

# “Honor List” and “Shame Roll”: Quasi-Experimental Evidence of the Effect of Performance Feedback under Political Control

Wenchi Wei,<sup>\*</sup> Chengwei Wang,<sup>\*,†</sup> Wenkang Zhai,<sup>\*</sup> Wenzhao Li<sup>\*</sup>

<sup>\*</sup>Renmin University of China, China

<sup>†</sup>City University of Hong Kong, China

Address correspondence to the author at [liwenzhao@ruc.edu.cn](mailto:liwenzhao@ruc.edu.cn).

## Abstract

This article examines how the use of an “honor list” and “shame roll” as a means of performance feedback can influence governments’ future performance improvement, focusing on a prominent performance management reform implemented in Chinese local governments. We draw upon classic behavioral theories of organizations to propose testable hypotheses. The empirical analyses use 3,300 observations based on the 333 grassroots governments of the capital city of China as the unit of analysis. Regression discontinuity design estimations show that entering the honor list reduces governments’ performance improvement in the next period, while entering the shame roll helps governments improve their performance. Moreover, the level of performance improvement is higher at the cutoff for both the honor list and shame roll if governments rely more on the higher authority’s fiscal resources. This study advances a more nuanced understanding of the performance feedback effect under political control and a rigid administrative hierarchy.

## Introduction

Performance management systems constitute a pivotal tool for shaping organizational performance (Gerrish and Spreen 2017; Koning and Heinrich 2013; Pasha, Kroll, and Ash 2021; Walker, Damanpour, and Devece 2011). Performance feedback is a key element of the performance management cycle that can substantially affect subsequent organizational decision-making and performance improvement (Hong 2019, 2020; Hong, Kim, and Son 2020; Meier, Favero, and Zhu 2015). Scholars of public administration and management have widely acknowledged that management matters for organizational performance in the public sector (e.g., Meier and O’Toole 2002; Moynihan and Pandey 2005; O’Toole and Meier 1999). However, the inverse loop from past performance to future management and performance outcomes, which is referred to as the *performance feedback effect*, has been understudied until recently (e.g., Flink 2019; Hong 2019, 2020; Hong, Kim, and Son 2020; Meier, Favero, and Zhu 2015; Nicholson-Crotty, Nicholson-Crotty, and Fernandez 2017). This scholarly omission is puzzling given considerable attention paid to public sector performance management and improvement in recent decades (Hong 2019, 2020). Thus, evidence concerning the performance feedback effect is of great importance in both academic and practical senses.

In applying the behavioral theory of organizations to the public sector (Cyert and March 1963; Simon 1955, 1972, 1997), this article focuses on a prominent performance management reform occurring in Chinese local governments and examines its effect on governments’ performance improvement. Over two years ago, the capital city of China substantially reformed its 12345 hotline, an online central

hub introduced to provide citizens with quick answers to inquiries regarding municipal services and allowing citizens to express demands and complaints about government services and community life. The political leader of the capital city government ascribes great significance to the 12345 hotline. In particular, an *honor list* and *shame roll* were introduced to publicize and rank the 333 subordinate governments’ performance in terms of dealing with citizens’ inquiries, demands, and complaints through the 12345 hotline. In this way, governments’ social aspirations are set up, making the negative gaps between actual performance and social aspirations a powerful incentive for local officials to adjust their managerial strategies and decision making for performance improvement.

This research contributes to the literature in the following ways. First, it provides additional evidence regarding the effect of performance feedback in the public sector. In particular, this research focuses on a political context that substantially differs from that of Western countries. We thus examine in depth a distinct way in which the performance feedback effect may take place under a regime featuring strict political control and a top-down administrative hierarchy (Cai 2008). Second, unlike most existing studies that focus on managerial decision making as the outcome of interest of performance feedback (e.g., Meier, Favero, and Zhu 2015; Nicholson-Crotty, Nicholson-Crotty, and Fernandez 2017; Salge 2011; Zhu and Rutherford 2019), this research attends to future performance improvement, about which existing evidence is relatively rare. Third, theoretically, this research scrutinizes whether and how the effect of performance feedback is contingent on governments’ fiscal capacity,

drawing upon organizational capacity and resource dependence theory. In particular, a higher degree of fiscal reliance on a higher authority may enhance the power position of the higher authority over local governments, forcing them to adhere to the higher authority's policy priorities. On the other hand, high reliance on fiscal transfers suggests local governments' weak fiscal capacity and, thus, low likelihood of future performance improvement. Last, methodologically, this article capitalizes on a specific case in which the forcing variable, cutoff, and treatment can be clearly identified and are valid for a regression discontinuity (RD) design, which allows us to make causal estimations of the performance feedback effect using administrative data.

The remainder of this article proceeds as follows. The second section "Literature Review: Performance Management and Feedback Effect" introduces the related literature about performance management and feedback effect in the public sector, which is followed by theoretical analyses and hypotheses in the third section "Theories and Hypotheses." The fourth section "Institutional Background and the 12345 Hotline" introduces the institutional and contextual background, the fifth section "Methodology: Model, Data, and Validity Tests" presents the methodology employed, and the sixth section "Empirical Findings" reports the empirical results. Finally, we discuss the contributions and potential limitations of the research, propose opportunities for future studies, and conclude the research.

## Literature Review: Performance Management and Feedback Effect

Existing studies concerning public performance management have thoroughly examined the factors leading to variations across public organizations in the effectiveness of performance management systems (see Ammons and Roenigk 2015; Chun and Rainey 2005; Kroll and Moynihan 2015; Lavertu and Moynihan 2013; Moynihan and Pandey 2010; Poister, Pasha, and Hamilton Edwards 2013). Scholars have also investigated the factors shaping the ways in which external stakeholders use performance information (see Askim, Johnsen, and Christophersen 2008; Bourdeaux and Chikoto 2008; Holm 2017; Julnes and Holzer 2001; Moynihan and Landuyt 2009; Moynihan and Pandey 2010). This section pays specific attention to three strands of literature in the field of public performance management with more relevance to our present research, including work on (1) the distinctions between objective and subjective performance measures as well as citizen involvement in public sector performance measurement, (2) the behavioral consequences of performance feedback, and (3) the use and effectiveness of rewards and punishments (so-called *carrot* and *stick*) for performance improvement.

For the first strand of literature, existing studies have long debated the merits of agency- and citizen-generated performance measures in the public sector (Schachter 2010). Scholars viewing performance measurement as a scientific process and holding that the evaluation of public sector performance should be isolated from political pressures generally prefer objective performance measures (Nicholson-Crotty, Theobald, and Nicholson-Crotty 2006; Schachter 2010), assuming that objective measures generated by administrators or professional experts are more valid and reliable (Brown

and Coulter 1983; Coe and Brunet 2006; Schachter 2010). In contrast, the subjective measures generated from citizen surveys are regarded as context dependent and prejudiced by respondents' sociodemographic backgrounds (e.g., Brown and Coulter 1983; Licari, McLean, and Rice 2005; Sanger 2008; Shingler et al. 2008). Recent studies on public performance measurement have called for a reformulation of the objective–subjective dichotomy and have claimed the importance of citizen-generated measures and citizen involvement for public performance measurement (e.g., Schachter 2010; Yang 2005; Yang and Holzer 2006). In addition, previous studies have shown that the subjective perception measures of public service performance by multiple stakeholders are consistent with objective archival performance indicators (e.g., Song and Meier 2018). This present study provides another example demonstrating the importance of citizens' subjective evaluation of public sector performance in understanding governments' future decision making and performance improvement.

For the second body of literature, despite its substantial academic and practical significance, the performance feedback effect—that is, how organizations respond to previous performance successes or failures, adjust their managerial strategies and behaviors, and ultimately enhance their future performance outcomes—was largely underexplored until recently (Hong 2019, 2020; Nicholson-Crotty, Nicholson-Crotty, and Fernandez 2017; Wang and Ma 2020). Previous studies on public sector performance management generally view organizational performance as the result of managerial structures, strategies, networking, and specific decision making (Andrews, Boyne, and Walker 2006; Boyne and Walker 2010; Meier and O'Toole 2003). In recent years, scholars have begun to regard organizational performance as the cause and to examine how previous performance outcomes can shape subsequent decision making, which in turn may affect organizational performance in the future (Hong 2019, 2020; Meier, Favero, and Zhu 2015; Nielsen 2014; Zhu and Rutherford 2019).

Previous studies on the performance feedback effect in the public sector primarily focus on how the gaps between actual performance and social or historical aspirations can influence organizations' decision making and behaviors (e.g., Flink 2019; Meier, Favero, and Zhu 2015; Zhu and Rutherford 2019). Theoretically, based on bounded rationality and Bayesian decision theory, Meier, Favero, and Zhu (2015) construct a comprehensive model of the effect of performance gaps on managerial decision making in a variety of respects. Salge (2011) extends the arguments deriving from the behavioral theory of the firm to the public sector context and argues that public organizations initiate search processes as a response to performance gaps and that such processes are conditional on organizational slack and regulatory endorsement. Empirically, drawing upon a combined dataset from multiple surveys of US hospitals, Zhu and Rutherford (2019) find that multiple managerial decisions are substantially shaped by historical and social aspiration gaps. In addition, Hong (2019) and Hong, Kim, and Son (2020) exploit the context of the Korean central government's annual evaluation of the performance of public organizations to examine the performance feedback effect, finding that both social and historical aspiration gaps can substantially influence organizations' performance improvement in the future.

For the third strand of literature, previous studies of human resource management in the private sector have long focused on the effectiveness of rewards and punishments in terms of enhancing employee cooperation, inspiring employee motivation, and promoting organizational performance (Andreoni, Harbaugh, and Vesterlund 2003; Carver and White 1994, also see Podsakoff et al. 2006 for a meta-analysis). The importance of rewards and punishments in providing incentives for human behaviors is rooted in two distinct psychological mechanisms that drive human motivations: the *behavioral activation system* and the *behavioral inhibition system* (Carver and White 1994; Gray 1987). The former is said to be sensitive to signals of reward and nonpunishment, while the latter is sensitive to signals of punishment and nonreward (Carver and White 1994). Both systems are key in shaping individual responses to the efforts of organizational leaders for performance improvement. Rewards and punishments (carrot and stick) in human resource management are the essential elements of organizational control needed to motivate employees (Liang, Xue, and Wu 2013) and are at the heart of the strategies used by transactional leaders. Empirically, although there is a plethora of research on the use of rewards or punishments in the field of organizational studies, the findings are largely mixed or inconsistent (Chen, Ramamurthy, and Wen 2012), with one key consensus pointing to a greater likelihood of achieving organizational goals through the combined use of rewards and punishments (Andreoni, Harbaugh, and Vesterlund 2003; Podsakoff et al. 2006).

In sum, the present research builds upon the expanding literature about citizen involvement in public sector performance measurement and the performance feedback effect and contributes additional evidence to the related literature by focusing on a prominent government performance management reform implemented in Chinese local governments. This research also extends theories and empirical evidence of the performance feedback effect by taking into account how organizational capacity may moderate this effect. In addition, we delve into a case involving a combined use of rewards and punishments for organizational performance improvement in the public sector, which corresponds to the long-standing tradition of studying carrot-and-stick approaches in dealing with different levels of organizational performance.

## Theories and Hypotheses

### Bounded Rationality, Satisficing, and Performance Stagnation

The rationality assumption of decision making holds that organizations as problem solvers “follow an optimization process that includes formulating a problem, listing the possible courses of action, and assessing the rewards (or penalties) and corresponding probabilities attached to each option” (Hong 2020, 934). Forester (1984, 23–24) summarizes the prerequisites of full rationality as “a well-defined problem, a full array of alternatives to consider, full baseline information, full information about the consequences of each alternative, full information about the values and preferences of citizens, and fully adequate time, skill, and resources.” Despite its significant value for theoretical model construction and analytic framework development, the full rationality assumption exists primarily as the “postulation and deduction characteristic of theoretical economics” (Jones 2003, 397) instead of as

normative or descriptive principles for organizational decision making in the real world (Simon 1955, 1997). Indeed, what organizations are generally confronted with includes (1) ambiguous and poorly defined problems; (2) incomplete information; and (3) limited time, skills, and resources (Forester 1984; Simon 1972). As a result, in the process of decision making, organizations usually need to alter their goals from meeting a maximum level to meeting a “satisfactory” level of performance (Cyert and March 1963; Simon 1955, 1972, 1997).

Whether organizations are satisfied with certain performance outcomes rests on the gap between their actual performance and their *aspiration*, which is defined as “a goal variable that must be achieved for decision makers to feel that an outcome is satisfactory” (Hong 2020, 934). Distinguished by the type of reference object, organizational aspirations are broadly categorized into *historical* and *social* aspirations (Cyert and March 1963), with the former referring to organizations’ own previous performance and the latter referring to the performance of peer groups (Flink 2019; Hong 2019). Organizations with performance falling below the social aspiration are presumably eager to close the gap between their performance and that of their peer groups. Nevertheless, as organizations’ performance reaches a level above the social aspiration, historical aspiration becomes more relevant for their future decision making (Bromiley 1991; Hong 2019).

The theoretical arguments here do not deny the importance of historical aspirations in terms of shaping organizations’ future performance improvement. The core of our arguments, as proposed by Bromiley (1991), is that social and historical aspirations have different degrees of impact at different levels of organizational performance. When organizational performance is below the social aspiration level, social aspiration gaps are more powerful in shaping future decision making and performance improvement. However, once their performance exceeds the social aspiration level, organizations are primarily motivated to improve their performance in relation to the historical aspiration level (Bromiley 1991; Flink 2019; Hong 2019). In a context in which the social comparison group is clearly defined, such as in our case where the higher authority ranks the performance of grassroots governments monthly and allocates rewards and punishments based on their relative performance, social aspiration is generally more relevant for the behavioral motivations of governments.

Of note, the theories underpinning our analyses include Simon’s (1955, 1972) theory of bounded rationality and Cyert and March’s (1963) behavioral theory of organizations, which both focus on the behaviors of organizations. As explicitly elaborated by Simon (1972, 162), “a theory of rational behavior may be concerned with the rationality of individuals or the rationality of organizations. In fact, the two bodies of theory are not wholly distinct.” Our analyses below focus on government leaders as the subjects of analysis. By doing this, we explicitly assume that the behavioral motivations of government leaders represent those of governments. This practice may “obscure the complex dynamics among various actors” (Hong 2019, 2). In our case, this concern is largely alleviated since in the specific political and social context of China, local government leaders generally hold ultimate authority over local affairs and therefore can fundamentally control the policymaking and implementation. In this vein, each grassroots government can be viewed as a monolithic organization, within which the government leader plays a dominant role. Therefore, it



is reasonable to use the behavioral motivations of government leaders to explain those of grassroots governments. This has the advantage of alleviating the concern of ecological fallacy, which uses statistical analyses based on data at the aggregate level to interpret behaviors of individuals (Robinson 1950).

In our research context, entering the honor list constitutes a powerful incentive and a key political task for local officials. Thus, the performance of peer grassroots governments becomes the social aspiration of each government. In particular, the power of social aspiration in motivating performance improvement becomes fairly strong as significant political rewards are linked to placement on the honor list amid the higher authority's strict political control over subordinate governments. Entering the honor list releases a clear signal of exceeding the social aspiration. As a consequence, social aspirations will be presumably less salient for governments' performance improvement in the next period (Hong 2019). Therefore, compared to governments not placed on the honor list, governments entering the honor list are expected to be fairly satisfied with their performance and have less motivation to improve their future performance. Based on the theoretical analyses presented above, the first hypothesis subject to empirical testing is as follows.<sup>1</sup>

### *Hypothesis 1 (Satisficing Hypothesis)*

Compared to governments not entering the honor list, governments placed on the honor list are less likely to improve their performance in the next period.

### **Negativity Bias, Blame Avoidance, and Performance Improvement**

Public officials are presumably eager to claim credit for positive policy outputs and to avoid blame for policy failures (McGraw 1991; Lau 1982, 1985; Weaver 1986). Previous theoretical and empirical studies have shown that compared to credit claiming, blame avoidance is more powerful in shaping officials' policy choices and managerial activities (Hood 2007). The fundamental rationale underpinning policymakers' blame avoidance inclination is rooted in humans' loss aversion instinct. As argued by Weaver (1986, 373), humans "who have suffered losses are more likely to notice the loss, to feel aggrieved and to act on that grievance, than gainers are to act on the basis of their improved state."

<sup>1</sup>The theory of the zone of indifference developed by Meier et al. (2015) and the model of relative risk aversion developed by Nicholson-Crotty et al. (2017) suggest that organizations with small performance gaps, either positive or negative, are expected to stay within the zone of indifference and to be less likely to take risk actions. In this sense, organizations will view slightly positive performance gaps and slightly negative performance gaps indifferently. This theoretical argument makes more sense in a context without an external authority's deliberate actions to designate a specific benchmark level of performance and to accordingly allocate rewards and punishments. In our case of study, the higher authority ranks grassroots governments' performance scores, sets an honor list accordingly, and rewards governments placed on the honor list. In this context, governments barely entering the honor list and those barely failing to enter the honor list are high-performing organizations and are confronted with very different consequences; namely, governments with positive performance gaps (relative to the cutoff for the honor list) are rewarded while those with negative performance gaps are not. In the context of a higher authority's strict political control and with significance attached to the honor list, grassroots governments are unlikely to be indifferent about slightly positive or negative performance gaps.

Humans' asymmetric responses to positive and negative information have been referred to as *negativity bias* (Hong 2019, 2020; Nielsen and Moynihan 2017; Olsen 2015; Rozin and Royzman 2001) and have been demonstrated by many studies from multiple research fields. For instance, the theory of loss aversion and prospect theory in economics suggest an asymmetric pattern of persons' responses to different types of information (Soroka 2006). In the political science field, it has been commonly recognized that negative information plays a more influential role in influencing citizens' voting behaviors (e.g., Klein and Ahluwalia 2005; Soroka and McAdams 2015). Provided that humans are more sensitive to losses than to gains, higher authority leaders are assumed to favor subordinates who have prevented policy losses or failures more than those who have brought about equivalent levels of gains. This psychological preference can substantially influence officials' policy choices and managerial behaviors, causing them to devote more effort to avoiding unpopular actions (Weaver 1986).

In our research context, entering the shame roll suggests that governments' performance falls below the social aspiration, causing a negative social aspiration gap to emerge. Grassroots governments placed on the shame roll face the higher authority's blame and, consequently, receive fewer opportunities for promotion or are even subject to punishments. As a result, government officials are compelled to actively search for solutions, aiming to improve future performance and close the performance gap (Bromiley 1991; Greve 2008; Joseph and Gaba 2015). Grassroots government leaders' motivation for future performance improvement may be remarkably strong given the considerable political significance attached to the 12345 hotline performance by the political leader of the BMG. Accordingly, we have the second hypothesis as follows.

### *Hypothesis 2 (Blame Avoidance Hypothesis)*

Compared to governments not entering the shame roll, governments placed on the shame roll are more likely to improve their performance in the next period.

### **Organizational Capacity, Resource Dependence, and Performance Improvement**

One of the underlying assumptions of the performance feedback effect elaborated on above is that as performance gaps appear, organizations maintain adequate capacity to adjust their managerial strategies and practices as a response. However, this assumption may not always hold in reality. When confronted with similarly sized performance gaps, organizations' reactions may be substantially different due to disparities in their capacity (e.g., Andrews and Boyne 2010; Bolton, Augustine Potter, and Thrower 2016; Brody, Eun Kang, and Bernhardt 2010; Huber and McCarty 2004; Huber and Shipan 2002; McGuire and Silvia 2010). Organizational capacity has been defined as a multidimensional concept that involves "various elements of an organization, providing the basis upon which organizations can implement programs and achieve goals" (Fredericksen and London 2000, 233). For instance, Fredericksen and London (2000) understand organizational capacity in terms of leadership and vision, management and planning, fiscal planning and practice, and operational support. Similarly, Brody, Eun Kang, and Bernhardt (2010, 171) interpret organizational



capacity from the perspectives of “financial resources, staffing, technical expertise, communication and information sharing, leadership...” Among the previous studies, the capacity for resource acquisition and management is always an essential dimension (Brody, Eun Kang, and Bernhardt 2010; Brown 2012; McGuire and Silvia 2010). In particular, for the government, the capacity to acquire sufficient revenues constitutes a useful proxy for general capacity (Fukuyama 2013).

In our research context, it is reasonable to postulate that grassroots governments’ responses to their previous 12345 hotline performance is contingent on their capacity to leverage fiscal resources. Specifically, when listed on the shame roll, motivated by the powerful incentive to avoid blame and punishments, grassroots government officials are expected to capitalize on their fiscal resources to conduct problem screening and solution searches. Even when they enter the honor list and a positive social aspiration gap appears, due to the generally positive effect of organizational capacity on performance (Andrews and Boyne 2010), governments with greater fiscal capacity are also more likely to achieve a higher degree of performance improvement in the next period. The present research operationalizes grassroots governments’ fiscal capacity using the ratio of fiscal transfers from the municipality and district government to their total revenues.<sup>2</sup> A high *fiscal transfer ratio* suggests great fiscal reliance of grassroots governments on the higher authority and a lack of adequate fiscal capacity. We thus have the following hypothesis.

### **Hypothesis 3a (Organizational Capacity Hypothesis)**

Reliance on fiscal transfers from the higher authority negatively influences the treatment effect of entering the honor list and shame roll on governments’ performance improvement in the next period.

Resource dependence theory has its roots in the classic works of Emerson (1962) and Pfeffer and Salancik (1978) and has been widely used to explain individual and organizational behaviors. The central theses of resource dependence theory include that “every organization occupies a social exchange position, which can be characterized in terms of both dependence and its inverse, power” (Emerson 1962; Malatesta and Smith 2011, 610; Pfeffer and Salancik 1978) and that “organizational survival hinges on the ability to procure critical resources from the external environment” (Casciaro and Piskorski 2005, 167). As a result, organizations possessing resources that others view as critical will play a dominant role in the power dependence dynamic of the relationship (Emerson 1962; Malatesta and Smith 2014; Pfeffer and Nowak 1976; Pfeffer and Salancik 1978). In the private sector, organizations overly relying on critical resources from the environment or other entities may use a variety of tactics to restructure their dependence on the resources (Casciaro and Piskorski 2005). These tactics include “reducing the interest in valued resources, cultivating alternative sources of supply, or forming coalitions” (Casciaro and Piskorski 2005, 167; Malatesta and Smith 2014). However, in the public sector, constrained by institutional rules, lower levels of government may have much less autonomy and capacity to restructure their dependence on resources allocated by the higher authority.

<sup>2</sup>The primary types of grassroots governments’ revenues include general revenues, government fund revenues, fiscal transfers from the municipality and district government, and previously unspent fiscal surplus.

In our case of the 12345 hotline, higher fiscal transfer ratios suggest greater dependence of grassroots governments on the higher authority for fiscal resources. As a result, as the fiscal transfer ratios increase, the higher authority’s power position and influence over grassroots governments escalate, as does the likelihood of grassroots governments being forced to improve their 12345 hotline performance, which is favored by the political leader of the BMG. Moreover, the combination of resource dependence with the institutional context featuring the higher authority’s strict political control over subordinate officials will substantially reinforce the power position of the higher authority.

### **Hypothesis 3b (Resource Dependence Hypothesis)**

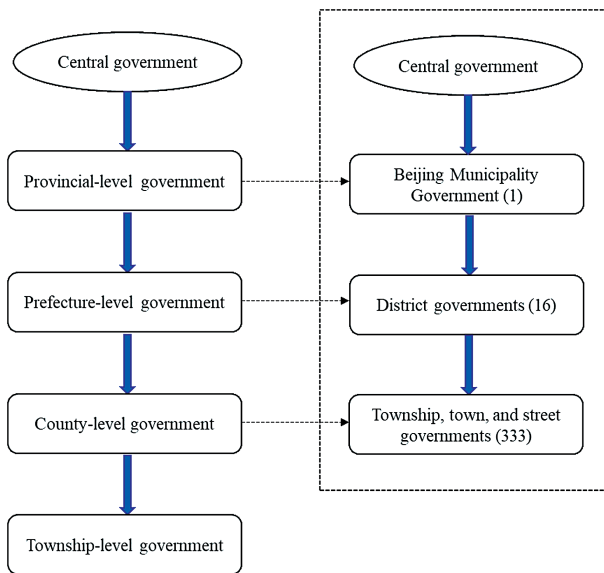
Reliance on fiscal transfers from the higher authority positively influences the treatment effect of entering the honor list and shame roll on governments’ performance improvement in the next period.

## **Institutional Background and the 12345 Hotline**

This section introduces the institutional and contextual background serving as the basis for our subsequent empirical analyses. As shown in the left panel of figure 1, China has a centralized administrative system composed of the central government and four levels of local government, including province-, prefecture-, county-, and township-level governments. Each level of government has a Communist Party secretary and a head of government serving as dual leaders. The career promotion of the leaders of subordinate governments is largely determined by the higher authority (Cai 2008; Xu 2011; Zhang 2006). The centralized personnel management system constitutes the basis for political control and a strict top-down hierarchical administrative structure (Xu 2011; Zhang 2006). Confronted with the political pressure exerted by the higher authority to maintain societal stability and prevent citizen dissatisfaction, local government officials are usually highly vigilant about citizen blame and are forced to enhance their responsiveness to citizen demands and complaints. The right panel of figure 1 shows the administrative structure of the Beijing Municipality Government (BMG), which corresponds to the national structure. The BMG is a province-level government subject to the direct administration of the central government and is composed of 16 prefecture-level districts, which govern 333 subordinate county-level township, town, and street governments (hereafter called *grassroots governments*) as of 2020. The Beijing municipality covers a geographic area of 16,410.54 square kilometers (6336.15 square miles) and has a population of 21.89 million.<sup>3</sup> Figure 2 illustrates the administrative area of the Beijing municipality, district, and grassroots governments.

Chinese local governments’ 12345 hotline is a nonemergency telephone hotline similar to the 311 system in the United States, which is intended to create a central hub for subscribers to access a variety of municipal services and to divert regular inquiries and nonemergency calls from the number reserved for emergency calls. Through the 12345

<sup>3</sup>The population data are derived from the results of the Seventh National Population Census released by the National Bureau of Statistics of China in May 2021: [http://www.stats.gov.cn/english/PressRelease/202105/t20210510\\_1817188.html](http://www.stats.gov.cn/english/PressRelease/202105/t20210510_1817188.html)



**Figure 1.** Administrative structure of the National and Beijing Municipality Government.

hotline, ordinary citizens have the opportunity to obtain quick answers to inquiries with regard to municipal services, to make comments and to report problems or difficulties they encounter. The 12345 hotline in China has become an essential means for facilitating citizen–government interaction and improving citizen satisfaction with government services. In January 2019, the BMG substantially reformed its 12345 hotline, building it into a comprehensive public platform for citizen consultations and service requirements and embedding strict political instructions and performance evaluations into the system. The core purpose of the reform was to provide prompt, efficient, and timely responses to citizens’ inquiries, demands, and complaints. The Beijing Municipality Citizen Hotline Service Center (hereafter called the *official call center*) contracts out call-receiving work to a professional telecommunication company to carry out this task. In 2020, the number of calls to the 12345 hotline exceeded 11 million, representing a 55% increase relative to that in 2019 levels.

After receiving citizens’ inquiries, demands, or complaints through the 12345 hotline, the official call center sorts them and sends the cases to the corresponding grassroots governments or agencies, which must deal with the issues properly. The official call center makes periodic follow-up telephone calls to selected citizens who have required services or expressed complaints through the 12345 hotline and asks whether the grassroots governments or agencies have responded to their demands or complaints, whether the reported issues have been properly addressed, and whether they are satisfied with the solution measures. The official call center integrates the follow-up interview results into grassroots government-level data. Based on these data, three key indicators of the 12345 hotline performance of each of the 333 grassroots governments—namely, the response rate, resolution rate, and satisfaction rate—are computed monthly. The average percentage value of the three indicators is employed as the final score measuring each grassroots government’s 12345 hotline performance. For instance, if a township government has a response rate of 100%, a resolution rate of 83.2%, and a

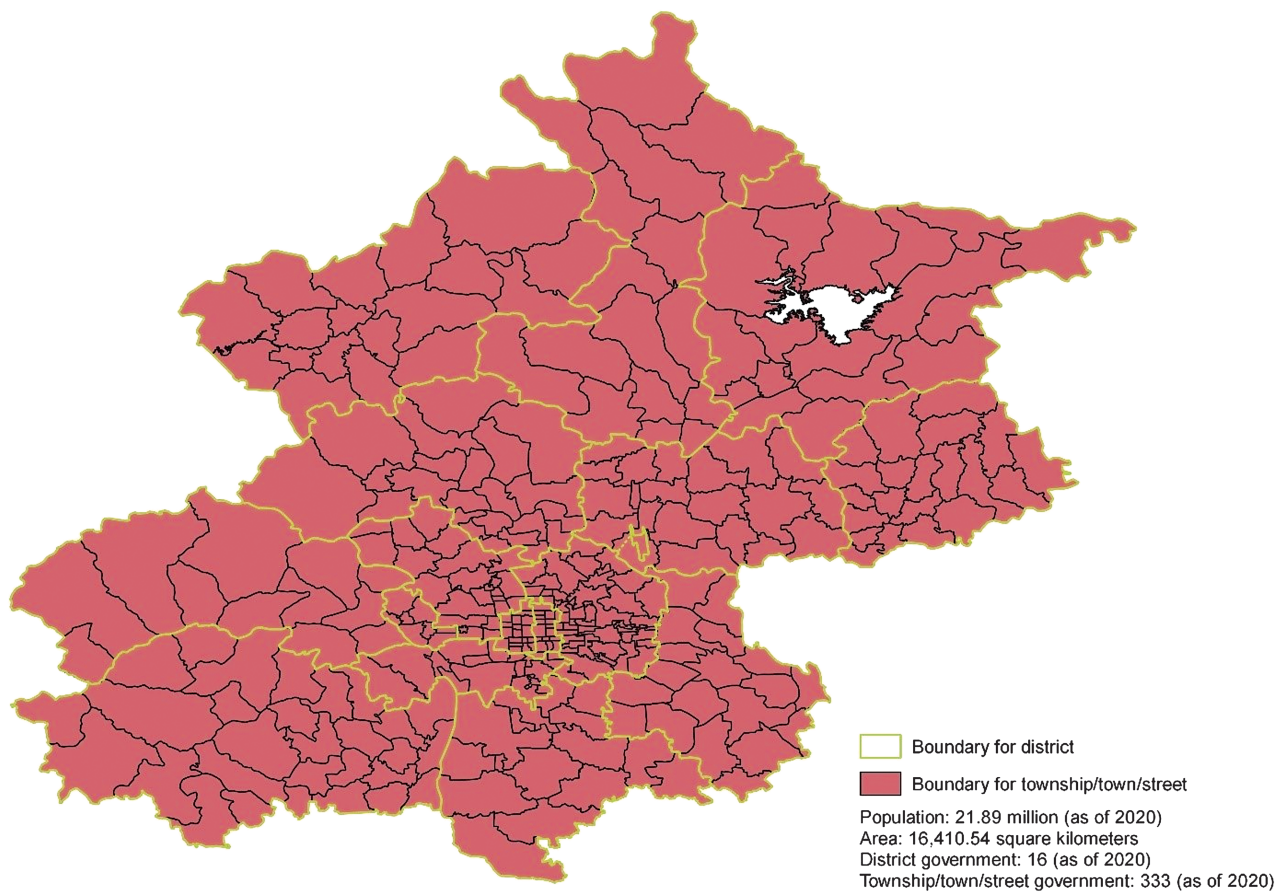
satisfaction rate of 80.5%, the average percentage value is 87.9, which is taken as the government’s 12345 hotline performance score.

The political leader of the BMG attaches great importance to the 12345 hotline performance of grassroots governments. In January 2019, the BMG began to convene monthly meetings (usually at the end of each month) to review and remark on grassroots governments’ 12345 hotline performance. The political leaders of district and grassroots governments are obligated to attend the meetings on-site, and other primary officials of all levels of government need to be present via internet video. In the meetings, an *honor list* that includes the 10 grassroots governments obtaining the highest performance scores and a *shame roll* that includes the 10 grassroots governments obtaining the lowest performance scores in the present month are released. The political leader of the BMG personally releases the honor list and shame roll in the meetings. In addition to regular ceremonies of list releasing, in each monthly meeting, the leaders of three selected district governments and the leaders of grassroots governments entering the shame roll are required to introduce the problems and difficulties in terms of operating the 12345 hotline as well as strategies for performance improvement. The political leader of the BMG then makes comments and deploys resources to cope with the reported issues. The consequences of being listed on the honor list and shame roll are introduced in the [supplementary appendix 1](#). Overall, in a regime featuring a higher authority’s strict political control over subordinate officials, performance evaluation conducted by the higher authority serves as a key means to incentivize and control the behaviors of lower-ranking officials (Chen, Li, and Zhou 2005; Choi 2012; Li and Zhou 2005). Therefore, as the political leader of the BMG attaches great importance to 12345 hotline performance, grassroots government officials do so as well.

## Methodology: Model, Data, and Validity Tests

### Formalized Framework of RD Design

An RD design is characterized by three fundamental components: forcing variable, cutoff, and treatment (Cattaneo, Idrobo, and Titiunik 2019; Cattaneo, Jansson, and Ma 2020, 2021). In a formal setting for an RD design, each unit of analysis receives a score, known as the forcing variable. A specific treatment is assigned to the units whose scores are above (or below) a specific value, known as the cutoff. Assuming that there are  $n$  units for analysis as indexed by  $i = 1, 2, 3 \dots n$ , each receiving a forcing variable denoted as  $X_i$ . A specific treatment is imposed on unit  $i$  if  $X_i \geq c$  (or  $X_i \leq c$ ), where  $c$  is the cutoff value. The key assumption underpinning an RD design is the randomized distribution of units in the neighborhood of the cutoff along the forcing variable. In other words, a successful RD design requires that the units barely receiving the treatment be equal to those barely failing to receive the treatment by expectation in terms of the observed characteristics. If that is true, in the neighborhood of the cutoff, the units on one side of the cutoff can be exploited as the counterfactuals of the units on the other side. The fundamental identification strategy of an RD design is to isolate the exogenous variation of the treatment status in the neighborhood of the cutoff and to examine the treatment effect by comparing the outcomes of interest of the units on either



**Figure 2.** Administrative area of the Beijing municipality, regional, and grassroots governments.

side of the cutoff. In this sense, RD design estimates the local average treatment effect (LATE) at the cutoff (Cattaneo, Idrobo, and Titiunik 2019; Cattaneo, Jansson, and Ma 2020, 2021; Lee and Lemieux 2010).

### Model and Data

In our RD design, the units of analysis are the 333 grassroots governments. The forcing variable each unit receives is the monthly generated 12345 hotline performance score. As introduced above, to promote the prompt responsiveness of grassroots governments to citizens' inquiries, demands, and complaints through the 12345 hotline, the BMG convenes a monthly review meeting, in which the *honor list* including the 10 grassroots governments with the highest performance scores and the *shame roll* including the 10 grassroots governments with the lowest performance scores are released by the political leader. According to our research setting, the treatment assigned is "entering the honor list or shame roll." The cutoff value determining units' treatment status varies each month on the grounds that the 12345 hotline performance of grassroots governments is evaluated in a monthly cycle. Thus, over the same timeline, the performance scores are computed, and the honor list and shame roll are generated. For instance, in the months when the scores of the well-performing units are generally high, the cutoff score for the honor list should be high. Similarly, in the months when the scores of the grassroots governments ranked at the bottom are generally low, the cutoff for the shame roll should be low. To facilitate

analysis, we center the cutoff for the honor list and shame roll in the empirical estimations. More specifically, the centered cutoff is set to zero, and the forcing variable is accordingly computed as the distance between the real performance score and the real cutoff. Of note, as introduced above, the units receiving a performance score below the cutoff for the shame roll are placed on the shame roll. To prevent confusion, when conducting estimations for the effect of entering the shame roll, we use the inverse value of the centered forcing variable in the analysis, making the positive centered forcing variable result in the treatment status of being placed on the shame roll. This is consistent with the condition for the honor list.

We draw upon the common method of previous studies to employ a global polynomial approach in the empirical estimations, including all grassroots governments as the units of analysis (e.g., Ferreira and Gyourko 2009; Folke and Snyder 2012; Hong 2019, 2020; Lee, Moretti, and Butler 2004). The outcome of interest—namely, performance improvement in terms of scores and rankings—is modeled as follows to test hypothesis 1 and 2.

$$PI_{i,m+1} = \alpha + \beta D_{i,m} + \delta X_{i,m} + \gamma_i + \theta_m + \varepsilon_{i,m} \quad (1)$$

where  $PI_{i,m+1}$  denotes the measured performance improvement of a certain grassroots government  $i$  in month  $m + 1$ , which is computed as the difference in government  $i$ 's performance between months  $m + 1$  and  $m$ . We measure  $PI_{i,m+1}$  using two indicators: performance score improvement and performance



**Table 1.** Summary Statistics

Variable	N	Mean	SD	Min	Max
Panel A: primary variables					
Performance score improvement	3,300	1.829	9.372	−42.780	48.300
Performance ranking improvement	3,300	−0.735	109.674	−332	332
Entering the honor list (treatment, dummy)	3,300	0.038	0.191	0	1
Entering the shame roll (treatment, dummy)	3,300	0.030	0.171	0	1
Centered forcing variable for the honor list	3,300	−13.078	11.351	−53.080	14.490
Centered forcing variable for the shame roll	3,300	−12.724	10.086	−49.140	21.030
Panel B: fiscal transfer and covariate variables					
Fiscal transfer ratio	3,252	0.007	0.057	0	0.818
Total revenues (log)	3,252	18.636	1.031	13.892	21.106
Total expenditures for general services (log)	3,108	17.345	0.932	11.940	20.216
Total “three public expenditures” (log)	3,120	12.338	1.055	9.401	17.849
Geographic area (km <sup>2</sup> , log)	3,300	3.049	1.451	−0.223	5.946
Total population (log)	3,145	10.623	0.965	7.813	12.792
Population density (population per km <sup>2</sup> , log)	3,145	7.538	2.084	2.786	10.704
Locally registered population ratio	2,815	0.574	0.233	0.102	0.969

*Note:* Panel A reports the primary variables for the RD estimations, including the dependent, forcing, and treatment variables. Panel B reports the fiscal transfer and covariate variables, which are used to test the moderating effect or conduct robustness checks. The unit of analysis is the 333 grassroots governments. The sample period runs from March 2019 to April 2020, during which the data for 4 months are unavailable, resulting in a final sample period of 10 months. Among 3,330 grassroots government-year observations, 30 have missing values for the necessary data. Therefore, the total number of observations for the baseline model is 3,300. Due to missing data, the numbers of observations for variables reported in Panel B may be below 3,300. The data sources and measures of some variables are introduced in the [supplementary appendix 5](#): table 1.

ranking improvement.  $D_{i,m}$  is a dummy variable indicating whether government  $i$  in month  $m$  is placed on the honor list or shame roll,  $X_{i,m}$  refers to the centered forcing variable for placement on the honor list or shame roll of government  $i$  in month  $m$ , and  $\gamma_i$  and  $\theta_m$  denote the unit and month fixed effects, respectively. Error term  $\varepsilon_{i,m}$  designates the part of the disturbance asymptotically uncorrelated with treatment variable  $D_{i,m}$ . Last,  $\beta$  is the estimated parameter indicating the LATE of entering the honor list or shame roll.

To test hypothesis 3a and 3b, which address the potential moderating effect of fiscal capacity on the effect of entering the honor list and shame roll on governments' performance improvement, we modify the empirical model presented above by including the fiscal capacity variable ( $C_{i,m}$ ) and its interaction with the treatment variable ( $D_{i,m}$ ) as follows.

$$PI_{i,m+1} = \alpha + \beta_1 D_{i,m} + \beta_2 C_{i,m} + \beta_3 D_{i,m} \times C_{i,m} + \delta X_{i,m} + \gamma_i + \theta_m + \varepsilon_{i,m} \quad (2)$$

The data on grassroots governments' 12345 hotline performance scores and rankings are primarily derived from official reports distributed by the BMG. The unit of analysis is the 333 grassroots governments. The sample period runs from March 2019 to April 2020, during which the data for four months are unavailable; thus, the final sample period is 10 months. Among the 3,330 government-year observations, 30 have missing values for the performance data, leaving in 3,300 observations as the sample for analysis. The data on the pretreatment covariates—namely, grassroots governments' basic socioeconomic, demographic, and financial traits, which are used to conduct the validity test of our RD design or to test moderating effects—are collected from the official websites of the municipality and district governments and complemented by several online sources. [Table 1](#) reports

the summary statistics and data sources of the variables used in the empirical estimation.

### Validity Tests of the RD Design

The common approaches of testing the validity of an RD design include a balance test of the pretreatment covariates of the units on either side of the cutoff and a density test of the forcing variable at the cutoff ([Cattaneo, Idrobo, and Titiunik 2019](#); [Imbens and Lemieux 2008](#)). The methods, procedures, and results (refer to [table 2](#) and [figures 3](#) and [4](#) for the results) of our validity tests are reported in the [supplementary appendix 2](#), all verifying the validity of our RD design.

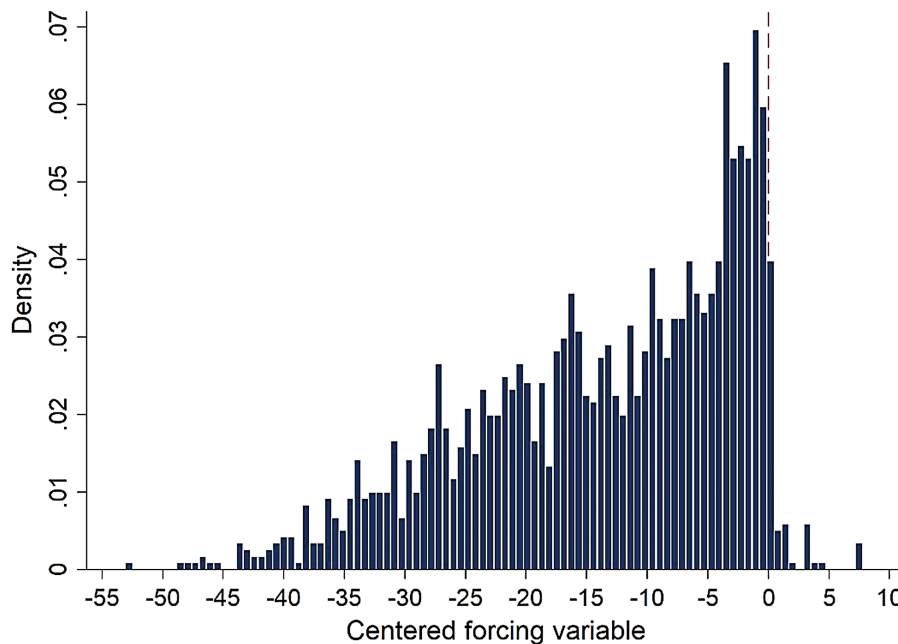
### Empirical Findings

This empirical section first implements formal regression analyses and reports the RD estimations of the effect of entering the honor list and shame roll on grassroots governments' performance improvements in the next period. We employ a global polynomial approach, which has the merit of delivering a good overall approximation of the relationship between the outcome of interest and the forcing variable based on the whole sample of analysis and, meanwhile, revealing the potential discontinuity of this relationship at the cutoff ([Calónico, Cattaneo, and Titiunik 2015](#)). Notably, to alleviate the potential ceiling effect (refer to the [supplementary appendix 3](#) for more details), we eliminate the observations that have a full performance score (100) from the sample for analysis in the RD estimation of the effect of entering the honor list on performance score improvement in the next period. [Tables 3](#) and [4](#) report the performance feedback effect of entering the honor list and shame roll, respectively. Column (1) and (2) of [table 3](#) report the results of regressing the baseline model specified by [equation \(1\)](#), which includes a first-order polynomial

**Table 2.** Local Linear Polynomial RD Estimations for Pretreatment Covariates

Pretreatment Covariates	Honor List				Shame Roll			
	MSE-Optimal Bandwidth	RD Estimator	Robust <i>p</i> -Value	Effective Number Observations	MSE-Optimal Bandwidth	RD Estimator	Robust <i>p</i> -Value	Effective Number Observations
Total revenues	4.280	-0.541	.074*	442	4.691	-0.115	.757	639
Total expenditures for general services	3.218	0.0662	.379	323	4.291	0.0379	.729	538
Total “three public expenditures”	5.210	0.283	.477	479	4.479	0.150	.490	567
Geographic area	4.782	0.0753	.703	484	3.739	-0.140	.796	467
Total population	4.167	-0.473	.165	419	4.979	-0.0639	.932	680
Population density	7.573	-0.619	.240	584	4.617	0.172	.631	594
Locally registered population ratio	3.960	0.0449	.215	374	5.561	0.0573	.281	761

Note: \*\*\*  $p < .01$ , \*\*  $p < .05$ , \*  $p < .1$  (for two-tailed tests).

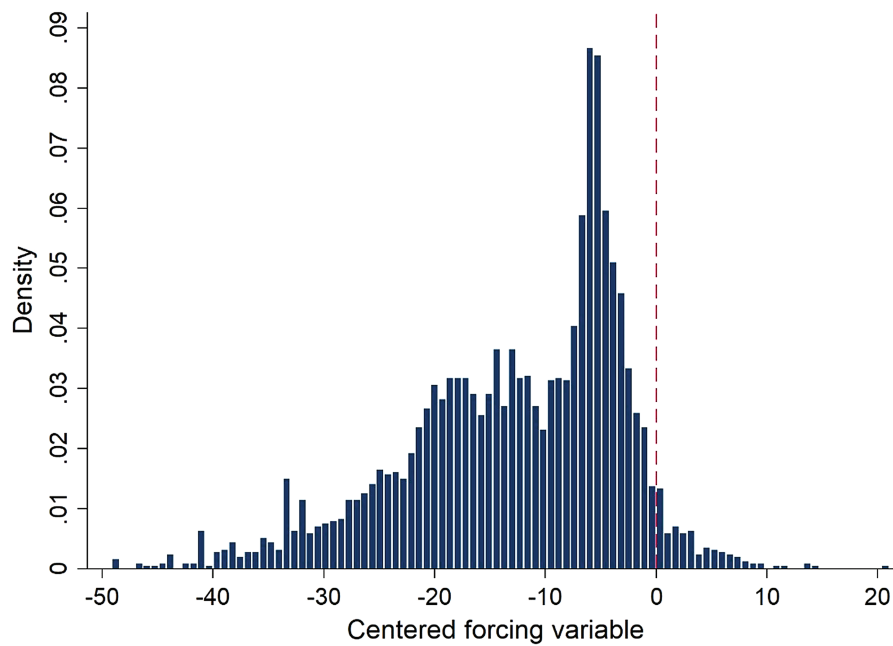
**Figure 3.** Histogram of the centered forcing variable for placement on the honor list.

of the function of the centered forcing variable and applies month and grassroots government fixed effects in the regression analyses. Column (2) of table 3 shows that compared to governments not placed on the honor list, governments entering the honor list show less performance improvement in the next month by 36 in performance rankings. The effect is statistically significant at a confidence level below  $p = .01$ . Column (1) of table 3 shows that the effect of entering the honor list on governments' performance score improvement in the next month is not statistically significant. Notably, the removal of observations achieving the full score from the sample of analysis generates a conservative estimation of the treatment effect of entering the honor list on governments' performance score improvement in the next month.

Column (3) and (4) of table 3 report the results of regressing the model specified by equation (2), showing that as grassroots governments' reliance on the higher authority's

fiscal transfers increases, they achieve more performance improvement in terms of scores and rankings after entering the honor list. In particular, a 1-standard deviation (SD) increase in the fiscal transfer ratio (0.057, according to the summary statistics presented in table 1) positively moderates the treatment effect of entering the honor list on performance score improvement by 1.546 and on performance ranking improvement by approximately 20. Based on the results presented in table 3, the satisficing hypothesis (hypothesis 1) and resource dependence hypothesis (hypothesis 3b) are supported. Of note, the satisficing hypothesis is primarily supported by the estimated results concerning grassroots governments' performance improvement in terms of rankings.

We next focus on entering the shame roll as the treatment. Column (1) and (2) of table 4 report the results of regressing the baseline model specified by equation (1), and column (3) and (4) report the results of regressing the model specified



**Figure 4.** Histogram of the centered forcing variable for placement on the shame roll.

**Table 3.** RD Estimation of the Effect of Entering the Honor List on Performance Improvement

Variables	(1) Score Improvement	(2) Ranking Improvement	(3) Score Improvement	(4) Ranking Improvement
Entering honor list	0.457 (1.608)	-36.000*** (13.07)	-0.723 (1.608)	-42.52*** (13.35)
Centered FV for honor list	-0.809*** (0.0220)	-8.138*** (0.252)	-0.810*** (0.0220)	-8.181*** (0.253)
Fiscal transfer ratio			-2.344 (3.703)	2.967 (20.11)
Entering honor list × Fiscal transfer ratio			27.12*** (5.463)	353.8*** (74.91)
Constant	-12.71*** (0.646)	-149.7*** (6.556)	-12.68*** (0.649)	-150.3*** (6.603)
Number of observations	3,207	3,300	3,162	3,252
R-squared	0.483	0.303	0.484	0.308
Number of units	333	333	330	330
Covariates	No	No	No	No
Month fixed effects	Yes	Yes	Yes	Yes
Unit fixed effects	Yes	Yes	Yes	Yes

Note: The treatment for analysis is “entering the honor list.” Column (1) and (2) report the results of regressing the baseline model specified by equation (1), and columns (3) and (4) report the results of regressing the model specified by equation (2). All columns include unit and time fixed effects. Robust standard errors are reported in parentheses.

\*\*\* $p < .01$ , \*\* $p < .05$ , \* $p < .1$  (for two-tailed tests).

by equation (2). The first two columns show that compared to governments not listed on the shame roll, governments entering the shame roll show more performance improvement in the next month by 4.382 in terms of scores and more performance improvement by approximately 27 in terms of rankings. Column (3) and (4) show that a one-SD increase in the fiscal transfer ratio positively moderates the treatment effect of entering the shame roll on performance score improvement by 1.352 and on performance ranking improvement by approximately 7. Thus, the blame avoidance hypothesis (hypothesis 2) and resource dependence hypothesis (hypothesis 3b) are supported by the results in table 4.

We next use RD plots to graphically illustrate the effect of entering the honor list and shame roll on grassroots governments’ performance improvements. In particular, we aggregate the data before plotting them and present two summaries in graphs: a global quadratic polynomial regression fit of the outcomes of interest on the centered forcing variable,

represented by a solid line, and local sample means, represented by dots.<sup>4</sup> Figures 5–8 show the binned means of the outcomes of interest against the centered forcing variable separately for the honor list and shame roll as well as a quadratic global polynomial fit estimated separately for the control and treatment groups, using the approach of mimicking variance RD plots with evenly-spaced bins.<sup>5</sup>

Figure 5 shows that the improvement in the performance score jumps at the cutoff for entering the honor list. More specifically, relative to those barely failing to enter the honor

<sup>4</sup>“The local sample means are created by first choosing disjoint (i.e., non-overlapping) intervals or ‘bins’ of the score, calculating the mean of the outcome for the observations falling within each bin, and then plotting the average outcome in each bin against the mid-point of the bin” (Cattaneo, Idrobo, and Titiunik 2019, 21).

<sup>5</sup>The results are consistent when different orders of polynomial fits (such as first- and third-order polynomial fits) and different methods of choosing the number (such as the integrated mean squared error method) and location of bins (such as the quantile-spaced method) are employed.

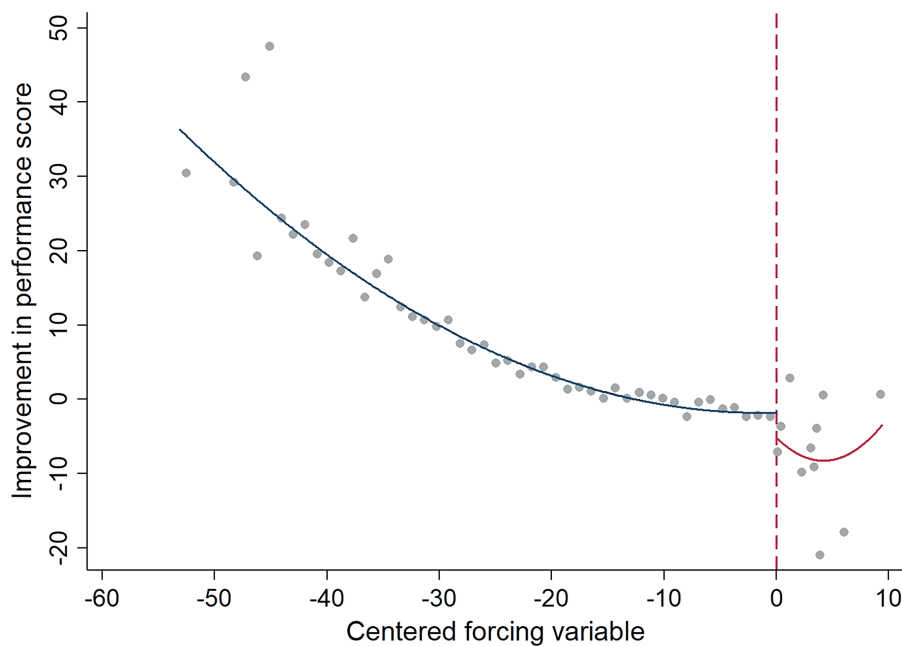


**Table 4.** RD Estimation of the Effect of Entering the Shame Roll on Performance Improvement

Variables	(1) Score Improvement	(2) Ranking Improvement	(3) Score Improvement	(4) Ranking Improvement
Enter shame roll	4.382*** (0.873)	27.41** (11.73)	4.146*** (0.868)	24.73** (11.72)
Centered FV for shame roll	0.767*** (0.0222)	8.114*** (0.259)	0.769*** (0.0223)	8.160*** (0.257)
Fiscal transfer ratio			-2.748 (3.773)	9.784 (31.95)
Enter shame roll × Fiscal transfer ratio			23.72*** (1.816)	124.8*** (18.98)
Constant	12.75*** (0.524)	109.7*** (5.479)	12.85*** (0.529)	110.9*** (5.474)
Number of observations	3,300	3,300	3,252	3,252
R-squared	0.496	0.302	0.498	0.305
Number of units	333	333	330	330
Covariates	No	No	No	No
Month fixed effects	Yes	Yes	Yes	Yes
Unit fixed effects	Yes	Yes	Yes	Yes

Note: The treatment for analysis is “entering the shame roll”. Columns (1) and (2) report the results of regressing the baseline model specified by [equation \(1\)](#), and columns (3) and (4) report the results of regressing the model specified by [equation \(2\)](#). All columns include unit and time fixed effects. Robust standard errors are reported in parentheses.

\*\*\* $p < .01$ , \*\* $p < .05$ , \* $p < .1$  (for two-tailed tests).

**Figure 5.** Mimicking variance RD plot with evenly-spaced bins (score, honor list).

list, grassroots governments barely making the honor list show less performance score improvement in the next period. Similarly, [figure 6](#) reveals that grassroots governments barely making the honor list in the current month show less performance ranking improvement in the next period. In addition, [figures 7 and 8](#) show that compared to grassroots governments positioned just to the left of the cutoff for the shame roll, those positioned just to the right of the cutoff and entering the shame roll exhibit higher performance score and ranking improvements in the next period.

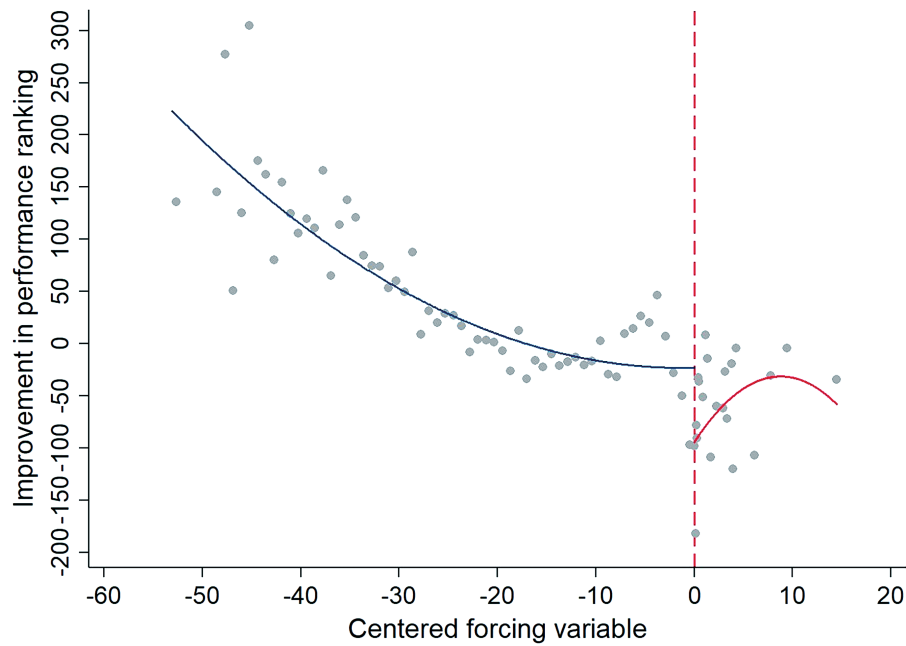
### Robustness Checks

We further conduct several robustness checks to verify the reliability of our empirical findings. The robustness checks include examinations of whether the empirical findings are sensitive to the slope of the fitted curve on either side of the

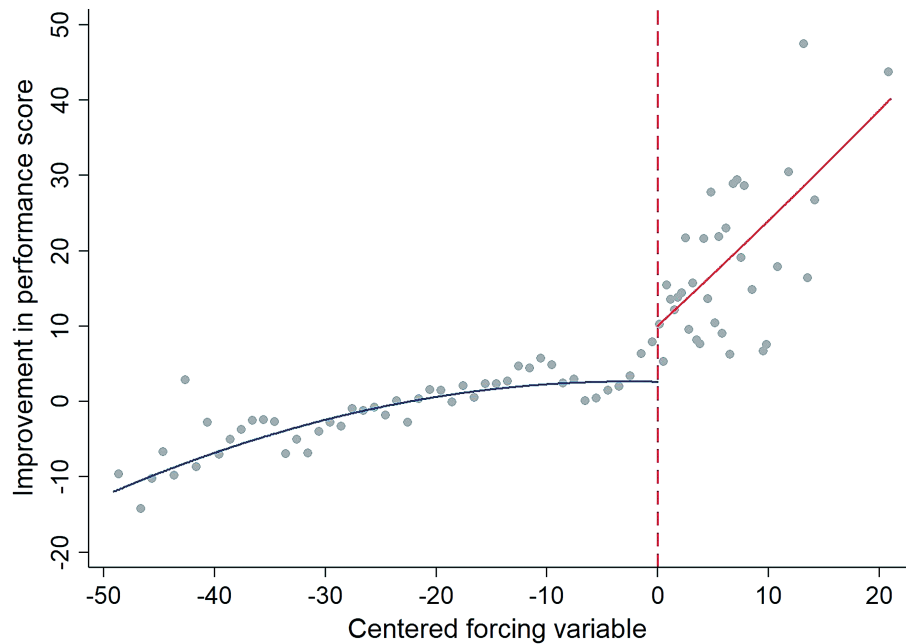
cutoff, whether the empirical findings are sensitive to the inclusion of pretreatment covariates in RD regressions, and whether the potential effect of entering the honor list and shame roll may carry through for more than one month. The procedures used for robustness checks and the corresponding empirical models and results are reported in the [supplementary appendix 4](#). The results of the robustness checks are consistent with those reported in [tables 3 and 4](#).

### Discussions

This article not only provides additional empirical evidence concerning the effect of performance feedback in the public sector but also examines a unique way in which the performance feedback effect may take place. In Western countries where political officials are chosen through competitive elections and residents are mobile across jurisdictions with



**Figure 6.** Mimicking variance RD plot with evenly-spaced bins (ranking, honor list).

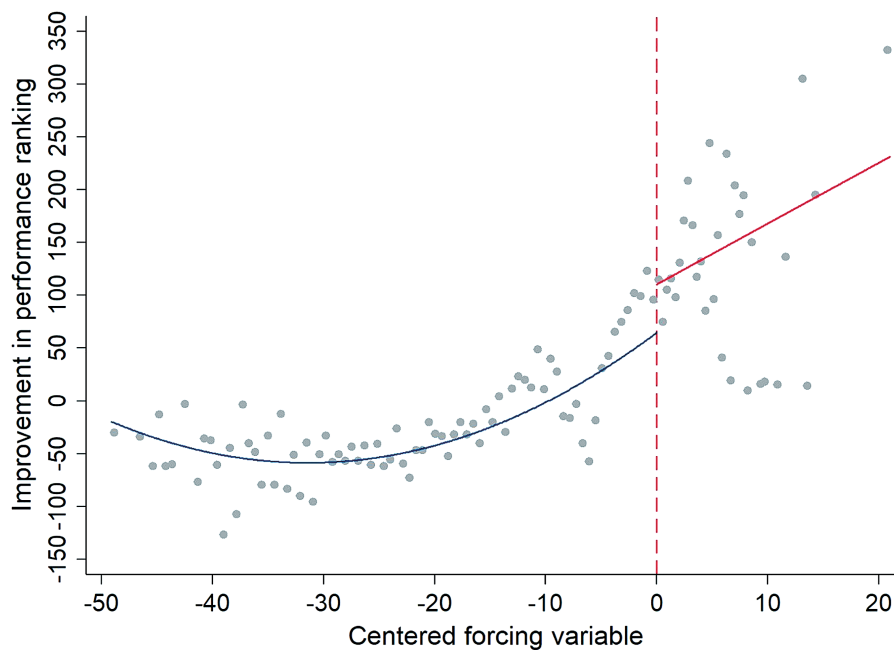


**Figure 7.** Mimicking variance RD plot with evenly-spaced bins (score, shame roll).

relatively low costs, citizens' voting with "hands and feet" constitutes the primary mechanism through which performance feedback may provoke government changes and learning and, consequently, bring about performance improvements (Flink 2019; Holm 2018; Hong 2019; Nielsen 2014). This present study advances a more nuanced understanding of the performance feedback effect in a context featuring a higher authority's strict political control over subordinate officials and a rigid top-down administrative hierarchy.

The present research also advances the understanding of performance management reforms undertaken by Chinese local governments in both theoretical and empirical respects.

Chinese local governments have borrowed substantially from the New Public Management reforms and practices of global trends, for which performance evaluation and management is the core theme (Walker and Wu 2010; Wescott and Jones 2007). Previous studies focusing on the performance management of Chinese local governments have long debated whether their reforms and practices are driven by international trends of marketization and decentralization or by specific political needs (Chan and Suizhou 2007; Walker and Wu 2010). The international literature generally views performance management as a pivotal means to strengthen bureaucrat accountability and improve government effectiveness



**Figure 8.** Mimicking variance RD plot with evenly-spaced bins (ranking, shame roll).

and efficiency, while the arguments featuring Chinese characteristics hold that performance management in China is intended “to manage central-local relations, to enhance state capacity, and to manage and control local officials” (Gao 2009; Walker and Wu 2010, 37S). In particular, as mentioned above, in a regime featuring a higher authority’s strict political control over subordinate officials, performance measurement is usually used to “ensure that local officials comply with higher-level policy priorities” (Chan and Gao 2008, 8). Indeed, the present research introduces a typical case of performance management reform employed under strict political control over local governments by a higher authority. Our findings thereby add evidence to the literature concerning the consequences of performance management reforms amid a social and political environment substantially distinct from that of the Western context.

One issue that deserves more discussions concerns the possibility of simple regression to the mean in our empirical estimation, which may constitute an alternative explanation of the empirical findings and is consistent with predictions based on isomorphism theory. Gerrish and Spreen (2017, 602) describe isomorphism as a phenomenon where “local governments will mimic the position of the average unit, resulting in a decline in the indicator’s variance without improvement in the average value.” This contradicts the circumstance of real performance improvement defined as an increase in the group mean value, regardless of the change in the SD. We follow the approach of Gerrish and Spreen (2017) by using fixed effects regressions and an ANOVA-variant, Bartlett’s test to test the possibility of isomorphism. Specifically, we regress grassroots governments’ 12345 hotline performance scores on dummies indicating time periods, including grassroots government fixed effects. The results suggest that governments’ performance scores show a trend of increasing over time during the sample period. In addition, Bartlett’s test of equal variance between groups fails to reject the hypothesis that the groups of grassroots governments have equal variances of performance scores. Therefore, the statistical evidence largely alleviates

the concern of isomorphism of organizational performance in our case. In addition, the RD estimations reported above show the LATE at the cutoff instead of the effect by average (Cattaneo, Idrobo, and Titiunik 2019), further mitigating the concern of isomorphism described by Gerrish and Spreen (2017).

Another issue being worthy of further discussions relates to the conclusion regarding the effects of centralization and punishments on organizational performance in the public sector. Our present study reveals positive effects of centralization and punishments on government performance; however, we do not intend to claim universal applicability of the evidence or regard centralization and punishments as panaceas for performance improvement. Existing findings about the effect of punishments and rewards on government performance are generally mixed or inconsistent (Andreoni, Harbaugh, and Vesterlund 2003; Chen, Ramamurthy, and Wen 2012; Podsakoff et al. 2006). In addition, theoretically, centralized structures have both benefits and harm to organizational performance. As pointed out by Andrews et al. (2009, 60), centralized authority and decision making can facilitate greater decision speed, provide firm direction and goals, establish clear lines of authority, and improve professional qualifications of bureaucrats (Elgin and Carter 2019). However, centralization may also harm organizational performance “by preventing middle managers and street-level bureaucrats from making independent decisions, enshrining inflexible rules and procedures, and undermining responsiveness to changing environmental circumstances” (Andrews et al. 2009, 60). Empirically, existing studies in both the private and public sector have failed to find a consistent relationship between centralized or decentralized structures and organizational performance (see Andrews et al. (2009) for a literature review).

The effects of centralization and punishments on government performance may rest on many other organizational characteristics or contextual factors, such as resource reliance on external entities (Pfeffer and Salancik 1978), administrative



capacity (Elgin and Carter 2019), and political scenarios (Beazer 2015). We speculate that in our case of study, at least two contextual circumstances lead to the findings of positive effects of centralization and punishments. First, amid a regime featuring a higher authority's strict political control over subordinate officials and a rigid top-down administrative hierarchy, grassroots governments' great reliance on the higher authority's political and budgetary resources compels local officials to follow the higher authority's directives (Chan and Gao 2008) and to prevent themselves from falling behind the peer governments in the 12345 hotline performance. Second, local officials' insufficient administrative capacity and weak intrinsic public service motivations may constitute another critical context for our findings. In particular, centralization and punishments can force grassroots government officials to invest substantially in the performance indicators to which the higher authority place more priorities. This helps ameliorate grassroots governments' capacity deficiency in dealing with citizens' demands and complaints through the 12345 hotline. In addition, external pressures through centralized directives and punishments can create extrinsic motivations for local officials to take prompt actions to improve the 12345 hotline performance.

It is also helpful to discuss the potential limitations of this research and point out directions for further study. First, the present research addresses grassroots governments' 12345 hotline performance scores and rankings as the primary focus of interest, neither of which directly reflect a socially desirable outcome (Dias and Maynard-Moody 2007; Hong 2017, 2019). Astute local officials may adopt various strategies to increase their performance scores and rankings in the short run; however, the real perceptions of local citizens have not been closely examined. Further studies may investigate how performance feedback can influence citizens' perceptions of and trust in government officials and public institutions. Second, this research examines the performance feedback effect in the short term (the next month), which is reasonable because governments' performance is evaluated and ranked on a monthly timeline. Nevertheless, local officials may need more than one month to fully adjust their managerial strategies and actions. Thus, further research on the long-term effect of performance feedback using a more extended sample period is warranted. Third, this article specifically attends to governments' fiscal capacity as the moderating factor of the performance feedback effect. Indeed, many more factors of the external environment and internal management can be considered for empirical tests, which will be valuable for both theoretical development and empirical inquiry regarding the mechanism through which the performance feedback effect is realized in the public sector (Meier, Favero, and Zhu 2015). In addition, by focusing on a recent reform of public performance management adopted in China, this research responds to the call for additional studies by Hong (2019, 2020) and provides evidence of external validity for existing conclusions. Future studies of the performance feedback effect in the public sector conducted in various political, social, and cultural contexts are still worthwhile.

Another potential limitation concerns the narrow construction of performance indicators based on available data at the expense of unmeasured outcomes (Koning and Heinrich 2013; Pasha, Kroll, and Ash 2021). In particular, among the three indicators of 12345 hotline performance, one indicator may be improved at the expense of another. Statistical tests

show that the performance indicators have Pearson correlation coefficients of close to 0.9, which is statistically significant at a confidence level below  $p = .001$ . This result largely alleviates the concern of trade-offs between multiple performance indicators. In addition, the records of calls to the 12345 hotline show that the most frequently covered topics include market management, traffic management, utility services, public safety, basic municipal services, environmental protection, and medical services. Although most of the demands and complaints from citizens are unlikely to have serious side effects, equity concerns remain because silent populations may be disadvantaged. Therefore, we cannot fully rule out that performance improvements in the measured areas were achieved by shifting attention and resources away from other important but unmeasured areas, thereby creating the potential to harm less-vocal stakeholders.

## Conclusion

This article focuses on a prominent public performance management reform implemented in the capital city of China that purports to promote grassroots governments' responsiveness to citizens' inquiries, demands, and complaints. To achieve this goal, the BMG releases and ranks the 333 grassroots governments' monthly 12345 hotline performance and introduces the honor list and shame roll as a means of extolling well-performing governments and denouncing poorly performing governments. Due to the great importance attached to 12345 hotline performance by the political leader, significant rewards and punishments are linked to placement on the honor list and shame roll, respectively. In this way, being placed on the honor list or shame roll serves as a specific type of performance feedback. This article examines whether and how entering the honor list and shame roll can affect grassroots governments' performance improvement in the next period and how governments' fiscal capacity can moderate this effect.

The theoretical inference draws insights from Simon's (1955, 1972) theory of bounded rationality and Cyert and March's (1963) behavioral theory of organizations, suggesting that an organization's past performance can influence its future decision making, which in turn shapes the resultant organizational performance (Cyert and March 1963; Gavetti et al. 2012; Greve 2008; Hong 2019, 2020; Nicholson-Crotty, Nicholson-Crotty, and Fernandez 2017). First, bounded rationality causes organizations to pursue performance satisfaction rather than maximization (Simon 1955, 1972). Whether a certain level of performance is satisfactory rests on a comparison between actual performance and social or historical aspirations. Entering the honor list suggests that governments have already achieved a level of performance exceeding the social aspiration, thus attenuating officials' desire to further improve their future performance (*satisficing hypothesis*). Second, given that blame avoidance serves as a powerful behavioral motivation for local officials (Hood 2007; Weaver 1986), governments placed on the shame roll are presumably compelled to actively conduct problem screening and solution searching to close the negative aspiration gap (*blame avoidance hypothesis*). Third, governments' fiscal reliance on the higher authority may either positively moderate the performance feedback effect (*resource dependence hypothesis*) or exert a negative moderating effect (*organizational capacity hypothesis*).

We use an RD design to test the proposed hypotheses. The empirical findings show that compared to governments not placed on the honor list, governments entering the honor list show less performance ranking improvement in the next period. In contrast, compared to governments not placed on the shame roll, governments entering the shame roll are more likely to improve their performance scores and rankings in the next period. In addition, as governments' reliance on the higher authority's fiscal transfers increases, the treatment effect of entering the honor list and shame roll on their performance improvement in the next period becomes more positive. In other words, the level of performance improvement is higher at the cutoff for both the honor list and shame roll if grassroots governments rely more on the higher authority's fiscal resources.

## Supplementary Material

Supplementary data are available at *Journal of Public Administration Research and Theory* online.

## Funding

This work was supported by the National Natural Science Foundation of China (72104242 to Wenchi Wei and 71874198 to Wenzhao Li) and the Beijing Social Science Foundation (20JCC023).

## Acknowledgments

We would like to thank the editors of *JPART* and three anonymous reviewers for highly constructive comments of previous versions of this paper.

## Data Availability

The data underlying this article are available in the Harvard Dataverse, at <https://doi.org/10.7910/DVN/766RSE>.

## References

- Ammons, D. N., and D. J. Roenigk. 2015. Benchmarking and interorganizational learning in local government. *Journal of Public Administration Research and Theory* 25(1): 309–35.
- Andreoni, J., W. Harbaugh, and L. Vesterlund. 2003. The carrot or the stick: Rewards, punishments, and cooperation. *American Economic Review* 93(3): 893–902.
- Andrews, R., and G. A. Boyne. 2010. Capacity, leadership, and organizational performance: Testing the black box model of public management. *Public Administration Review* 70(3): 443–54.
- Andrews, R., G. A. Boyne, J. Law, and R. M. Walker. 2009. Centralization, organizational strategy, and public service performance. *Journal of Public Administration Research and Theory* 19(1): 57–80.
- Andrews, R., G. A. Boyne, and R. M. Walker. 2006. Strategy content and organizational performance: An empirical analysis. *Public Administration Review* 66(1): 52–63.
- Askim, Jostein, Å. Johnsen, and K.-A. Christophersen. 2008. Factors behind organizational learning from benchmarking: Experiences from Norwegian municipal benchmarking networks. *Journal of Public Administration Research and Theory* 18(2): 297–320.
- Beazer, Q. H. 2015. Political centralization and economic performance: Evidence from Russia. *The Journal of Politics* 77(1): 128–45.
- Bolton, A., R. Augustine Potter, and S. Thrower. 2016. Organizational capacity, regulatory review, and the limits of political control. *The Journal of Law, Economics, and Organization* 32(2): 242–71.
- Bourdeaux, C., and G. Chikoto. 2008. Legislative influences on performance management reform. *Public Administration Review* 68(2): 253–65.
- Boyne, G. A., and R. M. Walker. 2010. Strategic management and public service performance: The way ahead. *Public Administration Review* 70: s185–92.
- Brody, S. D., J. Eun Kang, and S. Bernhardt. 2010. Identifying factors influencing flood mitigation at the local level in Texas and Florida: The role of organizational capacity. *Natural Hazards* 52(1): 167–84.
- Bromiley, P. 1991. Testing a causal model of corporate risk taking and performance. *Academy of Management Journal* 34(1): 37–59.
- Brown, M. 2012. Enhancing and measuring organizational capacity: Assessing the results of the US Department of Justice Rural Pilot Program Evaluation. *Public Administration Review* 72(4): 506–15.
- Brown, K., and P. B. Coulter. 1983. Subjective and objective measures of police service delivery. *Public Administration Review* 43(1): 50–58.
- Cai, Y. 2008. Power structure and regime resilience: Contentious politics in China. *British Journal of Political Science* 38(3): 411–32.
- Calonico, S., M. D. Cattaneo, and R. Titiunik. 2015. Optimal data-driven regression discontinuity plots. *Journal of the American Statistical Association* 110(512): 1753–69.
- Carver, C. S., and T. L. White. 1994. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology* 67(2): 319–33.
- Casciaro, T., and M. J. Piskorski. 2005. Power imbalance, mutual dependence, and constraint absorption: A closer look at resource dependence theory. *Administrative Science Quarterly* 50(2): 167–99.
- Cattaneo, Matias D., Nicolás Idrobo, and Rocío Titiunik. 2019. *A practical introduction to regression discontinuity designs: Foundations*. Cambridge: Cambridge Univ. Press.
- Cattaneo, M. D., M., Jansson, and X. Ma. 2020. Simple local polynomial density estimators. *Journal of the American Statistical Association* 115(531): 1449–55.
- . 2021. Local regression distribution estimators. *Journal of Econometrics*. Online first. doi:10.1016/j.jeconom.2021.01.006
- Chan, H. S., and J. Gao. 2008. Performance measurement in Chinese local governments: Guest editors' introduction. *Chinese Law & Government* 41(2–3): 4–9.
- Chan, H. S., and E. Li Suizhou. 2007. Civil service law in the People's Republic of China: A return to cadre personnel management. *Public Administration Review* 67(3): 383–98.
- Chen, Y., K. Ramamurthy, and K.-W. Wen. 2012. Organizations' information security policy compliance: Stick or carrot approach? *Journal of Management Information Systems* 29(3): 157–88.
- Chen, Ye, H. Li, and L. A. Zhou. 2005. Relative performance evaluation and the turnover of provincial leaders in China. *Economics Letters* 88(3): 421–25.
- Choi, E. K. 2012. Patronage and performance: Factors in the political mobility of provincial leaders in post-Deng China. *The China Quarterly* 212: 965–81.
- Chun, Y. H., and H. G. Rainey. 2005. Goal ambiguity and organizational performance in US federal agencies. *Journal of Public Administration Research and Theory* 15(4): 529–57.
- Coe, C. K., and J. R. Brunet. 2006. Organizational report cards: Significant impact or much ado about nothing? *Public Administration Review* 66(1): 90–100.
- Cyert, Richard M., and James G. March. 1963. *A behavioral theory of the firm*. Vol. 2, no. 4. Englewood Cliffs, NJ: Prentice-Hall.
- Dias, J. J., and S. Maynard-Moody. 2007. For-profit welfare: Contracts, conflicts, and the performance paradox. *Journal of Public Administration Research and Theory* 17(2): 189–211.
- Elgin, D. J., and D. P. Carter. 2019. Administrative (de) centralization, performance equity, and outcome achievement in rural contexts: An empirical study of US child welfare systems. *Governance* 32(1): 23–43.
- Emerson, R. M. 1962. Power-dependence relations. *American Sociological Review* 27(1): 31–41.

- Ferreira, F., and J. Gyourko. 2009. Do political parties matter? Evidence from US cities. *The Quarterly Journal of Economics* 124(1): 399–422.
- Flink, C. M. 2019. Predicting budgetary change: The effect of performance gaps. *Journal of Public Administration Research and Theory* 29(2): 227–37.
- Folke, O., and J. M. Snyder. 2012. Gubernatorial midterm slumps. *American Journal of Political Science* 56(4): 931–48.
- Forester, J. 1984. Bounded rationality and the politics of muddling through. *Public Administration Review* 44(1): 23–31.
- Fredericksen, P., and L. Rosanne. 2000. Disconnect in the hollow state: The pivotal role of organizational capacity in community-based development organizations. *Public Administration Review* 60(3): 230–39.
- Fukuyama, F. 2013. What is governance? *Governance* 26(3): 347–68.
- Gao, J. 2009. Governing by goals and numbers: A case study in the use of performance measurement to build state capacity in China. *Public Administration and Development* 29(1): 21–31.
- Gavetti, G., H. R. Greve, D. A. Levinthal, and W. Ocasio. 2012. The behavioral theory of the firm: Assessment and prospects. *Academy of Management Annals* 6(1): 1–40.
- Gerrish, E., and T. L. Spreen. 2017. Does benchmarking encourage improvement or convergence? Evaluating North Carolina's fiscal benchmarking tool. *Journal of Public Administration Research and Theory* 27(4): 596–614.
- Gray, Jeffrey Alan. 1987. *The psychology of fear and stress*. Cambridge, UK: Cambridge Univ. Press.
- Greve, H. R. 2008. A behavioral theory of firm growth: Sequential attention to size and performance goals. *Academy of Management Journal* 51(3): 476–94.
- Holm, J. M. 2017. Double standards? How historical and political aspiration levels guide managerial performance information use. *Public Administration* 95(4): 1026–40.
- . 2018. Successful problem solvers? Managerial performance information use to improve low organizational performance. *Journal of Public Administration Research and Theory* 28(3): 303–20.
- Hong, S. 2017. What are the areas of competence for central and local governments? Accountability mechanisms in multi-level governance. *Journal of Public Administration Research and Theory* 27(1): 120–34.
- . 2019. A behavioral model of public organizations: Bounded rationality, performance feedback, and negativity bias. *Journal of Public Administration Research and Theory* 29(1): 1–17.
- . 2020. Performance management meets red tape: Bounded rationality, negativity bias, and resource dependence. *Public Administration Review* 80(6): 932–45.
- Hong, S., S. H. Kim, and J. Son. 2020. Bounded rationality, blame avoidance, and political accountability: How performance information influences management quality. *Public Management Review* 22(8): 1240–63.
- Hood, C. 2007. What happens when transparency meets blame-avoidance? *Public Management Review* 9(2): 191–210.
- Huber, J. D., and N. McCarty. 2004. Bureaucratic capacity, delegation, and political reform. *American Political Science Review* 98(3): 481–94.
- Huber, John D., and Charles R. Shipan. 2002. *Deliberate discretion?: The institutional foundations of bureaucratic autonomy*. Cambridge: Cambridge Univ. Press.
- Imbens, G. W., and T. Lemieux. 2008. Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142(2): 615–35.
- Jones, B. D. 2003. Bounded rationality and political science: Lessons from public administration and public policy. *Journal of Public Administration Research and Theory* 13(4): 395–412.
- Joseph, J., and V. Gaba. 2015. The fog of feedback: Ambiguity and firm responses to multiple aspiration levels. *Strategic Management Journal* 36(13): 1960–78.
- Julnes, Patria de Lancer, and M. Holzer. 2001. Promoting the utilization of performance measures in public organizations: An empirical study of factors affecting adoption and implementation. *Public Administration Review* 61(6): 693–708.
- Klein, J. G., and R. Ahluwalia. 2005. Negativity in the evaluation of political candidates. *Journal of Marketing* 69(1): 131–42.
- Koning, P., and C. J. Heinrich. 2013. Cream-skimming, parking and other intended and unintended effects of high-powered, performance-based contracts. *Journal of Policy Analysis and Management* 32(3): 461–83.
- Kroll, A., and D. P. Moynihan. 2015. Does training matter? Evidence from performance management reforms. *Public Administration Review* 75(3): 411–20.
- Lau, R. R. 1982. Negativity in political perception. *Political Behavior* 4(4): 353–77.
- . 1985. Two explanations for negativity effects in political behavior. *American Journal of Political Science* 29(1): 119–38.
- Lavertu, S., and D. P. Moynihan. 2013. Agency political ideology and reform implementation: Performance management in the Bush administration. *Journal of Public Administration Research and Theory* 23(3): 521–49.
- Lee, D. S., and T. Lemieux. 2010. Regression discontinuity designs in economics. *Journal of Economic Literature* 48(2): 281–355.
- Lee, D. S., E. Moretti, and M. J. Butler. 2004. Do voters affect or elect policies? Evidence from the US House. *The Quarterly Journal of Economics* 119(3): 807–59.
- Li, H., and L.-A. Zhou. 2005. Political turnover and economic performance: The incentive role of personnel control in China. *Journal of Public Economics* 89(9–10): 1743–62.
- Liang, Huigang, Y. Xue, and L. Wu. 2013. Ensuring employees' IT compliance: Carrot or stick? *Information Systems Research* 24(2): 279–94.
- Licari, M. J., W. McLean, and T. W. Rice. 2005. The condition of community streets and parks: A comparison of resident and nonresident evaluations. *Public Administration Review* 65(3): 360–68.
- Malatesta, D., and C. R. Smith. 2011. Resource dependence, alternative supply sources, and the design of formal contracts. *Public Administration Review* 71(4): 608–17.
- . 2014. Lessons from resource dependence theory for contemporary public and nonprofit management. *Public Administration Review* 74(1): 14–25.
- McGraw, K. M. 1991. Managing blame: An experimental test of the effects of political accounts. *American Political Science Review* 85(4): 1133–57.
- McGuire, M., and C. Silvia. 2010. The effect of problem severity, managerial and organizational capacity, and agency structure on intergovernmental collaboration: Evidence from local emergency management. *Public Administration Review* 70(2): 279–88.
- Meier, K. J., N. Favero, and L. Zhu. 2015. Performance gaps and managerial decisions: A Bayesian decision theory of managerial action. *Journal of Public Administration Research and Theory* 25(4): 1221–46.
- Meier, K. J., and L. J. O'Toole, Jr. 2002. Public management and organizational performance: The effect of managerial quality. *Journal of Policy Analysis and Management* 21(4): 629–43.
- . 2003. Public management and educational performance: The impact of managerial networking. *Public Administration Review* 63(6): 689–99.
- Moynihan, D. P., and N. Landuyt. 2009. How do public organizations learn? Bridging cultural and structural perspectives. *Public Administration Review* 69(6): 1097–105.
- Moynihan, D. P., and S. K. Pandey. 2005. Testing how management matters in an era of government by performance management. *Journal of Public Administration Research and Theory* 15(3): 421–39.
- . 2010. The big question for performance management: Why do managers use performance information? *Journal of Public Administration Research and Theory* 20(4): 849–66.
- Nicholson-Crotty, S., J. Nicholson-Crotty, and S. Fernandez. 2017. Performance and management in the public sector: Testing a model of relative risk aversion. *Public Administration Review* 77(4): 603–14.



- Nicholson-Crotty, S., N. A. Theobald, and J. Nicholson-Crotty. 2006. Disparate measures: Public managers and performance-measurement strategies. *Public Administration Review* 66(1): 101–13.
- Nielsen, P. A. 2014. Learning from performance feedback: Performance information, aspiration levels, and managerial priorities. *Public Administration* 92(1): 142–60.
- Nielsen, P. A., and D. P. Moynihan. 2017. How do politicians attribute bureaucratic responsibility for performance? Negativity bias and interest group advocacy. *Journal of Public Administration Research and Theory* 27(2): 269–83.
- Olsen, A. L. 2015. Citizen (dis) satisfaction: An experimental equivalence framing study. *Public Administration Review* 75(3): 469–78.
- O'Toole, L. J. Jr., and K. J. Meier. 1999. Modeling the impact of public management: Implications of structural context. *Journal of Public Administration Research and Theory* 9(4): 505–26.
- Pasha, O., A. Kroll, and M. Ash. 2021. Assessing the equity and effectiveness of PerformanceStat systems. *International Public Management Journal*. Online first. doi:[10.1080/10967494.2021.1918300](https://doi.org/10.1080/10967494.2021.1918300)
- Pfeffer, J., and P. Nowak. 1976. Joint ventures and interorganizational interdependence. *Administrative Science Quarterly* 21(3): 398–418.
- Pfeffer, Jeffrey, and Gerald R. Salancik. 1978. *The external control of organizations: A resource dependence perspective*. New York: Harper & Row.
- Podsakoff, P. M., W. H. Bommer, N. P. Podsakoff, and S. B. MacKenzie. 2006. Relationships between leader reward and punishment behavior and subordinate attitudes, perceptions, and behaviors: A meta-analytic review of existing and new research. *Organizational Behavior and Human Decision Processes* 99(2): 113–42.
- Poister, T. H., O. Q. Pasha, and L. Hamilton Edwards. 2013. Does performance management lead to better outcomes? Evidence from the US public transit industry. *Public Administration Review* 73(4): 625–36.
- Robinson, W. S. 1950. Ecological correlations and the behavior of individuals. *American Sociological Review* 15(3): 351–57.
- Rozin, P., and E. B. Royzman. 2001. Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review* 5(4): 296–320.
- Salge, T. O. 2011. A behavioral model of innovative search: Evidence from public hospital services. *Journal of Public Administration Research and Theory* 21(1): 181–210.
- Sanger, M. B. 2008. From measurement to management: Breaking through the barriers to state and local performance. *Public Administration Review* 68(s1): S70–85.
- Schachter, H. L. 2010. Objective and subjective performance measures: A note on terminology. *Administration & Society* 42(5): 550–67.
- Shingler, J., M. E. Van Loon, T. R. Alter, and J. C. Bridger. 2008. The importance of subjective data for public agency performance evaluation. *Public Administration Review* 68(6): 1101–11.
- Simon, H. A. 1955. A behavioral model of rational choice. *The Quarterly Journal of Economics* 69(1): 99–118.
- . 1972. Theories of bounded rationality. *Decision and Organization* 1(1): 161–76.
- . 1997. *Administrative behavior*. New York, NY: The Free Press.
- Song, M., and K. J. Meier. 2018. Citizen satisfaction and the kaleidoscope of government performance: How multiple stakeholders see government performance. *Journal of Public Administration Research and Theory* 28(4): 489–505.
- Soroka, S. N. 2006. Good news and bad news: Asymmetric responses to economic information. *The Journal of Politics* 68(2): 372–85.
- Soroka, S., and S. McAdams. 2015. News, politics, and negativity. *Political Communication* 32(1): 1–22.
- Walker, R. M., F. Damanpour, and C. A. Devece. 2011. Management innovation and organizational performance: The mediating effect of performance management. *Journal of Public Administration Research and Theory* 21(2): 367–86.
- Walker, R. M., and J. Wu. 2010. Future prospects for performance management in Chinese city governments. *Administration & Society* 42(suppl. 1): 345–555.
- Wang, C., and L. Ma. 2020. How does performance feedback impact performance improvement—evidence from the case of “sue for action” in Beijing (in Chinese). *Chinese Public Administration* 11(425): 117–25.
- Weaver, R. K. 1986. The politics of blame avoidance. *Journal of Public Policy* 6(4): 371–98.
- Wescott, C. G., and L. R. Jones. 2007. Managing for results and performance in Asia: Assessing reform initiatives in the public sector. *International Public Management Review* 8(1): 56–102.
- Xu, C. 2011. The fundamental institutions of China's reforms and development. *Journal of Economic Literature* 49(4): 1076–151.
- Yang, K. 2005. Public administrators' trust in citizens: A missing link in citizen involvement efforts. *Public Administration Review* 65(3): 273–85.
- Yang, K., and M. Holzer. 2006. The performance–trust link: Implications for performance measurement. *Public Administration Review* 66(1): 114–26.
- Zhang, X. 2006. Fiscal decentralization and political centralization in China: Implications for growth and inequality. *Journal of Comparative Economics* 34(4): 713–26.
- Zhu, L., and A. Rutherford. 2019. Managing the gaps: How performance gaps shape managerial decision making. *Public Performance & Management Review* 42(5): 1029–61.