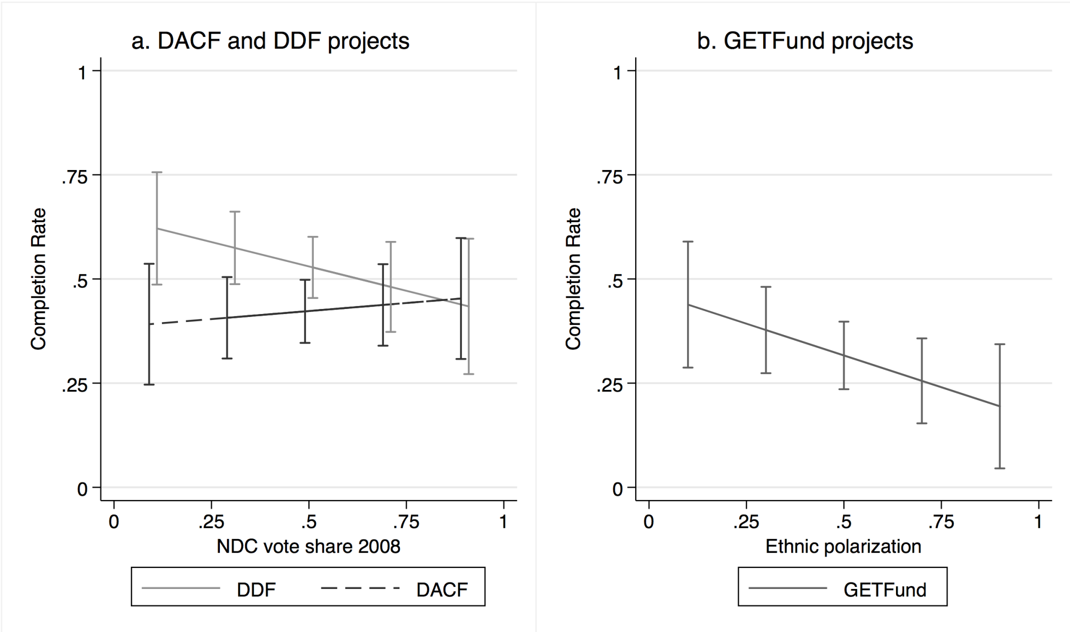
	(1)	(2)	(3)	(4)	(5)	(6)
	All projects	All projects	All projects	All projects	All projects	Schools only
DACF	0.0290	-0.110	-0.129	-0.0938	-0.114	-0.212
	(0.0271)	(0.0684)	(0.0806)	(0.0731)	(0.0988)	(0.113)
DDF	0.138**	0.149	0.0806	0.166*	0.0679	0.105
	(0.0306)	(0.0786)	(0.0889)	(0.0719)	(0.0926)	(0.132)
GETFund	-0.102**	-0.141	-0.172	-0.00867	-0.137	-0.274*
	(0.0340)	(0.0874)	(0.105)	(0.0806)	(0.108)	(0.117)
NDC vote share 2008	0.00769	-0.0579				
	(0.0732)	(0.114)				
Partisan alignment interactions						
DACF * NDC vote share		0.287*	0.269	0.156	0.0566	0.413
		(0.142)	(0.168)	(0.135)	(0.206)	(0.226)
DDF * NDC vote share		-0.0274	0.0864	-0.152	-0.184	-0.00288
		(0.166)	(0.191)	(0.148)	(0.190)	(0.275)
GETFund * NDC vote share		0.0785	0.0570	-0.226	-0.0913	0.323
		(0.176)	(0.216)	(0.147)	(0.199)	(0.230)
Constant	0.371**	0.411**	0.411**	0.612**	0.819**	0.368**
	(0.0868)	(0.0991)	(0.0594)	(0.105)	(0.135)	(0.0510)
DDF * NDC Voteshare - DACF * NDC Voteshare						
Difference		-0.314	-0.182	-0.308	-0.241	-0.416
[p-value]		[0.018]	[0.137]	[0.056]	[0.113]	[0.018]
Observations	13,339	13,339	13,339	10,420	8,801	4,204
R-squared	0.109	0.110	0.245	0.700	0.765	0.281
Fixed effects						
District			Yes		Yes	Yes
Location				Yes		
Contractor					Yes	
Year	Yes	Yes	Yes	Yes	Yes	Yes
Project type	Yes	Yes	Yes	Yes	Yes	Yes
Project characteristics	Yes	Yes	Yes	Yes	Yes	Yes

* p < 0.05, ** p < 0.01

Note: Dependent variable is project completion. Project characteristics are construction type (construction/ maintenance) and project year indicators (based on manual linking). NDC vote share 2008 is the ruling party's voteshare in the 2008 presidential elections in the district (or its antecedent district, for districts that split in 2012). Huber-White robust standard errors clustered at district level (location- and contractor-level in columns 4 and 5, respectively).

The magnitude of the estimated effects is economically important, and appears to explain a significant amount of the DDF-DACF completion rate gap. Moving a project from a district with a 32.2 percent ruling party vote share to a district with a 65.3 percent share (one standard deviation below and above the mean) reduces the DDF-DACF completion probability gap from 16.1 to 5.8 percentage points. Figure 8(a) illustrates this visually. This is consistent with the theoretical mechanism, which posits that fiscal institutions can help mitigate the negative consequences of distributive conflict for project completion.

Figure 8: Partisan Alignment Conditions Fund Source Effects



Note: Estimates are for construction of a new latrine in the mean location holding all other variables at means, based on specification from Table 3 Column 4. Note that marginal effects are measured relative to other fund sources, so these slopes do not indicate the overall effect of NDC vote share on project completion.

Interpretation and alternative explanations

One possible alternative explanation for the contrasting effects of partisan vote share on DACF and DDF projects is that the ruling party is intentionally using project completion to reward its partisans and punish its opponents, but is constrained in doing so on DDF projects by greater oversight. To assess this, we can examine the effect of vote share on GETFund projects, since implementation and completion decisions for GETFund projects are undertaken directly by the central government through a committee with opaque decision-making and reporting processes. If there were a deliberate political strategy to use project completion (or non-completion) to target or punish particular constituencies, it should be revealed by GETFund project completion patterns. Instead, the interaction term coefficients for the GETFund are closer to those of the DDF across most specifications, as shown visually by Figure 8(b).[99] While it is not possible to directly assess or rule out the intentional use of project completion as a targeting strategy without more fine-grained data, this is suggestive evidence that the different effects of partisan vote share on DACF and

[99] Column 6 in Table 3 is the exception to this pattern.

DDF projects are due to the institutions' effectiveness in mitigating intertemporal bargaining failures, rather than a deliberate tactic.

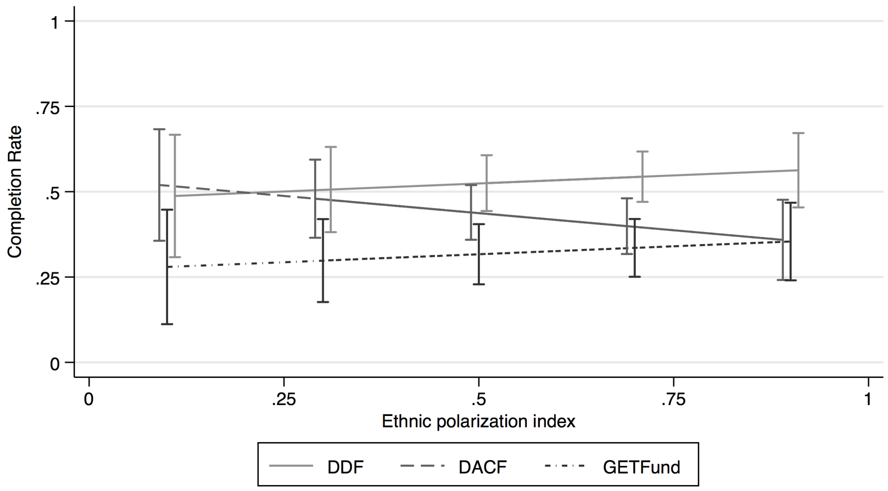
A second alternative theoretical explanation posits that the performance incentives of district-level politicians from the ruling party – especially the DCE – are weaker in opposition-leaning districts. This could be because there are fewer intra-party promotion opportunities in these districts: a common career advancement path for DCEs is to become Members of Parliament, but they are unlikely to be elected on the ruling party ticket in opposition-leaning districts. Lacking any career incentive to perform well, they may choose to use their terms in office to enrich themselves and their supporters through corruption, including through graft in project procurement and construction.

As an alternative test of the idea that intertemporal bargaining failure among actors with divergent distributive preferences can lead to incomplete projects, I therefore re-estimate Equation 2 with a different proxy for the ease of inter-temporal exchange among district-level political actors: ethnic polarization.[100] If a district is polarized between two large ethnic groups, then ethnic divisions are likely to be salient and this may increase distributive pressures and grievance and thus decrease actors' ability to sustain intertemporal bargains over project sequencing.

---

[100] Ethnic diversity is often argued to undermine public good provision (Easterly and Levine 1999; Alesina, Baqir, and Easterly 1999; Miguel and Gugerty 2005) and foster conflict (Montalvo and Reynal-Querol 2005), although other studies dispute this (Glennerster, Miguel, and Rothenberg 2013), and Ichino and Nathan (2013) argue that the political effects of ethnicity can depend on local ethnic geographies at the sub-district level, particularly with respect to the targeting of locally non-excludable public goods like infrastructure projects. Rather than weighing in on this debate, I use ethnic diversity as a proxy for more general forms of latent distributive conflict in polities; to the extent it is a weak proxy, this should bias any findings toward zero. I use ethnic polarization (calculated following Montalvo and Reynal-Querol 2005, using the Ghana Statistical Service's primary ethnic categories and data from the 2010 Population and Housing Census) rather than ethnic fractionalization because polarization maps more readily both onto the two-party dominant nature of Ghanaian politics; using fractionalization produces the same patterns, albeit somewhat weaker.

Figure 9: Fund Source Interaction with Ethnic Polarization



Note: Estimates are for construction of a new latrine in the mean location holding all other variables at means, based on specification from Table 3 Column 4. Note that marginal effects are measured relative to other fund sources, so these slopes do not indicate the overall effect of NDC vote share on project completion. Ethnic polarization is calculated following Montalvo and Reynal-Querol (2005), using the Ghana Statistical Service's primary ethnic categories and data from the 2010 Population and Housing Census.

Consistent with this, Figure 9 shows that the interaction between project fund source and ethnic polarization exhibits the same patterns as district vote share: the gap between DDF and DACF completion rates is largest when exchange is difficult (ethnic polarization is high), and the effect of polarization on GETFund projects is similar to that of the DDF. Although the precise mechanisms through which ethnic polarization affects actors' ability to sustain intertemporal exchange is likely to be different than those discussed above for partisan vote shares, this is nonetheless a useful robustness check, especially since ethnic polarization and NDC vote share are negatively correlated in the data and the results are thus based on different components of variation in the underlying construct.

3.6 CONCLUSION

This paper has sought to draw theoretical and empirical connections between the literatures on policy implementation, distributive politics, and fiscal institutions. While these literatures have each made major contributions to understanding public service delivery in developing countries, the relationship among them is little explored. This lacuna is surprising since the literature on the development of macro-historical institutions and state capacity has long emphasized that issues of distribution and efficiency are inseparable, and that the function of economic institutions is precisely to resolve the tensions between distribution and efficiency created by commitment problems (Acemoglu, Johnson, and Robinson 2005). This paper demonstrates that this is also true at the level of individual development projects, with implications for studies in these fields that neglect these relationships.

These findings also have relevance for the design of fiscal institutions and donor programs, as well as public sector reform more broadly. Not only do fiscal institutions have a significant influence on project outcomes, they also mediate – and in some circumstances can reduce – the potential negative efficiency consequences of distributive conflict on project allocation. This is a fertile area for future theory development and empirical work, and one with immense practical relevance. Reforms to improve bureaucratic capacity and independence or to increase political accountability have proved to be slow and uncertain, but this paper has shown that different fiscal institutions can lead to different project outcomes even when project selection and implementation is conducted by the same politicians and bureaucrats. These findings are hopeful, in that they suggest that efforts by donors and finance ministries alike to use their control of resource delivery to improve the operational efficiency of other government bureaucracies can be successful. While I focus on the effects of fiscal institutions rather than the factors that shape their emergence, this study also echoes the conclusion of the literatures on budget institution reform and soft sub-national budget constraints that the operation of fiscal institutions is deeply political (Rodden, Eskeland, and Litvack 2003; Rodden 2006; Andrews 2013). Understanding the relationship between these political and distributive pressures, the efficiency of policy implementation, and fiscal institutions is an important frontier for research on the political economy of public service delivery.

3.7 CONTRIBUTIONS

This paper's main theoretical contribution to the public management literature on infrastructure delivery is its development of a theoretical mechanism to explain unfinished projects that is rooted in distributive conflict. While a growing literature on infrastructure delivery examines how management practices, project characteristics, and institutional incentives can shape project implementation (Flyvbjerg, Holm, and

Buhl 2004; Flyvbjerg 2007, 2013; Budzier and Flyvbjerg 2013; Rasul and Rogger 2014), this paper focuses instead on the politics of project selection and implementation as a potential cause of implementation problems. In this respect it builds on the case study-based public administration literature on policy implementation, in which changing political and technical contexts and intertemporal bargaining during implementation feature prominently (Pressman and Wildavsky 1973, Grindle 1980, Cornell 2014). This theoretical emphasis also distinguishes it from the large body of work that focuses on corruption as a cause or mechanism of public service delivery failure (e.g. Reinikka and Svensson 2004; Olken 2007; Beekman, Bulte, and Nillesen 2014).

The paper also contributes to the budgeting and public finance literature on fiscal institutions, fiscal federalism, and fiscal decentralization, which examines how fiscal institutions affect fiscal outcomes such as aggregate spending and borrowing (Ferejohn and Krehbiel 1987; von Hagen and Harden 1995; Alesina, Hausmann, Hommes, and Stein 1999; Rodden, Eskeland, and Litvack 2003; Rodden 2006). In contrast, I focus on the operational effects of different fiscal institutions, thus helping to move this literature beyond its overwhelming focus on macro-fiscal outcomes (Wehner and de Renzio 2013). This also links with the economics and political science literature on decentralization that studies how variation in the source of public funds affects their management and public service outcomes at the sub-national level (Fisman and Gatti 2002, Gadenne and Singhal 2014, Gadenne 2015). Whereas the sparse empirical literature on this question has focused largely on identifying empirical differences in aggregate outcomes (e.g. district-level school enrollment rates), my empirical focus is at the project level and is more tightly linked to a specific theoretical mechanism.

From the perspective of comparative politics, the paper makes two theoretical contributions. The first is to introduce implementation considerations – specifically the multi-stage character of infrastructure projects – into the booming literatures on distributive politics, clientelism, and the political economy of public service provision in developing countries (Besley *et al* 2004, Banful 2011, Franck and Rainer 2012, Golden and Min 2013, Kramon and Posner 2013, Jablonski 2014, Burgess *et al* 2015, Harding 2015). Not only is poor project implementation substantively relevant, but it is also poorly explained by the foundational theories of distributive politics on which this literature is based (Cox and McCubbins 1986, Lindbeck and Weibull 1987, Dixit and Londregan 1996), which instead focus exclusively on policy choice. This theoretical insight also has important methodological implications: whereas the existing empirical literature takes the observed distribution of inputs into or outputs from public service delivery as revealed measures of politicians' tactical targeting strategies, I point out that the multi-stage implementation process associated with many types of public goods may introduce systematic differences between these two measures.

To make sense of the poor observed policy implementation outcomes, the paper introduces a second theoretical contribution: instead of modeling policy decisions as a tactical choice by a unitary incumbent who faces commitment problems with respect to voters (Robinson and Torvik 2005, Keefer 2007, Keefer and Vlaicu 2008, Keefer and Khemani 2009), I analyze them as the outcome of a collective choice process amongst multiple political actors (district assembly members and the DCE) who face commitment problems with respect to each other over the future distribution of policy choices. Combined with political instability – in this case arising from the inherently unstable nature of log-rolling coalitions (Riker and Brams 1973, Weingast and Marshall 1988) rather than the possibility of political alternation after an election that characterizes much of the other theoretical literature – these two theoretical elements explain how project implementation could be poor even when all politicians have strong incentives to deliver projects. Finally, by showing how fiscal institutions can mitigate the negative efficiency consequences of these intertemporal bargaining failures, this paper is related to the American politics literature that shows that bureaucracies mediate the outcome of distributive political dynamics (Bertelli and Grose 2009, Berry and Gersen 2010, Gordon 2011, Ting 2012), extending it into a new context and proposing a new theoretical mechanism.

REFERENCES

Abotsi, Pascal Kafu. 2013. "Contractors Blame Economic Hardship on Gov." *The Chronicle,* 18 September, <http://thechronicle.com.gh/contractors-blame-economic-hardship-on-gov/>, accessed 8 April 2015.

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2005. “Institutions as a Fundamental Cause of Long-Run Growth.” In Aghion, Philippe, and Steven N. Durlauf (eds), *Handbook of Economic Growth, Volume 1A*, 385-472.

Alesina, Alberto. 1988. “Credibility and Policy Convergence in a Two-Party System With Rational Voters.” *American Economic Review* 78(4): 796-805.

Alesina, Alberto, Baqir, Reza, Easterly, William, 1999. “Public goods and ethnic divisions.” *Quarterly Journal of Economics* 114 (4): 1243– 1284.

Alesina, Alberto, Ricardo Hausmann, Rudolf Hommes, and Ernesto Stein. 1999. “Budget Institutions and Fiscal Performance in Latin America.” *Journal of Development Economics* 59: 253-273.

Altonji, Joseph, Todd G. Elder, and Christopher E. Taber. 2005. “Selection on observed and unobserved variables: assessing the effectiveness of Catholic schools.” *Journal of Political Economy* 113 (1): 151–184.

Andrews, Matt. 2010. “How Far Have Public Financial Management Reforms Come in Africa?” HKS Faculty Research Working Paper Series, RWP10-018, John F. Kennedy School of Government, Harvard University.

Andrews, Matt. 2013. *The Limits of Institutional Reform in Development: Changing Rules for Realistic Solutions*. Cambridge: Cambridge University Press.

Baird, Sarah, Joan Hamory Hicks, and Edward Miguel. 2008. “Tracking, Attrition and Data Quality in the Kenyan Life Panel Survey Round 1 (KLPS-1).” University of California CIDER Working Paper.

Banful, Afua Branoah. 2011a. “Old Problems in the New Solutions? Politically Motivated Allocation of Program Benefits and the ‘New’ Fertilizer Subsidies.” *World Development* 39(7): 1166-1176.

Banful, Afua Branoah. 2011b. “Do formula-based intergovernmental transfer mechanisms eliminate politically motivated targeting? Evidence from Ghana.” *Journal of Development Economics* 96: 380-390.

Beekman, Gonne, Erwin Bulte, and Eleonora Nillesen. 2014. “Corruption, investments and contributions to public goods: Experimental evidence from rural Liberia.” *Journal of Public Economics* 115: 37-47.

Bellows, John, and Edward Miguel. 2009. “War and local collective action in Sierra Leone.” *Journal of Public Economics* 93, 1144–1157.

Berry, Christopher, and Jacob Gersen. 2010. “Agency Design and Distributive Politics.” Mimeo, October.

Bertelli, Anthony, and Christian R. Grose. 2009. “Secretaries of Pork? A New Theory of Distributive Public Policy.” *Journal of Politics* 71(3): 926-945.

Besley, Timothy, Rohini Pande, Lupin Rahman, and Vijayendra Rao. 2004. “The Politics of Public Good Provision: Evidence from Indian Local Governments.” *Journal of the European Economic Association* 2(2-3): 416-426.

Besley, Timothy, and Torsten Persson. 2009. "The Origins of State Capacity: Property Rights, Taxation, and Politics." *American Economic Review* 99(4): 1218-1244.

Briggs, Ryan C. 2014. “Aiding and Abetting: Project Aid and Ethnic Politics in Kenya.” *World Development* 64: 194-205.

Budzier, Alexander, and Bent Flyvbjerg. 2013. “Making-Sense of the Impact and Importance of Outliers in Project Management through the Use of Power Laws.” Mimeo, Oxford.

Burgess, Robin, Remi Jedwab, Edward Miguel, Ameet Morjaria, and Gerard Padró i Miquel. 2013. “The Value of Democracy: Evidence from Road Building in Kenya.” Mimeo, August.

Citi FM Online. 2014. "MMDAs to be forced to complete projects – PAC." <http://citifmonline.com/2014/11/01/mmdas-to-be-forced-to-complete-projects-pac/#sthash.KBZMV5fB.dpbs>, accessed 27 March, 2015.

Cornell, Agnes. 2014. "Why Bureaucratic Stability Matters for the Implementation of Democratic Governance Programs." *Governance* 27(2): 191-214.

Cox, Gary, and Mathew McCubbins. 1986. "Electoral Politics as a Redistributive Game." *Journal of Politics* 48(2): 370-389.

Daily Guide. 2015. "$817,000 Blown on abandoned Land Bills." <http://www.ghanaweb.com/GhanaHomePage/NewsArchive/artikel.php?ID=351212>, accessed 20 March, 2015.

de Figueiredo, Rui. 2002. "Electoral Competition, Political Uncertainty, and Policy Insulation." *American Political Science Review* 96(2): 321-333.

de Renzio, Paolo, Matt Andrews, and Zac Mills. 2011. "Does Donor Support to Public Financial Management Reforms in Developing Countries Work? An Analytical Study of Quantitative Cross-Country Evidence." Overseas Development Institute Working Paper 327, April.

District Assemblies Common Fund Administrator. 2014. The Common Fund Newsletter 1, June. <http://www.commonfund.gov.gh/DACF Newsletter.pdf>, accessed 25 March, 2015.

Dixit, Avinash, and John Londregan. 1996. "The Determinants of Success of Special Interests in Redistributive Politics." *Journal of Politics* 58(4): 1132-1155.

Dixit, Avinash, Gene M. Grossman, and Faruk Gul. 2000. "The Dynamics of Political Compromise." *Journal of Political Economy* 108(3): 531-568.

Easterly, William, and Ross Levine. 1997. “Africa’s growth tragedy: policies and ethnic divisions.” *Quarterly Journal of Economics* 112 (4): 1203–1250.

Ferejohn, John, and Keith Krehbiel. 1987. “The Budget Process and the Size of the Budget.” *American Journal of Political Science* 31(2): 296-320.

Fisman, Raymond and Roberta Gatti. 2002. “Decentralization and Corruption: Evidence from U.S. Federal Transfer Programs.” *Public Choice* 113(1/2): 25-35.

Flyvbjerg, Bent, Mette K. Skamris Holm, and Soren L. Buhl. 2004. “What Causes Cost Overrun in Transport Infrastructure Projects?” *Transport Reviews: A Transnational Transdisciplinary Journal* 24(1): 3-18.

Flyvbjerg, Bent. 2013. “Quality Control and Due Diligence in Project Management: Getting Decisions Right by Taking the Outside View.” *International Journal of Project Management* 31: 760-774.

Flyvbjerg, Bent. 2014. “What You Should Know About Megaprojects and Why: An Overview.” *Project Management Journal* 45(2): 6-19.

Franck, Raphael, and Ilia Rainer. 2012. “Does the Leader’s Ethnicity Matter? Ethnic Favoritism, Education, and Health in Sub-Saharan Africa.” *American Political Science Review* 106(2): 294-325.

Freiku, Sebastian. 2013. “Ahafo Ano North Assembly Takes Development To Constituents.” *The Chronicle*, 7 October, <http://thechronicle.com.gh/ahafo-ano-north-assembly-takes-development-to-constituents/>, accessed 8 April 2015.

Gadenne, Lucie. 2015. “Tax Me, but Spend Wisely? Sources of Public Finance and Government Accountability.” Mimeo, June.

Gadenne, Lucie, and Monica Singhal. 2014. “Decentralization in Developing Economies.” *Annual Review of Economics* 6: 581-604.

Ghana Audit Service. 2014. “Report of the Auditor-General on the Public Accounts of Ghana Ministries, Departments and Other Agencies (MDAs) for the Year Ended 31 December 2013.” Accra, 1 September.

Ghana News Agency. 2014. “Ghana @ 50 Project left to rot.” <http://www.ghananewsagency.org/print/75622>, accessed 27 March, 2015.

Ghana News Agency. 2015. “Political Party Interference Marks Upcoming Assemblies’ Elections – CODEO.” <http://www.ghanaweb.com/GhanaHomePage/NewsArchive/Political-party-interference-marks-upcoming-assemblies-elections-CODEO-376599>, accessed 14 September, 2015.

Glennerster, Rachel, Edward Miguel, and Alexander D. Rothenberg. 2013. “Collective Action in Diverse Sierra Leone Communities.” *The Economic Journal* 123: 285-316.

Golden, Miriam, and Brian Min. 2013. “Distributive Politics around the World.” *Annual Review of Political Science* 16: 73-99.

Gonzalez, Felipe, and Edward Miguel. 2015. “War and local collective action in Sierra Leone: A comment on the use of coefficient stability approaches.” *Journal of Public Economics* 128: 30-33.

Gordon, Sanford. 2011. “Politicizing Agency Spending Authority: Lessons from a Bush-era Scandal.” *American Political Science Review* 105(4): 717-734.

Grindle, Merilee (ed). 1980. *Politics and Policy Implementation in the Third World*. Princeton: Princeton University Press.

Harding, Robin. 2015. “Attribution and Accountability: Voting for Roads in Ghana.” *World Politics*, forthcoming.

Harding, Robin, and David Stasavage 2014. "What Democracy Does (and Doesn't Do) for Basic Services: School Fees, School Inputs, and African Elections." *Journal of Politics* 76(1): 229-245.

Harris, J. Andrew, and Daniel Posner. 2015. "(Under What Conditions) Do Politicians Reward Their Supporters? Evidence From Kenya's Constituencies Development Fund." Mimeo, 7 January.

Helmke, Gretchen, and Steven Levitsky. 2004. "Informal Institutions and Comparative Politics: A Research Agenda." *Perspectives on Politics* 2(4): 725-740.

Huber, John, and Nolan McCarty. 2004. "Bureaucratic Capacity, Delegation, and Political Reform." *American Political Science Review* 98(3): 481-494.

Ichino, Nahomi, and Noah Nathan. 2013. "Crossing the Line: Local Ethnic Geography and Voting in Ghana." *American Political Science Review* 107(2): 344-361.

International Monetary Fund. 2015. "Making Public Investment More Efficient." 11 June.

Jablonski, Ryan. 2014. "How Aid Targets Votes: The Impact of Electoral Incentives on Foreign Aid Distribution." *World Politics* 66(2): 293-330.

Keefer, Philip. 2007. "Clientelism, Credibility, and the Policy Choices of Young Democracies." *American Journal of Political Science* 51(4): 804-821.

Keefer Philip, and Stuti Khemeni S. 2009. "When do legislators pass on pork? The role of political parties in determining legislator effort." *American Political Science Review* 103(1): 99–112.

Keefer, Philip, and Razvan Vlaicu. 2008. "Democracy, Credibility, and Clientelism." *Journal of Law, Economics, and Organization* 24(2): 371-406.

Kramon, Eric, and Daniel Posner. 2013. “Who Benefits from Distributive Politics? How the Outcome One Studies Affects the Answer One Gets.” *Perspectives on Politics* 11(2): 461-474.

Kunateh, Masahudu. 2014. “Ghana: Projects At Standstill Amid Common Fund Issues.” *CAJ News Africa*, 10 April, <http://allafrica.com/stories/201404110307.html>, accessed 25 March 2015.

Jones, Bryan D., and Frank R. Baumgartner. 2005. *The Politics of Attention*. Chicago: University of Chicago Press.

Joy News. 2014. “President Mahama: I will continue all abandoned projects.” <http://www.myjoyonline.com/news/2014/September-10th/president-mahama-i-will-continue-all-abandoned-projects.php>, accessed 25 March, 2015.

Leaver, Clare. 2009. “Bureaucratic Minimal Squawk Behavior: Theory and Evidence from Regulatory Agencies.” *American Economic Review* 99(3): 572–607

Lindbeck, Assar, and Jörgen W. Weibull. 1987. “Balanced-Budget Redistribution as the Outcome of Political Competition.” *Public Choice* 52(3): 273-297.

Luna, Joe. 2015. “A Theory of Political Organization.” 19 April.

Miguel, Edward, and Mary Kay Gugerty. 2005. “Ethnic diversity, social sanctions, and public goods in Kenya.” *Journal of Public Economics* 89: 2325-2368.

Montalvo, José G. and Marta Reynal-Querol. 2005. “Ethnic Polarization, Potential Conflict, and Civil Wars.” *American Economic Review* 95(3): 796-816.

Nathan, Noah. 2015. “Does Participation Reinforce Patronage? Policy Preferences and Turnout in Urban Ghana.” Mimeo, 20 April.

National Commission on Civic Education (NCCE). 2015. “Assessing the Effectiveness of District Assemblies in Ghana’s Democracy.”

Nunn, Nathan and Leonard Wantchekon. 2011. "The Slave Trade and the Origins of Mistrust in Africa." *American Economic Review* 101(7): 3221-3252.

Olken, Benjamin. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115(2): 200-249.

Olken, Benjamin, and Rohini Pande. 2012. "Corruption in Developing Countries." *Annual Review of Economics* 2012(4): 479-509.

Oster, Emily. 2015. "Unobservable Selection and Coefficient Stability: Theory and Evidence." January 26.

Pressman, Jeffrey L., and Aaron Wildavsky. 1974. *Implementation.* Berkeley: University of California Press.

Rasul, Imran, and Daniel Rogger 2014. "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service." Mimeo, October.

Reinikka, Ritva, and Jakob Svensson. 2004. "Local Capture: Evidence from a Central Government Transfer Program in Uganda." *The Quarterly Journal of Economics* 119(2): 679-705.

Riker, William, and Steven Brams. 1973. "The Paradox of Vote Trading." *American Political Science Review* 67(4): 1235-1247.

Robinson, James A., and Ragnar Torvik. 2005. "White Elephants." *Journal of Public Economics* 89: 197-210.

Robinson, James A., and Thierry Verdier. 2013. "The Political Economy of Clientelism." *Scandinavian Journal of Economics* 115(2): 260-291.

Rodden, Jonathan A. 2006. *Hamilton's Paradox: The Promise and Peril of Fiscal Federalism*. Cambridge: Cambridge University Press.

Rodden, Jonathan, Gunnar S. Eskeland, and Jennie Litvack (eds). 2003. *Fiscal Decentralization and the Challenge of Hard Budget Constraints*. London: MIT Press.

Ting, Michael. 2012. "Legislatures, Bureaucracies, and Distributive Spending." *American Political Science Review* 106(2): 367-385.

Today Newspaper. 2015. "Election of MMDCEs: Ahwoi got it wrong – Nduom." 25 August. <http://www.ghanaweb.com/GhanaHomePage/NewsArchive/Election-of-MMDCEs-Ahwoi-got-it-wrong-Nduom-377412>, accessed 25 August 2015.

Tsai, Lily. 2007. *Accountability without Democracy: How Solidary Groups Provide Public Goods in Rural China*. Cambridge: Cambridge University Press.

TV3 Network. 2015. "Mahama calls for moderation in demands on gov't." <http://www.ghanaweb.com/GhanaHomePage/NewsArchive/artikel.php?ID=353332>, accessed 4 July 2015.

Von Hagen, Jürgen, and Ian Harden. 1995. "Budget Processes and Commitment to Fiscal Discipline." *European Economic Review* 39: 771-779.

Ward, Hugh, and Peter John. 1999. "Targeting benefits for electoral gain: constituency marginality and the distribution of grants to English local authorities." *Political Studies* 47: 32–52.

Weghorst, Keith R., and Staffan I. Lindberg. 2013. "What Drives the Swing Voter in Africa?" *American Journal of Political Science* 57(3): 717-734.

Wehner, Joachim, and Paolo de Renzio. 2013. "Designing Fiscal Institutions: The Political Economy of PFM Reforms." In *The International Handbook of Public Financial Management*, ed. Richard Allen, Richard Hemming, and Barry Potter. Basingstoke: Palgrave Macmillan, 84-97.

Weingast, Barry, and William Marshall. 1988. “The Industrial Organization of Congress; or, Why Legislatures, Like Firms, Are Not Organized as Markets.” *Journal of Political Economy* 96(1): 132-163.

APPENDIX 3.1: ANNUAL PROGRESS REPORTS

Figure A1: Annual Progress Report Sample Page

Progress Report **ADANSI SOUTH**

No	Projectdetail	Location	Fundingsource	Contractor	Originalcontractsum	Revisedcontractsum	Paymenttodate	DateAwarded	DateCommencement	DateExpectedComple	DateActualcompletion	Staged	Status	Rema
ADS0001	construction of 6 unit classroom block	AWOROSO	Get Fund	ESSIL CXONSTRUCTION	156,969.67	149,900.00		05-Aug-10	25-Aug-10	12-May-11		20	On-going	
ADS0002	construction of 6 unit classroom block	KOJO MANKRONG	Get Fund	FESAGE ENTERPRISE	157,575.83	149,900.00		05-Aug-10	26-Aug-10	19-May-11		20	On-going	
ADS0003	construction of 6 unit classroom block	EDWINASE	Get Fund	SUE ENGEERING	158,250.95	149,900.00		05-Aug-10	25-Aug-10	26-May-11		20	On-going	
ADS0004	construction of 3 unit classroom block (JHS)	SUBRISO	DDF	V FORGHAN LTD	59,947.36		59,947.36	18-Nov-10	29-Nov-10	19-May-11	18-Nov-11	100	completed	
ADS0005	construction of 20 seater vault toilet chamber	WURIYIE	DDF	FAUSBON CONSTRUCTION WORKS	52,847.95		52,035.46	18-Nov-10	23-Nov-10	17-May-11	18-Nov-11	100	completed	
ADS0006	construction of 3 unit classroom block	NKORANSA	DDF	HAFA-WELL WORKS	60,077.05		60,077.49	18-Nov-10	30-Nov-10	24-May-11	18-Nov-11	100	completed	
ADS0007	construction of 3 unit classroom block	ODUMASSE	DDF	ADUOPY CONSTRUCTION & SUPPLY LTD	61,985.00		61,985.00	18-Nov-10	29-Nov-10	18-May-11	18-Nov-11	100	completed	

Figure A2: Storage of APRs in National Development Planning Commission Library

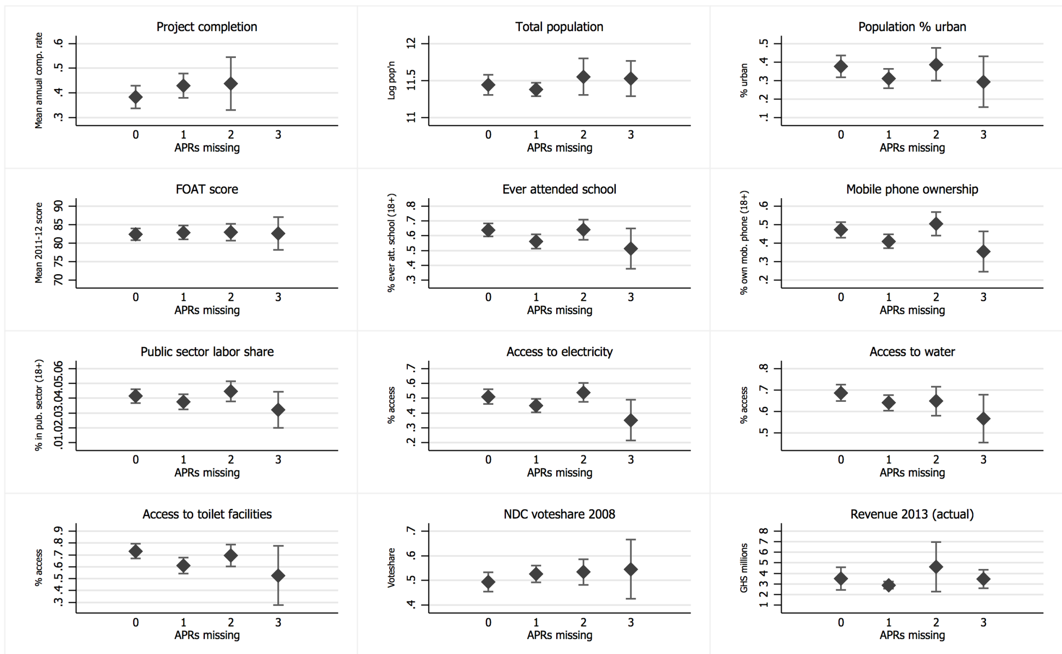
APPENDIX 3.2: APR SAMPLE BALANCE AND MISSING DATA

The coverage of the Annual Progress Report (APR) database is remarkably high, given that most had to be located in hard copy in the offices of the National Development Planning Commission (NDPC) in Accra or of the Regional Coordinating Councils (RCCs) in the ten regional capitals. Altogether 479 APRs were located. The maximum notional number of APRs for the period 2011-13 would be 602: 170 for 2011, 216 each for 2012 and 2013. Of these, 407 APRs contained project tables with sufficient information to be entered into the database. The final database thus covers 67.6 percent of possible district-year observations. Nevertheless, there are concerns that reporting could be correlated with other variables of interest, such as project completion rates.

As this Appendix shows, however, there is little evidence that reporting completeness is correlated with district characteristics. Figure A3 below plots the unweighted means and 95 percent confidence intervals of a wide range of variables, by the number of APRs that are missing for each district. The most important balancing test is for average annual project completion, this study's main dependent variable. Although it is not possible to calculate this for districts with all three APRs missing, there is no statistically significant difference in average completion rates across districts with different levels of reporting completeness; indeed districts with more incomplete reporting have, if anything, slightly higher completion rates, although this difference is not statistically significant. This alleviates the concern that estimated national project completion rates may be biased upwards due to reporting incompleteness.

The sample also appears to be balanced across the other variables reported in Figure A3. In addition to a wide range of demographic, social, and economic variables drawn from the 2010 Population and Housing Census, this includes: districts' scores on the Functional and Organizational Assessment Tool (FOAT) evaluation undertaken to assess districts' compliance with a set of procedural requirements as part of the allocation and disbursement procedure for DDF funds; the vote share in the district of the National Democratic Congress (NDC), which was the ruling party during the sample period, from the 2008 presidential elections; and budget size, as measured by the total revenue of the district in 2013. There are no apparent patterns across reporting completeness in any of the variables examined, so there is no evidence that the sample coverage of the APR database is biased.

Figure A3: APR Sample Balance on Selected Covariates



Note: Mean and 95% confidence intervals shown for each group of districts. Access to electricity is percentage of individuals reporting using mains electricity for lighting; access to water is percentage reporting using pipe-borne, borehole, or public tap water for drinking; access to toilet facilities is percentage using a WC, KVIP, pit latrine, or public toilet facility (all from Population and Housing Census 2010).

Unfortunately, due to the inconsistent reporting formats used by districts in producing their APRs, many observations are missing some important variables. Table A1 gives an indication of this for a selected number of variables. Although this affects the types of analysis that can be done on the data, there is no indication that the missing variables are anything other than a result of districts' use of different formats.

Table A1: Coverage of Key Variables in Dataset

	Non-missing			Non-missing	
Variable name	Obs.	Pct.	Variable name	Obs.	Pct.
Project title	14,246	100.0%	Contract sum	9,869	69.3%
Completion status	13,339	93.6%	Commencement date	5,518	38.7%
Fund source	11,226	78.8%	Completion date - expected	5,061	35.5%
Location	11,326	79.5%	Completion date - actual	1,424	10.0%
Contractor	9,319	65.4%	Expenditure to date	6,224	43.7%

Note: See Appendix 3 for full variable descriptions. Percentages are as percent of total core sample (n=14,246).

APPENDIX 3.3: VARIABLE CONSTRUCTION AND CODING

All APR database variables were coded algorithmically from text strings by defining a set of word or phrases corresponding to values; the particularities of this process for each variable, along with other relevant data and coding notes, are detailed below. Project numbers and statistics in this appendix are given at the point of coding, and thus may differ from those in the final database from which repeat observations and non-infrastructure projects have been excluded.

FundSource

Project fund source was constructed from APR entries for project's fund source for nearly all observations, although in a small number of cases (178, or 1.1 percent) there was no dedicated entry for fund source but fund sources were named in the project title. These were combined, and then coded into fund source categories according to a set of text strings commonly occurring in the data – e.g. for DACF, these were "DACF", "COMMON FUND", "D A C F", and "CF". All projects where the fund source was listed as the district itself were coded as using internally generated funds (IGF), together with those where IGF was directly identified as the fund source. The "Other donor" category comprises 51 sources, each of which accounts for only a small fraction of the total. Some of these are clearly identifiable – e.g. "USAID", "EU", "WORLD VISION" – while others are vague – e.g. "DONOR", "NGO", "CHINESE GRANT". "Other GoG" includes all fund sources associated with central government other than the GETFund and Road Fund: e.g. "MOE" (Ministry of Education), "GOG" (Government of Ghana – typically used to mean central government in the Ghanaian context), "SECTOR MINISTRY". The "Other/Unknown" category includes all projects for which no fund source was reported, as well as projects that could not be assigned to one of the other categories – most notably a small number of projects (143) funded by the local Member of Parliament (MP) using the minor allocation of the DACF which is disbursed to them as a constituency development fund, and 70 projects for which the "community" was listed as a fund source. All projects with more than one identifiable fund source (about 3 percent) were coded as "Multiple".

ProjectType

Project type was constructed using sets of commonly used text strings in the project title to first group projects into sixteen types of infrastructure projects:

- *Agriculture:* dams, irrigation, dug-outs;
- *Borehole:* boreholes, wells;
- *Clinic:* clinics, health centres, hospitals, wards;
- *Construction – other:* abattoirs, computer centers, libraries, taxi ranks, lorry parks, community centers, sports stadiums, light industrial areas, warehouses;
- *Culvert*: culverts, drains, ditches, gutters;
- *Electricity:* electrification, substations;
- *Latrine:* latrines, Kumasi ventilated improved pits (KVIPs), toilets, water closets;
- *Market:* market stalls, stores, sheds, meat shops;
- *Office:* administration blocks, assembly/town/council halls, courts, police stations, fire stations;
- *Road:* roads (paved, graveled, or dirt), bridges, spot improvements, speed humps, paving works;
- *School:* classroom blocks, kindergartens, nurseries, early childhood development centres;
- *School – other:* dormitories, dining halls, hostels, school feeding kitchens;
- *Staff housing:* bungalows, guest houses, accommodation blocks, residences, quarters;
- *Streetlights*
- *Waste management:* refuse dumps, rubbish storage; and
- *Water:* water systems, water harvesting, water supply, reservoirs and storage, pipe-borne water works, water distribution.

In addition, two categories of non-infrastructure projects were constructed but not included in the analysis:

- *Procurement:* purchase, supply, distribution, and furnishing (e.g. tractors, desks, computers), acquiring land for projects, equipment of facilities; and

- *Services:* a wide variety of activities related to service-provision and other non-infrastructure activities, e.g. training, vaccination campaigns, capacity building, tax collection, celebrations, monitoring, public education, sponsoring.

The guiding principle in distinguishing between infrastructure and non-infrastructure projects was that projects involving physical transformation were coded as infrastructure (e.g. building a classroom block), whereas projects consisting only of related activities that did not themselves involve physical transformation (e.g. acquiring land to build a classroom block, supplying a classroom block with textbooks) were coded as non-infrastructure.

This algorithmic coding resulted in unique project types for 74.4 percent of projects, while 12.8 percent were not assigned a type and another 12.8 percent were assigned two or more types. These 5,569 projects were manually inspected and disambiguated if possible, or if the project genuinely straddled two types it was coded as “multiple”.

Finally, the category “school” was sub-divided into six categories according to the size of the classroom block: five categories for 2-, 3-, 4-, 6-, and 12-unit classroom blocks, and a sixth residual category for classroom blocks of indeterminate size, or reported projects which actually involved more than one discrete structure (e.g. construction of two 3-unit classroom blocks). Number of units was coded algorithmically by defining a set of 41 common text string permutations used to denote construction of a single classroom block (e.g. “1NO 3-UNIT [CLASSROOM BLOCK]”, “[CONSTRUCTION] OF 3-UNIT [CLASSROOM BLOCK]”).

Prior to analysis, projects with missing type or coded as “services”, “procurement” were dropped. The project categories used in the analysis therefore comprise the fifteen non-school infrastructure types listed above; six types of schools (five according to classroom block size, and one residual category); and the type “multiple” comprising all projects that could not be manually coded into a unique type.

ConstructionType

Project titles often include a phrase that identifies whether the project constitutes new (greenfield) construction, or repair, maintenance, renovation, or rehabilitation of an existing project that had been completed previously. The former category was coded as projects including the general text string “CONSTRUCTION OF” and abbreviations or misspellings of this; project type-specific construction verbs such as “DRILLING OF”, “PAVING”, and “SPOT IMPROVEMENT”; and strings indicating that the project is a greenfield project in its second or subsequent year, such as “COMPLETION OF”, “CONTINUE”, and “CLADDING”. (The APRs are inconsistent in the extent to which they alter these prefaces for a given project across years (i.e. whether they change “CONSTRUCTION OF” in the first year of a project to “COMPLETION OF” in its second year), so these were coded together as greenfield projects.) Project titles containing general phrases such as “MAINTENANCE”, “REPAIR”, “RENOVATION”, and “REHABILITATION”, or project type-specific phrases such as “DESILTING”, “RE-ROOFING”, “RESURFACING”, and “RESHAPING” were coded as maintenance/repair/renovation projects. Altogether 76.4 percent of projects were coded as greenfield construction, 11.9 percent as maintenance/repair/renovation, and the remaining 11.6 percent could not be uniquely identified as either type.

ProjectCompletion

Project completion was coded as a binary variable by combining information from three raw variables, of which one or two are typically reported in each APR: *ProjectStatus* (e.g. “COMPLETED”, “INSTALLED AND IN USE,” “100 WORK DONE”), *Remarks* (similar), and *PercentWork* (on the scale 0-100; 100 coded as complete). Projects were coded as complete if they were at a stage where physical construction work had been completed, regardless of whether they had been formally handed over, furnished, commissioned, and put into use – for example “COMPLETED YET TO BE FURNISHED AND COMMISSIONED” was coded as complete. This yielded a unique completion coding for 91.6 percent of observations; the remainder were disambiguated by visual inspection if possible, and given a missing value if it was impossible to determine the project’s status conclusively.

Although the gap between physical completion and putting the facility into use is of potential interest, physical completion was chosen as a cutoff point for the purposes of the APR database because: 1) the status of post-construction activities like furnishing, commissioning, and use are reported inconsistently in the APRs; and 2) the analytical focus of this paper is on infrastructure project construction, not subsequent service provision using those facilities.

Contractor

A total of 6,798 unique contractor names are listed in the APR database for 10,701 infrastructure projects. However, many of these are clearly the same contractor but with different spellings (e.g. "WRKS" for "WORKS"), abbreviations (e.g. "LTD." for "LIMITED"), or omissions (e.g. dropping "LIMITED" or "INC."). In order to combine these, contractor names were stripped of these and other generic elements of company names (e.g. "ENTERPRISE", "TRADING", "MESSRS.", "M/S", "COMPANY"), as well as punctuation marks and spaces. This reduced the number of unique contractor names from 6,798 to 5,113. Using these corrected contractor names rather than the raw names slightly changes the point estimates on fund source regression coefficients, but not the differences between these coefficients, which are the quantities of interest.

District

In mid-2012, 45 of Ghana's 170 districts were split to create 46 new districts (one district was split into three), leaving a total of 216 districts. The 46 new districts were all entirely contained within a single parent district, so there was no realignment of borders between districts. The 2011 and 2013 APRs thus reflect the 170 and 216 districts, respectively. For 2012 districts reported according to the new (216) district names, although many of the newly created districts did not report as they had only been in existence for approximately six months and were still waiting offices, personnel, etc. This creates some concern about duplications or omissions in the reporting of projects in split districts that started prior to the split, and it is unclear how consistently these matters were handled across districts. However, restricting the sample to districts that did not split in 2012 does not affect any of the results presented above, and the regression results are robust to the inclusion of district-year fixed effects that would capture any disruption caused by these administrative splits, so the potential data concerns created by the district splits do not appear to affect the analysis.

For purposes of project linking and fixed effects, the post-split "parent" district (the one that maintained the existing district capital, political leadership, and the majority of its administrative staff) is treated as the same district as the pre-split combined district, regardless of whether it changed its name, while the new "child" district is treated as a new district.

The other secondary data sources drawn on by this paper differ in whether they report the old 170 or new 216 districts for 2012. This means that in some cases (e.g. with budget data) APR data from a post-split 2012 district is matched to other secondary data from a pre-split 2012 combined district. Of data sources that are time invariant over 2011-13, the District Medium Term Development Plans (DMTDPs) and pre-2012 electoral data both use the 170 districts, while the 2010 Population and Housing Census initially used the 170 districts but has been recoded to correctly reflect the new 216 districts.

Classroom Block Additional Facilities

For all classroom blocks for which it was possible to identify the number of units (2, 3, 4, 6, or 12), three indicator variables representing additional facilities included in the project were defined: latrines and toilets (project titles including the strings "LATRINE", "TOILET", "KVIP", etc.); offices/stores/libraries ("OFFICE", "STORE", "COMMON ROOM", "LIBRARY"); and general ancillary facilities (various spellings and abbreviations of "ANCILLARY"). These variables were not coded as mutually exclusive, although it is not common for one project to combine multiple types of ancillary facilities. A residual variable was defined for the 38.0 percent of projects that do not appear to include any of these ancillary facilities.

APPENDIX 3.4: ATTRITION IN PROJECT LINKING

Because very few districts assign unique tracking numbers to projects, linking projects across years had to be done manually. For each district for which all three years of data were available, records of projects coded as being in the same location (e.g. village, neighborhood) in different years were visually inspected according to their project title, fund source, completion status, contract sum, and other potentially identifying information, and linked if they were obviously the same project. Conditional on being incomplete in 2011 or 2012, only 33.8 percent of projects could be identified in the following year, indicating a high degree of attrition in reporting and linking. This gives rise to two concerns: first, differential attrition rates across fund sources could bias the within-district estimates of fund sources completion rates.[101] Second, attrition is likely to be correlated with project completion (if bureaucrats stop reporting unfinished projects that have been abandoned) and thus poses a challenge for estimating the overall completion rate.

To investigate the first possibility, I construct an attrition indicator variable equal to one if a project that is incomplete in 2011 or 2012 can be linked to the same project's record in the following year (2012 or 2013, respectively), and zero otherwise. This variable is defined only for projects in districts that have three years of APR data. I then use this as the dependent variable in an attrition probability model, estimated as a linear probability model, where the key variables of interest are fund source indicator variables.

The results are presented in Table A1. Column 1 estimates the model with no controls and indicator variables only for the three major fund sources; Column 2 adds the baseline set of district, year, and project type fixed effects, plus project characteristics; Column 3 adds indicator variables for the remaining minor fund sources; and Column 4 estimates the model for school buildings only. The differences among the coefficients on the three major fund sources are small and are not statistically

[101] Attrition rates also vary across districts, and it is possible that districts interpret the reporting mandate in slightly different ways: some may report all projects that were underway in the district, whether or not they were active during the year, while others may only report projects that were active or included in annual budgets. However, this does not pose a major threat to this paper's main analyses, as the district fixed effects would cancel out district-level differences in attrition.

significant in any of the specifications. Among the minor fund sources, projects funded by districts' own Internally Generated Funds (IGF) have the highest attrition rates, together with those with multiple fund sources. There is slightly more variation across coefficients in the schools only specification in Column 4, but with a reduced sample size that creates a great deal of statistical uncertainty about the parameter values. Overall, none of these results create cause for concern that this paper's main results are driven by differential reporting attrition rates across project fund sources.

Table A2: Project Attrition by Fund Source

	(1)	(2)	(3)	(4)
	All projects	All projects	All projects	Schools only
Project fund source				
DACF	-0.137	-0.004	0.007	-0.016
	(0.030)**	(0.044)	(0.065)	(0.119)
DDF	-0.107	0.008	0.018	0.061
	(0.034)**	(0.066)	(0.079)	(0.123)
GETFund	-0.086	-0.033	-0.015	-0.006
	(0.028)**	(0.067)	(0.100)	(0.142)
IGF			0.148	0.093
			(0.113)	(0.160)
Other donor			0.004	0.105
			(0.092)	(0.147)
Other GoG			-0.034	0.297
			(0.083)	(0.208)
Road Fund			0.000	0.000
			(0.000)	(0.000)
Multiple			0.132	0.019
			(0.117)	(0.157)
Constant	0.712	0.613	0.595	0.735
	(0.014)**	(0.142)**	(0.155)**	(0.275)*
Coeff. equality tests (prob > F)				
DACF = DDF	0.457	0.823	0.847	0.369
DACF = GETFund	0.143	0.604	0.701	0.885
DDF = GETFund	0.585	0.581	0.668	0.424
Observations	2033	2033	2033	780
R-squared	0.01	0.02	0.02	0.02
Fixed effects				
District		Yes	Yes	Yes
Year		Yes	Yes	Yes
Project type		Yes	Yes	Yes
Project characteristics		Yes	Yes	Yes

* significant at 5%; ** significant at 1%

Note: Dependent variable is project attrition - whether a project that is unfinished in a given year can be located in the following year's report, conditional on the report being contained in the APR database. Other/unknown is omitted fund source category. Project characteristics are construction type (construction/maintenace) and project year indicators (based on manual linking). Huber-White robust standard errors clustered at district level.

To address the second concern, I therefore estimate three different sets of completion rates, which are almost identical for projects' first year but diverge thereafter:

- Upper bound. Projects are classified into years (1-3) according to their reported year of commencement (e.g. a 2012 observation of a project that started in 2011 is in its second year). No correction is made for attrition. Sample is all projects with non-missing commencement year.
- Middle estimate. Projects are classified into years according to manual linking (see above). Incomplete projects that cannot be traced to the subsequent year are treated as missing in the subsequent year (i.e. no correction is made for attrition). Projects that have not been linked to an observation from previous year are assumed to be in their first year. Sample is all projects from districts for which all three years of data are available. This is the method used in the main body of the paper.
- Lower bound. Same as middle estimate, but incomplete projects that cannot be traced to the subsequent year are assumed *not* to have been finished (i.e. attriting observations are treated as incomplete).

The upper bound estimate will be biased upward if unfinished/abandoned projects are more likely to attrite from the dataset than completed projects, which is probable. Likewise, the lower bound estimate will be biased downward if untraceable projects are actually completed in the subsequent year but not reported, or if the projects were completed and reported but not linked by the manual tracing methodology. The middle estimate is situated between these two but may also be biased, although the direction of this bias is unclear *a priori*. To the extent that the middle and lower estimates incorrectly group projects that are in their second or subsequent years but are appearing in the dataset for the first time as first-year projects, the first-year completion rates may be biased; in practice however this bias appears to be small, as the first-year completion rates are very similar under all three estimates.

Figure A4: Three Completion Rate Estimates

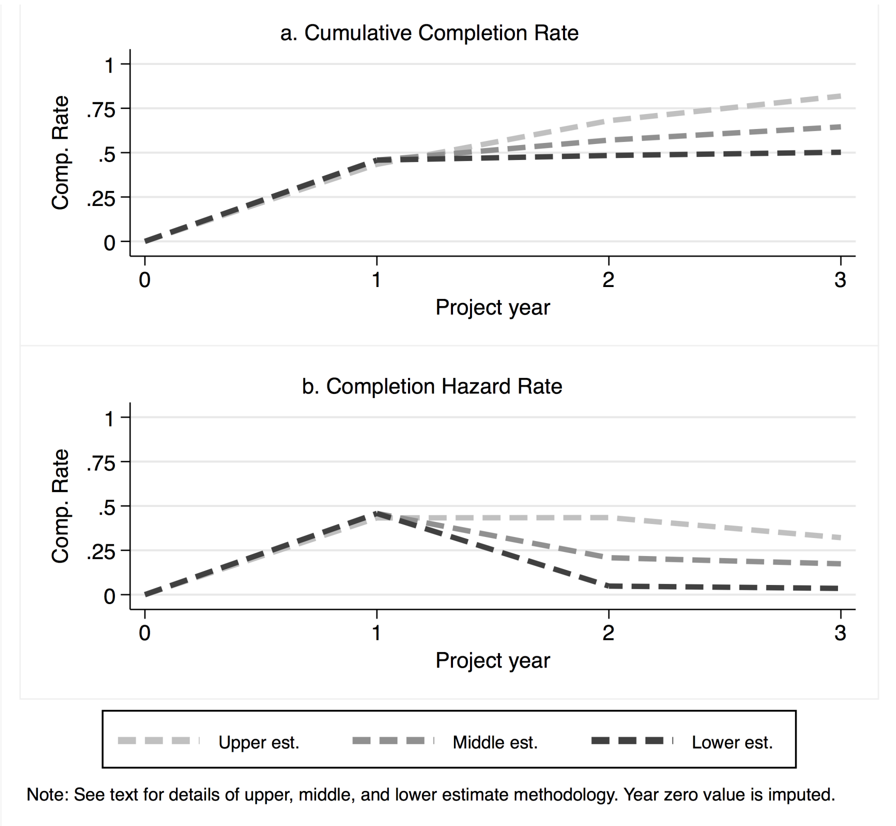
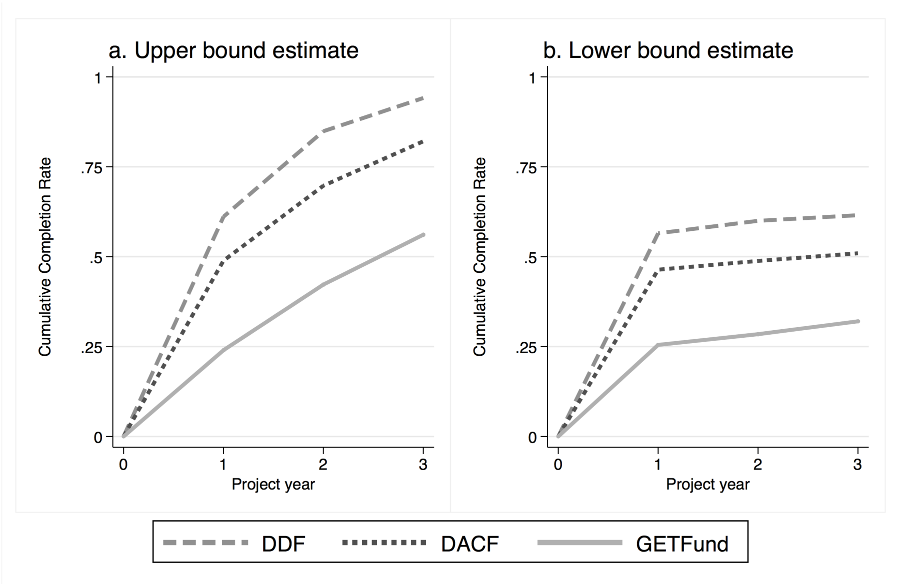


Figure A4 presents the three-year completion rates using these different estimation methods. The differences in the second and third years among the methods are large, although the key qualitative finding (that projects' completion hazard rate decreases after the first year) remains. However, Figure A5 (analogous to Figure 7 in the main text) shows that the differences across fund sources vary little across the estimation methodologies. This provides further reassurance that the main findings of the paper are not affected by concerns over attrition.

Figure A4: Project Completion by Fund Source – Alternative Estimates



APPENDIX 3.5: SAMPLE PROJECT PHOTOS

Figure A6: Incomplete Staff Bungalow

Figure A7: Abandoned Excavation for a Classroom Block, with an Additional Unfinished Classroom Block in the Background

Figure A8: Unfinished Classroom Block

Figure A9: Borehole at a School – In Progress

Chapter Four

ONE SIZE DOES NOT FIT ALL:

BUDGET INSTITUTIONS AND PERFORMANCE IN GHANA'S SPENDING MINISTRIES

Chapter Abstract

A large literature investigates the link between budget institutions and performance across countries, but little attention has been paid to variation across spending ministries *within* countries. Since these ministries operate under the same formal budget institutions and national-level context, they are the ideal sites in which to study the gaps between *de jure* and *de facto* budget practices, and between budget allocations and actual expenditures. These gaps can be substantial, especially in developing countries, but are poorly understood. This article demonstrates that there is large and systematic variation among spending ministries in Ghana in expenditure outturns and volatility, compliance with formal budget processes, and other outcomes such as extra-budgetary spending. The findings pose a challenge to theories of budget institutions and their reform: not only is there no correlation between procedural compliance and budget outcomes, but the drivers of ministries' budget performance are heterogeneous and often idiosyncratic. The existence of substantial variation in the quality of budget execution among ministries also has implications for the validity of analyses that use either budget allocations *or* actual expenditure to measure the outcomes of policy processes.

One Size Does Not Fit All:
Budget Institutions and Performance in Ghana's Spending Ministries

Research on budget institutions and performance has largely focused on the role of finance ministries, legislatures, and national-level political factors, with the behavior and performance of spending ministries treated as a consequence of these higher-level institutions. This approach is logical when budgets are well executed and political principals have strong control of the bureaucracy and expenditure, but outside wealthy democracies these conditions are the exception rather than the norm: in a sample of 45 mostly developing countries, for instance, Addison (2014) finds that nearly 70 percent of ministry budget heads are over- or underspent by 5 percent or more. Even in developed and transition countries, budget execution is a complex managerial task (Pitsvada 1983; Thompson 1993; Dougherty, Klase, and Song 2003) and budget overruns are common for certain types of public organizations (Kornai, Maskin, and Roland 2003; Kornai 2009) and projects (Flyvbjerg 2013; Engel, Fischer, and Galetovich 2009). Yet there is little theory or systematic evidence on budget execution in spending ministries (also referred to as line ministries or spending units) and its determinants. What scarce literature does exist (Stasavage and Moyo 2000, Simson and Welham 2014) suggests that there may be dramatic differences in budget execution among spending ministries even within the same country, but does not investigate this within-country variation systematically.

This article seeks to shed new light on these questions by examining variation in the quality of budget preparation and execution among spending ministries in Ghana's central government. This variation is particularly interesting because spending ministries are all subject to the same formal budget institutions and processes, as well as the same national-level political and bureaucratic context. Holding these constant brings into focus two prominent themes from recent research on public financial management (PFM) reform in developing countries: the widespread existence of significant gaps between *de jure* and *de facto* budget practices (Andrews 2010a; de Renzio, Andrews, and Mills 2011; Wehner and de Renzio 2013), and the importance of context and informal institutions in determining performance and reform success (Hallergberg, Scartascini, and Stein 2009; Andrews 2013; Cangiano, Curristine, and

Lazare 2013).[102] These studies have largely relied on country case studies and cross-national comparisons, with the result that the discussion of *de facto* practices remains somewhat impressionistic, raising doubts about whether broad generalizations about informality might sometimes conceal as much as they reveal. Killick, for example, calls Ghana's expenditure budget a "façade", "essentially ritualistic, with little bearing on reality" (2005, 1) – is this extreme emphasis on informality empirically justified? By examining variation in performance and practices among spending ministries within one country, this article adds empirical precision to the study of the relationship between the design and implementation of budget systems, and demonstrates that contextual factors and informal practices can vary dramatically even within the same country.

Empirically, the article begins by using fiscal data to confirm that ministry-level expenditure outturns in Ghana are characterized by large deviations from allocations, but that the extent of expenditure overruns and volatility varies widely across ministries. To understand the mechanisms and causes behind this variation in a context where the availability of administrative data is very limited (as is the case in most developing countries), the article then draws on structured interviews with 23 budget schedule officers in Ghana's Ministry of Finance (MoF) to construct two budget performance indices for each spending ministry: a Processes Index, reflecting how well the ministry fulfills the procedural requirements of budget preparation and execution; and an Outcomes Index, reflecting the incidence of extra-budgetary requests, accumulation of avoidable arrears, and strategic manipulation of activity costings. Both indices reveal significant variation among ministries. On the procedural side, some ministries execute formal budget processes almost perfectly, while others hardly carry them out at all. The same is true of budget outcomes: some ministries exhibit all the shortcomings of spending over and outside the budget which have been challenges for Ghana's PFM system as a whole, while other ministries are models of fiscal discipline. The variation across ministries is not well explained by structural factors such as budget size and complexity, expenditure composition, or revenue sources. Furthermore, the Processes and Outcomes indices are uncorrelated

[102] Although these themes have gained increased prominence and traction recently, they have also been pointed out by earlier authors such as Caiden and Wildavsky (1980) and Schick (1998).

with each other: the ministries that execute budget processes well are not necessarily the ministries with good outcomes, and vice versa.

Digging deeper into the mechanisms and causes underlying this disconnect reveals that, rather than sharing a common explanation, the drivers of ministries' budget performance can be quite idiosyncratic to organizations' particular circumstances. The article draws on qualitative interview data on extra-budgetary spending in three social-sector ministries to show that it has distinct causes in each case: poor technical planning in one ministry, a highly unpredictable sectoral context in another, and pressure stemming from political competition in a third. Much as the same symptom in a patient can be due to different diseases, the same problem in different ministries can have different causes. This diversity of practices, performance, and contexts across spending ministries poses the question of the extent to which it can be optimal to apply the same set of budget institutions and processes to all ministries within a given country. There is, of course, a need for some level of consistency and harmonization across different organizations in government, but the absence of any theoretical or empirical literature that directly addresses this question suggests that the homogenizing, one-size-fits-all approach to spending ministries in budgeting is by default rather than conscious design.

While budgeting and spending decisions have been the focus of key theoretical discussions of bureaucratic behavior (Wildavsky 1964, Niskanen 1971, Dunleavy 1991), this article is (to the author's knowledge) the first study to systematically document and explore *variation* in budget performance and management among spending ministries in a single country.[103] A handful of articles note some variation in expenditure outturns across ministries in developing countries but do not explore it in depth (Stasavage and Moyo 2000, Simson and Welham 2014). The most commonly used international PFM assessment tool, the Public Expenditure and Financial Accountability (PEFA) review, includes only one performance indicator (PI-2) that

[103] A large literature examines political control of executive agencies through administrative procedures (e.g. McCubbins, Noll, and Weingast 1987). I do not discuss this literature here because it focuses mainly on policy decisions rather than budget management, and its applicability in developing countries is limited by the poor enforcement of administrative rules. The relationship between administrative constraints and budget performance would be an interesting area for future research.

focuses on differences among spending ministries, and this is only in the context of aggregate compositional deviations in expenditure (PEFA Secretariat 2012). Likewise, the existence of systematic differences in budget execution across ministries has important implications for the validity of theories of policy processes that rely on budget *allocations* to measure policy change (e.g. Jones *et al* 2009; Breunig, Koski, and Mortensen 2010).

The remainder of the article proceeds as follows. The next section discusses theoretical perspectives on spending ministries in PFM, while Section 4.2 discusses budget institutions in Ghana and examines the distribution of expenditure outturns across ministries. Section 4.3 presents the interview methodology, discusses variation in budget processes and outcomes across ministries, and demonstrates that the two are uncorrelated. Section 4.4 draws on qualitative evidence to investigate the different underlying causes of one budget problem: extra-budgetary spending. The article concludes by discussing the implications of these findings for theories of the design, implementation, and reform of budget institutions.

4.1 CHALLENGES FOR THEORY

Because spending ministries typically appear as secondary, undifferentiated actors in theory and empirical research on budget institutions, there is little theoretical guidance for analyzing *variation* in budget performance among spending ministries. A large literature has examined how formal institutions and processes affect expenditure budgets and outcomes at the national level (Weingast, Shepsle, and Johnsen 1981; Ferejohn and Krehbiel 1987; von Hagen and Harden 1995; Alesina, Hausmann, Hommes, and Stein 1999), but these factors do not vary across spending ministries within a country. There is also a burgeoning literature on the distribution of changes in budget allocations in Western democracies and how this distribution is shaped by institutional and political features (Jones *et al* 2009). Some of this literature bears on within-country, organization-level variation by disaggregating among sectors, government functions, or expenditure types (Breunig and Koski 2006 & 2012) or analyzing differences across similar organizations, such as school districts (Robinson, Flink, and King 2014). However, studies of budget allocations have limited relevance for developing country contexts where large deviations of actual expenditures from allocations are the norm rather than the exception (Ablo and Reinikka 1998, Addison 2014).

In response to the perceived shortcomings of reforms aimed at formal budget structures and processes, a recent body of literature has documented and sought to explain the commonly observed gap between *de jure* and *de facto* budget practices in many developing countries (Schick 1998; Hallerberg, Scartascini, and Stein 2009; Andrews 2010b & 2013). The analytical focus of this work has also shifted emphasis from formal institutions as drivers of outcomes to country context, informal institutions, and political economy. While these perspectives provide a useful entry point for analysis of ministry-level variation in budget performance by demonstrating that the execution of formal rules and procedures cannot be taken for granted and by conceiving of budget outcomes through an organizational (rather than purely political or institutional) lens, their analytical focus has nonetheless been on country-level variation, with little discussion of the existence or potential causes of variation at the organizational level.[104]

While existing theory on budget institutions is thus of little help in explaining spending ministry-level variation, some broad sets of hypotheses can be derived from the limited empirical literature on spending ministry budgeting and other adjacent literatures. Stasavage and Moyo (2000) and Simson and Welham (2014) both document variation in expenditure outturns across ministries in several countries in Africa, and speculate that this variation could be attributable to differential political clout among ministries. Wildavsky (1975) suggests that uncertainty is a key variable for understanding budgetary processes, albeit in a country-level sense, and Dougherty, Klase, and Song (2003) describe "managerial necessity" in the face of uncertainty as

---

[104] The irregular enforcement of formal rules in cases of persistently poor budget execution bears a strong resemblance to the soft budget constraint syndrome, which has been influential in the literature on fiscal decentralization (Rodden, Eskeland, and Litvack 2003). Although it has not previously been applied to the context of spending ministries (to the author's knowledge), the soft budget constraint provides a useful way of conceptualizing the *ex ante/ex post* inconsistency in poor budget execution, and an alternative (although not mutually exclusive) explanation to Andrews's (2013) invocation of isomorphic mimicry for why developing country governments might adopt formal structures but not implement them.

motivating within-year rebudgeting in West Virginia cities. Dunleavy (1991) classifies different types of agencies according to their budget structures, and other authors examine challenges associated with particular types of spending, such as infrastructure (Flyvbjerg 2013), which differ in their incidence across ministries. A growing literature demonstrates significant variation in operational management quality and performance among public sector organizations in developing (Leonard 2010, Gingerich 2013) and developed (Ingraham, Joyce, and Donahue 2003; Bloom, Propper, Seiler, and Van Reenen 2015) countries; financial management quality and budget performance could be characterized by the same range.

While the following empirical analysis investigates the role of these broad sets of factors in explaining variation among spending ministries, it may also be the case that there is no single generalizable explanation for variation in ministries' budget performance – that the diversity of budget problems may be matched by a diversity of causes. Indeed, sector-specific approaches to budgeting often emphasize the idiosyncratic challenges associated with budget management in particular sectors: Johnson (2013), for example, reports that Ghana's Ministry of Food and Agriculture regularly faces difficulties because of misalignment between the national budget cycle and the growing seasons for various crops. To the extent that this particularity predominates over more general explanations for variation, it poses an even stronger theoretical challenge to the application of homogenous budget institutions and processes to all spending ministries. Engaging with the operational idiosyncrasies of specific sectors is untidy but also increasingly necessary as PFM research and theory struggles to move beyond its overwhelming emphasis on fiscal discipline (Wehner and de Renzio 2013) to develop a stronger understanding of the linkages between budget institutions and operational efficiency.

4.2 BUDGET EXECUTION IN GHANA

Budget execution has historically been a problem for Ghana, as for many developing countries, even though Ghana's formal budget institutions are conducive to a high degree of control by the MoF and have been the subject of over 15 years of donor-backed reform efforts. Budget formulation is highly centralized within the MoF, with Parliament's role being limited in practice to approving the budgets proposed by

Government and *ex post* monitoring. Ghana's use of cash budgeting means that releases for budgeted activities must be requested and approved during budget execution, under general monthly warrants for goods and services and project-specific warrants for investment or asset expenditures. These releases are often subject to significant delays because of cash flow constraints, and this gives the MoF an effective veto and thus an even higher degree of discretion in budget execution.[105] Internal structures and processes for budgeting, planning, and monitoring within ministries are also subject to a rigid and highly prescriptive legal and administrative framework, aimed at ensuring uniformity across ministries. However, there is a significant gap between how these institutions work on paper and in practice, as noted by the country's 2012 PEFA performance review: "While the rules and procedures are generally clearly established and described in the regulatory framework and internal guidelines, they are often not adhered to. Compliance to formal procedures and/or use of informal procedures seem to be a challenge across many areas in the budget execution." (ECORYS 2013, 15) On the surface, then, Ghana seems a textbook example of a PFM system that is plagued by chronic overspending as well as a gap between *de jure* and *de facto* budget practices.

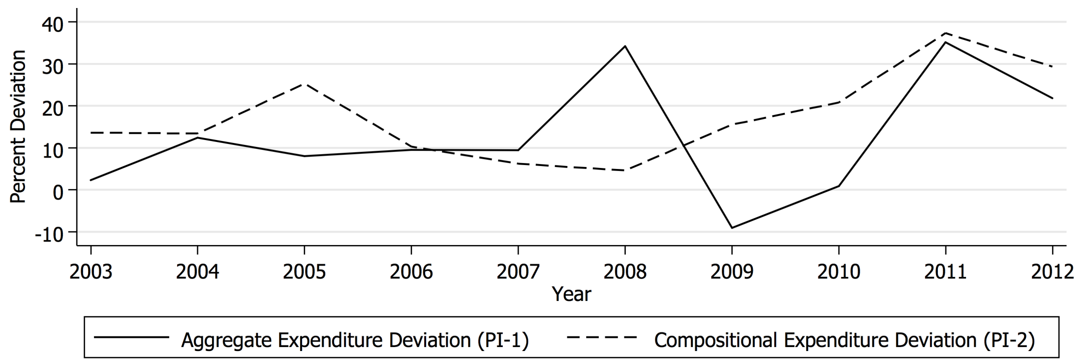
Figure 1 confirms Ghana's problems with expenditure control. From 2003-12, Ghana's aggregate primary expenditure (PEFA PI-1) exceeded the budget in every year except 2009 (a post-election year of fiscal consolidation), overrunning by eight percent or more in seven of those years.[106] The budget head-level compositional

[105] Salary levels are centrally determined, and are paid directly by the Controller and Accountant-General's Department. Ghana changed its budget classification scheme in 2012. For the purposes of this article, non-personnel expenditure is treated as spending in the categories of Services and Investment for 2009-11, and in the categories of Goods and Services and Assets for 2012. The budget procedures described here applied to the 2009-12 period. In 2011 Ghana began a reform to transition to programme based budgeting, but this did not take full effect until 2013. Programme budgets were prepared in a small number of ministries in 2011 and 2012 as a pilot, but these ministries also prepared the standard activity-based budgets and these served as the effective basis for budget management during these years.

[106] These figures are based only on appropriations and outturns for Ghana's Consolidated Fund, as the official expenditure reports produced by Ghana's Controller and Accountant General do not report on spending from donor projects or programs, ministries' internally generated funds (IGF), or statutory funds. In some cases ministries may be able to use these resources to compensate for unpredictability

deviation in expenditure (PEFA PI-2) averaged 17.63 percent, and was within ten percent in only two of these years.[107] The combination of aggregate overspending with high compositional deviation leads to the perverse situation that while the country as a whole suffers from chronic overspending, individual spending ministries experience unpredictability and frequent shortfalls of resources.

Figure 1: Aggregate and Compositional Expenditure Deviations in Ghana, 2003-12



Note: Figures for 2003-11 from World Bank (2006) and ECORYS (2010 & 2013); figures for 2012 calculated by author from Controller and Accountant-General (2013). Covers Consolidated Fund expenditure only. PI-1 is the percentage deviation of actual from budgeted aggregate expenditure from the Consolidated Fund. PI-2 is the percentage deviation of actual from budgeted expenditure at subhead level for the top 20 expenditure headings (ministries), after adjusting the subhead-level budgeted expenditure to account for aggregate expenditure deviations.

Figure 2 shows this unpredictability more vividly, using the distribution of ministry outturns for non-personnel expenditure from 2009-12 expressed as a ratio of budget allocations (i.e., zero indicates that a ministry's actual expenditure exactly matched its allocation, and one indicates that the ministry overspent its budget by 100 percent).[108]

---

of Consolidated Fund releases, but these alternative sources of funds are distributed highly unequally across ministries and (with the exception of IGF) are usually earmarked for specific purposes. The Consolidated Fund represents approximately 88% of overall central government operations, excluding donor funds and transfers to local government (ECORYS 2013, 42).

[107] The formulas for PEFA PI-1, aggregate expenditure deviation, and PEFA PI-2, compositional expenditure deviation, are:

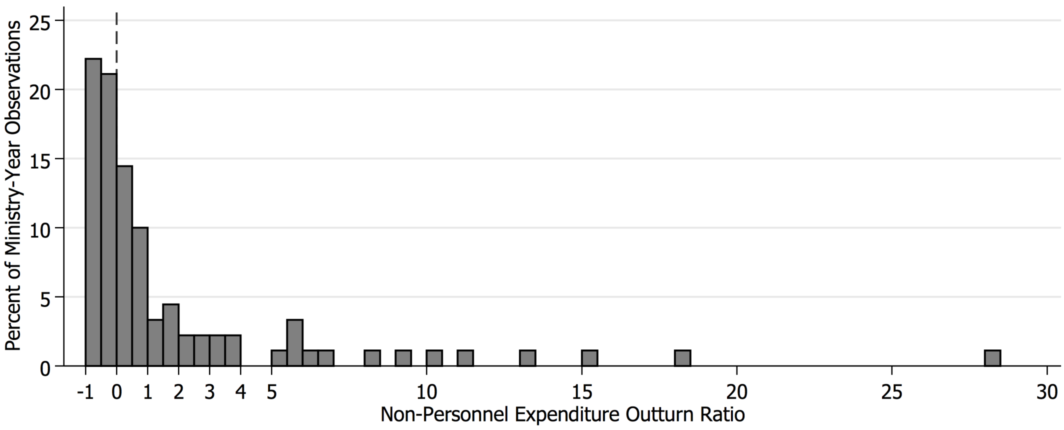
(PI-1) $\frac{\sum_i^I Actual_i - \sum_i^I Allocation_i}{\sum_i^I Allocation_i}$ (PI-2) $\frac{\sum_i^I |Actual_i - Allocation_i|}{\sum_i^I Actual_i}$

for all ministries i.

[108] Throughout this paper, I calculate each ministry's mean expenditure outturn ratio as:

The focus is on spending on activities rather than personnel because personnel spending is centrally determined and paid, and thus largely out of the control of individual ministries. Despite high aggregate budget deficits, ministries' outturns were less than their budget allocation for over 40 percent of the observations. At the same time, one third of ministry-year observations overspent their allocations by 100 percent or more. There is a long tail of extremely high outturns, with one in ten ministry-year observations showing overspending of greater than 674 percent – up to a maximum of 20,461 percent. These figures actually *understate* the unpredictability of releases from the perspective of spending ministries, since the year-end aggregates mask project-level variation as well as delays in the timing of releases.

Figure 3: Average Non-Personnel Outturn Ratio by Ministry, 2009-12



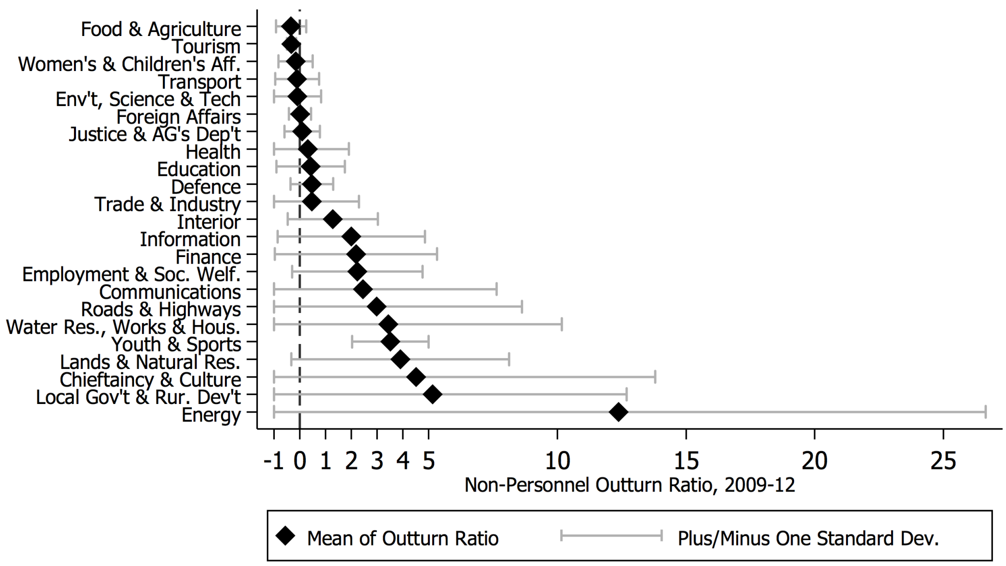
Note: Author's calculations from CAGD (2010, 2011, 2012, 2013). Non-personnel expenditure calculated as Services plus Investment for 2009-11, and Goods and Services plus Assets for 2012. Covers Consolidated Fund expenditure only. Ministry of Energy 2011 outturn ratio of 204.6 (20,461%) is excluded as an outlier.

$$\frac{\sum_t^T \frac{Actual_t - Allocation_t}{Allocation_t}}{T}$$

for years *t* from 2009-2012. For allocations I use original budget allocations for 2010-12, since the original allocations better reflect the information that spending ministries use to plan for the year. However I use the supplementary budget allocations for 2009, since the previous budget was passed under a different government prior to the December 2008 election and thus reflected a different set of priorities.

However, the extent of these problems varies dramatically across ministries. Figure 3 shows each ministry's mean outturn on non-personnel expenditure for 2009-12, as well its standard deviation. Some ministries exceed their budget by huge margins while others receive less than their allocation on average, and some ministries also experience more resource volatility than others. The variation across ministries is not driven by a handful of outlier years – recalculating the mean outturn ratio after dropping the highest year for each ministry returns a very similar ranking, and the two means for each ministry have a correlation of 0.90.

Figure 3: Average Non-Personnel Outturn Ratio by Ministry, 2009-12



Note: Author's calculations from CAGD (2010, 2011, 2012, 2013). Non-personnel expenditure calculated as Services plus Investment for 2009-11, and Goods and Services plus Assets for 2012. Covers Consolidated Fund expenditure only. Minus one standard deviation range is left-truncated at -100%, equivalent to a 0% outturn. Ministry of Energy 2011 outturn ratio is excluded from all calculations as an outlier.

This dramatic variation in expenditure outturns is not inconsistent with the perspective that emphasizes informal budget institutions and the gap between *de facto* and *de jure* budget practices, but it does demand further explanation. Given the exact same formal budget institutions and processes, why do some ministries massively overspend their budgets while others spend below their allocations? Budget outturns in themselves are uninformative about the roots of these problems because they are the joint outcome of actions by the MoF and spending ministries. Overspending could be due to poor planning by the spending ministry or to a deliberate political decision from the core executive, for example. Many different factors could be at work, and simply noting that there is a gap between *de jure* and *de facto* budget practices neither helps us understand which of these factors is driving outcomes nor explains variation across ministries. This interpretational ambiguity motivates the novel measures of budget performance introduced in the next section.

4.3 DIFFERENT MINISTRIES HAVE DIFFERENT PROBLEMS

Systematic studies of *de facto* budget practices have been hampered by the difficulty of measuring informal practices and the lack of administrative data in many developing countries. Aside from actual budget outturns, publicly available data on spending ministries' budget performance is scarce; in Ghana, even the MoF itself does not routinely track other indicators of spending ministries' performance. To get around this, this section draws on structured interviews with 23 budget schedule officers in the MoF to construct two budget performance indices for each of Ghana's 23 spending ministries: a Processes Index and an Outcomes Index.

Measuring Processes and Outcomes

Each budget schedule officer is responsible for one spending ministry and processes all budget transactions for that ministry throughout the budget cycle. They interact with their schedule ministries on a daily basis and are therefore very well informed not only about the formal aspects of that ministry's budget performance, but also about informal and contextual factors. At the time of the interviews, officers had been handling their schedule ministry for an average of 2.6 years, had worked in the Budget Division for an average of 11.5 years, and had all served on the schedule of at least one other ministry previously, so they were knowledgeable about their schedule ministry and well positioned to make comparative judgments about their performance over a multi-year period.

The Processes Index is composed of six formal, procedural elements of budgeting with which all ministries are required to comply:

- Annual budget submission
    1) **Timeliness**
    2) **Completeness**
    3) **Quality of costing**
- In-year release requests
    4) **Timeliness**
    5) **Completeness**
- In-year financial reports
    6) **Regularity of submission**

The Outcomes Index comprises schedule officer assessments of the extent to which budget implementation in the ministry is characterized by three negative outcomes:

1) **Extra-budgetary requests**. How frequently the ministry makes in-year requests for significant expenditures outside of the annual budget.
2) **Strategic cost manipulation**. The extent to which the ministry deliberately underestimates activity/project costs in the annual budget in order to include additional projects on the budget, under the assumption that this will allow them to start the project and then lobby for additional funds during the year.[109]
3) **Avoidable payment arrears from unapproved/unbudgeted commitments**. Whether accumulation of significant payment arrears is a common occurrence in the ministry, either due to the ministry entering into commitments before receiving approval or to making extra-budgetary requests that receive the MoF's approval but for which funds are not available since the expenditure was not budgeted for.[110]

[109] This phenomenon is similar to Engel, Fischer, and Galetovich's (2009) description of how governments and firms instrumentalize soft budget constraints in the context of public-private partnerships in infrastructure to increase spending, with resultant distortions in project procurement, planning, and management. Studying the connection between ministry-level PFM processes and project-level outcomes would be a fruitful area for future study; see Flyvbjerg (2013) for one such effort.

[110] In both of these cases, the arrears are primarily attributable to actions and decisions of the ministry rather than the MoF. Arrears arising from genuine commitments (budgeted for and approved by the MoF) for which cash is simply not available are excluded, since this is out of the ministry's control.

The design of the indices and interviews incorporated a number of features intended to reduce various sources of bias. The indices focus on aspects of budget management that are easy for schedule officers to discuss objectively, and thus exclude processes and outcomes that either require highly subjective evaluations or about which schedule officers may not be fully informed. Because there are well known biases associated with directly asking respondents for subjective assessments (Schwarz 1999, Meier and O'Toole 2014), and following recent methodological innovations in the empirical management literature (Bloom and Van Reenen 2010), the interviews discussed each ministry's performance in objective terms, starting with open questions followed by probing and requests for examples in order to establish the actual state of a practice or incidence of a particular problem in the ministry.

Responses to each item were transcribed by hand and subsequently post-coded by the author on the scale [0 = Not a problem, 1 = Minor problem, 2 = Major problem]. For the timeliness of annual budget requests, for example, a response of "They are on time, but sometimes with a day or two lag - we have to chase them" was coded as 1, "Other [ministries] try to bring them on time, but [this ministry] always delays, I have to chase them to bring it" was coded as 2, and "They always meet the deadlines we give them" was coded as 0. To minimize concerns of potential interviewer bias and ensure consistency across ministries, the codings were anchored to the qualitative interview transcriptions. This was considered preferable to asking schedule officers directly for a subjective or numerical rating since the interpretation of the scale and evaluation criteria would likely be inconsistent across individuals, possibly in ways that would be correlated with the ministry's performance. [111]

Importantly, all interviews were conducted anonymously both with respect to the schedule officer and the ministry being discussed. This step was necessary to elicit frank responses, but has the consequence that the analysis in the following sections cannot reveal the performance of specific ministries. To aggregate the items into

[111] The schedule officer interviews focused on the schedule officer's entire experience with the ministry, not just the past year, in order to capture persistent patterns rather than one-off fluctuations. Schedule officers did not have any difficulty in making this distinction.

indices, each item was transformed into a normalized z-score and averaged, with any missing items excluded. The resulting index was then reconverted to a z-score and reversed so that the index is centered on zero with standard deviation of one, and positive scores indicate better performance.

The coding scale for the two indices is coarse but nevertheless informative. It is difficult to validate the indices against other data sources, as the creation of the indices was motivated precisely by the lack of existing data. However, the Outcomes Index has a strong (-0.44) and statistically significant correlation with non-personnel expenditure outturns, and this is in the expected direction – worse scores on the Outcome Index are correlated with higher levels of overspending. While this correlation would not be expected to be perfect, since the Outcomes Index includes elements of budget performance other than simply fiscal discipline, it is nonetheless reassuring that these two measures have the expected correlation.

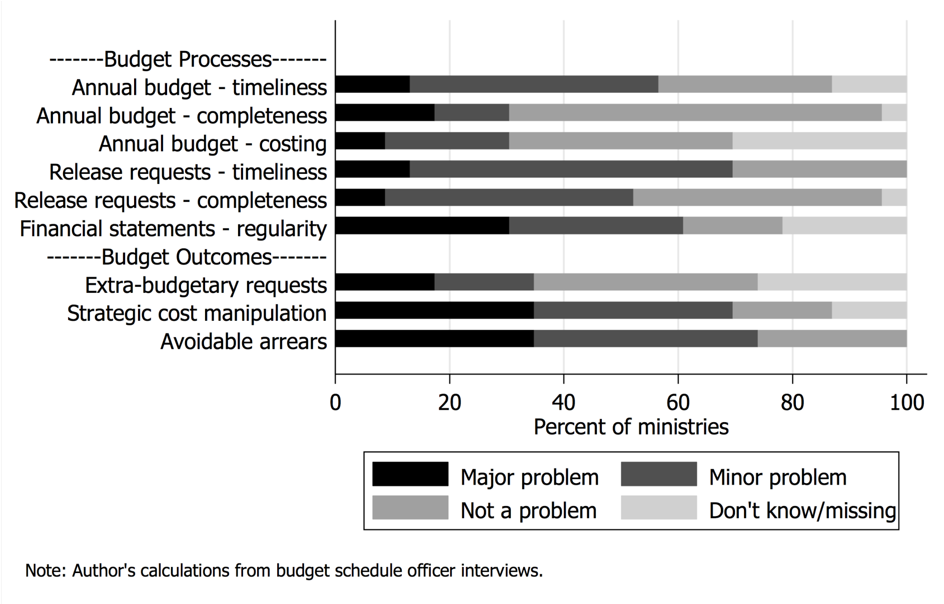
While the item scores in the Outcomes Index may be less precise than actual expenditure outturns, they are in many ways more informative because they cover dimensions of performance that are not necessarily reflected in raw outturns. Extra-budgetary requests, for example, cause cash flow problems for the MoF and undermine budget credibility, and strategic cost manipulation often has negative consequences for project planning and delivery. Furthermore, the items in the Outcomes Index can all be attributable to actions or inactions by the spending ministry (at the technical or political level), whereas budget outturns are a joint outcome of actions taken by the spending ministry and the MoF. In the absence of detailed administrative data – as is the case for most developing countries – these interview-based measures provide the most rigorous possible measure of spending ministry budget performance.

Analyzing Variation in Processes and Outcomes

Figure 4 shows that there is a significant amount of variation across every item in the two indices. Even though submission of in-year financial reports (for example) is a minor or major problem for most ministries, some ministries do submit them regularly. The same holds true for the Outcomes items: while extra-budgetary requests and arrears are problems for most ministries, other ministries avoid them.

Each item is a problem for some ministries but not for others; sometimes *de facto* actually does line up with *de jure*, and sometimes it does not.

Figure 4: Performance on Processes and Outcomes Items



Overall some ministries perform better than others, but it is not the case that some ministries do everything well and others do everything badly. Table 1 presents the item scores for ministries representing the 25th, 50th, and 75th percentiles of each index. The 25th-percentile ministry on the Processes Index, for example, has major problems with its annual budget costing and in-year release request timeliness, minor problems with annual budget timeliness and in-year request completeness, but no issues with annual budget completeness or submission of in-year financial reports. The 75th-percentile ministry, by contrast, has a major problem with timeliness of in-year release requests, but no other problems. The incidence of negative outcomes also varies: the ministry at the 25th percentile of the Outcomes Index has major or minor problems on all three items, whereas the 75th-percentile ministry only has a minor problem with extra-budgetary requests. Different ministries have different problems.

Table 1: Processes and Outcomes in Selected Ministries

Budget Processes	25th-percentile ministry	50th-percentile ministry	75th-percentile ministry
Annual budget submissions			
Timeliness	Minor	Not	Not
Completeness	Not	-	Not
Costing	Major	Major	Not
In-year release requests			
Timeliness	Major	Minor	Major
Completeness	Minor	Not	Not
In-year financial reports			
Regular submission	Not	-	-
Processes Index score	-0.74	-0.07	0.58
Budget Outcomes	25th-percentile ministry	50th-percentile ministry	75th-percentile ministry
Extra-budgetary requests	Major	Minor	Minor
Strategic cost manipulation	Minor	-	Not
Arrears	Major	Minor	Not
Outcomes Index score	-1.04	0.19	0.94

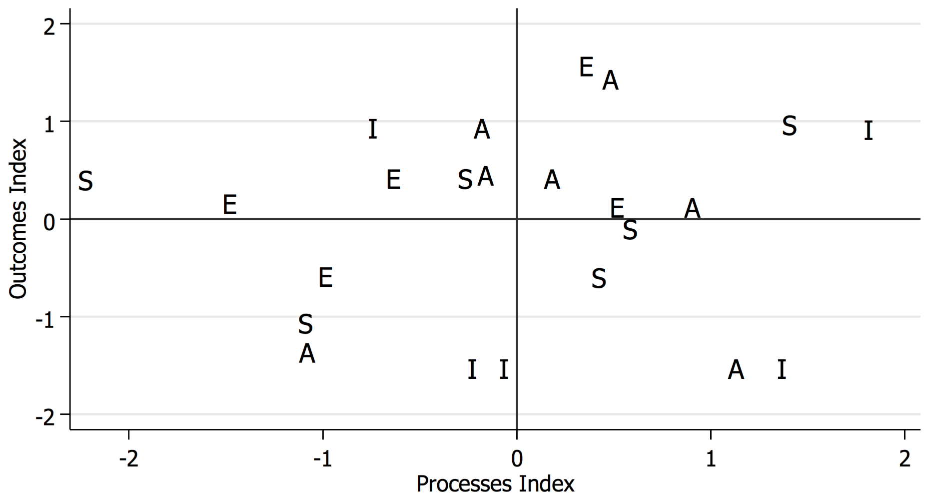
Note: Author's calculations from budget schedule officer interviews. Dashes indicate Don't know/missing data.

Equally surprisingly, there is no correlation between ministries' performance on the Processes and Outcomes Indices, as shown by Figure 5. Some ministries do well on Processes and also have good Outcomes, while others that also score highly on Processes have very poor Outcomes. There is a similar range of Outcomes scores for ministries with poor Processes. These patterns are not attributable to differences inherent in the sectoral task characteristics of ministries: while infrastructure-sector ministries score better on Processes and worse on Outcomes than other ministries (on average), these differences are not statistically significant and most of the variation is within rather than between sectors.[112] More generally, Appendix 4.1 demonstrates that the structural characteristics of ministries – expenditure composition, revenue sources, budget size and complexity, and sector – are only weakly (if at all) correlated with

[112] Sector classifications are based on Government of Ghana's classification in budget statements, but with the Ministry of Energy reclassified as Infrastructure sector rather than Economic due to its focus on public investment activities. Treating it as an Economic sector agency further weakens these slight differences across sectors.

budget performance.[113] While these structural characteristics may matter for budget performance, their impact is far from deterministic.

Figure 5: Budget Processes and Budget Outcomes



Note: Author's calculations from budget schedule officer interviews. Both indices calculated as z-scores. Positive scores indicate better performance. Sector labels: A = Administration & Public Safety, E = Economic, I = Infrastructure, S = Social.

Again, the lack of an obvious relationship between budget processes and outcomes does not necessarily indicate the absence of a causal relationship, since there are many other mediating factors. However it does confirm that, at the ministry level, adherence to formal procedural requirements is neither a necessary nor sufficient condition for good budget outcomes. It also paints a more nuanced picture of the interaction of *de jure* and *de facto* practices than is often portrayed by authors that emphasize the role of informality in developing country budget processes: formal procedures are not always followed but sometimes they are, and some ministries have poor budget

[113] The weak correlation between budget performance and ministries' expenditure profiles is particularly surprising. As is common worldwide, Ghana applies different budget processes to these expenditure categories, with the implicit assumption that the necessary control and disbursement procedures vary according to the type of expenditure, rather than the spending organization – that ministries as organizations "matter" only through their spending profiles. These results suggest otherwise.

outcomes while others do not. Understanding this diversity of performances is key to designing and implementing realistic reforms.

4.4 SAME SYMPTOM, DIFFERENT DISEASES

This section takes the analysis of diversity among spending ministries a step further by showing how the same problem can arise in different ministries for different reasons. Not only do different ministries have different problems, but what appears to be the same problem can have different causes. This section demonstrates this by supplementing the quantitative indices with qualitative interview data to illustrate how three social sector ministries which have a major problem with extra-budgetary requests each have a different reason for doing so. While the preceding analysis suggested that factors other than formal structures and processes are driving variation in budget performance across ministries, this section cautions against assuming that these “other” factors – political economy, informal institutions, and context – are the same for all ministries within a country.

Ministry X has very poor planning and financial management, with the lowest Processes score of any ministry. Their annual budget submissions are always late – the schedule officer has to “chase” them to submit it. Even when it is submitted it does not conform to the guidelines given by the MoF, and the costing is “sometimes not proper”. Their in-year release requests are also consistently late and incomplete, and their financial reports are not filed. The schedule officer for Ministry X reports that the ministry also has a major problem with extra-budgetary requests, not because of political pressure or unpredictable events, but simply because each year they make requests for activities that they had put in their annual budget submission but which did not make it into the budget.

Another ministry, Ministry Y, is above average on the technical aspects of budget management, but also consistently makes significant extra-budgetary requests. In their case, however, they are responsible for providing funding for activities whose occurrence and timing is entirely out of their control, so extra-budgetary requests are almost inevitable. These unexpected requests can present significant cash flow problems for the MoF, especially since resource constraints and aggregate overspending prevent the accumulation of an adequate central contingency reserve. In these cases the MoF typically releases part of the money for the activity, the ministry somehow sources the remaining funds to enable the activity to take place – often off-budget – and the difference ends up as payment arrears.

In a third social-sector ministry, Ministry Z, extra-budgetary requests are also a major problem, due not to poor administration or unpredictable events but to political pressure. The ministry is large and high profile so politicians frequently make off-the-cuff promises to communities which then have to be catered to immediately, through extra-budgetary requests. (Yet political salience alone does not appear to be sufficient to cause these problems – the other social-sector ministry which is similarly large and prominent does not experience this type of problems with extra-budgetary spending.)

This illustrates why it is important for reforms to take into account not just the variation in budget performance among ministries, but also its causes. And just as a doctor's treatment must focus on curing the disease rather than merely treating the symptoms, improving budget performance requires identifying and dealing with the different underlying problems in each ministry. Technical fixes, like stricter procedural controls or training, may be effective at reducing extra-budgetary requests in Ministry X, but would be useless for Ministries Y and Z. Greater commitment to fiscal discipline by politicians could help for Ministry Z, but would do little for Ministries X and Y. And in the absence of an adequate contingency fund, Ministry Y is likely to continue to accumulate arrears due to extra-budgetary requests regardless of anything done at the technical or political levels. The same point – that a common symptom may have different causes – can also be made about other aspects of budget performance discussed above, such as strategic cost manipulation, arrears, or even execution of basic budget procedures.

4.5 DISCUSSION AND IMPLICATIONS

This article has sought to demonstrate that different spending ministries in Ghana vary in their budget performance, in terms of overall budget execution, other budget outcomes, and the implementation of procedural requirements, and that the underlying drivers of budget performance also differ across ministries. In doing so, it suggests that integrating spending ministries into positive theories of budget performance and reform is a crucial task for future research, especially in developing countries where budget execution and effective control of the bureaucracy cannot be taken for granted. It argues that this will require engaging with the operational and contextual idiosyncrasies of specific organizations. The article also sheds new light on the relationship between *de jure* and *de facto* budget institutions and their link to budget performance. There is widespread agreement in recent literature on PFM and governance more generally about the importance of informal institutions and context, but most existing research treats them as an amorphous residual category that is occasionally invoked on an *ad hoc* basis to explain why reality does not match what is on paper. The challenge for research now is to define and measure these phenomena precisely enough to integrate them rigorously into theory and analysis, so that the role of context in reform design can move beyond caveats and vague warnings.

The article also takes some preliminary steps towards addressing budget outcomes beyond simple fiscal discipline. Linking research on financial and operational management is a crucial area for future study, since access to resource inputs is a critical element in public good delivery of public goods and operational efficiency is an important goal of budgeting. In highlighting the uneasy fit between diverse ministries and a homogenizing set of budget institutions, this article illustrates the theoretical and empirical challenge that this will pose. The article also contributes to this by introducing a structured interview methodology that addresses multiple levels of budget performance in an integrated (although by no means comprehensive) way, and is particularly suited to developing country contexts where availability of administrative data is limited.

Variation among ministries also has significant normative implications for theories of the implementation, design, and reform of budget systems. As a matter of implementation, responding to the diversity among ministries would mean recognizing that ministries (like countries) have different arrays of contextual factors and baseline performance, and reforms will therefore face different sets of challenges and require different interventions. Potential responses to this challenge could include defining administrative indicators of procedural performance, fiscal discipline, or planning and monitoring processes, and conditioning certain aspects of budget institutions (such as levels of discretion in expenditure) on compliance with minimum requirements – an institutionalization of the “basics first” approach (Schick 1998).

This may be an area where practice has unconsciously crept ahead of theory: for example, the UK Treasury delegates different spending powers to different departments (“Treasury consents”), taking into account not only the characteristics of particular expenditure types but also the past performance of the department (Her Majesty’s Treasury 2013); Schick (2003, 78) reports that the Thai government has designed a “hurdle” system that grants broader discretion to agencies that meet certain minimum standards. The US federal budget process is known for Congress’s extensive use of earmarks (which are of course agency-specific), but these coexist with a multitude of authorization forms that give agencies more flexibility than traditional one-year direct budget appropriations. Likewise, in Ghana political and organizational differences among ministries (e.g. in technical capacity or political salience) are often taken account of informally through internal discussions within the MoF and the core executive. However, because there is little deliberate consideration of how best to do this, it is done on an informal or *ad hoc* basis. A more systematic approach could help provide guidance about how budget institutions can best recognize and accommodate the diverse sets of organizations that they serve.

As a matter of budget system design, researchers and reformers alike need to ask whether one-size-fits-all approaches to budgeting might be as problematic within countries as they are between countries, given the diverse contexts within which ministries operate. The need for some amount of harmonization and consistency at the national level is clear, but is the current homogenizing approach optimal or merely a taken-for-granted relic of a purely accounting approach to budgeting? In practice the distinction between design and implementation is of course blurred, since most reforms take place over the medium- to long-term. Criticizing the tendency to impose overdetermined, multi-stage, long-term reform plans, Allen (2013) argues that PFM reform efforts at the country level should focus on “selectivity” rather than “sequencing” – picking the most important and winnable battles to fight, rather than trying to impose a unified blueprint – and this may also be true *within* countries. Similarly, Andrews (2006) recommends that reformers should move beyond the binary debate about “basics first” or “best practices” approaches: recognizing variation among ministries may allow countries to prioritize “budget basics” where necessary while still allowing space to go “beyond” basics to more sophisticated or flexible processes in other ministries.

4.6 CONTRIBUTIONS

This paper’s most direct theoretical contribution is to the theoretical and empirical literature on budget institutions. Whereas the vast majority of this literature has studied the design and consequences of budget institutions at the national level (Weingast, Shepsle, and Johnsen 1981; Ferejohn and Krehbiel 1987; von Hagen and Harden 1995; Alesina, Hausmann, Hommes, and Stein 1999; Hallerberg, Scartascini, and Stein 2009), this paper focuses on measuring and explaining variation across organizations within the same government. This is an area where both theory and evidence are significantly more limited – Stasavage and Moyo (2000) and Simson and Welham (2014) are the rare exceptions that do document and discuss variation in budget execution and performance across organizations within the same government. Although there is little well-established theory to explain variation across ministries, this paper synthesizes and examines some plausible hypotheses and finds that the observed variation is not well explained by any single one of these. Instead, different causes seem to be driving outcomes in different ministries, often in ways that are quite idiosyncratic to particular sectoral characteristics or contexts. While the implications of this finding are not simple or straightforward, they are nonetheless theoretically consequential and present a significant challenge for future theory generation and further empirical studies.

The paper also makes secondary contributions to three related literatures on budgeting and public financial management. First, it extends the sparse but important literature on budget execution in developing (Ablo and Reinikka 1998, Addison 2014) and developed (Pitsvada 1983; Thompson 1993; Dougherty, Klase, and Song 2003) countries. In doing so it also presents a new theoretical argument, and introduces an interview-based mixed-method data collection methodology to engage more deeply with the informality and contextual uncertainty that are salient features of real-world budget execution.

Second, this focus on organization-level budget execution suggests that the literatures on soft budget constraints among sub-national governments (Rodden, Eskeland, and Litvack 2003; Rodden 2006) and state-owned enterprises and other public organizations (Kornai, Maskin, and Roland 2003; Kornai 2009) could also be usefully extended to the context of budget execution in central government ministries (although this exercise is beyond the current scope of this paper).

Third, the empirical documentation and theorizing of systematic differences in budget execution among ministries has methodological implications for theories of political decision-making that use budget allocations as measures of policy change (e.g. Jones *et al* 2009; Breunig, Koski, and Mortensen 2010). If actual disbursements differ systematically from allocations, then the latter may not be a valid indicator of policymakers' true policy intentions. Alternatively, and more interestingly, allocations and actual disbursements may each reflect the policy preferences of *different* sets of actors within government. Exploring this question would be a fruitful avenue for future research.

In examining variations in management and performance among organizations that operate within identical formal budget institutions, this paper also helps shed light onto the sector-specific contextual and political economy factors that are highlighted by the policy-oriented literature on the reform of budgets and public sector institutions more broadly (Schick 1998; Killick 2005; Andrews 2010b & 2013; de Renzio, Andrews, and Mills 2011; Cangiano, Curristine, and Lazare 2013; Wehner and de Renzio 2013). As discussed above, the principal conclusion of this paper is that the operation of these factors is important, yet often quite idiosyncratic to particular sectoral contexts. While future research should seek to construct and test more systematic theories about how these factors are determined and affect ministries' budget performance, an alternative (or possibly complementary) approach to academic investigation as well as practical reform might start by engaging with the particular problems of particular organizations rather than assuming a common diagnosis and common prescription.

REFERENCES

Twenty-three anonymous interviews with budget schedule officers in Ghana's Ministry of Finance.

Ablo, Emmanuel, and Ritva Reinikka. 1998. "Do Budgets Really Matter? Evidence From Public Spending on Education and Health in Uganda." World Bank Policy Research Working Paper Series 1926.

Addison, Douglas. 2013. "The Quality of Budget Execution and Its Correlates." World Bank Policy Research Working Paper 6657, October.

Alesina, Alberto, Ricardo Hausmann, Rudolf Hommes, and Ernesto Stein. 1999. "Budget Institutions and Fiscal Performance in Latin America." *Journal of Development Economics* 59: 253-273.

Allen, Richard. 2013. "Challenges of Reforming Budgetary Institutions in Developing Countries." In *Public Financial Management and Its Emerging Architecture*, eds. Marco Cangiano, Teresa Curristine, and Michel Lazare. Washington, DC: International Monetary Fund, 411-430.

Andrews, Matt. 2006. "Beyond 'Best Practice' and 'Basics First' in Adopting Performance Budgeting Reform." *Public Administration and Development* 26: 147-161.

Andrews, Matt. 2010a. "How Far Have Public Financial Management Reforms Come in Africa?" HKS Faculty Research Working Paper Series, RWP10-018, John F. Kennedy School of Government, Harvard University.

Andrews, Matt. 2010b. "Good Government Means Different Things in Different Countries." *Governance* 23 (1): 7-35.

Andrews, Matt. 2013. *The Limits of Institutional Reform in Development: Changing Rules for Realistic Solutions*. Cambridge: Cambridge University Press.

Bloom, Nicholas, and John Van Reenen. 2010. "New Approaches to Surveying Organizations." *American Economic Review Papers and Proceedings* 100 (May): 105-109.

Bloom, Nicholas, Carol Propper, Stephan Seiler, and John Van Reenen. 2015. "The Impact of Competition on Management Quality: Evidence from Public Hospitals." *Review of Economic Studies* 0: 1-133.

Breunig, Christian, and Chris Koski. 2006. "Punctuated Equilibria and Budgets in the American States." *Policy Studies Journal* 34(3): 363-379.

Breunig, Christian, and Chris Koski. 2012. "The Tortoise or the Hare? Incrementalism, Punctuations, and Their Consequences." *Policy Studies Journal* 40(1): 45-67.

Breunig, Christian, Chris Koski, and Peter B. Mortensen. 2010. "Stability and Punctuations in Public Spending: A Comparative Study of Budget Functions." *Journal of Public Administration Research and Theory* 20: 703-722.

Caiden, Naomi, and Aaron Wildavsky. 1980. *Planning and Budgeting in Poor Countries*. New Brunswick: Transaction Press.

Cangiano, Marco, Teresa Curristine, and Michel Lazare. 2013. "The Emerging Architecture of Public Financial Management." In *Public Financial Management and Its Emerging Architecture*, eds. Marco Cangiano, Teresa Curristine, and Michel Lazare. Washington, DC: International Monetary Fund, 1-17.

Controller and Accountant-General's Department. 2010. *Report and Financial Statements of the Controller and Accountant-General of the Public Accounts on the Consolidated Fund for the Period Ended December 31, 2009.* Accra.

Controller and Accountant-General's Department. 2011. *Report and Financial Statements of the Controller and Accountant-General of the Public Accounts on the Consolidated Fund for the Period Ended December 31, 2010.* Accra.

Controller and Accountant-General's Department. 2012. *Report and Financial Statements of the Controller and Accountant-General of the Public Accounts on the Consolidated Fund for the Period Ended December 31, 2011.* Accra.

Controller and Accountant-General's Department. 2013. *Report and Financial Statements of the Controller and Accountant-General of the Public Accounts on the Consolidated Fund for the Period Ended December 31, 2012.* Accra.

de Renzio, Paolo, Matt Andrews, and Zac Mills. 2011. "Does Donor Support to Public Financial Management Reforms in Developing Countries Work? An Analytical Study of Quantitative Cross-Country Evidence." Overseas Development Institute Working Paper 327, April.

Dougherty, Michael John, Kenneth A. Klase, and Soo Geun Song. 2003. "Managerial Necessity and the Art of Creating Surpluses: The Budget-Execution Process in West Virginia Cities." *Public Administration Review* 63(4): 484-497.

Dunleavy, Patrick. 1991. *Democracy, Bureaucracy, and Public Choice: Economic Explanations in Political Science.* London: Prentice-Hall.

ECORYS. 2010. *Republic of Ghana Public Expenditure and Financial Accountability 2009*. Rotterdam.

ECORYS. 2013. *Ghana: Public Expenditure and Financial Accountability Performance Review.* Rotterdam.

Engel, Eduardo, Ronald Fischer, and Alexander Galetovich. 2009. "Soft Budgets and Renegotiations in Public-Private Partnerships." National Bureau of Economic Research Working Paper 15300.

Ferejohn, John, and Keith Krehbiel. 1987. "The Budget Process and the Size of the Budget." *American Journal of Political Science* 31(2): 296-320.

Flyvbjerg, Bent. 2013. "Quality Control and Due Diligence in Project Management: Getting Decisions Right by Taking the Outside View." *International Journal of Project Management* 31: 760-774.

Gingerich, Daniel. 2013. "Governance Indicators and the Level of Analysis Problem: Empirical Findings from South America." *British Journal of Political Science* 43 (3): 505-540.

Hallerberg, Mark, Carlos Scartascini, and Ernesto Stein, eds. 2009. *Who Decides the Budget? A Political Economy Analysis of the Budget Process*. Cambridge: Harvard University Press.

Her Majesty's Treasury (HMT). 2013. *Managing Public Money*. July.

Ingraham, Patricia W., Philip G. Joyce, and Amy Kneedler Donahue (2003). *Government Performance: Why Management Matters*. London: The Johns Hopkins University Press.

Johnson, Michael. 2013. "Understanding the Real Budget Process: The Case of Ghana." International Food Policy Research Institute Discussion Note 029.

Jones, Bryan D., Frank R. Baumgartner, Christian Breunig, Christopher Wlezien, Stuart Soroka, Martial Foucault, Abel François et al. 2009. "A General Empirical Law of Public Budgets: A Comparative Analysis." *American Journal of Political Science* 53 (4): 855–873.

Killick, Tony. 2005. "The Politics of Ghana's Budgetary System." Center for Democratic Development/Overseas Development Institute Policy Brief No. 2, November.

Leonard, David. 2010. "'Pockets' of Effective Agencies in Weak Governance States: Where Are They Likely and Why Does It Matter?" *Public Administration and Development* (30): 91-101.

McCubbins, Matthew 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.

Meier, Kenneth J., and Laurence J. O’Toole Jr. 2014. “Subjective Organizational Performance and Measurement Error: Common Source Bias and Spurious Relationships.” *Journal of Public Administration Research and Theory* 23: 429-456.

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

PEFA Secretariat. 2012. “‘Field Guide’ for Undertaking an Assessment Using the PEFA Performance Measurement Framework.” Washington, DC.

Pitsvada, Bernard T. 1983. “Flexibility in Federal Budget Execution.” *Public Budgeting and Finance* 3(2): 83-101.

Pollitt, Christopher, Colin Talbot, Janice Caulfield, and Amanda Smullen. 2005. *Agencies: How Governments Do Things Through Semi-Autonomous Organizations*. Basingstoke: Palgrave Macmillan.

Robinson, Scott E., Carla M. Flink, and Chad M. King. 2014. “Organizational History and Budgetary Punctuation.” *Journal of Public Administration Research and Theory* 24: 459-471.

Rodden, Jonathan, Gunnar S. Eskeland, and Jennie Litvack (eds). 2003. *Fiscal Decentralization and the Challenge of Hard Budget Constraints*. London: MIT Press.

Schick, Allan. 1998. “Why Most Developing Countries Should Not Try New Zealand’s Reforms.” *World Bank Research Observer* 13(1): 123-31.

Schick, Allan. 2003. “The Performing State: Reflection on an Idea Whose Time Has Come but Whose Implementation Has Not.” *OECD Journal on Budgeting* 3(2): 71-103.

Schwarz, Norbert. 1999. “Self-Reports: How the Questions Shape the Answers.” *American Psychologist* 54 (2): 93-105.

Simson, Rebecca, and Bryn Welham. 2014. “Incredible Budgets: Budget Credibility in Theory and Practice.” Overseas Development Institute Working Paper 400.

Stasavage, David, and Dambisa Moyo. 2000. “Are Cash Budgets a Cure for Excess Fiscal Deficits (and at What Cost)?” *World Development* 28 (12): 2105-2122.

Thompson, Fred. 1993. “Matching Responsibilities with Tactics: Administrative Controls and Modern Government.” *Public Administration Review* 53(4): 308-318.

Von Hagen, Jürgen, and Ian Harden. 1995. “Budget Processes and Commitment to Fiscal Discipline.” *European Economic Review* 39: 771-779.

Wehner, Joachim, and Paolo de Renzio. 2013. “Designing Fiscal Institutions: The Political Economy of PFM Reforms.” In *The International Handbook of Public Financial Management*, ed. Richard Allen, Richard Hemming, and Barry Potter. Basingstoke: Palgrave Macmillan, 84-97.

Weingast, Barry, Kenneth Shepsle, and Christopher Johnsen. 1981. “The Political Economy of Benefits and Costs: A Neoclassical Approach to Distributive Politics.” *Journal of Political Economy* 89 (4): 642-664.

Wildavsky, Aaron. 1964. The politics of the budgetary process. Boston: Little-Brown.

Wildavsky, Aaron. 1975. Budgeting: A Comparative Theory of Budgetary Processes. Boston: Little, Brown and Co.

World Bank. 2006. *Ghana – Public Finance Management Performance Report and Performance Indicators: 2006 External Review of Financial Management*. Washington, DC: World Bank.

APPENDIX 4.1: STRUCTURAL CHARACTERISTICS OF SPENDING MINISTRIES AND BUDGET PERFORMANCE

The variation in measures of budget performance among ministries prompts the question: is this variation driven by unobserved informal, contextual, and organizational factors, or is it merely a function of the different structural characteristics of spending ministries? While the main body of this article focuses on the former, the analysis of these factors would be less pressing if variation was largely due to differences across ministries in "fundamentals" such as budget size and complexity, expenditure composition, and revenue sources. Ministries that predominantly undertake investment spending could be more prone to bad outcomes or difficult technical processes than ministries whose main expenditure is salaries, for example. This Appendix investigates this question. It demonstrates that while there are substantial differences in these variables across ministries, they are only weakly (if at all) correlated with measures of budget performance and explain only a small fraction of the variation across ministries. While these results must be interpreted with care, especially given the relatively small number of observations (although this sample does represent the universe of ministries within Ghana), they nonetheless suggest that variation in budget performance across ministries is not primarily attributable to simple structural differences.

Table A1 demonstrates the immense diversity of ministries' budgets in terms of size, complexity, expenditure profiles, and revenue sources. The average budget allocation of the smallest ministry from 2009-12 was 273 times smaller than that of the largest ministry, the number of pages in the budget ranges from 17 to 2117, and ministries' staff size varies by a factor of over 200.[114] Some ministries spend 87.9% of their budget on personnel, while others spend the same amount on investment and assets projects, and ministries range from 97.5% funded by the Government of Ghana's Consolidated Fund to 75.6% funded by donors.

[114] Staff size figures are for staff of the ministry headquarters, plus those supervised departments and agencies which form part of the Civil Service. (According to Ghana's administrative system, some agencies are part of the Public Service but not the Civil Service.) Consistent figures for the entire Public Service are not available, but these nevertheless give an indication of the variation across ministries.

Table A1: Structural Variation Across Ministries

	Obs.	Mean	Std. Dev.	Minimum	Maximum
Size and complexity					
Total budget size (GH¢ millions)	23	184.6m	359.0m	6.1m	1,666.5m
Budget complexity (no. pages)	22	252.5	576.7	17	2117
Staff size	23	1224.8	1849.3	38	7822
Expenditure composition					
% Personnel	23	59.5%	25.5%	10.8%	87.9%
% Services	23	16.3%	15.4%	1.3%	70.5%
% Investment	23	24.2%	23.1%	3.2%	87.9%
Revenue sources					
% Government of Ghana	23	47.9%	31.2%	1.1%	97.5%
% Internally generated	23	11.3%	13.4%	0.0%	40.1%
% Statutory	23	17.6%	22.0%	0.0%	74.9%
% Donor	23	23.2%	27.7%	0.0%	75.6%

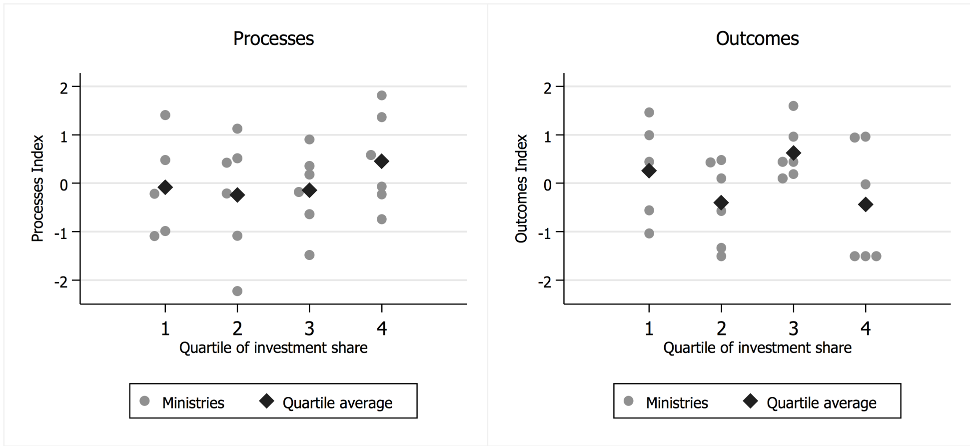
Note: Expenditure composition and budget size are calculated on mean 2009-12 allocations, revenue source is calculated on 2012 allocations only due to inconsistent reporting in earlier years, and budget complexity is the number of pages in the ministry's 2013 detailed budget allocation. Staff size is the number of civil service personnel working in the ministry and its departments and agencies as of May 2013 (earlier years not available). Services expenditure is calculated using the Services budget category for 2009-11 and Goods and Services for 2012, and Investment is calculated using the Investment category for 2009-11 and Assets for 2012.

The variation in expenditure composition across ministries is particularly interesting – as is common worldwide, Ghana's expenditure budget applies different processes to these spending categories, with the implicit assumption that the necessary control and disbursement procedures vary according to the type of expenditure, rather than the spending organization – that ministries "matter" only through their spending profiles. Likewise, in previous studies expenditure composition has been used as a basis to classify spending units (Dunleavy 1991) or distinguish among them (e.g. Breunig and Koski 2012). The share of investment spending may be thought to be especially important in determining budget performance, as infrastructure projects are notorious for delays, cost overruns, and other budgetary problems (Flyvbjerg 2013). Figure A1 examines this assumption by plotting the Processes and Outcomes indices by quartiles of investment spending as a share of ministries' total budget allocations.[115] There is

[115] Aggregating the ministries into quartiles of investment share is necessary to maintain the anonymity protections afforded to budget schedule officers, since plotting these indices directly against publicly available fiscal data would make it possible to identify individual ministries.

no clear relationship for either index: a wide range of good and bad performers can be found in the top, bottom, and intermediate quartiles.

Figure A1: Budget Performance by Investment Share in Expenditure

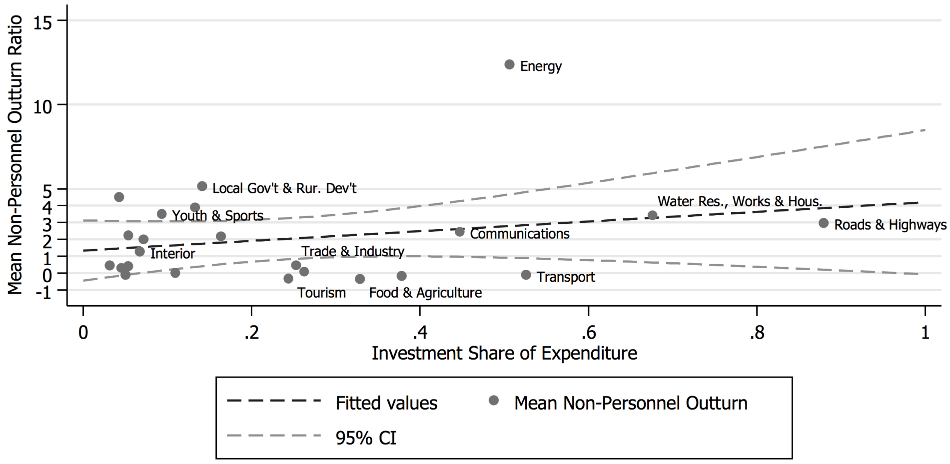


Note: Investment share is the unweighted mean of capital investment allocation as a percentage of the ministry's total budget allocation. It is calculated using the Investment category for 2009-11 and Assets for 2012. Overlapping markers have been shifted horizontally for clarity.

Figure A2 plots mean non-personnel expenditure outturns against ministries' investment share in expenditure, with selected ministries labeled. The overall correlation is positive, but not statistically significant. Among the five infrastructure-sector ministries with high investment shares, Energy is a significant positive outlier (even after excluding its extremely high outturn for 2011), while Transport's overspending is significantly below predicted and the other infrastructure ministries fall almost exactly on the trend line. Together, Figures A1 and A2 demonstrate that while infrastructure spending does entail its own challenges, its incidence across ministries is not driving the observed variation in budget performance. The bilateral correlations between other structural characteristics and budget performance (not shown) are similarly weak.

To examine these correlations in a multivariate setting, Table A2 conducts descriptive regressions of the three measures of budget performance – Budget Processes, Budget Outcomes, and mean non-personnel outturn ratio – on the sets of structural characteristics. The lack of explanatory power that these structural characteristics have over the three measures of budget performance is striking. The only coefficients that are even marginally statistically significant are between budget size and the Processes index in columns 1 and 4 (and even that is not robust to the inclusion of other variables in column 7), and between budget complexity and budget outturns in column 3. Even considered together, these structural characteristics are not good predictors of budget performance – a Wald test that all coefficients jointly equal zero is not rejected in any of the specifications.

Figure A2: Budget Outturns by Investment Share in Expenditure



Note: Mean non-personnel expenditure outturn calculated as an unweighted mean based on the categories Services plus Investment for 2009-11, and Goods and Services plus Assets for 2012. Covers Consolidated Fund expenditure only. Ministry of Energy 2011 outturn ratio of 204.6 (20,461%) is excluded as an outlier. Investment share is the unweighted mean of capital investment allocation as a percentage of the ministry's total budget allocation, calculated using the Investment category for 2009-11 and Assets for 2012.

These results must be interpreted cautiously due to the small sample size, which also restricts the possibility of adopting a more sophisticated modeling approach. Likewise, these (non-)relationships should of course be interpreted as correlations rather than causal claims, and this analysis is not presented in order to claim that structural characteristics are unimportant. Rather, the analysis simply suggests the more humble conclusion that the observed variation in budget performance is not well explained by these structural characteristics alone. This makes the task of understanding how informal, contextual, and other usually unobserved factors affect budget performance even more pressing.

Table A2: Structural Characteristics and Budget Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Processes	Outcomes	Outturns	Processes	Outcomes	Outturns	Processes	Outcomes	Outturns
Size and complexity									
Ln budget size	0.351	-0.172	0.72	0.382	-0.171	0.793	0.204	-0.131	0.409
	(0.184)*	(0.197)	(0.510)	(0.195)*	(0.209)	(0.547)	(0.208)	(0.240)	(0.582)
Budget complexity	-0.216	0.005	-1.079	-0.192	-0.046	-1.019	-0.098	-0.034	-0.870
	(0.210)	(0.224)	(0.581)*	(0.219)	(0.234)	(0.614)	(0.226)	(0.261)	(0.632)
Ln staff size	-0.113	-0.031	0.460	-0.079	-0.038	0.538	-0.113	-0.018	0.453
	(0.146)	(0.156)	(0.403)	(0.155)	(0.166)	(0.435)	(0.148)	(0.170)	(0.413)
Expenditure composition									
% Services				1.427	-0.661	3.310			
				(1.575)	(1.683)	(4.409)			
% Investment				0.861	-1.211	2.047			
				(0.987)	(1.054)	(2.761)			
Revenue sources									
% Internally generated							-1.197	1.092	0.708
							(1.709)	(1.971)	(4.782)
% Statutory							1.657	-0.348	4.990
							(1.091)	(1.257)	(3.051)
% Donor							0.390	-0.368	1.586
							(0.854)	(0.984)	(2.388)
Constant	-4.671	3.249	-8.998	-6.011	3.901	-12.109	-2.810	2.631	-5.628
	(2.981)	(3.186)	(8.248)	(3.367)*	(3.596)	(9.421)	(3.249)	(3.746)	(9.089)
R-squared	0.19	0.06	0.20	0.25	0.14	0.25	0.33	0.10	0.33
F test: Coefficients jointly = 0	1.38	0.40	1.54	1.06	0.50	1.06	1.24	0.29	1.24
Prob > F	0.28	0.75	0.24	0.42	0.77	0.42	0.34	0.93	0.34
Observations	22	22	22	22	22	22	22	22	22

* significant at 10%; ** significant at 5%; *** significant at 1%

Note: Standard errors in parentheses. Dependent variables are the mean non-personnel outturn ratio (2009-12), the Budget Outcomes index, and the Budget Processes index. Expenditure composition and budget size are calculated on mean 2009-12 allocations. Revenue source is calculated on 2012 allocations only due to inconsistent reporting in earlier years. Budget complexity is the natural logarithm of pages in the ministry's 2013 detail budget allocation. Staff size is the number of civil service personnel working in the ministry and its departments and agencies as of May 2013 (earlier years not available). The omitted categories are % Personnel for expenditure composition, and % Government of Ghana Consolidated Fund for revenue sources. Services spending is calculated using the Services budget category for 2009-11 and Goods and Services for 2012, and Investment is calculated using the Investment category for 2009-11 and Assets for 2012.

Chapter Five

CONCLUSION:

ORGANIZATIONS, INSTITUTIONS, AND REFORM

Chapter Abstract

This chapter ties together the diverse threads of this dissertation's three papers and considers their collective theoretical, methodological, and empirical contributions. It then sets out this dissertation's potential practical relevance for reformers inside and outside of government. I conclude with a somewhat more speculative discussion of the role of organizations in the long-term development of state capacity.

Conclusion:
Organizations, Institutions, and Reform

This dissertation began with the observation that the administrative apparatus of the state is composed of a set of organizations, and suggested that there may be advantages to thinking about states and state capacity in developing countries through an explicitly organizational lens. The introduction laid out a conceptual framework to array the various levels of determinants of organizational performance, and put forth two motivating questions that each paper addresses:

1. What is the extent of within-government variation in performance?
2. What mechanisms generate this variation?

While these two central questions are far too broad and rich for any study to answer comprehensively, each of the three papers in this dissertation has sought to provide novel empirical evidence of substantial organization-level variation in a different context, and to illustrate different mechanisms that generate this variation. The first paper focused explicitly on the question of within-government variation in management quality among organizations, and proposed a mechanism grounded in relational contract theory to explain this variation within a common institutional context. The second paper showed that fiscal institutions – meso-institutions in Ménard's (2014) terminology, and organizations in their own right – can have significant impacts on the performance of the service delivery units they fund, and demonstrated a political economy mechanism through which this effect occurs. Finally, the third paper showed that there is also substantial variation in organizations' budget management and performance despite operating within a formally identical set of budget institutions, structures, and processes, and argued that the mechanisms that generate performance are heterogeneous and idiosyncratic across ministries. The diversity of empirical contexts and theoretical mechanisms examined in these three papers reinforces the importance – as well as the potential richness – of within-government variation in performance as a topic for research.

In this brief conclusion, I first draw some connections among the theoretical, methodological, and empirical contributions of the three papers. Each paper already contains a discussion of its specific contributions, so to avoid repetition I instead focus on their broader collective contributions. I then consider – somewhat more speculatively – the implications of these contributions for theories of reform, beginning with normative recommendations for reform design and moving on to deeper questions about the extent to which the long-term development of state capacity might be led by organizational rather than institutional change.

5.1 TOWARDS AN ORGANIZATIONAL APPROACH TO GOVERNMENT PERFORMANCE

At the broadest level, this dissertation has aimed to demonstrate that examining the organizational basis of government bureaucracies in developing countries is both a necessary and productive task for studies of public service delivery, governance, and state capacity. It is necessary, because within-government variation in performance at the organization-level is empirically significant and unexplained by prevailing theories that focus on government-wide institutional and political determinants. And it is productive, because it opens up connections to several rich and highly developed bodies of theory and empirical methods in other fields or sub-fields where organizations have long been a key unit of analysis. For example, "From Institutions to Organizations" drew on relational contracts theory from organizational economics (Hart and Moore 2008, Gibbons and Henderson 2013) as well as the literatures on the management of discretion and on organizational performance, two key themes for public administration and management in developed countries (Mintzberg 1983; Wilson 1989; Ingraham, Joyce, and Donahue 2003). These literatures complement and extend the comparative politics and governance literatures on informal institutions that have been influential in studying developing states but nonetheless have significant limitations (Helmke and Levitsky 2004; Andrews 2013). Likewise, in a parallel to the American politics literature on how agencies mediate distributive politics (Bertelli and Grose 2009, Berry and Gersen 2010, Gordon 2011, Ting 2012), "Policy Implementation, Distributive Politics, and Fiscal Institutions" showed that fiscal institutions can play a similar role in a developing country setting and with respect to implementation as well as distribution.

Further exploring the connections between these bodies of theory could be useful not only to studies of government in developing countries, but also in developed ones. Although informality, weak enforcement of administrative rules, and poor policy implementation certainly tend to be more salient features of states in poor countries than in wealthy ones, they are by no means absent, as discussed for example by the literatures on the practices of street-level bureaucrats (Lipsky 1980) and policy implementation (Pressman and Wildavsky 1974). Imperfect contracting and relational contracts between organizations have also been studied in developed countries in the context of outsourcing and policy networks (Bertelli and Smith 2009), and integrating a relational view of management *within* organizations is an area in which significant lessons could be learned from studies of developing country governments.

One reason why there have been few medium- or large-N studies of organizations in developing country governments is methodological: whereas government statistics offices produce and publish national- and regional-level data that scholars can easily draw on, little such data is available at the organizational level. This problem is exacerbated by the difficulty of measuring and comparing the outputs of many public organizations, and by the importance of informal and non-verifiable internal processes in their operation. Given these data constraints, this study's adaptation and use of structured interview techniques for integrated mixed-method analysis is perhaps its main methodological contribution. This approach provides something of a compromise between small-N case studies, which provide rich qualitative data but a small and possibly unrepresentative sample, and large-N studies (either those that use administrative data, such as this dissertation's second paper, or closed-ended surveys) that provide statistical power and representativeness but may be misleading in contexts where informality and discretion are important and the gaps between *de jure* and *de facto* practices are large.[116] This approach also has some drawbacks: like qualitative studies it requires some discretion and judgment on the part of the interviewer or researcher; as with quantitative studies there is a tradeoff between comparability and flexibility; and analytical methods for small-N and large-N studies

[116] With Imran Rasul and Daniel Rogger, I am currently implementing a follow-up study in which the same interview method is being used by 45 civil servant enumerators to conduct a management survey of all 4,000 senior civil servants in Ghana. Although this is methodologically more similar to the original Bloom and Van Reenen (2007) study and Rasul and Rogger's (2014) adaptation, we have nevertheless attempted to integrate aspects of qualitative data collection.

are both more standardized and commonly accepted than those for analyzing medium-N data. Nevertheless, this method can be another useful tool for researchers interested in the functioning of government bureaucracies in both developing and developed countries.

Empirically, each paper of this dissertation has made a novel empirical contribution: "From Institutions to Organizations" showed the range of within-government variation in management practices and demonstrated the qualitative alignment of this variation with that predicted by relational contract theory; "Policy Implementation, Distributive Politics, and Fiscal Institutions" provided the most comprehensive estimate to date on the extent and costs of project non-completion in any country, and demonstrated that fiscal institutions can not only affect project completion but also mediate the impact of distributive pressures on implementation; and "One Size Does Not Fit All" is (to my knowledge) the first study to systematically document and explore variation in budget performance and management among spending ministries in a single country.

Collectively, these three papers make a strong case for scholars beyond the public administration sub-field to engage with the mundane bureaucratic practices and iterated organizational processes through which states translate policies into outputs. While political scientists and economists alike are showing an increased interest in policy implementation and service delivery, a great deal of this work has either focused on corruption (Reinikka and Svensson 2004; Olken 2007; Beekman, Bulte, and Nillesen 2014), even though this is merely one cause or mechanism of implementation failure, or has studied the impact of a one-time intervention or policy change (Banerjee, Deaton, and Duflo 2004; Reinikka and Svensson 2011; Banerjee *et al* 2014; Yanez-Pagans and Machicado-Salas 2014) even though these effects may dissipate as agents and organizations settle into long-run equilibrium (Olken and Pande 2012). Lasting and longer-term improvements in performance are likely to require continuous and iterated improvement – and thus better management and improved organizational cultures – rather than one-off external interventions. Understanding when, where, and how these improvements occur requires scholars to think of bureaucracies as organizations, and design their research strategies accordingly.

5.2 AN ORGANIZATIONAL PERSPECTIVE ON REFORM

Thinking about state bureaucracy as a set of organizations rather than as an institution also has implications for the design of reforms. In particular, the diagnoses of government performance on which large-scale public sector reforms are based are often homogenizing, in that they assume that the same problems exist across the whole of government. Yet this homogenous view of the performance landscape within government is contradicted by this dissertation's empirical findings on the extent of variation in performance at the organizational level. Homogenizing, one-size-fits-all approaches to governance reform have been much criticized for trying to impose "best practice" solutions that may not fit with countries' diverse contexts, with differences in technical capacity treated as one such contextual variable (Grindle 2004, Andrews 2013). Due to their focus on cross-country variation, however, these critiques have often conflated *horizontal* heterogeneity – in institutional, political, and socio-cultural context – with *vertical* heterogeneity in levels of performance or management quality. This dissertation focuses attention on the latter by examining variation within government at the organization level, thus holding constant the main country-level drivers of horizontal heterogeneity, and showing that one-size-fits-all reforms can become one-size-fits-almost-none even when abstracting from the horizontal dimension of context. The diversity of management quality in a single government – and the resulting implications for reform elaborated herein – suggest that treating differences in management quality as more than just one element of country context can be one important way of moving beyond the mantra that "context matters" towards a specific understanding of when, where, and how it matters.

The existence of systematic within-country variation in the performance of government organizations demands a change in the conceptual frameworks applied to improving management and performance. This article has argued that a significant amount of this variation is of a vertical nature rather than horizontal – differences in quality rather than context (although the latter doubtless also exists). To the extent these differences are a reflection of poor management, rather than constrained optimization or optimization of a different set of goals, it suggests that *incremental* approaches to reform are both possible and meaningful. At a practical level, an incremental approach requires a different mindset when defining reform objectives and content. In their discussion of persistent performance differences among private sector firms, Gibbons and Henderson remark: "…one needn't aspire to move a 10th percentile firm to the 90th; moving up a quartile would be a big deal, so we need to know whether (and, if so, how) it can be done" (2013, 684). Analogously, public reforms need not aim to improve an organization's performance from the 10th percentile to the 90th, or some other lofty standard of "best practice"; simply moving up a quartile would be a big deal, and so reform efforts should focus on making meaningful yet realistic steps towards better performance.

The recognition of vertical heterogeneity among government bureaucracies implies that improving management means different things in different organizations, both in terms of the ambition of reform (reform objectives) and the specific steps necessary to improve performance (reform content). Instead of defining a reform by specifying the set of management practices that are desired for all organizations *ex post*, reform should start with a diagnosis of where each organization actually is *ex ante* and then identify reform interventions that would improve on them incrementally yet meaningfully. This perspective complements an existing body of work on policy and management reform that emphasizes incrementalism as a response to context (Andrews 2013) and complexity (Hausmann 2008) – both elements of horizontal diversity. Elaborating the connections between these two perspectives is an important area for empirical study and theory development. In terms of reform content, such an approach would recognize that different organizations have different problems, perhaps driven (as in this dissertation's third paper) by different causes, and therefore differentiate among the immediate steps necessary for each organization's improvement. This differentiation could either be explicit, or achieved by laying out graduated steps as done (for example) by the United Kingdom's Departmental Capability Review program (Cabinet Office 2009). This is, of course, easier said than done, as illustrated by the debates in the budget reform literature on "basics first" versus best practices approaches (Andrews 2006) and on selectivity versus sequencing (Allen 2013).

In this sense, recognizing the need to move away from homogenizing diagnoses and prescriptions does not answer all the difficult questions of reform design, but it does push reformers to deal with a more productive set of questions. To return to the "growth diagnostics" analogy put forward in the introduction, it pushes reformers not to ask "What are the problems of government on average?" but instead to ask "What are the specific binding constraints to performance in each organization?" Instead of using this appreciation of heterogeneity to design reforms that are more complex and nuanced but equally rigid in their application, reform and organizational improvement should be conceived of as ongoing, iterative processes – more in the spirit of Andrews's (2013) "problem-driven iterative adaptation" or the "continuous improvement" management philosophy that has proven so influential in the private sector and organizational economics literature on organizational performance (Liker and Franz 2011). Although public sector reforms are often imagined as "once and for all" reforms that will simultaneously "fix" every bureaucratic problem, the continued existence of significant variations and shortcomings in performance even within developed country governments and private sectors should signal to reformers in developing countries that any reform is but one step of many in a long journey.

Finally, this dissertation has shed new light onto the potentials and limitations of externally and internally driven attempts to improve organizational performance. First, the conceptual framework presented in the introduction points out that the determinants of performance are simultaneously external (both government-wide and organization-specific) and internal to each organization. While many reform efforts are driven by forces external to the organization – either from higher-level government authorities or by international donors – the non-codifiable, non-verifiable, discretion-intensive nature of many aspects of public bureaucracies' operations limits the extent to which these external parties can observe, enforce, and incentivize important aspects of performance. To return to the language of relational contracts, external parties and interventions may be able to force organizations to achieve perfunctory performance – compliance with the letter of the law, execution of formal management processes – but are unlikely to be able to force them to achieve consummate performance – compliance with the spirit of the law or process. Improving performance may thus entail qualitatively different steps for different organizations: focusing on compliance with formal processes for struggling organizations (again reminiscent of the "basics first" mantra in public financial management) but then on the complementary informal aspects of performance for those that are already executing formal processes well. However, the dynamics of reform may not be as straightforward as this simple heuristic – for example, Chassang's (2010) theoretical model of the development of relational contracts in organizations argues that first achieving perfunctory performance may actually make reaching consummate performance more difficult – and the empirical literature on changing relational contracts within organizations is in its infancy (Bloom *et al* 2013; Atkin *et al* 2015; Blader *et* al 2015; Martinez *et al* 2015), so this is an urgent and exciting area for future study.

5.3 ORGANIZATIONS AND THE DEVELOPMENT OF STATE CAPACITY

Adopting an organizational perspective on the design and mechanisms of these short- to medium-term reform processes also poses a challenge to macro-historical theories of the long-term development of bureaucratic capacity. While existing theoretical and empirical studies have focused almost exclusively on explaining cross-country (or cross-region) variation or national-level changes in state capacity driven by broad political economy pressures and institutional development (North 1990; Tilly 1990; Evans and Rauch 1999; Acemoglu 2005; Greif 2007; Besley and Persson 2009; Acemoglu, Ticchi, and Vindigni 2011), they have largely overlooked the mechanisms through which these broad political and institutional forces are translated into routine bureaucratic practices in specific organizations. The closest this literature comes to discussing these mechanisms is Besley and Persson's (2009) model of the process through which rulers "invest" in state capacity, but even here it is assumed that there exists an investment technology that is perfectly predictable and known to rulers, as if the process of bureaucratic reform were as simple as collecting interest on a savings deposit. The simplicity of this investment function contrasts with theoretical and empirical studies of actual public sector reform processes, which emphasize the challenging and uncertain nature of these reforms. This raises the question: in the long-term of historical state development, are the challenges of organizational reform mere footnotes to larger trajectories driven by broader political and institutional forces? Or might there be a macro-historical parallel to the microeconomic theory and evidence that organizational performance is not fully determined by external contextual factors, and that public organizations' political principals cannot always manipulate agents to achieve their desired outcomes?

Within-government variation in organizational performance poses another related question of the macro-historical literature: when state capacity increases in response to broad political and institutional forces, does it do so simultaneously across the whole of government? Or does it increase organization by organization, through contemporaneous but semi-independent organizational processes? This may seem like a semantic question, but it has deep consequences for understanding the causes and mechanisms of both long- and near-term theories of bureaucratic reform. If bureaucratic development is driven primarily by forces whose operation is government-wide – social norms, judicial systems, average education levels, meritocratic service-wide personnel systems, etc. – then one would expect this rising tide to lift all boats more or less simultaneously. But as the conceptual framework suggests, these forces could also affect government organizations unevenly, through organization- or sector-specific channels. Besley and Persson's (2009) bellicist model of the development of state capacity is one of the rare formal models that does disaggregate the concept of state capacity by distinguishing between rulers' investments in the state's legal capacity and fiscal capacity, but even this only disaggregates along two dimensions. More generally, however, these "investments" may be spread across government agencies in different patterns in cases where state-building is motivated more by economic modernization drives than external threats, and in different patterns still where politicians' incentives are shaped more by electoral accountability than by external threats. This idea that state-building can take different shapes is consistent with the case study literature that has emphasized the role of particular government organizations in countries' historical development, such as the professionalism of the British navy during the 18th century, when the British army was essentially a patronage organization for aristocrats (Fisher and Lundgren 1975), or Japan's Ministry of Trade and Industry during the post-war boom (Johnson 1982). State capacity could also develop unevenly due to sector- or organization-specific constellations of interests, for example depending on the composition of business coalitions and their relationships with the state as suggested by the literature on business-state relations (Bräutigam, Rakner, and Taylor 2002; Opoku 2010). In the conceptual framework from this dissertation's introduction, these would be examples of external, organization-specific factors that could drive uneven development of state capacity.

An equally interesting question for future study is whether significant advances in state capacity can occur endogenously, through internal developments in particular organizations. While the literature on state capacity has largely focused on how factors external to the state drive improvements in bureaucratic capacity, this dissertation has argued theoretically and empirically that these structural and contextual factors may not be as determinative of organizational performance as they are often thought to be. Institutionalist treatments of organizations tend implicitly to treat organizational variation as a stochastic process that essentially cancels itself out over time. Even if these developments are as-random, however, they can still be consequential if leading organizations play a role in shaping the development of other organizations, for example through the diffusion of organizational innovations (Guler, Guillen, and Macpherson 2002), development of professional organizations (Greenwood, Suddaby, and Hinings 2002) or isomorphism within institutional fields (DiMaggio and Powell 1983). Tolbert and Zucker (1983) find some evidence of this in their classic study of civil service professionalization in American municipalities: structural characteristics of municipalities predicted adoption of civil service reforms among early adopters but these factors lose their explanatory power for later adopters, suggesting that once these reforms became seen as legitimate their spread was driven by isomorphism rather than structural factors. Stochastic organizational performance improvements could also lead to (positive or negative) endogenous development of state capacity through network effects; Acemoglu, García-Jimeno, and Robinson (2015) find empirical evidence of such network effects in Colombia. The idea that small random events can lead to persistently different long-term equilibria due to the dynamics of path dependence is also at the heart of the theoretical literature on the development of relational contracts within organizations (Chassang 2010).

This dissertation also points to a broader need to reconsider the role of organizations in institutional change. Although previous discussions of institutional change have recognized that organizations' activities can lead to incremental changes in institutions (North 1990) which may or may not reinforce that institution (Greif and Laitin 2004) or subvert its original intent (Mahoney and Thelen 2010; Hacker, Pierson, and Thelen 2015), they focus solely on the interaction between institutions and organizations rather than engaging with the internal dynamics of organizations that drive this interaction. Through management and hierarchy, organizations act calculatively and continuously to shape the formal and informal constraints and incentives that their members face when carrying out their day-to-day activities – in this sense the organization is a nexus of institutions. And if these organizational processes play a significant role in the long-run development of state capacity, might they also play a significant role in the long-run development of other societal institutions? This dissertation has argued theoretically and empirically that organizations may be less constrained by institutional and political equilibria than is often assumed by institutional and political economy approaches to government performance; this leeway should lead to a corresponding increase in our estimation of organizations' potential role in effecting institutional change, whether deliberately or "as a by-product of [their] maximizing activities" (North 1990, 78). Yet the direction of this change, and organizations' effectiveness in achieving it, depends on the outcome of internal organizational processes that this dissertation has argued are complex and subject to multiple equilibria. To understand the role of organizations in institutional change thus requires a deeper engagement with these internal processes and dynamics.

Finally, there are pragmatic reasons for adopting an organizational approach to the study of state bureaucracies. Even if institutions and political economy factors are more powerful determinants of performance than purely organizational factors, the available levers for reform – for politicians, bureaucrats, donors, and citizens alike – often take the form of organizations or organizational processes. Even when reforming an abstract and diffuse informal institution or rewriting a constitution is beyond the power of any actor or group, the same actors may nonetheless be able to achieve more humble steps such as improving organizational culture in one government agency, instituting a new monitoring system in a fiscal institution, or simply writing out a divisional to-do list so that staff know what to do when they come to work on Monday morning. This dissertation has demonstrated empirically that these marginal yet collectively significant variations in performance exist contemporaneously within the same government, and has developed theoretical explanations for the mechanisms underlying this variation. While the power of these organizational mechanisms may often be overshadowed by higher-level institutional and political factors, this does not detract from the potential of organizations to be meaningful vehicles for reform. To paraphrase Gibbons and Henderson, we needn't aspire to move a 10^{th}-percentile bureaucracy up to the 90^{th}, either across or within governments; moving up a quartile would be a big deal, so we need to know whether – and how – it can be done.

REFERENCES

Acemoglu, Daron (2005). "Politics and Economics in Weak and Strong States." *Journal of Monetary Economics* 52: 1199–1226.

Acemoglu, Daron, Davide Ticchi, and Andrea Vindigni. 2011. "Emergence and Persistence of Inefficient States." *Journal of the European Economic Association* 9(2): 177-208.

Acemoglu, Daron, Camilo García-Jimeno, and James Robinson. 2015. "State Capacity and Economic Development: A Network Approach." *American Economic Review* 105(8): 2364-2409.

Allen, Richard. 2013. "Challenges of Reforming Budgetary Institutions in Developing Countries." In *Public Financial Management and Its Emerging Architecture*, eds. Marco Cangiano, Teresa Curristine, and Michel Lazare. Washington, DC: International Monetary Fund, 411-430.

Andrews, Matt. 2006. "Beyond 'Best Practice' and 'Basics First' in Adopting Performance Budgeting Reform." *Public Administration and Development* 26: 147-161.

Andrews, Matt. 2013. *The Limits of Institutional Reform in Development: Changing Rules for Realistic Solutions*. Cambridge: Cambridge University Press.

Atkin, David, Azam Chaudhry, Shamyla Chaudhry, Amit Khandelwal, and Eric Verhoogen. 2015. "Organizational Barriers to Technology Adoption: Evidence from Soccer-Ball Producers in Pakistan." NBER Working Paper 21417, July.

Banerjee, Abhijit, Angus Deaton, and Esther Duflo. 2004. “Wealth, Health, and Health Services in Rural Rajasthan.” *American Economic Review* 94(2): 326–330.

Banerjee, Abhijit, Raghabendra Chattopadhyay, Esther Duflo, Daniel Keniston, and Nina Singh. 2014. “Can Institutions be Reformed from Within? Evidence from a Randomized Experiment with the Rajasthan Police.” Mimeo.

Beekman, Gonne, Erwin Bulte, and Eleonora Nillesen. 2014. “Corruption, investments and contributions to public goods: Experimental evidence from rural Liberia.” *Journal of Public Economics* 115: 37-47.

Besley, Timothy, and Torsten Persson. 2009. “State Capacity, Conflict, and Development.” *Econometrica* 78(1): 1-34.

Berry, Christopher, and Jacob Gersen. 2010. “Agency Design and Distributive Politics.” Mimeo, October.

Bertelli, Anthony, and Christian R. Grose. 2009. “Secretaries of Pork? A New Theory of Distributive Public Policy.” *Journal of Politics* 71(3): 926-945.

Bertelli, Anthony, and Craig Smith. 2009. “Relational Contracting and Network Management.” *Journal of Public Administration Research and Theory* 20: i20-i44.

Blader, Steven, Claudine Gartenberg, Rebecca Henderson, and Andrea Prat. 2015. “The Real Effects of Relational Contracts.” *American Economic Review Papers and Proceedings* 105(5): 452-456.

Bloom, Nicholas, and John Van Reenen. 2007. Measuring and Explaining Management Practices Across Firms and Countries. *Quarterly Journal of Economics,* 122**,** 1351-1408.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. “Does Management Matter? Evidence from India.” *Quarterly Journal of Economics* 128(1): 1-51.

Bräutigam, Deborah, Lise Rakner, and Scott Taylor. 2002. "Business Associations and Growth Coalitions in Sub-Saharan Africa." *Journal of Modern African Studies* 40(4): 519-547.

Cabinet Office. 2009. *Capability Reviews: An Overview of Progress and Next Steps*. Cabinet Office, December.

Chassang, Sylvain. 2010. "Building Routines: Learning, Cooperation, and the Dynamics of Incomplete Relational Contracts." *American Economic Review* 100(1): 448-465.

DiMaggio, Paul, and Walter Powell. 1983. "The Iron Cage Revised: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review* 48(2) :147-160.

Evans, Peter, and James Rauch. 1999. "Bureaucracy and Growth: A Cross-National Analysis of the Effects of 'Weberian' State Structures on Economic Growth." *American Sociological Review 64 (5)*: 748-765.

Fisher P, Lundgren P. 1975. "The Recruitment and Training of Administrative and Technical Personnel." In *The Formation of National States in Western Europe*, Charles Tilly (ed), Princeton: Princeton University Press: 527–543.

Gibbons, Robert, and Rebecca Henderson. 2013. "What Do Managers Do?" In *The Handbook of Organizational Economics*, ed. Robert Gibbons and John Roberts. Princeton: Princeton University Press.

Gordon, Sanford. 2011. "Politicizing Agency Spending Authority: Lessons from a Bush-era Scandal." *American Political Science Review* 105(4): 717-734.

Greenwood, R, R. Suddaby, and C. R. Hinings. 2002. "Theorizing Change: The Role of Professional Associations in the Transformation of Institutionalized Fields." *Academy of Management Journal* 45:58–80.

Greif, Avner. 2007. "The Impact of Administrative Power on Political and Economic Development: Towards a Political Economy of Implementation." Mimeo, December.

Greif, Avner, and David Laitin. 2004. "A Theory of Endogenous Institutional Change." *American Political Science Review* 98(4): 633-652.

Grindle, Merilee. 2004. "Good Enough Governance: Poverty Reduction and Reform in Developing Countries." *Governance* 17 (4): 525-548.

Guler, I., M. Guillen, and J. Macpherson. 2002. "Global Competition, Institutions, and the Diffusion of Organizational Practices: The International Spread of ISO 9000 Quality Certificates." *Administrative Science Quarterly* 47:207–32.

Hacker, Jacob, Paul Pierson, and Kathleen Thelen. 2015. "Drift and Conversion: Hidden Faces of Institutional Change." In Mahoney, James, and Kathleen Thelen (eds), *Advances in Comparative-Historical Analysis*, Cambridge: Cambridge University Press, pp. 180-210.

Hart, Oliver, and John Moore. 2008. "Contracts as Reference Points." *Quarterly Journal of Economics* 123(1): 1-48.

Hausmann, Ricardo. 2008. "The Other Hand: High Bandwidth Development Policy." CID Working Paper No. 179, September.

Helmke, Gretchen, and Steven Levitsky. 2004. "Informal Institutions and Comparative Politics: A Research Agenda." *Perspectives on Politics* 2(4): 725-740.

Ingraham, Patricia W., Philip G. Joyce, and Amy Kneedler Donahue (2003). *Government Performance: Why Management Matters*. London: The Johns Hopkins University Press.

Johnson, Chalmers. 1982. *MITI and the Japanese Miracle: The Growth of Industrial Policy, 1925-1975*. Stanford: Stanford University Press.

Liker, Jeffrey K., and James Franz, 2011. *The Toyota Way to Continuous Improvement*. New York: McGraw-Hill.

Lipsky, Michael. 1980. *Street Level Bureaucracy: Dilemmas of the Individual in Public Services*. New York: Russell Sage Foundation.

Mahoney, James, and Kathleen Thelen (eds). 2010. *Explaining Institutional Change: Ambiguity, Agency, and Power*. Cambridge: Cambridge University Press.

Martinez, Elizabeth, Nancy Beaulieu, Robert Gibbons, Peter Provonost, and Thomas Wang. 2015. "Organizational Culture and Performance." *American Economic Review Papers and Proceedings* 105(5): 331-335.

Ménard, Claude. 2014. "Embedding Organizational Arrangements: Towards a General Model." *Journal of Institutional Economics* 10(4): 567-589.

Mintzberg, Henry. 1983. *Structure in Fives*. Englewood Cliffs, NJ: Prentice Hall.

North, Douglass C. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.

Olken, Benjamin. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115(2): 200-249.

Olken, Benjamin, and Rohini Pande. 2012. "Corruption in Developing Countries." *Annual Review of Economics* 2012(4): 479-509.

Opoku, Darko Kwabena. 2010. *The Politics of Government-Business Relations in Ghana, 1982-2008*. New York: Palgrave Macmillan.

Pressman, Jeffrey L., and Aaron Wildavsky. 1974. *Implementation.* Berkeley: University of California Press.

Rasul, Imran, and Daniel Rogger. 2014. "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service." University College, London, June.

Reinikka, Ritva, and Jakob Svensson. 2004. "Local Capture: Evidence from a Central Government Transfer Program in Uganda." *The Quarterly Journal of Economics* 119(2): 679-705.

Reinikka, Ritva, and Jacob Svensson. 2011. "The Power of Information in Public Services: Evidence from Education in Uganda." *Journal of Public Economics* 100(7-8): 956-966.

Tilly, Charles. 1990. *Coercion, Capital and European States, AD 990-1992.* Cambridge: Blackwell.

Ting, Michael. 2012. "Legislatures, Bureaucracies, and Distributive Spending." *American Political Science Review* 106(2): 367-385.

Tolbert, Pamela, and Lynne Zucker. 1983. "Institutional Sources of Change in the Formal Structure of Organizations: The Diffusion of Civil Service Reform, 1880-1935." *Administrative Science Quarterly* 28(1): 22-39.

Wilson, James. 1989. *Bureaucracy*. New York: Basic Books.

Yanez-Pagans, Monica, and Gustavo Machicado-Salas. 2014. "Bureaucratic Delay, Local-Level Monitoring, and Delivery of Small Infrastructure Projects: Evidence from a Field Experiment in Bolivia." *World Development* 59: 394-407.