

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Meritocracy in Autocracies: Origins and Consequences

Permalink

<https://escholarship.org/uc/item/4h31j156>

Author

Li, Weijia

Publication Date

2018

Peer reviewed|Thesis/dissertation

Meritocracy in Autocracies: Origins and Consequences

by

Weijia Li

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Gérard Roland, Chair

Professor Ernesto Dal Bó

Associate Professor Fred Finan

Associate Professor Yuriy Gorodnichenko

Spring 2018

Meritocracy in Autocracies: Origins and Consequences

Copyright 2018

by

Weijia Li

Abstract

Meritocracy in Autocracies: Origins and Consequences

by

Weijia Li

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Gérard Roland, Chair

This dissertation explores how to solve incentive problems in autocracies through institutional arrangements centered around political meritocracy. The question is fundamental, as merit-based rewards and promotion of politicians are the cornerstones of key authoritarian regimes such as China. Yet the grave dilemmas in bureaucratic governance are also well recognized. The three essays of the dissertation elaborate on the various solutions to these dilemmas, as well as problems associated with these solutions. Methodologically, the dissertation utilizes a combination of economic modeling, original data collection, and empirical analysis.

The first chapter investigates the puzzle why entrepreneurs invest actively in many autocracies where unconstrained politicians may heavily expropriate the entrepreneurs. With a game-theoretical model, I investigate how to constrain politicians through rotation of local politicians and meritocratic evaluation of politicians based on economic growth. The key finding is that, although rotation or merit-based evaluation alone actually makes the holdup problem even worse, it is exactly their combination that can form a credible constraint on politicians to solve the hold-up problem and thus encourages private investment. An extension of the model also demonstrates that rotation and merit-based evaluation reduces politicians' entrenched interests in existing firms. This allows new firms to enter the market, which sustains Schumpeterian "creative destruction" and long-term growth. In other words, the combination of rotation and merit-based evaluation achieves both commitment and strong flexibility, a property rarely satisfied by other commitment devices. Firm-level panel data from China are further consistent with the main predictions from my model.

The second chapter focuses on another critical dilemma in an autocracy, the loyalty-competence trade-off. An autocrat usually refuses to appoint a competent governor with a broad discretionary power because the governor can use his competence and discretion to challenge the autocrat. Through a game-theoretical model, I show that one-party state can potentially solve the dilemma by appointing both a party secretary and a governor to co-rule a province. The party secretary controls political power, while the governor commands the provision of the public good. The arrangement forestalls local attempts to challenge the autocrat and establishes the autocrat's confidence to promote meritocracy and decentralization. I also characterize the optimal party-government relationship: the secretary should sometimes dominate over the governor in public good provision but not always the case. This is very different from canonical theories on the separation of powers in a democracy. The model is motivated by the experience of China's party state, whose elements of meritocracy and dual leadership are modern incarnations of key institutions in Imperial China. From key historical records, I construct variables measuring political institutions over 1,300 years in Chinese history via textual analysis techniques. The statistical analysis uncovers a long-run correlation between meritocracy and dual leadership, showing the first order relevance of the theory.

In the first chapter, I show how a meritocratic government with strong bureaucratic rules can contribute to a thriving market economy. In the third chapter, I also look at the other direction: when and how private economy contributes to bureaucratic capacity. This helps answer the question why some regimes maintain a persistent meritocracy, while aristocrats capture others. I show that in an environment with weak property rights, a meritocratic government and a thriving private economy reinforce each other. This matches important historical episodes such as the "Tang-Song Transition" in Medieval China. In an extension, I also show that legal property rights sufficiently strong can eliminate multiple equilibria, and that stronger property rights cause a more dynastic government. The extended model is employed to illustrate the difference between meritocracies in Imperial China and Ottoman Empire. It also implies a causal relationship between comparative law and state building.

To my family

Contents

Contents	ii
List of Figures	iv
List of Tables	v
1 Rotation, Meritocracy, and Property Rights	1
1.1 Introduction	1
1.2 Literature	3
1.3 A Workhorse Model	5
1.4 Rotation, Performance-based Rewards, and “Creative Destruction”	17
1.5 Extension: Multiple Rounds of Bargaining	23
1.6 Empirical Evidence	26
1.7 Conclusion	33
2 Meritocracy, Decentralization, and Local Dual Leadership	39
2.1 Introduction	39
2.2 Literature	43
2.3 Party-state Structure in Chinese Provinces	45
2.4 The Benchmark Model	47
2.5 Delegation, Decentralization, and Political Meritocracy	55
2.6 Weak Delegation: The Secretary’s Complete Dominance over the Governor	60
2.7 Dynamic Concern and Ambiguity about the Party-Government Relationship	62
2.8 Case Studies of Modern Chinese Politics	64
2.9 Case Studies from Imperial China	66
2.10 Conclusion	74
3 Aristocracy, Meritocracy, and Property Rights	76
3.1 Introduction	76
3.2 The Workhorse Model	79
3.3 Patronage, Legal Property Rights, and Meritocratic Government	91
3.4 Conclusion	97

Bibliography	98
A Appendices of <i>Rotation, Meritocracy, and Property Rights</i>	106
A.1 Proofs for Propositions in the Text	106
A.2 The Model when the New Politician is never Informed	118
A.3 General Functional Forms for Performance-based Rewards	119
B Appendices for <i>Meritocracy, Decentralization, and Local Dual Leadership</i>	137
B.1 Proofs for Propositions in the Text	137
B.2 Micro-foundation for Focusing on Equilibria with Pure Strategy for Section 3.5 and Section 3.6	156
B.3 A Model with Alternative Timeline	159

List of Figures

1.1	The Benchmark Model	8
1.2	Rotation Helps Performance Rewards	12
1.3	The Workhorse Model that Illustrates the Interdependence	16
1.4	The Model with Firm Dynamics	22
1.5	Robustness to Multiple Rounds of Bargaining	24
1.6	Rotation and Promotion in Local Politics	27
1.7	Proxying Rotation and Promotion Events	28
2.1	Meritocracy and Dual Local Leadership	69
2.2	with Alternative Proxy for Dual Leadership	70
2.3	Rebellion, Dual Leadership, and Meritocracy	70
2.4	Rebellion and Dual Leadership	71
2.5	Rebellion and Dual Leadership with Alternative Measure	71
3.1	Multiple Equilibria: Aristocracy vs. Meritocracy	83
3.2	Multiple Equilibria with Personnel Control	87
3.3	Better Legal Property Rights lead to a More Entrenched Bureaucracy	94
A.1	General Functional Forms – fixed α	123
A.2	General Functional Forms – $\alpha_1 > \alpha_2$	123

List of Tables

1.1	Summary Statistics Part I	34
1.2	Summary Statistics Part II	35
1.3	Relevance of Others' Retirement on Own Turnover	36
1.4	Anticipated Job Vacancy on Capital Composition	37
1.5	Timing of Anticipated Job Vacancy on Capital Composition	38
2.1	An Example of Normalization	68
2.2	The Dickey-Fuller Test of Cointegration	70
2.3	Meritocracy and the First Measure of Dual Leadership	73
2.4	Meritocracy and the Second Measure of Dual Leadership	73

Acknowledgments

I must express my deep debt to my primary advisor Professor Gérard Roland. He has been the guiding star in my career development and truly a father figure to not only me but to many other students. His lectures on comparative economics are the best examples of the grand vision and analytic sophistication an economist can and should have. They kindled my enthusiasm for my fields, and my whole research agenda on political meritocracy engenders from the comparative economics course: the second chapter of this dissertation, for example, starts as its term paper. As to his office hours, attending them has always been the high points of my graduate studies. There Professor Roland lends unreserved support that helps me overcome difficulties in both work and life. His sharp insights and sparks of wisdom over economics and the real society out there are as unmatched and educational as his lectures. I would be extremely honored if I could grow into such a scholar who both thinks and cares deeply.

I also learned so much from Professor Dal Bó. Attending his lectures on historical governments is an unforgettable experience, akin to attending best dramas with the invaluable bonus to interact with the narrator. Professor Dal Bó is always eager to teach us his formidable power to synthesize from a vast number of qualitative sources and convert them into a coherent narrative and a deeply thoughtful formal model. The third chapter of the dissertation is my first effort to practice what I learn from him. Professor Dal Bó also carefully read the early drafts of the first chapter and helped me drastically improve its presentation.

I have also benefited from the extensive and unreserved support from Professor Fred Finan and Professor Yuriy Gorodnichenko. Professor Fred Finan has always been very supportive of students who aspire to be theorists. His lectures on empirical political economy show thoughtful approaches to bridging theory and empirics, which I find immensely helpful in writing my own dissertation. Professor Yuriy Gorodnichenko has been guiding me to navigate through the labyrinth of the literature on the planned economy that helps me write a work in progress that complements the second chapter, along with the literature on time-series econometrics that allows me to write the empirical section of the second chapter.

Professor Peter Lorentzen has also played a pivotal role in my career development. He shows me how the application of game theory provides such a refreshing perspective to understand political authoritarianism. The modeling strategies of both first and second chapters are influenced by works of Professor Lorentzen through my research assistantship for him. Professor Lorentzen is unusually considerate to his students, and I am so privileged to be one of them. Professor Yang Xie has been a best friend and a very best mentor one can possibly ask for. He lends crucial support that helps me go through the job search process and dissertation writing. His intellectual prowess and deeply caring personality show what a young scholar should look like. In a word, Yang sets up a brilliant model for me to follow.

I also benefited a lot from the very best faculty group in economic history. I thank especially Professor Marty Olney for the biweekly meetings she kindly organizes for students in the economic history group. I also thank Professor J. Bradford DeLong for his extremely insightful comments on the second chapter that help me improve it a lot.

In addition, Professor Barry Eichengreen and Professor Jan de Vries taught me economic history and Professor Kevin O'Brien taught me Chinese politics; these two courses have also been important in my career development.

The opportunity to study at UC Berkeley's Ph.D. Program was generously supported by the recommendations of Professor Yingyi Qian, Professor Emmanuel Saez, Professor Chris Shannon, and Professor Jing Cao. They all have a lasting intellectual influence on me throughout my graduate school years. For example, Professor Yingyi Qian's foundational works on comparative economics provide numerous inspirations for me, and my enthusiasm for Professor Emmanuel Saez's work has been a key reason behind my persistent interest in studying the role of government in the economy.

I also want to express my deep gratitude to my undergraduate advisors Professor Xue Sherry Qiao and Professor Jun Wang. What's more, I learn so much from Professor Ping He and Professor Hongbin Li, who act as faculty advisors to the undergraduate research program at School of Economics and Management, Tsinghua University. Professor Chong'en Bai and Professor Yingyi Qian are also extremely caring for students who aspire to become research economists. All of them are always ready to guide us and share their invaluable insights on research. One cannot possibly ask for more intellectual, responsible, and caring mentors.

Shaoda Wang has been a great friend and colleague throughout my graduate school years, and I appreciate his support and thoughtful comments on my work. I also thank Dagan Lu and Minsong Zhou for excellent research assistance. The dissertation has also benefited particularly from the comments of Efthymios Athanasiou, Chong'en Bai, Gillian Brunet, Arthur Campbell, Douglas Campbell, Jiahua Che, Jidong Chen, Kang Chen, Zeger Van Der Wal, Xinyu Fan, Gunes Gokmen, Junyi Hou, Sergei Izmalkov, Katerina Jensen, Dimitri Koustas, Ruixue Jia, Erik Johnson, Claudio Labanca, Matthew Leister, Lixing Li, Wei Li, Xi Lu, Yi Lu, Pushkar Maitra, Andrei Markevich, Marta Troya Martinez, Vai-Lam Mui, Martha Olney, Laura Puzzello, Danny Quah, Ramkishan Rajan, Michelle Rendall, David Schonholzer, Susan Shirk, Zheng (Michael) Song, Ferenc Szucs, Michele Valsecchi, Brian Viard, Yijiang Wang, Yuqian Wang, Zenan Wang, Michael Ward, Binzhen Wu, Tianyang Xi, Guo Xu, Chenggang Xu, Yiqing Xu, Yang Yao, Leslie Young, Noam Yuchtman, and Hosny Zoabi. I am also grateful for comments from the participants of 2016 Tsinghua/UC San Diego Conference on Analysis of China's Institutions, UC San Diego Young Scholar Conference, Western Political Science Association Mini-Conference on Non-democratic Politics, and seminars in UC Berkeley, Cheung Kung Graduate School of Business, Monash University, the National University of Singapore, New Economic School, Peking University, and Tsinghua University. I thank Shuo Chen's effort in making the personnel data on city-level

officials in China publicly available. The dissertation is supported by the Helen Gan and Richard Aston Fellowship for China Studies and the John L. Simpson Memorial Research Fellowship for Comparative Studies. All errors are my own.

Finally, no words could express my gratitude to my parents, Zhaoyun Chen and Guoqiang Li, whom I owe so much for their unreserved understanding and persistent encouragement that endlessly lends confidence and support for me. My aunt Zhaojun Chen has always been very proud of me, and her strong sense of social responsibility motivates me all the time. My beloved grandfather Keyin Chen had always believed me in everything. I therefore dedicate this dissertation to my family.

Chapter 1

Rotation, Meritocracy, and Property Rights

1.1 Introduction

The holdup problem is destructive to economic growth, especially in environments without well-defined property rights (Williamson, 1985; Grossman and Hart, 1986). When constitutional constraints on politicians are weak, entrepreneurs anticipate excessive *ex post* extraction from politicians (Persson and Tabellini, 2000; Acemoglu, 2003). Thus, authoritarian regimes have to rely on alternative constraints on politicians to solve the holdup problem against entrepreneurs. Many papers show that politicians can be disciplined by performance-based rewards that compensate (local) politicians for economic growth, especially meritocratic promotion for growth (Maskin et al. 2000; Lazear and Oyer, 2012). However, models of performance-based rewards usually shy away from the holdup problem (Laffont, 2001), the very dilemma behind the urgency of strong property rights (Grossman and Hart, 1986).

Alternatively, politicians can enter a reciprocal relationship with the entrepreneurs so that reputation concerns can constrain the politicians. Among the many problems of reputation-based solutions (Hart, 1995), politicians inevitably develop entrenched interests in their cronies so that the politicians block the entrance of new firms and the associated “creative destruction”, endangering sustainable growth (Aghion and Howitt, 1995; Acemoglu et al. 2006). In other words, reputation-based solutions cannot extend credible protection to new firms.

I propose that political rotation is at the heart of credible constraints on politicians in authoritarian regimes, especially local politicians. Namely, political rotation provides *de facto* property rights for entrepreneurs that guard against politicians’ extraction. The starting point of this proposition is that a politician with prolonged tenure can become very knowledgeable about local conditions, which endows her with formidable power. To formalize the

idea, I allow a politician to acquire detailed information about an entrepreneur's project, especially the profitability of the project. Equipped with such information, the politician wields enormous *ex post* bargaining power against the entrepreneur when they bargain on the division of the surplus. Notice that power and information are interchangeable in my model. Political rotation plays a decisive role in alleviating the acute holdup problem. If a politician is rotated to a new jurisdiction, the information about the last jurisdiction becomes completely useless. This discourages the politician from power/information acquisition, establishing credible property rights to the entrepreneur.

The bulk of the paper investigates the interaction between rotation and performance-based rewards. I show that stronger performance-based rewards exacerbate the holdup problem, requiring more intense political rotation. Note that information acquisition helps the politician to avoid bargaining breakdown with the entrepreneur. This is because a politician fully informed about the project can extract rent based on the realized profitability of the project so that the entrepreneur always accepts the extraction. Consequently, an informed/empowered politician can ensure that the project will always be finished, thus reaping the full benefit of performance-based rewards. By contrast, a politician uninformed about the project cannot condition her rent extraction on the realized profitability. An entrepreneur whose project yields fewer profits than the proposed extraction will reject cooperation with the politician, engendering substantial likelihood of bargaining breakdown. Thus, an uninformed/disempowered politician can capture only a partial benefit from performance-based rewards. To summarize, stronger rewards based on economic performance raise the return for a politician to empower herself, a temptation that has to be discouraged by more intense rotation. *When performance-based rewards are more intense, political rotation has to be more frequent.*

On the other hand, excessive political rotation itself can be disastrous. This is the traditional view that political rotation creates "roving bandits" (Olson, 1993). In my model, excessive rotation induces the politician to confiscate the capital invested by the entrepreneur into the project. The investment is necessary for the entrepreneur's project to be productive, so capital confiscation denies future production. However, a politician with short tenure loses little by pursuing such shortsighted policy. The roving-bandit concern puts an upper bound on rotation frequency, while the prior analysis on information acquisition identifies the lower bound. Importantly, the upper bound also increases with performance-based rewards, which raises the stake of future production opportunity and the cost of capital confiscation. Stronger performance-based rewards make frequent rotation possible by counterbalancing the temptation to confiscate capital. *Thus, rotation and performance rewards are supporting each other. Rotation discourages information acquisition tempted by performance-based rewards. Performance rewards disincentivize shortsighted policy, which allows frequent rotation.*

Political rotation also eliminates entrenched interests and facilitates "creative destruction". Thus, the solution to the holdup problem by rotation and performance-based rewards over-

comes the standard trade-off between commitment and flexibility (Levin, 2003; Chassang, 2010). Suppose that a new project better than the old one may arrive. The politician has no time to learn about the new project; she can support either the old project or the new project, but not both. Without rotation, the informed politician can take all rents away from the old project, but as she is uninformed about the new project, she has to award significant information rents to the new entrepreneur. Thus, even if the new project is better than the old project, the politician will probably stick with the old one. Performance-based rewards induce politicians to focus on productivity rather than economic rents, so they can partially reduce entrenched interests. With sufficient rotation, the politician is equally uninformed about the old and new projects. This situation completely eliminates entrenched interests, so the politician will endorse the new project that is on average “better” than the old one. Rotation complemented by performance-based rewards provides both flexibility and *de facto* property rights that render it possible to sustain long-term growth.

I confront predictions of the theory with Chinese firm-level data matched with politician characteristics. Specifically, the theory should apply especially well to investment in equipment and buildings that are difficult to move. Better political incentives should not boost firms’ holdings in liquid assets such as cash and intellectual property. A firm with large liquid assets can simply move to another city when predation is looming. The data shows that a firm indeed holds more physical capital when it anticipates rotation and promotion of the mayor of the city where the firm locates. The effect is especially strong for young mayors new in office, who face much stronger personnel incentives. For different types of ownership, private firms are especially responsive to anticipated rotation and promotion while state-owned enterprises (SOEs) are much less affected. This is consistent with the theory because SOEs already enjoy substantial bargaining power against local politicians. Taken together, the empirical evidence lends high credibility to the theory.

1.2 Literature

Starting from the influential work of Olson (1993), the political economy literature usually treats the “roving-bandits” created by political rotation as a formidable hurdle to reliable property rights (Rose-Ackerman and Palifka, 2016). My paper shows that the traditional view is appropriate when it abstracts political rotation as a stand-alone institution, but less so when rotation is one of the building blocks in an institutional cluster. My analysis demonstrates that political rotation and performance-based rewards interact in an intriguingly symbiotic way. The two institutions can be destructive as stand-alone arrangements, but their interaction helps each other restore desired disciplinary effects.

There is a relatively separate literature in personnel economics on rotation. Most of these studies take a human capital approach. They show that lateral moves of employees help 1) the employer learn the ability of the subordinates (Ortega, 2001; Eriksson and Ortega, 2006) and 2) the employees accumulate a diverse set of skills that improves their eligibility for

promotion (Eriksson and Ortega, 2006; Friebel and Raith, 2014; Jin and Waldeman, 2017). My paper takes a distinct political approach to understanding rotation. The perspective of political economy proves to be very fruitful in generating fresh insights on the role of rotation. In addition, Friebel and Raith (2014) and Jin and Waldeman (2017) focus on the interaction between rotation and narrowly-defined promotion, while performance-based rewards in my paper are easily applicable to many personnel phenomena.

Performance-based rewards are a key topic in political economy and economics of organizations (Roland, 2000; Lazear and Oyer, 2012). Researchers advocate and formalize many ideas why performance-based rewards may misfire and create unintended consequences (for example, see Gibbons, 1987 and Holmstrom and Milgrom, 1991). To the best of my knowledge, my paper is the first attempt to understand how performance rewards may exacerbate the temptation to accumulate information-based power. By focusing on the holdup problem, my theory explains why high-powered incentive, not carefully designed, can be especially destructive in autocracy. The insight is further applicable to generic organizations, where a paramount concern of headquarters is the potential loss of control over divisions (Qian, 1994). Apart from highlighting this dilemma, my paper also systematically explores the institutional solution that can limit the negative side of performance-based rewards.

My paper is closely related to the large literature on political connection and economic outcomes. Many influential papers demonstrate various ways that political connection can increase business value and economic efficiency (Khwaja and Mian, 2008; Ferguson and Voth, 2008; Bai et al. 2014). However, it is puzzling that private business can reap such huge benefit from political connection, as by definition politicians command formidable coercive power that should enable them to capture the bulk of the surplus (Weingast, North 1989). Moreover, political connection seems to matter most where constitutional constraints on politicians are weak. This makes it especially thought-provoking to investigate the alternative constraints on authoritarian politicians that justify the value of political connections for entrepreneurs.

The literature on state capacity (Besley and Persson, 2011) focuses on how fiscal capacity and legal capacity engender economic growth together. Less attention has been paid to how bureaucratic capacity is decisive for economic outcomes, which is the focus of the state capacity literature in its original contributions (such as Mann, 1986 and Evans and Rauch, 1999). My research offers a concrete micro-foundation on the importance of bureaucratic capacity, defined as impersonal rules and controls over bureaucrats and politicians (Evans and Rauch, 1999).

A huge literature shows that Schumpeterian “creative destruction” engenders sustainable growth (Aghion and Howitt, 1992; Aghion et al. 2013). It is well known that the key threat to “creative destruction” is the entrenched interests of powerful politicians who benefit from their cronies’ firms (Olson, 1983; Aghion and Howitt, 1996; Acemoglu et al. 2006). This dilemma constitutes the foundational force behind the “middle-income trap” of many

economies that rely on crony capitalism to initiate industrialization (Acemoglu et al. 2006; Acemoglu and Robinson, 2012). My analysis shows that the problem is solvable by institutionalized personnel control over politicians using strong rotation and performance-based rewards. This argument is consistent with the empirical literature showing that “creative destruction” can be active in emerging economies with little legal capacity but high state capacity (Brandt et al., 2012).

1.3 A Workhorse Model

Performance-based Rewards in the Benchmark Model

In a reduced-form manner, this section illustrates the basic insights of models on performance-based rewards. It serves as the key building block to formalize my own ideas in later sections. There are three players: a principal, a local politician, and an entrepreneur. The principal is the central authority who implements personnel policy. The game has two periods. The timeline and strategies of players are as follows:

At $t = 1$, the entrepreneur arrives and decides whether to invest in a project that costs him $k > 0$. The project’s output y is uncertain. It has a distribution $F(y)$ continuously distributed on $[\underline{y}, \bar{y}]$ with the cumulative distribution function $F(\cdot)$, $F' \equiv f \cdot \frac{f(y)}{1-F(y)}$ increases in y ; in other words, the distribution satisfies the monotone hazard rate property. If the entrepreneur does not invest, the game ends and all players get 0. If he pays k , the project begins but remains unfinished.

After the entrepreneur pays k , the profitability of the project y is revealed to the entrepreneur. The principal and the local politician do not know y , but they both know $F(y)$.

At the end of $t = 1$, the principal rotates the politician with probability π and appoints another one, who will take office in $t = 2$. If the politician in $t = 1$ is rotated, she gets an exogenous payoff of \tilde{U} by serving in another jurisdiction.

At $t = 2$, the politician (either the same one as in $t = 1$ or a newly appointed politician) proposes to extract w from the entrepreneur. If the entrepreneur accepts, the politician lobbies the central government, and the value of the project y is fully realized. The entrepreneur gets $y - w - k$, as he cares about the net profit. The principal wants a large total output, so she gets y . For the politician, she gets $w + Ry$. The politician’s payoff has two parts: w is the economic rent, and Ry represents the performance-based rewards, where the exogenous parameter R measures the intensity of performance rewards. A large R means that performance rewards are highly valued by the politician.

R is the key parameter in the model. A preferred interpretation is that R measures how much the politician values promotion opportunities. If R is large, higher political positions

are prized by the local politician. In this case, the politician strongly dislikes bargaining breakdown, which will demolish the valuable opportunity of promotion. If R is small, higher political positions are not very lucrative. The politician cares less about her career path than the current opportunity of rent extraction. It is straightforward to provide a micro-foundation that validates the interpretation of R as the value of promotion opportunity, as shown in Appendix 3. I also demonstrate that R can be interpreted as the likelihood of promotion for a given output level. Thus, a polity enjoys high R if the personnel turnover is very active, which can occur with mandatory retirement rule that releases many political positions regularly (Svolik, 2012). In this case, the prospect of promotion is realistic and reasonable likely, so promotion opportunity constitutes strong incentive imposed the politician. If the polity is a gerontocracy with little hope of promotion for young local politicians, the politician's career concern cannot form effective constraints on the politician. The micro-foundation in Appendix 3 focuses on the interpretation of Ry as promotion opportunity, as promotion is far more important in motivating politicians than direct monetary compensation (Maskin et al. 2000; Svolik, 2012). In general, it is still very useful to organize the basic story of performance-based rewards in a reduced-form manner. The insights can be applied to many forms of performance-based rewards, and all theoretical results will be robust regardless of the specific way performance rewards are implemented. It is worth noting that the linearity of Ry is purely for expositional purposes. Appendix 3 shows that all key results are unchanged under very general functional forms¹.

If the entrepreneur does not accept the extraction, the politician will not help the entrepreneur and the entrepreneur has to abandon the project. Although the project's potential productivity is y , the realized output is 0. The entrepreneur gets $-k$. The principal gets 0. The politician gets 0. The solution concept is sequential equilibrium.

The necessity for the entrepreneur and the politician to strike a bargain is a key assumption in the model. What is the reason that the entrepreneur cannot finish his project by himself? The rationale is that markets in many economies are highly regulated and fragmented. On the one hand, the central government installs many entry barriers to protect firms that directly benefit the central government (e.g., Gordon and Li, 2009). On the other hand, politicians have a strong interest in enforcing trade barriers so that non-local firms can be excluded from the local market (e.g., Young, 2000). The consequence is that an entrepreneur can invest in capacity to produce plenty of goods. However, it is very difficult to sell them and realize the profit. Local politicians are playing a decisive role here. They have the necessary political capital to lobby the central government for a permit so that the local

¹In Appendix 3, the politician gets the performance rewards with probability $V(y, \alpha)$, and she values the performance rewards at R . Thus, the expected value of performance rewards is $V(y, \alpha)R$. I assume that $\frac{\partial V}{\partial y} > 0$, $\frac{\partial V}{\partial \alpha} > 0$: the politician is more likely to reap the reward if the output is higher, and α is a parameter of the function V . We can see that R and α correspond to the two interpretations: R measures the value of promotion opportunities, and α affects the likelihood of promotion for any output level. Importantly, the only assumptions I need to impose on $V(y, \alpha)$ is that $\frac{\partial V}{\partial y} > 0$, $\frac{\partial V}{\partial \alpha} > 0$, and $V(0, \alpha) = 0$, and all but one key results in this paper are preserved.

firm can enter the national market (Wedeman, 2011; Bai et al. 2014). They can also try to break down trade barriers by lobbying the central government or negotiating with other politicians. The full value of the entrepreneur's investment can be realized only through the mobilization of a politician's capacity and connection.

Regulations can be even more strict in markets of production factors. For example, regulations over labor market are especially punishing in most countries. In many cases, it is more effective for the entrepreneur to mediate labor disputes and to address labor unions through politicians. Otherwise, the entrepreneur's firm can be paralyzed and produces nothing. This further validates the notion that politician's inputs are indispensable for a successful firm, especially in developing economies.

Let us analyze the simple game. At $t = 2$, the entrepreneur accepts any extraction $w \leq y$. The politician, either a newly appointed one or the same one as in $t = 1$, proposes to extract:

$$w^* = \operatorname{argmax}_w [1 - F(w)] \{w + E[Ry|w \leq y]\}. \quad (1.1)$$

$[1 - F(w)] \{w + E[Ry|w \leq y]\}$ is the expected payoff to a politician who proposes to extract w from the entrepreneur. The politician faces some risk of bargaining breakdown: an entrepreneur with $y < w$ will reject the rent extraction rather than collaborating with the politician. In this case, the politician gets 0 from rent extraction and also gets 0 from performance rewards as the project will not be finished. The probability that there will be an agreement is $1 - F(w)$. The politician gets w from the rent extraction, and $E[Ry|y \geq w]$ from performance rewards. Thus, the utility of an uninformed politician who proposes to extract w is $[1 - F(w)] \{w + E[Ry|w \leq y]\} + F(w) * 0 = [1 - F(w)] \{w + E[Ry|w \leq y]\}$. In this case, the expected surplus captured by the entrepreneur is $[1 - F(w)] \{E[y|y \geq w] - w\}$. This is the probability that the project will be finished (which is $1 - F(w)$) times the net surplus to the entrepreneur (which is $E[y|y \geq w] - w$).

The first order condition for (1) with respect to w characterizes w^* , the optimal rent extracted by an uninformed politician:

$$\operatorname{max}_w [1 - F(w)] \{w + E[Ry|w \leq y]\} = \operatorname{max}_w \left\{ w \int_w^{\bar{y}} f(z) dz + R \int_w^{\bar{y}} z f(z) dz \right\}$$

$$w^* \int_{w^*}^{\bar{y}} f(z) dz - w^* f(w^*) - R w^* f(w^*) = 0$$

$$\frac{w^* f(w^*)}{1 - F(w^*)} = \frac{1}{1 + R}. \quad (1.2)$$

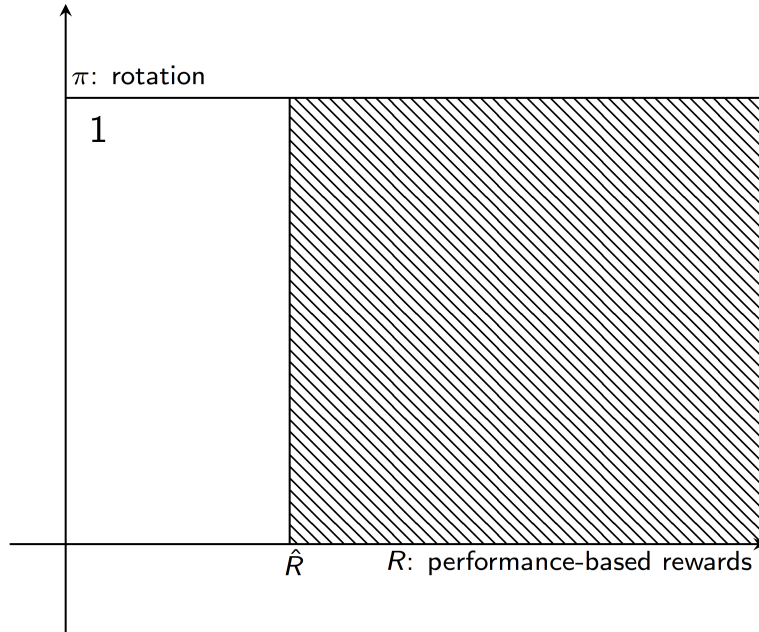


Figure 1.1: The Benchmark Model

The monotone hazard rate property guarantees that w^* is unique. It also ensures that w^* decreases with R , so the uninformed politician reduces rent extraction with stronger performance rewards. Basically, the politician faces a trade-off between more rents and a higher likelihood of bargaining breakdown. Performance-based rewards raise the cost of bargaining breakdown, which induces the politician to extract fewer rents from the entrepreneur.

In $t = 1$, the entrepreneur will invest if the expected surplus $[1 - F(w^*(R))]\{E[y|y \geq w^*(R)] - w^*(R)\}$ outweighs the investment cost k . I impose the following assumption:

Assumption 1:

$$[1 - F(w^*(0))]\{E[y|y \geq w^*(0)] - w^*(0)\} < k \text{ and } k < E[y].$$

The assumption expresses that a politician who wants only to maximize economic rents (which is the case when $R = 0$) extracts too much from the entrepreneur. Thus, the entrepreneur refuses to initiate investment. Performance-based rewards can solve the problem, as a politician who obtains large rewards for economic growth dislikes bargaining breakdowns. She extracts less from the entrepreneur, which increases the probability of a successful project and enables the entrepreneur to cover his *ex ante* cost of investment. We have the following result.

Proposition 1: *In the benchmark model, the entrepreneur will invest under sufficiently strong performance-based rewards.*

Algebraically, $\exists \hat{R} \in (0, \infty)$ such that $1 - F(w^(\hat{R}))\{E[y|y \geq w^*(\hat{R})] - w^*(\hat{R})\} = k$. The entrepreneur expects a non-negative profit if $R \geq \hat{R}$.*

Proofs of all propositions are in Appendix 1. Note that political rotation is irrelevant. Whether the politician is the same one as in $t = 1$ or newly appointed, she behaves exactly the same in $t = 2$ and extracts $w^*(R)$. In Figure 1, the horizontal axis is the intensity of performance-based rewards, while the vertical axis is the frequency of rotation. The shaded area is the parametric range in which the entrepreneur will invest. The entrepreneur will invest as long as the politician is sufficiently rewarded for economic growth.

Rotation and Performance-based Rewards when Politician can Accumulate Power

In this section, the key difference is that the politician can accumulate power. Specifically, after the entrepreneur has invested, the politician can pay a cost c such that the politician in $t = 2$ is fully informed about the project's value y . Everything else is the same as in the benchmark model. This formalizes the idea that a politician with a long tenure can become very knowledgeable about her jurisdiction, and consequently enormously powerful. As in classical models of information economics (Bolton and Dewatripont, 2005), power and information are interchangeable. A politician who knows everything about the entrepreneur's project wields formidable bargaining power against the entrepreneur, who will not be able to reap benefits from his own project. Rotation plays a key role here in discouraging the politician from information acquisition.

What are the practical ways to improve a politician's knowledge about local enterprises? She can rely on her office of assistants to implement inspection and networking for more sensitive information about private enterprises. She can boost the capacity of government bureaus that register and review private enterprises (Besley and Persson, 2011). In the more extreme form, she can build her coercive power so that the local firms have to surrender such information. Given that unelected local politicians can be "petty dictators" in authoritarian regimes (Lieberthal, 2005), there are so many ways for her to become immensely knowledgeable about local economic conditions.

The setup implicitly assumes that the newly appointed politician automatically inherits the information acquired by the politician at $t = 2$. The rationale is that the politician initiates the learning process through mobilizing the bureaucracy. It is her bureaucrats and assistants who actually implement the investigation. When the politician is rotated, the bureaucrats stay and work for the newly appointed politician, who can readily harvest the knowledge

embedded in the bureaucracy. This assumption helps highlight the key mechanism of the model. In Appendix 2, I show that my qualitative results do not rely on this assumption.

We need to derive restrictions on π to ensure that the politician in $t = 1$ does not pay c . It will guarantee some information rents to the entrepreneur, which is a necessary condition for the entrepreneur to gain some surplus from his project. Under the jurisdiction of an uninformed politician, the entrepreneur earns $[1 - F(w^*(R))]\{E[y|y \geq w^*(R) - w^*(R)]\}$. The entrepreneur will invest if the politician does not pay c and $[1 - F(w^*(R))]\{E[y|y \geq w^*(R) - w^*(R)]\} > k$ (the expected benefit from the project outweighs the cost).

At $t = 2$, the entrepreneur accepts any $w \leq y$. An informed politician extracts $w = y$, and the entrepreneur gets 0 surplus. An uninformed politician extracts w^* that satisfies Equation (1.1), where $\frac{w^*f(w^*)}{1-F(w^*)} = \frac{1}{1+R}$. The extraction proposed by an uninformed politician is the same as the case in the benchmark model.

The politician in $t = 1$ does not pay c if:

$$-c + (1 - \pi)(1 + R)E[y] + \pi\tilde{U} \leq (1 - \pi)[1 - F(w^*)]\{w^* + E[Ry | w^* \leq y]\} + \pi\tilde{U}.$$

The left-hand side is the expected payoff to a politician who plans to pay c . If so, with probability $1 - \pi$ the politician continues her tenure. She sets $w^* = y$ and extracts all surplus y away, without worrying about bargaining breakdown. As the politician knows the productivity of the firm, she can always calibrate rent extraction based on the realized productivity, something that an uninformed politician cannot achieve. Hence, an agreement between the entrepreneur and the fully informed politician is guaranteed, and the politician also reaps the full benefit of performance-based rewards Ry . The expected payoff with extended tenure is $(1 + R)E[y]$. With probability π , the informed politician is rotated, and she gets her exogenous payoff \tilde{U} .

The right-hand side is the expected payoff to a politician who plans to not pay c . With probability $1 - \pi$, she gets the same expected payoff as an uninformed politician, which is $[1 - F(w^*)]\{w^* + E[Ry | w^* \leq y]\}$, and with probability π she gets \tilde{U} . The inequality identifies the lower bound for rotation frequency π :

$$\pi \geq 1 - \frac{c}{(1 + R)E[y] - [1 - F(w^*)]\{w^* + E[Ry | w^* \leq y]\}} \equiv \underline{\pi}. \quad (1.3)$$

Thus, rotation has to be sufficiently frequent to forestall information acquisition. To guarantee that minimum frequency $\underline{\pi} \in (0, 1)$, I have the following assumption.

Assumption 2:

$$(1 + R)E[y] - [1 - F(w^*)]\{w^* + E[Ry | w^* \leq y]\} > c.$$

Recall that w^* is a function of R : $\frac{w^* f(w^*)}{1 - F(w^*)} = \frac{1}{1 + R}$. Assumption 2 is a restriction on exogenous parameters R and c . Suppose that Assumption 2 holds. Without rotation, the politician would pay c if the entrepreneur invests k , as the net benefit $(1 + R)E[y] - [1 - F(w^*)]\{w^* + E[Ry | w^* \leq y]\}$ outweighs the cost c . The entrepreneur refuses to initiate the investment since he predicts that he will get $-k$ by doing so. Thus, $\underline{\pi} > 0$ is guaranteed by Assumption 2. Without rotation, the politician in $t = 1$ will pay c and extract all rents.

In addition, it is clear that $(1 + R)E[y] - [1 - F(w^*)]\{w^* + E[Ry | w^* \leq y]\} > 0$:

$$\begin{aligned} & (1 + R)E[y] - [1 - F(w^*)]\{w^* + E[Ry | w^* \leq y]\} \\ &= (1 + R) \int_{\underline{y}}^{\bar{y}} z f(z) dz - w^* \int_{w^*}^{\bar{y}} f(z) dz - R \int_{w^*}^{\bar{y}} z f(z) dz \\ &= \int_{w^*}^{\bar{y}} (z - w^*) f(z) dz + (1 + R) \int_{\underline{y}}^{w^*} z f(z) dz > 0. \end{aligned}$$

Thus, $\underline{\pi} = 1 - \frac{c}{\int_{w^*}^{\bar{y}} (z - w^*) f(z) dz + (1 + R) \int_{\underline{y}}^{w^*} z f(z) dz} < 1$. We have $0 < \underline{\pi} < 1$.

Suppose that the principal chooses π that satisfies (3). As argued before, the entrepreneur pays k if the expected benefit from the project outweighs k :

$$k < [1 - F(w^*)]\{E(y | y \geq w^*) - w^*\}.$$

With Assumption 1 and Assumption 2, I can prove the main results of the paper.

Proposition 2: 1. *Sufficiently strong rotation and performance-based rewards incentivize the entrepreneur to invest.*

In algebra, denote $\hat{R}(k)$ and $\underline{\pi}(R, c)$ such that $[1 - F(w^*(\hat{R}))]\{E(y | y \geq w^*(\hat{R})) - w^*(\hat{R})\} = k$ and $\underline{\pi}(R, c) = 1 - \frac{c}{(1 + R)E[y] - [1 - F(w^*(R))]\{w^* + E[Ry | w^*(R) \leq y]\}}$. If $R > \hat{R}(k)$ and $\pi \geq \underline{\pi}(R, c)$, the politician does not pay c , and the entrepreneur will reap a non-negative return.

2. $\frac{\partial \underline{\pi}(R, c)}{\partial R} > 0$: *the minimum rotation frequency increases when performance-based rewards are stronger. If the minimum rotation frequency does not change, more intense performance-based rewards incentivize the politician in $t = 1$ to pay c , and the politician in $t = 2$ will fully predate the entrepreneur.*

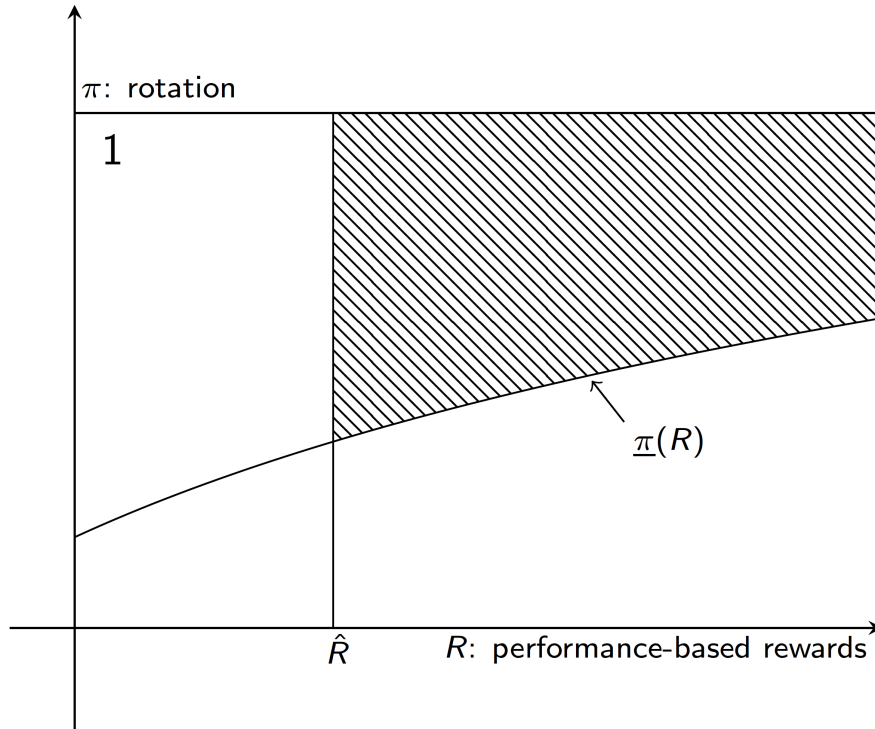


Figure 1.2: Rotation Helps Performance Rewards

Corollary 1: *Suppose $\pi = 0$. $\forall R \geq 0$, the entrepreneur does not invest.*

Propositions 1 and 2 and Corollary 1 can be summarized by Figure 1 (on page 7) and Figure 2. The shaded area is the parametric range such that the entrepreneur will invest. Figure 1 depicts Proposition 1, while Figure 2 illustrates Proposition 2 and Corollary 1. Figure 1 shows the effect of performance rewards when there is no severe holdup problem. The politician cannot invest c in Figure 1, so performance-based rewards alone can induce investment from the entrepreneur. Moreover, the probability of project completion $1 - F(w^*(R))$ also increases with R , as $w^*(R)$ decreases with R . Hence, economic performance will improve with a higher level of performance-based rewards. By contrast, Figure 2 shows the situation when the politician can acquire local information. In this case, we need a second constraint $\pi \geq \underline{\pi}(R, c)$ to ensure that the politician gives up the opportunity of information acquisition. Otherwise, an empowered politician would have extracted all surplus so that the entrepreneur never invests in the first place.

The key result of Proposition 2 is that $\frac{\partial \pi(R, c)}{\partial R} > 0$. Notice that:

$$\underline{\pi}(R, c) = 1 - \frac{c}{(1+R)E[y] - [1 - F(w^*(R))]\{w^* + E[Ry | w^*(R) \leq y]\}} \equiv 1 - \frac{c}{\Delta(R)}.$$

where $\Delta(R) \equiv (1+R)E[y] - [1 - F(w)]\{w + E[Ry | y \geq w]\}$ is the difference between the expected payoffs to an informed politician and an uninformed one. In other words, $\Delta(R)$ measures the benefit of local knowledge, and the temptation for the politician to be informed. Hence, $\underline{\pi}(R, c)$ increases with R because $\Delta(R)$ increases with R :

$$\frac{d\Delta(R)}{dR} = \int_{\underline{y}}^{\bar{y}} z f(z) dz - [1 - F(w)] \frac{\int_{w^*}^{\bar{y}} z f(z) dz}{[1 - F(w)]} = \int_{\underline{y}}^{w^*} z f(z) dz > 0.$$

Performance-based rewards exacerbate the temptation to learn. This finding is a key insight of the paper that directly contributes to the complementarity between rotation and performance-based rewards. If unrotated, the informed politician can reap all the benefit from performance-based rewards, as she can calibrate the rent extraction based on the realized productivity. Doing so ensures that there will be no bargaining breakdown, so the project will always be finished, and the politician will always get her performance-based rewards.

By contrast, if the politician does not pay the cost of learning, she faces a substantial risk of bargaining breakdown. Apart from the lost economic rent, the rewards for economic growth also fail to materialize when the politician and the entrepreneur cannot settle upon an agreement. An uninformed politician can only reap a partial benefit from performance-based rewards, while an informed politician captures the full benefit. When rewards for economic performance increase, it becomes even more tempting to invest in information acquisition because such information is more valuable under stronger performance rewards.

Such temptation to acquire local knowledge can be discouraged by rotation, as shown by the curve that represents $\underline{\pi}(R)$. With stronger performance rewards, the minimum rotation frequency has to increase to counterbalance the temptation to learn. Assumptions 1 and 2 guarantee that for any intensity of performance rewards $R \in [0, \infty)$, there is a minimum rotation frequency $\underline{\pi}(R, c) \in (0, 1)$. With the rotation prospect, the benefit from the detailed information realizes with such a small probability that it cannot justify the investment in local information.

$\pi \geq \underline{\pi}(R, c)$ such that the politician does not learn is a necessary condition for the entrepreneur to invest. As in Proposition 1, performance-based rewards also cannot be too low. In the extreme case where $R = 0$, the uninformed politician's problem reduces to $\max_w [1 - F(w)]w$; thus, the politician prefers to maximize expected rent. As $w^*(R)$ is

a decreasing function of R , $w^*(R = 0)$ is large. Without performance-based rewards, the uninformed politician does not care too much about bargaining breakdown because she is not rewarded for economic achievements. Thus, she is willing to risk the high probability of bargaining breakdown in exchange for a higher rent. This results in small *ex ante* surplus for the entrepreneur because of the high rent extraction attempted by the politician and the associated high risk of bargaining breakdown. By Assumption 1, $[1 - F(w^*(0))]\{E(y|y \geq w^*(0)) - w^*(0)\} < k$, so the entrepreneur does not invest when an uninformed politician is not politically rewarded for growth. A moderate degree of performance-based rewards \hat{R} such that $[1 - F(w^*(\hat{R}))]\{E(y|y \geq w^*(\hat{R})) - w^*(\hat{R})\} = k$ coupled with $\pi \geq \pi(\hat{R}, c)$ finally allows the entrepreneur to break even.

To summarize, if $\pi < \underline{\pi}(R, c)$, strong performance-based rewards cannot induce the entrepreneur to invest because the entrepreneur expects full extraction of his surplus in the future. A higher level of performance-based rewards, rather than alleviating the predicament, further justifies information acquisition and exacerbates the holdup problem. With frequent rotation such that $\pi \geq \underline{\pi}(R, c)$, the desired disciplinary effects of performance-based rewards are completely restored. The politician finds it unprofitable to invest in information acquisition. In this case, stronger rewards based on performance induce the uninformed politician to care a lot about striking a bargain with the entrepreneur. Frequent rotation limits the potentially destructive force of performance-based rewards and unleashes their role in aligning the incentive of the politician with the principal's objective. At the same time, if $R < \hat{R}$, frequent rotation achieves nothing, as the uninformed politician cares too little about the economy to behave benevolently. Thus, the effectiveness of rotation also relies on sufficiently strong rewards for economic growth. Performance-based rewards support rotation in another important manner in the next section.

The Interdependence of Rotation and Performance Rewards

A key problem with rotation is that it can encourage shortsighted policy. Indeed, surrendering the opportunity of information acquisition is itself a shortsighted policy, but it is a particular category that improves welfare. However, there are many other shortsighted policies that can drastically reduce welfare. Specifically, the model in Section 1.3 requires the politician to wait until period 2 when the entrepreneur can use the invested capital k to produce output. If rotation is too frequent, the politician may find it optimal to steal the invested capital in period 1. This is the famous "roving-bandit" problem articulated by Olson (1993). The policy instrument to discourage capital confiscation is performance-based rewards. Even if the politician anticipates a high likelihood of rotation, protection of private capital remains desirable under strong performance rewards. Basically, the reduced stake of future economic performance due to frequent rotation can be compensated by stronger performance rewards.

Assume $\pi \geq \underline{\pi}(R)$ so that the politician does not learn the value of the project. Suppose before that, the politician can decide whether to steal the capital k away. If the politician

does, she can resell the capital and gain a value of ηk from it, $\eta \leq 1$. All capital depreciates away in $t = 2$.

The politician will not steal the capital if:

$$\pi \tilde{U} + (1 - \pi) * 0 + \eta k < \pi \tilde{U} + (1 - \pi)[1 - F(w^*)]\{w^* + E[Ry | y \geq w^*]\}. \quad (1.4)$$

The left-hand side is the expected payoff to a politician who steals capital. With probability π , she is rotated and gets \tilde{U} . With probability $1 - \pi$, she continues her tenure. If the capital has been stolen, the entrepreneur cannot produce anything, so the politician gets no rents and no performance-based rewards. In either case, the politician always gets ηk . The right-hand side is the expected payoff to a politician who does not steal capital or acquires local information. With probability $1 - \pi$, she gets the same payoff as an uninformed politician $[1 - F(w^*)]\{w^* + E[Ry | y \geq w^*]\}$. With probability π , she gets \tilde{U} from serving in another jurisdiction. Equation (1.4) gives us:

$$\pi < 1 - \frac{\eta k}{[1 - F(w^*)]\{w^* + E[Ry | y \geq w^*]\}} \equiv \bar{\pi}(R).$$

So the “roving-bandit” concern puts an upper bound on rotation frequency. If rotation is excessively frequent, the politician will steal the capital and resell it, and the project will yield no surplus in the future. The key observation is that the upper bound $\bar{\pi}(R)$ is also an increasing function of R :

$$\frac{\partial \bar{\pi}(R)}{\partial R} = \frac{\eta k}{\{[1 - F(w^*)]\{w^* + E[Ry | y \geq w^*]\}}^2} [1 - F(w^*)] E[y | y \geq w^*] > 0.$$

Inequality (4) shows that the problem of rotation is that it reduces the desirability to protect private capital. Even if the politician steals the private capital today, she will not be affected by the destruction of growth opportunity if she gets rotated to another city. Thus, she does not care much about the consequence brought by her confiscation of private capital. Performance-based rewards discourage the temptation by restoring the stake in protecting private capital for the politician. Thus, a higher level of performance-based rewards allows more frequent rotation.

Now the question becomes: are there any parametric ranges so that

$$\pi \equiv 1 - \frac{c}{(1 + R)E[y] - [1 - F(w^*)]\{w^* + E[Ry | y \geq w^*]\}} \leq 1 - \frac{\eta k}{[1 - F(w^*)]\{w^* + E[Ry | y \geq w^*]\}} \equiv \bar{\pi}?$$

I have an additional assumption:

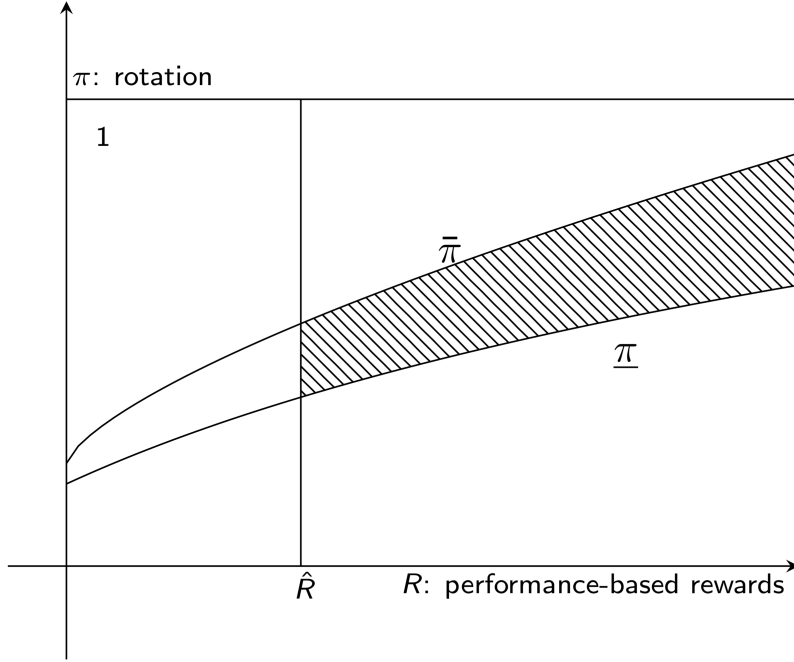


Figure 1.3: The Workhorse Model that Illustrates the Interdependence

Assumption 3:

$$\frac{c}{\eta k} > \max \left\{ \frac{E[y]}{[1 - F(w^*(0))]w^*(0)} - 1, \frac{F(w^*(0))E[y|y \leq w^*(0)]}{[(1 - F(w^*(0)))E[y|y \geq w^*(0)]]} \right\}$$

Thus, we can prove:

Proposition 3: Denote $\underline{\pi}(R) \equiv 1 - \frac{c}{(1+R)E[y] - [1 - F(w^*(R))]\{w^*(R) + E[ry|y \geq w^*(R)]\}}$,

$\bar{\pi}(R) \equiv 1 - \frac{\eta k}{[1 - F(w^*(R))]\{w^*(R) + E[ry|y \geq w^*(R)]\}}$. For $R \geq 0$, $\exists \pi \in [\underline{\pi}(R), \bar{\pi}(R)]$, such that the local politician neither acquires information nor steals capital.

Moreover, if $R' < R''$, then $\underline{\pi}(R') < \underline{\pi}(R'')$, and $\bar{\pi}(R') < \bar{\pi}(R'')$: Stronger performance-based rewards complement and are complemented by more frequent rotation.

A proper selection of π makes sure that the politician neither steals capital nor acquires information to capture all rent from the project. The key conclusion is that the complementarity between rotation and performance-based rewards persists, as the interval $[\underline{\pi}(R), \bar{\pi}(R)]$

is moving upward with a higher R . This is shown in the graph below. There is now an upper limit on the frequency of rotation, and the upper limit also increases with stronger performance-based rewards.

The two limits vividly characterize that rotation and performance-based rewards are supporting each other. Rotation discourages the temptation to dominate entrepreneurs, allowing higher-powered performance rewards. In turn, performance-based rewards make it less desirable to steal private capital, enabling more frequent rotation.

1.4 Rotation, Performance-based Rewards, and “Creative Destruction”

Rotation also creates active adaptation and “creative destruction”. As emphasized in the introduction, reputation-based solutions to the holdup problem rely on relational enforcement, which induces rigidity (Levin, 2003; Chassang 2010) and chocks “creative destruction” (Acemoglu et al. 2006). In my model, suppose that at the end of period 1 another entrepreneur arrives with a new project that is better than the old one. The question is whether the politician adapts to the new scenario and endorses the more productive project instead of the old one. Without rotation, the politician pays the cost of learning and can capture most surplus from the old project. In other words, the politician has considerable entrenched interest in the old project. If instead, the politician endorses the new project, she will have to pay large information rent. On average, the new project produces more surplus than the old one; but most likely, the politician cannot extract more surplus or obtain more performance-based rewards from the new project. As a consequence, the politician will be very conservative and endorse new project only if the realization of old project’s value is sufficiently low. With rotation, the politician has no entrenched interests with the old project. For either the new or old project, the politician needs to pay information rent to an entrepreneur. As the new project is more productive than the old one, the politician will always endorse the new project.

Setup

The timeline is similar to Section 1.3; but in $t = 2$, with probability p , a new project arrives that is “better” than the old one. There are still two periods. The old project’s profitability still follows $F(\cdot)$ with support $[y, \bar{y}]$, $F' \equiv f$. At $t = 1$, we have exactly the same timeline except for the last stage. Specifically, the profitability of the project y is revealed to the entrepreneur but not the politician. The politician can pay a cost c to learn the realization of $y \sim F(\cdot)$, which will be revealed to the politician next period.

At the end of $t = 1$, with probability $1 - p$, there is no opportunity for a new project. The politician proposes to extract w from the entrepreneur. If the entrepreneur accepts, he will

finish the project and payoffs are realized. If not, the politician gets 0 and the entrepreneur gets $-k$.

With probability p , another entrepreneur arrives with a new project such that the output $y \sim G(\cdot)$ with support $[y, \bar{y}]$, $G' \equiv g$. $G(\cdot)$ first order stochastically dominates $F(\cdot)$. The politician can choose to endorse the old project or the new one but not both. The rationale is that a local politician must devote all her political capital and connections to lobby for one firm. Dilution of the politician's resources means that firms in her jurisdiction cannot compete successfully with firms allotted exclusive support in other jurisdictions. The assumption can be relaxed, but I need at least some capacity constraint in the numbers of firms the politician can support.

Another difference with the workhorse model is that with probability $q \in (0, 1)$, the sunk cost investment $k = 0$. This is very important to generate the problem of entrenched interests: if the old entrepreneur never invests under an informed politician, the politician has no available project to form entrenched interests at all. Of course, we want to avoid q to be too small or too large. This ensures strong tensions regarding both the commitment problem and entrenched interests. There are other ways to impose the tensions, but this is the most parsimonious one.

Notice that for the new project, there is too little time to do effective learning. Then, she proposes to extract w from the endorsed project (either the new one or the old one). If the entrepreneur accepts, he will finish the project, and payoffs are realized. If not, the politician gets 0.

Players' payoffs with a completed project are very similar to Section 1.3. Specifically, the informed politician gets $w + Ry - c$ if a project (either the old or the new one) is finished with productivity y and if the politician extracts w from it. The uninformed politician gets $w + Ry$.

Adaptability with or without Learning

Let us analyze the modified game using "backward induction". At $t = 2$, suppose that a new project arrives and the politician did not pay c in $t = 1$. Then, she will get $U_1 \equiv \max_w [1 - F(w)] \{w + E_f[Ry | y \geq w]\}$ from endorsing the old project, and $U_2 \equiv \max_w [1 - G(w)] \{w + E_g[Ry | y \geq w]\}$ from the new one. Denote $w^* = \operatorname{argmax}_w [1 - F(w)] \{w + E_f[Ry | y \geq w]\}$ and $\tilde{w} = \operatorname{argmax}_w [1 - G(w)] \{w + E_g[Ry | y \geq w]\}$.

If c has been paid, she gets $(1 + R)y$ from supporting the old project in the case where the old project has been invested (the case where the old project has not been invested is trivial). We can prove the following result:

Proposition 4: *Suppose the politician paid c and is thus informed about the old project that has been invested.*

1. *She will endorse the new project with probability $F(\frac{U_2}{1+R}) = F(\frac{1}{1+R}[1 - G(\tilde{w})]\{\tilde{w} + E_g[Ry|y \geq \tilde{w}]\})$, which increases with R : performance-based rewards encourage adaptation.*

2. $\forall R < \infty$, $F(\frac{1}{1+R}[1 - G(\tilde{w})]\{\tilde{w} + E_g[Ry|y \geq \tilde{w}]\}) < F(E_g(y))$, *the first-best probability of adaptation. An informed politician holds entrenched interests in the old project, no matter how strong performance-based rewards are.*

The first part of Proposition 4 is simple to prove. The politician informed about the old project endorses the new project if:

$$(1 + R)y' < U_2 = [1 - G(\tilde{w})]\{\tilde{w} + E_g[Ry|y \geq \tilde{w}]\},$$

where y' is a realized draw from $F(\cdot)$. The left-hand side is the payoff to the informed politician if she sticks with the old project. The right-hand side is the payoff to the politician if she endorses the new project instead. As $y' \sim F(\cdot)$, the probability that the politician does so is $F(\frac{1}{1+R}[1 - G(\tilde{w})]\{\tilde{w} + E_g[Ry|y \geq \tilde{w}]\})$. Notice that even if $G(y)$ first order stochastically dominates $F(y)$, the politician supports the old project in most cases. This is because the politician can extract all surplus from the old project and ensures an agreement with the entrepreneur at the same time. If the politician supports the new project instead, she suffers from a significant difficulty of rent extraction, as well as a substantial risk of bargaining breakdown. Thus, although the new project is “on average” much better than the old project, the politician will most likely support the old project. The politician has entrenched interests in the old project.

Take the derivative of $F(\frac{1}{1+R}[1 - G(\tilde{w})]\{\tilde{w} + E_g[Ry|y \geq \tilde{w}]\})$ with respect to R : we find that the derivate is positive. Moreover, $\lim_{R \rightarrow \infty} F(\frac{1}{1+R}[1 - G(\tilde{w})]\{\tilde{w} + E_g[Ry|y \geq \tilde{w}]\}) = F(E_g(y))$, which is the first-best probability of adaptation. Hence, $\forall R < \infty$, $F(\frac{1}{1+R}[1 - G(\tilde{w})]\{\tilde{w} + E_g[Ry|y \geq \tilde{w}]\}) < F(E_g(y))$: we confirm that the informed politician *must* have entrenched interests, no matter how strong performance-based rewards are.

Performance-based rewards can partially correct entrenched interests, as $F(\frac{1}{1+R}[1 - G(\tilde{w})]\{\tilde{w} + E_g[Ry|y \geq \tilde{w}]\})$ increases with R . However, entrenched interests always persist, as $U_2 = [1 - G(\tilde{w}(R))]\{\tilde{w}(R) + E_g[Ry|y \geq \tilde{w}(R)]\}$ always entails some risk of bargaining breakdown, while endorsement of the old project completely avoids it. Bargaining breakdown under the new project vanishes only when $R \rightarrow \infty$. In this case, the payoff from rent extraction is completely dwarfed by the payoff from performance rewards, so the politician wants to ensure the

completion of the project. Thus, she extracts $\tilde{w}(R) \rightarrow \underline{y}$ as $R \rightarrow \infty$, so we have $U_2 = E_g(Ry)$. By contrast, endorsement of the old project gives a payoff of $Y' + Ry'$; y' is a realized draw from $F(\cdot)$. Again, as $R \rightarrow \infty$, the politician only cares about performance-based rewards. She simply compares $E_g(Ry)$ and Ry' , and endorses the new project if $E_g(y) \geq y'$. This utilizes all information possibly available to the politician in a manner perfectly aligned with the principal's interests, thus achieving first best. However, this asymptotic ideal can never work unless R is unrealistically large.

What about the politician who did not learn? She has no entrenched interests with the old project, and thus will support the new project that is “on average” better than the new project:

Proposition 5: *Suppose the politician didn't pay the cost of learning and a new project arrives with $G(y) \leq F(y)$, $\forall y$. The politician will always endorse the new project.*

For an uninformed politician, her payoffs from the new or the old project both entail a substantial risk of bargaining breakdown. In other words, it is equally challenging for her to extract from the old and the new projects. Given that the new project is “on average” more productive, the uninformed politician will support the new project that can offer more rents and more performance-based rewards. Indeed, the proof of Proposition 5 shows that both $[1 - G(\tilde{w})]\tilde{w} > [1 - F(w^*)]w^*$ (more rents from the new project) and $[1 - G(\tilde{w})]E[Ry|y \geq \tilde{w}] > [1 - F(w^*)]E[Ry|y \geq w^*]$ (more performance-based rewards from the new project).

Notice that the optimal degree of adaptation is to support the new project with probability $F(E_g(y))$, while an uninformed politician always supports the new project. Rotation solves the problem of entrenched interests, although at a cost. A politician who is constantly rotated has no incentive to acquire local information, so she can abandon an old project that turns out to be especially valuable and support the new project precisely because she does not know the value of the old project. This generates excessive adaptation that is undesirable from the perspective of the principal and the society. But, of course, this is a “necessary evil”; otherwise, an informed and unchecked politician will just take everything and the entrepreneur has no incentive to invest in period $t = 1$.

Rotation, Performance Rewards, and Adaptation

In $t = 1$, the politician does not pay the cost of learning if:

$$(1 - \pi)\{(1 - p)U_1 + pU_2\} + \pi\tilde{U} \geq$$

$$(1 - \pi) \left\{ (1 - p)(1 + R)E(y) + p \{ \Pi U_2 + (1 - \Pi)E[(1 + R)y | (1 + R)y \geq U_2] \} \right\} + \pi \tilde{U} - c. \quad (1.5)$$

where $U_1 \equiv \max_w [1 - F(w)] \{ w + E_f[Ry | y \geq w] \}$, $U_2 \equiv \max_w [1 - G(w)] \{ w + E_g[Ry | y \geq w] \}$, $\Pi = F\left(\frac{U_2}{1+R}\right)$.

$(1 - \pi) \{ (1 - p)U_1 + pU_2 \} + \pi \tilde{U}$ is the payoff to a politician uninformed about the old project. With probability $(1 - \pi)(1 - p)$, the politician continues her tenure in $t = 2$ and no new projects arrive. The politician gets $U_1 = \max_w [1 - F(w)] \{ w + E_f[Ry | y \geq w] \}$. With probability $(1 - \pi)p$, the continuing politician sees the arrival of a new project. Proposition 5 tells us that the uninformed politician always supports the new project, which gives her a payoff of $U_2 \equiv \max_w [1 - G(w)] \{ w + E_g[Ry | y \geq w] \}$.

$(1 - \pi) \left\{ (1 - p)(1 + R)E(y) + p \{ \Pi U_2 + (1 - \Pi)E[(1 + R)y | (1 + R)y \geq U_2] \} \right\} + \pi \tilde{U} - c$ is the payoff to a politician informed about the old project. With probability $1 - \pi$, the politician continues her term. When the new project does not arrive (with probability $1 - p$), the politician gets $(1 + R)E[y]$ from the old project. When the new project arrives, the politician supports the new project if $U_2 = \max_w [1 - G(w)] \{ w + E_g[Ry | y \geq w] \} > (1 + R)y'$, which occurs with probability $\Pi = F\left(\frac{U_2}{1+R}\right)$. With probability $1 - \Pi$, the politician still supports the old project and gets $E[(1 + R)y | (1 + R)y \geq U_2]$. We want to see whether the results in Proposition 2 are robust with the possible arrival of the new project.

Proposition 6: 1. *Sufficiently intense rotation and performance-based rewards induce an adaptive and benevolent local politician. The old and the new entrepreneurs invest accordingly.*

In algebra, denote \hat{R} such that $\hat{R}(k) = \max\{R_1, R_2\}$, where R_1 and R_2 satisfy:

$$(1 - p)[1 - F(w^*(R_1))] \{ E_f(y | y \geq w^*(R_1)) - w^*(R_1) \} = k$$

$$1 - G(\tilde{w}(R_2)) [E_g(y | y \geq \tilde{w}(R_2)) - \tilde{w}(R_2)] = k.$$

Also,

$$\underline{\pi} = 1 - \frac{c}{(1 - p) \{ (1 + R)E(y) - U_1 \} + p \{ (1 - \Pi)[E_f[(1 + R)y | (1 + R)y \geq U_2] - U_2] \}}$$

Then if $R > \hat{R}$ and $\pi \geq \underline{\pi}$, the politician does not pay c , and both the old and the new entrepreneurs reap non-negative returns.

2. $\frac{\partial \pi(R, c, p)}{\partial R} > 0$: the minimum rotation frequency increases when performance-based rewards are stronger. If the minimum rotation frequency does not change, more intense

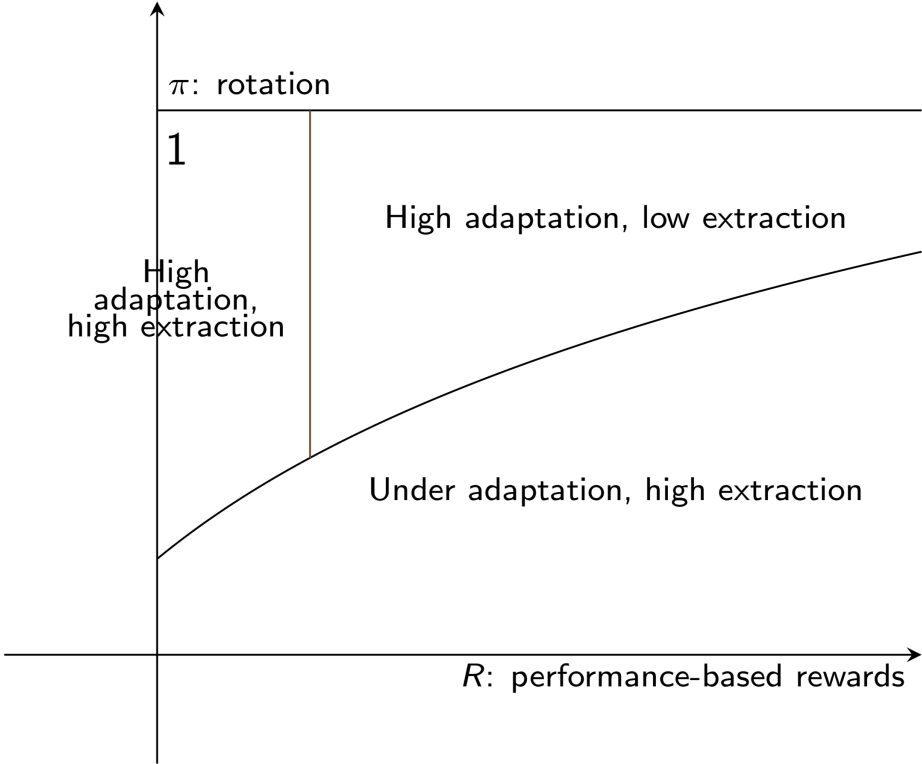


Figure 1.4: The Model with Firm Dynamics

performance-based rewards incentivize the politician in $t = 1$ to pay c , and the politician in $t = 2$ will fully predate the old entrepreneur.

The results are summarized in Figure 4, which shows that the complementarity between rotation and performance-based rewards persist under adaptation concerns. Specifically, with frequent rotation and strong performance-based rewards, the entrepreneur will invest. The local politician will facilitate “creative destruction” if a more productive project arrives. With stronger performance-based rewards, the equilibrium of high investment and high creative destruction can only be sustained with more frequent rotation. Thus, we formalize the notion that rotation and performance-based rewards provide property rights and facilitate “creative destruction” simultaneously, a remarkable feature rarely satisfied by *de facto* property rights based on reputation.

1.5 Extension: Multiple Rounds of Bargaining

A critique of above models is that rotation raises the bargaining power of the entrepreneur simply because the politician and the entrepreneur can only bargain in $t = 2$. If the politician has a chance to engage in multiple rounds of bargaining, she may gradually learn the productivity of the entrepreneur's project by inferring from the entrepreneur's strategy to reject or accept the politician's proposed rent extraction. The politician may not need to invest in local knowledge. If so, anticipated rotation in the future is not helpful at all. In this section, I deal with this critique by showing that all intuitions in Section 1.3 are robust even if the players can bargain in $t = 1$.

The timeline is the same as that in the workhorse model of Section 1.3 except that the politician can propose to extract w_1 in $t = 1$ before she decides whether to learn about y . If the entrepreneur accepts w_1 , the politician helps the entrepreneur to finish the project in $t = 2$ and gets her rent w_1 . The project's value is still realized in $t = 2$ because it is a long-term investment. If the politician and the entrepreneur settle an agreement to extract w_1 , the politician has no incentive to pay c and learn about the entrepreneur's project anymore. If w_1 is rejected, the politician can pay c to learn y , and the "subgame" is exactly the same as in the workhorse model of Section 1.3 following a rejection of w_1 : the principal rotates the politician with probability π , the politician proposes to extract w_2 , and the entrepreneur accepts or rejects the proposal.

Assumption 4: *Suppose that $w_1 = w_2$. If an entrepreneur with \hat{y} accepts w_1 , then any entrepreneur with $y \geq \hat{y}$ also accepts w_1 .*

The assumption is crucial to reduce the complexity of the problem when we consider the politician's strategy to offer $w_1 = w_2$. Without Assumption 4, there are so many possible distributions of the entrepreneur that reject w_1 when $w_1 = w_2$. With Assumption 4, the politician in $t = 2$ faces a distribution censored from the right, which significantly simplifies the problem of the politician in $t = 2$.

I can prove the following key result:

Proposition 7: *Denote:*

$$\underline{\pi}_1(R) = 1 - \frac{c}{(1+R)E[y] - \max_w [1 - F(w)]\{w + E[Ry | y \geq w \}]}$$

$$\underline{\pi}_2(R) : (1 - \underline{\pi}_2(R)) \max_{w_2} [1 - F(w_2)]\{w_2 + E[Ry | y \geq w_2 \]} =$$

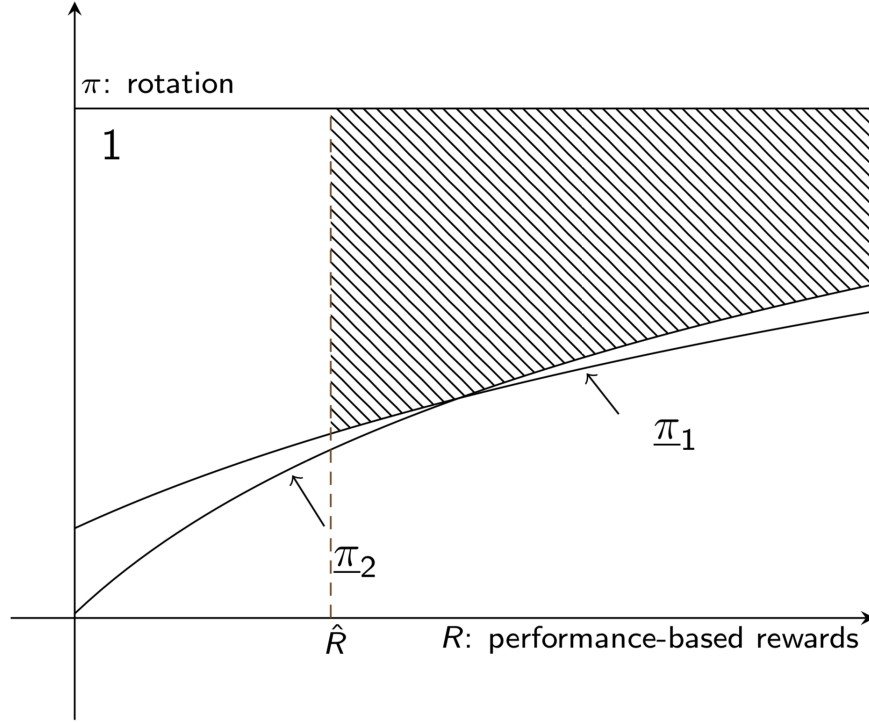


Figure 1.5: Robustness to Multiple Rounds of Bargaining

$$\max_{w_1} \left\{ F(w_1) \left\{ (1-\pi) \left\{ (1+R) E[y|y \geq w_1] \right\} + \pi \tilde{U} - c \right\} + [1-F(w_1)] \left\{ (1-\pi) \left\{ w_1 + E[Ry | y \geq w_1] \right\} + \pi \tilde{U} \right\} \right\}.$$

$$\hat{R} : k = [1 - F(w^*(\hat{R})) \{ E[y|y \geq w^*(\hat{R}) - w^*(R)] \}, \frac{w^*(\hat{R}) f(w^*(\hat{R}))}{1 - F(w^*(\hat{R}))} = \frac{1}{1 + \hat{R}}.$$

1. If $\pi \geq \underline{\pi}_1(R)$, $\pi \geq \underline{\pi}_2(R)$, and $R \geq \hat{R}$, the politician does not pay c to learn about y , and the entrepreneur makes the investment k .

2.

$$\frac{\partial \underline{\pi}_1(R)}{\partial R} > 0, \frac{\partial \underline{\pi}_2(R)}{\partial R} > 0$$

So with stronger performance rewards, the minimum rotation frequency $\underline{\Pi}(R) = \max\{\underline{\pi}_1(R), \underline{\pi}_2(R)\}$ also rises to guarantee the entrepreneur's ex ante investment.

Although the algebra becomes much more complex, the core intuition of the paper remains unchanged. As shown in Figure 5, with stronger performance-based rewards, the minimum rotation frequency still rises. The first constraint $\pi \geq \underline{\pi}_1(R)$ is the same constraint as in Section 1.3. It guarantees the sequential rationality of no information acquisition after w_1^* has been rejected. For the second constraint $\pi \geq \underline{\pi}_2(R)$, it guarantees that the payoff as an uninformed politician is higher than even the highest payoff that detailed information can deliver, which is:

$$\max_{w_1} \left\{ F(w_1) \{ (1-\pi)(1+R)E[y|y \geq w_1] + \pi\tilde{U} - c \} + [1-F(w_1)] \{ (1-\pi) \{ w_1 + E[Ry | y \geq w_1] \} + \pi\tilde{U} \} \right\} \quad (1.6)$$

$$= \max_{w_1} \left\{ -F(w_1)c + \pi\tilde{U} + (1-\pi) \{ [1-F(w_1)]w_1 + F(w_1)E[y|y \leq w_1] + E[Ry] \} \right\}. \quad (1.7)$$

To understand Equation (1.6), note that if the politician proposes to extract w_1 in $t = 1$, the entrepreneur gets $y - w_1$ by accepting w_1 . If he rejects w_1 , the politician becomes informed in $t = 2$ and extracts all surplus. Thus, the entrepreneur rejects w_1 if $y - w_1 \leq 0$, which occurs with probability $F(w_1)$. In this case, the politician pays c in $t = 1$. If not rotated, the politician extracts all surplus, so the expected payoff is $U^1 \equiv (1-\pi) \{ (1+R)E[y|y \geq w_1] \} + \pi\tilde{U} - c$, where the unrotated politician gets $(1+R)E[y|y \geq w_1]$ and the rotated politician gets \tilde{U} . If $y - w_1 \geq 0$, the entrepreneur accepts w_1 , which occurs with probability $1 - F(w_1)$. The payoff to the politician is $U^2 = (1-\pi) \{ w_1 + E[Ry | y \geq w_1] \} + \pi\tilde{U}$: the unrotated politician gets $w_1 + E[Ry | y \geq w_1]$. Thus, the payoff to a politician planning to pay c is $F(w_1)U^1 + [1 - F(w_1)]U^2 = F(w_1) \{ (1-\pi)(1+R)E[y|y \geq w_1] + \pi\tilde{U} - c \} + [1 - F(w_1)] \{ (1-\pi) \{ w_1 + E[Ry | y \geq w_1] \} + \pi\tilde{U} \}$.

Simple algebra gives us Equation (1.7), which shows that an unrotated and informed politician still reaps full benefit from performance-based rewards with bargaining opportunity in $t = 1$. Again, there will be no bargaining breakdown at all: either the players strike a bargain in $t = 1$, which delivers a non-negative surplus to the entrepreneur, or the players strike a bargain in $t = 2$, when the entrepreneur gets 0. But in either case, there will be no bargaining breakdown under any circumstance. By contrast, a substantial risk of bargaining breakdown persists even with multiple rounds of bargaining. If the uninformed politician is not rotated, the maximum utility is still $\max_w [1 - F(w)] \{ w + E[Ry | y \geq w] \}$. It seems counter-intuitive that the additional bargaining opportunity cannot improve the payoff of an uninformed politician. She cannot learn anything from the bargaining opportunity in $t = 1$. This is because it is never optimal for a forward-looking entrepreneur to reveal

information about his type early in the bargaining process. Given that he is patient, he can wait and exhaust all such opportunities for the politician. Thus, the politician cannot benefit from the additional bargaining opportunity and gets exactly the same utility as the case with only one opportunity to bargain. The key implication is that the unrotated and uninformed politician still only reaps partial benefit from performance-based rewards, which is $[1 - F(w^*)]E[Ry|y \geq w^*]$. In general, even if the players are not fully patient, the politician cannot learn everything from multiple rounds of bargaining (Sobel and Takahashi, 1986). This means that the politician who plans to not pay c remains less informed than the entrepreneur so that bargaining breakdown is always a possibility. Thus, in more general settings where the players are not fully patient, stronger performance-based rewards can still tempt the politician to invest in information acquisition, which has to be balanced by more frequent rotation.

1.6 Empirical Evidence

I examine how anticipated personnel events on local politicians affect firm's decisions in China. Beforehand, I need to introduce additional institutional background for the empirical analysis. Local governments have four tiers in China. There are 31 provincial-level units, with 10 to 15 cities/prefectures in each province. A city usually has jurisdiction over several counties. The two key politicians for a city are the party secretary and the mayor. The party secretary is the chief politician in the city. The mayor is formally subordinate to the secretary because the mayor always holds the concurrent position of a deputy party secretary. However, there is a distinctive division of labor between the party secretary and the mayor. The party secretary wields political power unmatched by any other politician in the city. She guarantees such power through direct control of the organizational department and propaganda department, the two most powerful departments in city party committee. The organizational department in a city controls all political appointment at the county level. The propaganda department directs state-owned media and controls censorship of commercial media. However, everyday management of the city government is the sole responsibility of the mayor. Specifically, all economic departments, such as revenue, construction, and commerce, are under the direct leadership of the mayor. By contrast, the city party committee does not include any economic departments. This arrangement is in stark contrast to local politics in the Soviet Union, where the party secretary also wields direct economic power through economic departments in the party committee. The empirical strategy heavily relies on variation from the dual leadership. Figure 6 illustrates the definition of rotation and promotion that I will use in the empirical analysis. Rotation is defined as a lateral transfer for a mayor to another mayor position. A promotion event occurs when a mayor is appointed as the secretary of her own city. A large literature documents that promotion in China is strongly correlated with performance (Maskin et al. 2000; Li and Zhou, 2005; Jia et al. 2017), so a promotion opportunity is a special form (and the most common form) of performance-based rewards in China. Rotation *and* promotion occur when a mayor is

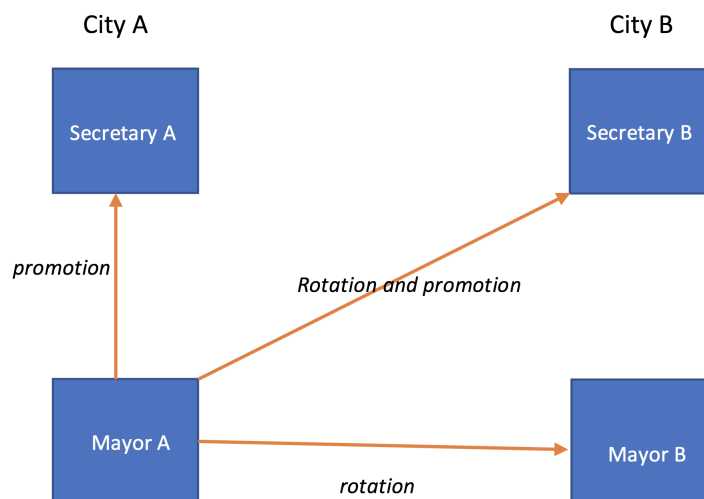


Figure 1.6: Rotation and Promotion in Local Politics

promoted as a secretary in another city.

The provincial organizational department devises and implements personnel management of all city-level politicians, including the appointment of both the party secretary and the mayor of a city. Hence, transfer of city politicians across provinces is very rare, which has important implications for the empirical analysis.

Testable Implications from the Theory

The main testable implication is the complementarity of rotation and performance-based rewards. The theory predicts that a firm invests more when the prospect of rotation and promotion are strong for the mayor of the city where the firm locates. The effect should be robust even after I account for the individual effects of rotation and promotion. The ideal experiment is to randomly assign mayors into four groups. The first group serves as the control group; the second group receives only promotion prospect; the third group receives only rotation prospect; the fourth group receives both. As such random assignment is rare, I want to utilize variation in anticipated rotation and promotion that are arguably exogenous.

Specifically, I extensively explore the “jackknife” or “leave-one-out” variation (see Figure 7). A natural proxy for future rotation events is the anticipated retirement of mayors in *other* cities². City-level officials in China face a mandatory retirement age of 60. Before that, they are usually transferred to an honorary and powerless position in the legislature

²I only look at mayors in other cities, but in the same province. This is because, as mentioned, personnel management at city-level is controlled by the provincial organizational department. So inter-provincial transfers of city-level politicians are extremely rare.

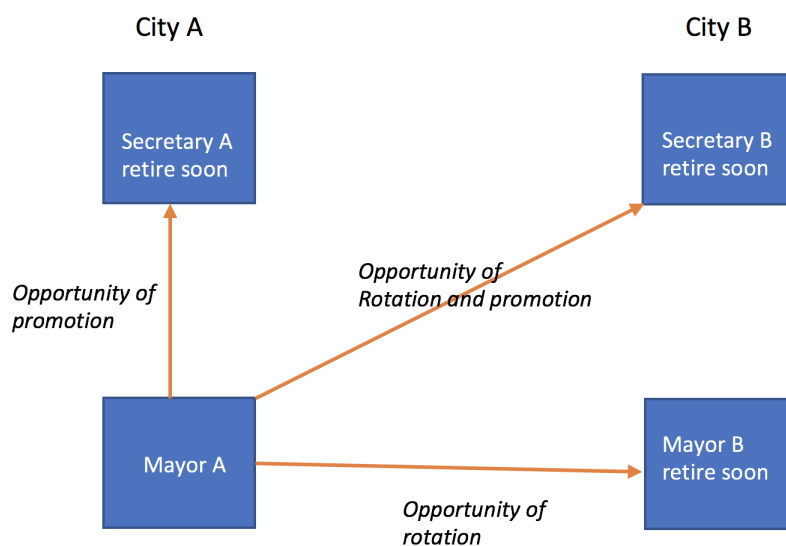


Figure 1.7: Proxying Rotation and Promotion Events

as a transition to full retirement at 60 (Wang, 2016; Xi et al. 2017). This event can happen any time after the city-level official turns 56. Thus, if the mayor of City B turns 56, all politicians in the province expect a job vacancy in the next four years. When the mayor of City B is actually assigned the honorary position or even retires from on her job, the mayor in City A is among the likely candidates to fill up the mayorship in City B. Hence, I proxy anticipated rotation by the fraction of mayors in *other* cities that are more than 56 years old.

For a promotion event, I look at whether the secretary in a specific city (e.g., City A) is more than 56 years old. In such a case, the mayor in city A anticipates a high likelihood of promotion if the city's economy grows fast, as the mayor of City A is the most likely candidate for the secretary position in City A.

We have a few proxies for the complementarity between rotation and promotion. The interactions between the above proxies for rotation and promotion can reasonably measure the complementarity effect. In addition, the anticipated retirement of secretaries in *other* cities improves the prospect of both rotation and promotion for a mayor, providing another variable that approximates the complementarity. Figure 7 illustrates the empirical variation I used to construct the proxies for anticipated personnel events.

To summarize: if my theory captures an important dimension of China's political economy, coefficients should be positive and precisely estimated for the interaction terms between

proxies for rotation and promotion and for the fraction of secretaries in other cities who are more than 56 years old.

The effect should also be different for private firms vs. state-owned enterprises (SOEs). We should observe a much more pronounced response from private firms to better-protected property rights, but the effect should be smaller for SOEs that have already enjoyed substantial bargaining power against the mayor.

Relevance of Retirement in Other Cities on a Mayor’s Own Tenure

I obtain personnel data on politicians from Chen (2016). For each city-year observation, the dataset identifies the governing secretary and mayor along with their age, gender, ethnicity, education, and work experience. The dataset covers all cities in the 27 provinces and autonomous regions between 2000 and 2010³. For firm covariates, I obtain them from Annual Survey of Industrial Production (ASIP). The survey contains all firms that have annual sales above 5 million RMB (equivalent to 800,000 USD). For each firm-year observation, the survey records the location, sales, inventory, number of employees, total assets, fixed assets (physical capital), accumulated depreciation, liquid assets (such as cash and account receivables), intangible assets (such as intellectual property), liability, industry code, and ownership by paid-in capital. I use the balance sheet data to construct the measurement of physical capital and also many other firm characteristics as control variables. I merge the ASIP data with politician personnel data so that each firm-year observation is identified with politicians governing the city where the firm locates. Summary statistics for key variables are provided in Table 1.

Before testing the theoretical predictions, I document the relevance of retirement in other cities for mayor turnover in a specific city. The regression I run is:

$$n_{jt} = \mu_j + PC_t + \text{retirement of other secretaries}_{jt} + \text{retirement of other mayors}_{jt} + s_{jt} + m_{jt} + \varepsilon_{jt}.$$

n_{jt} is whether the mayor j ’s term terminates in the year t . μ_j is mayor fixed effects. PC_t is the well-documented effect of “party congress”: turnover rate is high in years before party congress (Xi et al. 2017). The retirement of other secretaries/mayors records the fraction of secretaries/mayors who are 59 years old in the province in the year $t - 1$. These politicians

³In other words, it only excludes the four cities that enjoy provincial status: Beijing, Shanghai, Tianjin, and Chongqing. It is reasonable to exclude these cities, as their mayors and party secretaries are (exceptionally important) politicians with provincial ranks. It is the Central Organizational Department (rather than the provincial branch) that manages such prestigious positions. In addition, the availability of these positions is irrelevant to the mayor and the party secretary of a typical city, who only enjoys city ranks. This is because even the strongest candidate with a city rank can only be promoted to a position with a deputy provincial rank.

will all retire within the year t . In the process, their retirement generates many vacancies for the provincial organizational department to fill up.

s_{jt} denotes whether the secretary co-ruling with the mayor j is older than 56, and m_{jt} is whether mayor j is older than 56. Robustness checks with different age cutoffs are implemented, showing similar results. Alternative specifications using logit models produce similar estimates. They are omitted here. All standard errors are clustered at the mayor level.

Table 2 shows that retirement of other secretaries or mayors is strongly correlated with a mayor’s turnover. When secretaries or mayors in other cities retire, it strongly predicts termination of a mayor’s term. Other covariates also make economic sense. If the mayor or the secretary is old, turnover is more likely. An old mayor faces high “risk” of a transfer to an honorary position, while an old secretary increases the likelihood of promotion for the mayor. For the party congress effects, the omitted years are “party congress in 5 years”. As the Party Constitution requires a meeting of the party congress every 5 years, the omitted category represents the year immediately following a party congress. This year comes with the lowest turnover probability. We can see that the two years before the party congress, along with the year when party congress assembles, have pronounced higher likelihood of turnover. This finding is consistent with prior studies (Xi et al. 2017). In Columns (4) and (5), the retirement of other secretaries is defined as the fraction of other secretaries that were 58 (57) two (three) years ago. A similar definition applies to the retirement of other mayors. The results are very similar to Column (3).

Anticipated Retirement on Composition of Capital

The main empirical specification is:

$$y_{it} = \alpha_0 + \mu_i + \lambda_t + X_{it}\rho_0 + \rho_1 s_{it} + \rho_2 R_{it} + \beta s_{it} * R_{it} + \gamma C_{it} + \varepsilon_{it}, \quad (1.8)$$

where y_{it} is the ratio of fixed assets to total assets. In the Chinese Accounting Standards, fixed assets are defined as physical capital with long-term returns. μ_i is firm fixed effects, λ_t is year fixed effects, and s_i is a proxy for performance-based rewards. As said, studies on Chinese politics document that an age cutoff works very well to proxy for promotion opportunities (Wang, 2016; Xi et al. 2017). A secretary older than 56, who will retire in the next four years, is a precursor to a promotion opportunity for the mayor. Hence, I denote $s_{it} = 1$ if the secretary is older than 56 and 0 otherwise. R_{it} is the fraction of mayors in other cities older than 56. C_{it} records the fraction of secretaries in other cities older than 56. We are mainly interested in the coefficients on $s_{it} * R_{it}$ and C_{it} . X_{it} is a vector of control variables in all specifications of Tables 3 and 4. It includes rich covariates measuring characteristics about the firm, as well as the mayor and the party secretary of the city

where the firm locates. For firm characteristics, I control (lag and logarithm of) output, the number of employees, value added, profit, management fee, inventory, firm age, and debt. For politician characteristics, I control (for both the mayor and the party secretary) age, gender, ethnicity, education, and work experience (includes whether the politician used to work in the Communist Youth League and whether the politician used to work as a personal assistant or the director of the office for a senior politician)⁴.

Table 3 lists the main results of the paper. The first two rows are proxies for the complementarity of rotation and performance-based rewards. Column (1) is restricted to mayors who are in their first or second year in office. Notice that we have precisely estimated coefficients for the fraction of old secretaries in other cities, but the coefficient on $s_{it} * R_{it}$ is not precisely estimated. Columns (2) and (3) split the sample into young mayors and old mayors, with the cutoff at 53 years old⁵. We can see that the effects of anticipated rotation and promotion are much stronger for young mayors than for old mayors. In Column (2), the two proxies are both precisely estimated and economically significant for young mayors. If the fraction of old secretaries in other cities increase by 1%, a firm raises the ratio of fixed assets to total assets by 0.114%. The two proxies become either imprecisely estimated or have the wrong sign for old mayors in Column (3). The results are reasonable, as anticipated job vacancy is much less relevant for old mayors than for the young mayors. As a consequence, the anticipated rotation and promotion induced by expected job vacancy have a much stronger impact on young mayors than on old mayors. In other words, Column (3) serves as a placebo test, while Column (2) is the preferred specification because anticipated job vacancy provides the strongest incentives for young mayors new in office.

Column (4) is another placebo test, as it examines young mayors serving in office for more than three years. As the average term of a mayor is 3.8 years, these mayors face a higher hazard of immediate turnover. Presumably, they have less incentive to promote long-term investment, which generates payoffs after these mayors finish their tenures. The results in Column (4) show that the coefficient on $s_{it} * R_{it}$ is similar in scale but imprecisely estimated, while the coefficient on C_{it} has a wrong sign.

Column (5) and Column (6) split the sample of young mayors new in office to test heterogeneous effects on private firms vs. SOEs. We can see that the coefficient on C_{it} is very precisely estimated for both private firms and SOEs, but the point estimator for private firms is about 40% higher than that for SOEs. The second proxy for the complementarity effect $s_{it} * R_{it}$ is precisely estimated for private firms, but it is imprecisely estimated with the wrong sign for SOEs. Taken together, Columns (5) and (6) lend additional support to the theory.

⁴Whenever possible, I also include whether the mayor is older than 56, and its interaction with R_{it} . They are always insignificant with very small point estimates. They are also dropped from most specifications as the sample is usually restricted to either young or old mayors.

⁵This cutoff allows me to cluster standard errors at city level for Column (3). If I use cutoffs larger than 54, there are too few cities to cluster for Column (3). The qualitative results are quite similar, however, if I employ different cutoffs.

We do find that the effects of anticipated rotation and promotion are smaller for SOEs, presumably because SOEs already enjoy large bargaining power against local politicians.

Timing of Anticipated Retirement on Composition of Capital

Table 4 summarizes the results to test whether there are any dynamically heterogeneous effects. Specifically, I define three separate C_{it} : C_{it}^1 is the fraction of secretaries in other cities that are more than 58 years old; C_{it}^2 is the fraction of secretaries in other cities between 56 and 57 years old; and C_{it}^3 is the fraction of secretaries in other cities between 53 and 55 years old. R_{it}^1 , R_{it}^2 , R_{it}^3 are defined in a similar manner. Thus, I run:

$$y_{it} = \alpha_0 + \mu_i + \lambda_t + X_{it}\rho_0 + \rho_1 s_{it} + \rho_2 R_{it} + \beta_1 s_{it} * R_{it}^1 + \beta_2 s_{it} * R_{it}^2 + \beta_3 s_{it} * R_{it}^3 + \gamma_1 C_{it}^1 + \gamma_2 C_{it}^2 + \gamma_3 C_{it}^3 + \varepsilon_{it}.$$

β_1 to β_3 and γ_1 to γ_3 capture the dynamic heterogeneity of anticipated personnel events, if any. Column (1) of Table 4 again includes all mayors who are new in office. We can see that the coefficient on C_{it}^2 is precisely estimated, while coefficients on C_{it}^1 and C_{it}^3 are either very small or have a negative sign. None of the coefficients on $s_{it} * R_{it}^1$, $s_{it} * R_{it}^2$, or $s_{it} * R_{it}^3$ are precisely estimated. In the preferred specification, Column (2), both the coefficients on C_{it}^2 and $s_{it} * R_{it}^2$ are precisely estimated, while all other coefficients are either imprecisely estimated or have negative signs. Columns (3) and (4) are placebo tests using old mayors and mayors in their 3rd or 4th year in office. Columns (5) and (6) again split the sample into private firms and SOEs, and we can see that the precisely estimated coefficients on C_{it}^2 and $s_{it} * R_{it}^2$ are mostly driven by private firms.

Hence, the most relevant personnel events are those during 3 to 4 years from now. Those that will happen within 2 years have few effects, while those that will happen in 5 to 7 years have, at most, negative effects. Anticipated personnel events within 2 years cannot be influenced much by long-term investment today, which usually generates payoffs in the more distant future. In other words, it is largely determined who will fill up the job vacancy available in 2 years. Thus, mayors have less incentive to promote long-term investment to bid for job vacancies available almost immediately. For those personnel events that occur more than 5 years from today, it is also less relevant for the mayors who have on average 3.8 years of tenure. The presence of many mayors and secretaries between the age 53 and 55 can even intensify competition for the most lucrative positions for the next few years, as these politicians are fully qualified for personnel shifts. As these politicians have one last big chance for a meaningful promotion before retirement and as they are also the most experienced politicians that can be promoted, they constitute the most formidable competitors for higher positions. This explains why we find a negative sign for β_3 and γ_3 .

By contrast, job vacancy available 3 to 4 years from now will be released at exactly the time when the new mayor, with an average tenure of 3.8 years, anticipates that her tenure will be finished. Thus, the availability of jobs elsewhere precisely 3 to 4 years from now is

especially meaningful for the mayors who just took office. This explains why we see especially pronounced effects for job vacancies anticipated in the “intermediate” future.

1.7 Conclusion

The insights of this paper are applicable to generic organizations, as we can interpret the politician as a division manager and the entrepreneur as her subordinate. Thus, the models integrate key elements of personnel management, such as rotation, performance-based rewards, and authority relationships within a division. Starting from the simple intuition that a politician/manager can accumulate power over her tenure, the analysis reaches the surprising yet intuitive conclusion that strong performance-based rewards can be destructive. This is precisely because performance-based rewards encourage the politician/manager to accumulate power. The theory gives a simple rationale behind the management practice to implement intense rotation and performance-based rewards at the same time, which restores their desired disciplinary roles. Rotation-based solutions also deliver both flexibility and commitment, enabling organizations to achieve static and dynamic efficiencies.

There are several directions for future research. The paper presupposes high bureaucratic capacity with impersonal rules over politicians and shows that institutionalized rotation and performance-based rewards can energize the private economy. Where does such bureaucratic capacity come from? Specifically, is there any feedback from the private economy to bureaucratic capacity? The traditional view is that bureaucratic capacity is hurt by a strong private economy, as the best talents are recruited by the private sector (Caselli and Morelli, 2004). However, a strong market economy and impersonal bureaucracy arose together in many episodes (Fukuyama, 2011; Tackett, 2014). This begs the question of whether there is any positive feedback from a strong private sector to bureaucratic capacity. We also have many other episodes where strong bureaucracy lags significantly behind economic achievements (Fukuyama, 2011). It will be fruitful to construct a unified model and use the model for two exercises: 1) deriving conditions that determine the sign of the feedback’s direction 2) mapping key elements of the theory to real-world episodes.

Highly related to the above research agenda, it is intriguing to understand the historical origins and evolutions of bureaucratic capacity. I am now pursuing this line of research through a textual analysis of historical records of Imperial China. Specifically, a large number of historical bureaucratic positions are identified from a dictionary of them, and correlations of these bureaucratic positions in a large database of historical records are being constructed. This provides a rich empirical platform to investigate institutional complementarity and substitution that can inspire future theoretical research on the origins and evolutions of bureaucratic capacity.

Table 1.1: Summary Statistics Part I

	count	mean	sd	min	max
Firm characteristics					
ratio of fixed assets	2211700	.3529688	.228051	0	.9999979
fixed assets	2211700	8.574127	1.751689	0	19.15294
"flexible" assets	1810983	8.817254	1.753437	0	18.03701
output	2051817	10.24835	1.414884	0	19.70661
debt	2200074	9.178852	1.746334	0	18.85835
inventory	2097618	7.728052	1.879163	0	17.45033
management	2191117	6.900411	1.634475	0	17.41659
#employees	2058589	4.800024	1.148959	0	13.25281
firm age	2211601	12.15931	11.67262	1	363
ownership:					
state	2571858	.0718267	.2582007	0	1
private	2571858	.7132773	.452231	0	1
foreign	2571858	.2148731	.4107343	0	1
Mayor characteristics					
gender	2167305	1.950869	.2161407	1	2
ethnicity	2152417	.9724881	.1635695	0	1
age	2165908	50.48117	4.196176	36	60
age \leq 56	2179957	.9093711	.2870807	0	1
Work experience					
Youth league	2150124	.3670756	.4820075	0	1
Director of office	2149246	.5895435	.4919167	0	1

Table 1.2: Summary Statistics Part II

	count	mean	sd	min	max
Secretary characteristics					
gender	1782536	1.980095	.1396751	1	2
ethnicity	1734758	.9361456	.2444935	0	1
age	1770591	52.07351	3.766545	39	60
Secretary > 56	2179957	.2959815	.4564828	0	1
Work experience:					
Youth League	1781359	.345736	.4756078	0	1
Director of office	1790342	.6562059	.4749735	0	1
Anticipated retirement					
Fraction of other secretaries older than 56	1888847	.1735954	.1465631	0	1
Fraction of other mayors older than 56	1888847	.117085	.1096847	0	1

The following variables are in log scale: fixed assets, "flexible" assets, output, debt, inventory, management, number of employees. "Flexible" assets sums up liquid assets and intangible assets. Ownership variables are indicators: whether the firm is an SOE, private firm, or foreign firm. The ownership characterization is based on paid-in capital. For gender, 1 indicates female, 2 indicates male. For ethnicity, 0 indicates ethnic minority, 1 indicates *Han* Chinese. Work experience includes two dummies: whether the politician used to serve in Communist Youth League; whether the politician used to work as director of the office.

Table 1.3: Relevance of Others' Retirement on Own Turnover

	(1)	(2)	(3)	(4)	(5)
	Mayor Turnover				
Retirement of other secretaries	0.629** (0.301)	1.222*** (0.389)	0.830*** (0.317)	0.615*** (0.212)	0.784*** (0.185)
Retirement of other mayors	1.838*** (0.462)	2.702*** (0.634)	2.133*** (0.627)	1.808*** (0.509)	1.550*** (0.377)
Secretary age > 56	0.0737*** (0.0181)	0.298*** (0.0337)	0.255*** (0.0370)	0.254*** (0.0370)	0.257*** (0.0370)
Mayor age > 56	0.176*** (0.0395)	0.400*** (0.0680)	0.398*** (0.0647)	0.398*** (0.0649)	0.396*** (0.0672)
Party congress in: this year			0.152*** (0.0322)	0.163*** (0.0316)	0.162*** (0.0315)
2 years			0.157*** (0.0271)	0.170*** (0.0267)	0.164*** (0.0263)
3 years			0.0232 (0.0213)	0.0248 (0.0216)	0.00781 (0.0217)
Mayor FE	No	Yes	Yes	Yes	Yes
<i>N</i>	2884	2884	2884	2884	2884
adj. <i>R</i> ²	0.022	0.088	0.112	0.112	0.116

Standard errors are in parentheses and are clustered at mayor level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is whether the mayor's term is terminated in year t .

Table 1.4: Anticipated Job Vacancy on Capital Composition

	(1)	(2)	(3)	(4)	(5)	(6)
	fixed assets	fixed assets	fixed assets	fixed assets	fixed assets	fixed assets
fraction of other secretaries old (C_{it})	0.120*** (0.0232)	0.114*** (0.0231)	0.0591 (0.0496)	-0.0322 (0.0570)	0.118*** (0.0242)	0.0737*** (0.0178)
own secretary old* other old mayors ($s_{it} * R_{it}$)	0.0611 (0.0778)	0.157** (0.0773)	-0.581* (0.328)	0.202* (0.104)	0.184** (0.0782)	-0.0560 (0.0807)
fraction of other mayors old (R_{it})	-0.00108 (0.0468)	-0.00411 (0.0373)	0.107* (0.0579)	0.0974* (0.0548)	-0.00393 (0.0393)	-0.0548 (0.0416)
own secretary old (s_{it})	0.000457 (0.0106)	-0.0104 (0.0103)	-0.0183 (0.0196)	0.00745 (0.00890)	-0.0128 (0.0103)	0.00994 (0.0137)
Sample:						
Mayor age	all	young	old	young	young	young
Mayor term	early	early	early	late	early	early
Firm ownership	all	all	all	all	private+foreign	SOEs
N	376654	279576	78940	110072	258531	16117
adj. R^2	0.659	0.667	0.703	0.702	0.659	0.731

Standard errors are in parentheses and are clustered at city level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the ratio of fixed assets to total assets. All specifications control firm fixed effects, year fixed effects, the firm's characteristics, and the politicians' characteristics. For firm characteristics, I control (lag and logarithm of) output, the number of employees, value added, profit, management fee, inventory, firm age, and debt. For politician characteristics, I control (for both the mayor and the party secretary) age, gender, ethnicity, education, work experience (includes whether the politician used to work in the Communist Youth League and whether the politician used to work as a personal assistant or the director of the office for a senior politician).

Table 1.5: Timing of Anticipated Job Vacancy on Capital Composition

	(1)	(2)	(3)	(4)	(5)	(6)
	fixed assets	fixed assets	fixed assets	fixed assets	fixed assets	fixed assets
fraction of other secretaries	0.00644	0.0663	0.0790	0.00853	0.0737	0.0655*
retire in 2 years (C_{it}^1)	(0.0456)	(0.0488)	(0.0802)	(0.0689)	(0.0526)	(0.0376)
own secretary old * fraction of other	-0.0170	-0.150	1.560*	-0.0592	-0.134	-0.261***
mayors retire in 2 years ($s_{it} * R_{it}^1$)	(0.0950)	(0.128)	(0.853)	(0.219)	(0.136)	(0.0912)
fraction of other secretaries	0.117***	0.111***	0.0247	-0.0426	0.114***	0.0820***
retire 3 to 4 years from now (C_{it}^2)	(0.0296)	(0.0287)	(0.0475)	(0.0681)	(0.0301)	(0.0263)
own secretary old * fraction of other	0.0680	0.280**	-1.032***	0.101	0.307**	0.157
mayors retire 3 to 4 years ($s_{it} * R_{it}^2$)	(0.127)	(0.134)	(0.244)	(0.125)	(0.134)	(0.181)
fraction of other secretaries	-0.0331**	-0.0202	-0.0361	-0.00935	-0.0201	-0.0141
retire 5 to 7 years (C_{it}^3)	(0.0163)	(0.0202)	(0.0414)	(0.0294)	(0.0211)	(0.0201)
own secretary old * other	-0.0382	-0.107**	0.386	-0.158*	-0.104**	-0.133**
mayors retire 5 to 7 years ($s_{it} * R_{it}^3$)	(0.0383)	(0.0473)	(0.302)	(0.0802)	(0.0478)	(0.0659)
Sample:						
Mayor age	all	young	old	young	young	young
Mayor term	early	early	early	late	early	early
Firm ownership	all	all	all	all	private+foreign	SOEs
N	376654	279576	78940	110072	258531	16117
adj. R^2	0.660	0.668	0.705	0.703	0.661	0.732

Standard errors in parentheses and are clustered at city level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the ratio of fixed assets to total assets. All specifications control firm fixed effects, year fixed effects, the firm's characteristics, and the politicians' characteristics. For firm characteristics, I control (lag and logarithm of) output, the number of employees, value added, profit, management fee, inventory, firm age, and debt. For politician characteristics, I control (for both the mayor and the party secretary) age, gender, ethnicity, education, work experience (includes whether the politician used to work in the Communist Youth League and whether the politician used to work as a personal assistant or the director of the office for a senior politician).

Chapter 2

Meritocracy, Decentralization, and Local Dual Leadership

(Chinese regime) uses the higher ranked official to monitor the lower ranked one. At the same time, (the regime) also uses the lower ranked to divide the power of the higher ranked. The lower ranked cannot execute his power freely because he is monitored by the higher ranked. The high ranked is also constrained and cannot do whatever he wants. This is because his power is divided and delegated to the lower ranked. As a consequence, the central government exerts effective control.

The checks between the higher ranked and the lower ranked is a fundamental principle of statecraft in Qing China.

Luo Er'gang, *A Record of Green Standard Army*

2.1 Introduction

The two cornerstones of Chinese polity are political meritocracy and economic decentralization. Local politicians are promoted based on competence and performance, and economic policymaking is delegated to the province. They solve many incentive problems and constitute the foundation of China's reform and growth (Roland, Qian, 1998; Maskin, Qian, Xu, 2000; Li, Zhou, 2005; Xu, 2011; Bai, Hsieh, Song, 2014).

However, meritocracy and decentralization contradict the basic logic of autocracy, the loyalty-competence trade-off (Glazer, 2002; Egorov, Sonin, 2011; Svolik, 2012; Bai, Zhou, 2014). In authoritarian regimes, the leader does not want to appoint very competent subordinates with large policymaking power: the subordinate can challenge the leader through competence and policy discretion. As a consequence, the leader usually appoints mediocre

subordinates and centralizes policymaking. Svobik (2012) identifies the trade-off as the main dilemma of autocracy. The research question is: what is the Chinese institution that overcomes the dilemma and implements meritocracy and decentralization?

I propose that the institution is the appointment of both a party secretary and a governor to co-rule a province. The party secretary is the first ranked politician and wields formidable political power. He has a huge capacity to mobilize the population through his control of party branches, mass organizations, and propaganda apparatus. The governor serves as the head of government and deputy party secretary simultaneously, so he is second ranked and a subordinate of the party secretary. However, there is substantial economic delegation to the governor: the local party committee does not have any economic departments. All key economic departments are under the direct leadership of the governor. Consequently, the secretary does not have direct access to economic power.¹ The economic delegation to the governor makes it extremely difficult for a province to challenge central authority. This solves the loyalty-competence dilemma and gives the central authority full confidence to decentralize policymaking and promote competent politicians.

The obvious problem is that the secretary sometimes dominates over the governor and takes over discretion of economic policymaking. The standard theory of the separation of powers stresses strong checks and balances between politicians, but the Chinese regime contradicts the theory and supports the bias toward the party secretary. For example, the Chinese Constitution guarantees the leadership of the party over the government, yet at the same time, the regime also emphasizes that the secretary's power should be constrained and he should not take control of everything. At first glance, it is unclear why the regime sometimes supports the secretary's advantage over governor but not always so. Moreover, prevalent collusion opportunities between the secretary and the governor endanger the effectiveness of economic delegation. These puzzles render it necessary to explore the mechanism in depth.

To achieve this goal, I adopt a reputation/signaling model. There are two types of secretary, a benevolent type and a normal/selfish type. The benevolent secretary cares about the population intrinsically, while the normal secretary only cares about his own utility. In the benchmark model, there is no economic delegation to the governor. Thus, the secretary can provide public good to signal his benevolence. *The signal is informative, as it is costly for the secretary to provide public good.* Then, the secretary can use his political power to mobilize the population and challenge the central authority. The population is happy to

¹This is a major difference between China and the Soviet Union: a local party committee in the Soviet Union had many economic departments, such as agriculture, education, construction, industry, transportation, light and food industry, trade and financial organs (Nough, 1969). This gives the Soviet secretary direct access to economic power. In a companion paper, I analyze the economic origins of the divergence in delegation patterns. I propose that market economy in China, which was established de facto in the 1980s through dual-track price liberalization, empowers governors in several ways that renders it necessary to delegate economic policymaking. In a planned economy such as the Soviet Union, delegation of economic power will generate large efficiency loss, which can be easily avoided in a market economy.

join such collective action: they infer that the secretary is probably benevolent from public good provision, and a benevolent secretary will award all benefit from collective action to the population.² Thus, discretion over public good is a powerful signaling device that solves the collective action problem. The higher the discretion delegated to a province is, the more informative the signal is. Moreover, the benefit from collective action is larger for more competent secretaries, so a more competent secretary is more likely to challenge the central authority. In turn, the central authority appoints mediocre secretaries and centralizes policy discretion over public good.

With economic delegation, signaling effort will fail. *Now the governor bears the cost of public good provision.* This makes the signaling device too cheap for the secretary. A cheap signal is very uninformative, so the population finds it unattractive to join collective action. In the equilibrium with a balance of power between the secretary and the governor, only a benevolent governor provides public good, which gives the population full information about the type of governor. *However, it tells the population absolutely nothing about the secretary.* Nevertheless, mobilization capacity still lies with the secretary, as he still controls key political power. The secretary will abstain from any opportunity to challenge the central authority, knowing that the population would never join it. Freed from loyalty concern, the central authority appoints very competent politicians and decentralizes economic policymaking.

The design is robust to the two concerns I mentioned above: the secretary's dominance and collusion opportunity. Even if the secretary takes over decision of public good provision or credibly shares benefit from collective action with the governor, it is still the governor who bears the cost of public good. The loyalty concern will be reduced drastically; as a consequence, meritocracy and decentralization prevail.

We will see that with the secretary's dominance, there is still a positive probability of collective action. Recall that there is no collective action under the balance of power. Thus, the model predicts that central authority should enforce a balance of power. However, the analysis ignores an important dimension: the governor has some reasonable chance to be promoted as secretary. Under a balance of power, the governor fully reveals his benevolence. He can easily organize collective action when he becomes the new secretary. This is extremely risky, and the central authority strictly prefers the secretary's dominance over the governor.

²The tactic to buy loyalty through public good provision has been very popular in Chinese politics. The Chinese Communist Party (CCP) itself won victory in the Chinese Civil War over the ruling Kuomintang through the extensive support of peasants. Peasants supported the CCP enthusiastically because the CCP enforced aggressive land reform in CCP-controlled areas (Pepper, 1999). Recently, the Bo Xilai Saga illustrates the model very well: the Bo administration initiated phenomenal building of public good, which led to enormous popularity of the administration. However, such popularity failed to transform into real support from the population when Bo was in trouble (Zhao, 2012). As we will see, this is a natural result of the elaborated secretary-governor duality. A more systematic empirical analysis is provided by Persson, Zhuravskaya (2015). They found that secretaries who can more easily buy loyalty indeed provide more public good such as education and health care to please local elites, and those secretaries are less aligned with the objectives of the central government.

The central authority can do better by creating uncertainty regarding the secretary's power *vis-a-vis* the governor. Specifically, denote ξ as the probability that the secretary dominates over the governor and $1 - \xi$ as the probability that there is a power balance. Note that if ξ is close to 0 or 1, there is not too much uncertainty, but if ξ is bounded away from 0 and 1 (so there is uncertainty), the population will refuse to join collective action led by either the current secretary or future one (who is the current governor). The intuition is that ξ *adds noise to the signal*. When the population tries to join current collective action launched by the current secretary, the population does not know whether the public good is provided by a benevolent secretary or a benevolent governor; when the population tries to join future collective action to be launched in the future by the current governor, the population does not know whether the *current* public good is provided by a benevolent governor or a strong secretary who forces the governor to do so. The uncertainty "confuses" the population and forestalls Bayesian learning. The analysis gives a concrete foundation for the central authority's contradictory statements and reluctance to clarify the party-government relationship. Such ambiguity and uncertainty is actually optimal for the central authority.

Several things to note: 1. ξ is manipulatable: the central authority can influence people's belief about the secretary's strength through the powerful Central Propaganda Department. In addition, the central authority can calibrate the powers endowed to the secretary and the governor so that the secretary may take control of public good provision but that is not always the case. 2. The uncertainty can only be sustained with secretary-governor duality. 3. The governor's dominance over the secretary is never optimal. If the governor takes over mobilization capacity, effectively, we return to the case of no delegation. This can explain why the secretary is first ranked and the governor is appointed as a subordinate to the secretary: the central authority must prevent the governor's dominance over the secretary at all cost.

The analysis of uncertainty over the party-state relationship echoes the literature on strategic belief manipulation in Chinese media censorship (e.g., Lorentzen, 2014). Together with the workhorse model, they provide a game theoretic foundation for regime stability, political meritocracy, and economic decentralization in China and explain many seemingly contradictory puzzles and dilemmas in the Chinese political economy.

China's meritocracy and dual leadership is a modern incarnation of key institutions in Imperial China. Meritocracy is a prominent feature of the political process in Imperial China. It is also a common practice that the emperor appoints several politicians to co-rule a jurisdiction. I construct time-series variables measuring political institutions for over 1,300 years from historical records. Cointegration analysis shows that the strong correlation between meritocracy and dual leadership is not spurious. I also use a Vector Error Correction Model to show that dual leadership Granger causes meritocracy. In other words, meritocracy arises only after the consolidation of dual leadership. To summarize, the prediction of the

theory is consistent with China's experience over many centuries. The time-series analysis also shows the first-order relevance of my theory.

2.2 Literature

The paper relates to a large amount of literature. First, there is influential literature on the separation of powers in democracy (e.g., Persson, Roland, Tabellini, 1997, 2000; Laffont, 2000; Dragu, Chen, Kuklinski, 2014). The classical contribution of Persson, Roland, Tabellini (1997) shows that a conflict of interest between politicians solves the accountability problem. Persson, Roland, Tabellini (2000) discuss the fiscal implications of separation of powers with checks and balances. Laffont (2000) investigates the contract design problem associated with the separation of powers, in particular collusion issues. Dragu, Chen, Kuklinski's (2014) work is a complete characterization of Madisonian checks and balances. My research contributes to this line of research by showing that "separation of powers" also works in autocracy, but through a drastically different channel with one side being weaker than the other side. Taking the collusion concern of Laffont (2000) seriously, I also show how "separation of powers" in Chinese style achieves collusion-proofness.

Delegation is a key topic in both economics and political science. The standard argument is that delegation utilizes the information advantage of agent (Bendor, Meirowitz, 2004). The key conflict is that the agent's objective usually differs from the principal's. Many papers discuss institutions to prevent the "bureaucratic drift" of agents, such as discretion limits (Epstein, O'Halloran, 1999), menu law (Gailmard, 2009), and administrative procedures (Spiller, Tiller, 1997). My research shows how the delegation of economic power from the secretary to the governor benefits a third party, the central authority.

Loyalty-competence tradeoff was recently established as a key dilemma in autocracy. Indeed, Svobik (2012) advocates it as the key threat for any autocrat in the introductory chapter. Glazer (2002) studies the dilemma in a private firm, and he emphasizes the tradeoff between external and internal rent-seeking. Egorov, Sonin (2011) formalize the tradeoff in a contract theory model, which is supported empirically by Bai, Zhou (2014). In terms of loyalty concerns associated with decentralization, Bardhan (2002) notes that decentralization can erode the authority and power of the central government.

A large literature engages in the debate of political meritocracy in China (e.g., Maskin, Qian, Xu, 2000; Li, Zhou, 2005; Shi, Adolph, Liu, 2012; Jia, 2014; Persson, Zhuravskaya, 2014; Bai, Hsieh, Song, 2014; Jia, Kudamatsu, Seim, 2015). The fundamental contribution of Maskin, Qian, Xu (2000) shows that M-Form in China provides a platform for yardstick competition that strongly promotes political meritocracy, and the theory finds empirical support from Li, Zhou (2005). The meritocracy story backfired recently. Shi, Adolph, Liu (2012) show that connection rather than competence plays a fundamental role in promotion, but Jia, Kudamatsu, Seim (2015) find that political connection and performance are

equally important. Jia (2014), with the title “pollution for promotion”, identifies the dark side of meritocracy. Persson, Zhuravskaya (2014) find that the promotion mechanism is weakened by politicians’ social ties, but Bai, Hsieh, Song (2014) re-established the meritocracy story through a channel that was different from Maskin, Qian, Xu, (2000). They argue that “crony capitalism with Chinese characteristics” supports meritocratic promotion and economic growth. Promotion and political meritocracy is one of the most active fields of Chinese political economy. One cannot help taking one step back and contemplating whether Chinese institutions provide a concrete foundation for the meritocracy story. This is where my contribution lies. Complementing the canonical argument that meritocracy works in China because of the organizational structure of M-Form (Maskin, Qian, Xu, 2000), I show that meritocracy also has a deep political foundation: the subtle and carefully managed interaction between party and government strongly promotes the central authority’s confidence to enforce meritocracy.

Decentralization is also a key topic in the Chinese economy. Fiscal decentralization kindled the initial reform and growth in China (Weingast, 1995; Qian, Weingast, 1997; Qian, Roland, 1998; Berkowitz, Li, 1999; Jin, Qian, Weingast, 2005). In general, recent efforts to decentralize government service have received much attention in the developing country context (Bardhan, 2002). The tide of Chinese fiscal decentralization was reversed in the mid-1990s, but expenditure decentralization never lost its momentum. Xu (2011) notes that “the total expenditure of Chinese sub-national governments accounted for approximately 70 percent of the national total, which was far larger than that of the world’s largest federal countries such as the United States (46 percent), Germany (40 percent), and Russia (38 percent)”. The large spending discretion delegated to provinces allows them to fiercely compete against each other in infrastructure and public service. My paper uses a single foundation to explain both meritocracy and decentralization, the two big forces behind China’s economic achievement. We see that loyalty concern brought by either competence or decentralization is eliminated under delegation. The two key lines of the literature that used to develop on their own paths have become unified under a single framework.

The party-government relationship is at the heart of Chinese political institutions (Shirk, 1992, 1993), but surprisingly, it has received at best sparse attention from academia. Shirk (1993) is the most important contribution to the topic. My theory is quite consistent with her narrative. She finds that after the Communist victory in the Chinese Civil War, China adopted the party-government structure of the Soviet Union and pushed it to the extreme form: before the 1980s, the party fully took over economic management in China. Beginning in the 1980s, the Chinese Communist Party (CCP) abolished the economic departments in party committees and delegated most economic affairs to the government. The reform has survived the conservative attack after 1989, and it established the current form of party-government relationship in China, which is the main object of my study. This observation was also mentioned and articulated by Shambaugh (2008) and Guo (2009), and the thesis is basically the same. Shirk (1993) argues that the party’s delegation to the government

utilizes the latter's information advantage and professional knowledge in managing the economy, leading to successful economic reforms in China. To the best of my knowledge, my paper is the first analysis of the party-government relationship in China using formal models. I propose a different mechanism that emphasizes how the party-government relationship constrains local officials. The field is currently too small compared to its central importance in Chinese politics, and more research on the topic will help us understand the inner mechanism of the Chinese party-state.

Political scientists always have a keen interest in one-party states. However, most of the important contributions focus on electoral authoritarian regimes (e.g., Magaloni, 2006), where elections are far more substantive than Chinese ones. Magaloni, Kricheli (2010) note that the one-party state is more stable and growth-friendly than other forms of authoritarian regimes. Svobik (2012) provides a fantastic argument of how the party is a good device for co-optation. Because the distribution of party benefits depends on the rank of the party member, many benefits can only be realized when the party member ascends over the party hierarchy. This provides strong incentive for party members to support the regime: if the regime falls, all current and *future* benefits associated with party membership will be lost. By devoting my attention to the largest one-party state that has enjoyed remarkable stability and phenomenal growth for the past four decades, I hope to provide new insights into the mechanism of authoritarian resilience.

2.3 Party-state Structure in Chinese Provinces

My focus is local governance, so in this section, I will introduce provincial-level politics in China to provide a concrete background for my modeling exercise. The model fits the provincial-level politics well, and it can be easily extended to the sub-provincial level.

In mainland China, there are thirty-one provincial-level territories, which are the highest units of local jurisdictions. I call all of them province for simplicity. In 2014, their median population was 37 million (equivalent to Poland), and the median GDP was 287 billion US dollars (slightly higher than Chile). Thus, a province is immensely powerful even vis-a-vis the central authority. This is particularly the case for two additional reasons. First, Chinese provinces are very self-contained; each of them has a relatively complete array of industry sectors (Qian, Roland, Xu, 2006). This means that a province has sufficient resources to support itself if it tries to disobey central authority. As I have emphasized, there is also a very high degree of economic decentralization in China (Xu, 2011), which reinforces the province's economic independence, which could seriously cause a threat to central authority. Indeed, during the 1990s, there was a real concern that China might break up into dozens of states as a consequence of the provinces' strong economic power.

The two big players in provincial-level politics are the provincial secretary and the governor. The provincial secretary is the first-ranked politician in a province. He is the head of the

provincial party committee, which encompasses departments of organization, propaganda, united front, plus the committee of law and politics. The party also directly controls mass organizations such as the Labor Union, the Women's Union, and the Communist Youth League. Among the departments, the organization department is in charge of all major personnel decisions. It also controls the huge network of party branches that reside in every social organization. The propaganda department controls newspapers and TV stations and conducts extensive censorship of the Internet. The united front department directs and communicates with "democratic parties", which are very small parties under the CCP's leadership yet with quite elite members. The committee of law and politics was immensely powerful and controlled the court, the police, and the procuratorate until quite recently (China launched a comprehensive reform of the legal system in 2014 and undermined the committee of law and politics, but the effect remains to be seen). Thus, the provincial secretary enjoys huge political power and has formidable mobilization capacity.

The governor is the second-ranked provincial politician and is subordinate to the secretary. Formally, the governor serves as the head of the provincial government and the first-ranked deputy party secretary simultaneously. The most important feature of the governor is that all major economic departments are under the leadership of the governor, so the secretary does not directly manage the economy. For example, the governor controls the departments of education, industry, agriculture, business, construction, communication, public finance, science and technology, reform and development, and human resource and social welfare. By comparison, there is no economic department within the party committee's jurisdiction at all.

To summarize, the governor does work under the leadership of the secretary since the governor also serves as the deputy secretary. Thus, in my model, I assume that the secretary retains all bargaining power with regard to a contract, but it is important to note that the governor has some discretion in economic policymaking. This is because the secretary does not have access to economic departments directly, which earns the governor an information advantage. The point is particularly clear when we compare the party-state structure in China with its counterpart in the Soviet Union. The Soviet secretary directly controlled many economic departments within the party committee, which significantly undermined the governor's information advantage. Thus, a governor in the Soviet Union had much less discretion in policymaking. In the benchmark model, I am going to ignore the governor as an active player, as in the Soviet Union, the secretary directly managed the economy.

The secretary-governor relationship in contemporary China has dual origins. It is a descendant of the Soviet secretary-governor relationship, but the Chinese governor commands more economic power. At the same time, it also originates from the relationship between the governor and lieutenant governor in the Ming and Qing Dynasties (1368 CE- 1912 CE). The prestigious historian Luo Er'gang, whose writing I cited at the beginning, summarizes their relationship as "checks between higher ranked and lower ranked". A governor in the Qing Dynasty resembles a party secretary in current China: he controlled comprehensive power and

wielded formal authority over the lieutenant governor, but the lieutenant governor directly controlled fiscal policy and civil governance. Luo Er'gang contemplates that the lieutenant governor was closely monitored by the governor, so the lieutenant governor could not use his economic power freely. The governor was the boss of a province in the Qing Dynasty, but he was still strongly constrained because his real power was divided and delegated to the lieutenant governor. Luo Er'gang further emphasizes that checks between higher ranked and lower ranked was a main principle of statecraft in Qing China, and it clearly also applied to the military and other civilian organizations. The contemporary secretary-governor relationship is an incarnation of the Qing Institution, but it operates in the framework of the Communist Party, so it is more institutionalized than its Qing ancestor.

2.4 The Benchmark Model

Setup

The benchmark model has three players, the principal (central authority), the provincial secretary, and the population. There are four stages in the benchmark game.

1. Appointment stage: The central authority chooses the competence of secretary W , which measures the economic surplus the secretary can produce. There are two ways to conceptualize the capacity of central authority to choose competence. The central authority in China usually faces a large pool of candidates. The Central Organizational Department collects and keeps extensive information about these candidates. Suppose that the central authority plans to appoint a provincial secretary. In theory, all politicians ranked as provincial officials are in the pool. The central authority has very detailed records of the candidates' backgrounds and achievements as well as many interviews with their colleagues and subordinates. Thus, the central authority can estimate candidates' competence quite precisely.

The central authority can also adjust the way to recruit politicians, thus changing the competence of the entire candidate pool. For example, China consolidated civil service exam approximately one thousand years ago, thus dramatically improving the competence of the entire politician group. Before the Economic Reform, Communist China did not institutionalize the civil service exam, which was only re-established in the 1980s. The new civil service exam again improves the competence of politicians. If the central authority reduces the intellectual barrier to become politicians (as China did during the Cultural Revolution), we can conceptualize it as reducing the competence of politicians.

The central authority also chooses the degree of decentralization k , which measures how much public good can be provided by the local authority. I will discuss this in more detail at the signaling stage.

Nature determines whether the secretary is normal or benevolent (with probability μ). Intuitively, a normal secretary cares only about his own rent, while a benevolent secretary cares intrinsically about the welfare of the population. If the secretary does not provide public good, status-quo payoff to the secretary is $Q \equiv \lambda W$, $0 < \lambda < 1$, λ being the bargaining power of the local politician vis-a-vis the central authority. Payoff to the principal is $R \equiv (1 - \lambda)W + S$ and payoff to the population is normalized to 0. S is the exogenous rent of the autocrat.

Note that the type has two dimensions. The first dimension is competence chosen by the principal and observed by everyone in the game. The second dimension is benevolence, which is private information to the secretary. *This is the only source of asymmetric information in the benchmark model.*

2. Signaling stage: The secretary decides whether to provide a public good. The population values the public good as $e(k)$, while it generates a cost of k to the secretary. The normal secretary does not value public good, and the benevolent secretary values it at $\gamma e(k)$ with $\gamma > 1$. In the benchmark model, the secretary controls the fiscal power, so he has the capacity to provide public good.

3. Mobilization stage: Nature decides whether there is a chance for collective action with probability π . If there is no chance, the normal secretary's payoff is $Q - k$, and the population obtains $e(k)$. For a benevolent secretary, the payoff is $\gamma e(k) - k + Q$, as he cares intrinsically about the population.

The secretary decides whether to launch a collective action; then, the population decides whether to join with a cost of c . If the secretary launches a collective action and the population does not join, the normal secretary will obtain $-k$, and the benevolent secretary will obtain $\gamma e(k) - k$. Note that the secretary loses status-quo payoff because the principal sacks the secretary.

4. Divide the pie: The secretary decides whether to award collective action benefit R to the population. The normal secretary obtains $R + Q + S - k$ if he captures all benefits and obtains $Q + S - k$ if he awards R to the population. The benevolent secretary obtains $R + Q + S + \gamma e(k) - k$ if he captures all benefits and obtains $Q + \gamma R + S + \gamma e(k) - k$ if he awards R to the population. In the event that a collective action succeeds, apparently a normal secretary wants to capture all benefits, while a benevolent one wants to award everything to the population. This drastic difference in *ex post* behavior provides a strong incentive for the population to learn about the benevolence of the secretary. Ideally, they only want to join collective action with a benevolent secretary.

Now I introduce three assumptions that I am going to maintain throughout the paper:

Assumption 1

$$c > \max\{\mu R, \lambda R\}$$

The assumption $c > \mu R$ is very standard. Intuitively, it guarantees that the population will not join collective action unless they update their belief on the secretary's benevolence. The population will obtain $\mu(R - c) + (1 - \mu)(-c) = \mu R - c$ if they join collective action without receiving any signal and 0 if not. Thus, $c > \mu R$ makes sure that the secretary has to send some costly signal. The assumption $c > \lambda R$ is more subtle. It implies that the central authority does not want the benevolent secretary to fully reveal his identity. I will discuss more on this in Section 4.3.

Assumption 2

$$\gamma e(k) - k > 0$$

The assumption says that public good provision always generates positive utility for the benevolent secretary. The assumption is akin to the assumption of "commitment type" in the reputation literature. Given that public good is severely under-provided in developing countries, this assumption also matches reality very well.

With Assumption 2, I can prove Lemma 2. It confirms the intuition that the benevolent secretary will always provide public good in any equilibrium.

The solution concept I am going to employ is Perfect Bayesian Equilibrium (PBE). Now I introduce the criterion to restrict off-equilibrium belief.

Assumption 3 (Off-equilibrium Belief) *Denote population's belief about the secretary's type immediately After the {launch of a collective action} as $\hat{\mu}$. Denote the population's belief about the secretary's type immediately after a chance of collective action emerges as $\bar{\mu}$.*

If $\hat{\mu}$ is off equilibrium, then restrict $\hat{\mu} = \bar{\mu}$.

Assumption 3 rules out equilibria that are trivially free of collective action because crucial beliefs are on off-equilibrium paths. Thus, they are not constrained by Bayes' Rule. This is a major problem with the analysis of dynamic games with incomplete information, and many authors have contributed to rule out equilibria that are supported by implausible beliefs (e.g., Cho, Kreps, 1982).

Specifically, it rules out the equilibrium that the benevolent secretary already provided public good and reveals his benevolence, but the population abstains from collective action because the population's belief $\hat{\mu}$ about the secretary's type after the secretary chooses

{launch} is not high enough. However, because all secretaries abstain from collective action, the event {launch} happens with probability zero. This means that any belief $\hat{\mu} \in [0, 1]$, however implausible it is, does not violate Bayes' Rule, but we will see that $\bar{\mu}$ is always well defined by Bayes Rule, because the benevolent secretary always provides public good.

The equilibrium is trivially free of collective action because the population insists on a very low belief on the secretary's benevolence, although only benevolent secretaries provide public good. It can be ruled out because the population reasons the following: only the benevolent type provides public good, so if the secretary perturbs their behavior a little bit and chooses {launch} with extremely small probability, I should know for sure that the secretary is benevolent if I saw someone leading a collective action. This is the case if the secretary "trembles hand" in the same manner regardless of his type. Note that this is closely related to the idea of Sequential Equilibrium (Kreps, Wilson, 1982). Formally, we have a Perfect Bayesian Equilibrium that only benevolent secretaries provide public good, both secretaries abstain from launching a collective action, and the population does not join any collective action. However, suppose the equilibrium strategy is perturbed by a sequence of small value $\{\epsilon^n\} > 0$ with $\lim_{n \rightarrow \infty} \epsilon^n = 0$. Then, Assumption 3 can be rationalized as:

$$\hat{\mu} = \lim_{n \rightarrow \infty} \frac{\mu(1 - \epsilon^n)\epsilon^n\pi}{\mu(1 - \epsilon^n)\epsilon^n\pi + (1 - \mu)(\epsilon^n)^2\pi} = \lim_{n \rightarrow \infty} \frac{\mu(1 - \epsilon^n)\pi}{\mu(1 - \epsilon^n)\pi + (1 - \mu)\epsilon^n\pi} = 1$$

Note that:

$$\bar{\mu} = \frac{\mu\pi}{\mu\pi + (1 - \mu) * 0 * \pi} = 1$$

So $\hat{\mu} = \bar{\mu}$. However, this means that the population should join. Thus, the above PBE does not satisfy Assumption 3.

Readers with more interest on how Assumption 3 works can consult the proofs of claim 2.1 in Proposition 2.

Characterizing Behaviors of the Secretary and Population

In this section, I am going to characterize PBEs given W and k . In next section, I am going to discuss principal's optimal choice of W^* and k^* . As an overview, we will see that for k large enough, we will have a unique pure strategy equilibrium. Otherwise, we will have a unique mixed strategy equilibrium. However, whatever equilibrium the principal wants to induce, he always faces an acute loyalty-competence tradeoff.

Let us first look at mixed strategy equilibria. I first introduce a few helpful lemmas.

Lemma 1

There is no PBE in which either normal or benevolent secretaries mix between {leading a collective action} and {not leading a collective action}.

All proofs are in the appendix. Lemma 1 allows me to simplify my analysis greatly. It rules out a major complication and induces my model to be consistent with canonical papers studying reputation effects. Next, I introduce two additional lemmas that are useful to prove the uniqueness of PBE.

Lemma 2 (Commitment of Benevolent Secretary)

In any PBE, the benevolent secretary always provides public good.

Lemma 2 formalizes the idea that the benevolent secretary is the “commitment type”: the benevolent secretary is committed to providing public good. This reduces the number of possible strategies we need to rule out for equilibrium uniqueness. I prove an additional lemma to further reduce the space of possible strategies we need to check.

Lemma 3 (Impossible Strategies)

The following strategy will not appear in any PBE: 1. {provide} and {not lead} for normal secretary; 2. population chooses {not join} while secretary chooses {launch}; 3. population chooses {join} while secretary chooses {not launch}

I analyze the mixed strategy equilibrium first:

Lemma 4

If $k < \pi(R + S)$, then only the mixed equilibrium exists, and it is unique. The normal secretary mixes between providing public good and not, and the population mixes between joining the collective action and not.

Denote that the normal secretary provides public good with probability \hat{x} ; the population chooses {join} with probability \hat{y} . We have:

$$\hat{x} = \frac{\mu}{1 - \mu} \left(\frac{R}{c} - 1 \right) = \frac{\mu}{1 - \mu} \left\{ \frac{(1 - \lambda)W}{c} - 1 \right\}$$

$$\hat{y} = \frac{Q + \frac{k}{\pi}}{Q + R + S} = \frac{\lambda W + \frac{k}{\pi}}{W + S}$$

In addition, the secretary chooses {launch} after providing public good.

From the above lemmas, I can prove:

Proposition 1 (Loyalty-competence tradeoff in Mixed-strategy Equilibrium)

If $k < \pi(R + S)$, the probability of collective action is:

$$Pr(\text{collective action}) = \mu \frac{R}{Q + R + S} \frac{\pi Q + k}{c} = \mu \frac{(1 - \lambda)W}{W + S} \frac{\pi W + k}{c}$$

Specifically, $Pr(\text{collective action})$ increases with competence W . It also increases with k , μ , π , and decreases with c .

Proposition 1 formalizes the competence-loyalty trade-off in mixed strategy equilibrium. A more competent secretary produces more economic surplus for the principal; but at the same time, he is also more likely to launch a collective action. There are two mechanisms: a more competent secretary means that the return to collective action is higher, so the secretary and the population find it more attractive to organize collective action. This is formalized by the term $\frac{(1-\lambda)W}{W+S}$. Meanwhile, a more competent secretary has more to lose if the collective action fails. As a consequence, the public good signal is more costly for the more competent secretary. This strengthens the informativeness of the signal, as the normal and competent secretary is less likely to provide public good and lead a collective action.

Next, let us turn to pure strategy equilibrium:

Proposition 2 (Loyalty-competence Tradeoff in Pure-strategy Equilibrium)

If $k \geq \pi(R + S)$, then only pure strategy equilibrium exists.

2.1 Collective action risk in equilibrium

If $R > c$, then in the unique PBE, {provide public good, not provide, lead, lead, join} is the strategy profile.

2.2 No collective action risk in equilibrium with low competence

If $R \leq c$, then in the unique PBE, {provide, not provide, not lead, not lead, not join} is the strategy profile.

Proposition 2 says that if the principal chooses a high degree of decentralization, then there are two possible outcomes. In the first one, the benevolent secretary fully reveals his identity

through public good provision, and the population joins collective action. In the second outcome, the competence of the secretary is so low that the population finds it undesirable to join, which eliminates all collective action risk. The second case of loyalty-competence tradeoff emerges here: the principal can eliminate collective action by choosing a very low competence $(1 - \lambda)WR = c$. If he chooses any competence such that $R > c$, he will face a risk of collective action with probability $\mu\pi$. Obviously, because the risk is constant at $\mu\pi$, the principal will choose $\bar{R} = (1 - \lambda)\bar{W}$ if $(1 - \mu)\bar{R} \geq c$ and $R = c$ otherwise. Thus, the principal faces either a competent but disloyal secretary or a loyal but mediocre one.

Principal's Problem

Given our analysis before, the principal's problem is:

$$\begin{aligned} \max_{W, k} & 1\{k < \pi((1 - \lambda)W + S)\} \max\left\{\left\{1 - \mu \frac{(1 - \lambda)W}{W + S} \frac{\pi\lambda W + k}{c}\right\}[(1 - \lambda)W + S], c + S\right\} \\ & + 1\{k \geq \pi((1 - \lambda)W + S)\} \max\{(1 - \mu)W + S, c + S\} \end{aligned}$$

Intuitively speaking, the principal should avoid the case where the benevolent secretary fully reveals his identity and the population gains full confidence to join collective action. This is formalized by Assumption 1, with which I can prove Proposition 3 that rules out the pure strategy equilibrium with high competence and high collective action risk:

Proposition 3 (Implementation of Mixed Equilibrium with Loyalty-competence Tradeoff)

The principal wants to implement the mixed strategy equilibrium; $k^ = 0$, and optimal competence is an interior solution.*

$k^* = 0$ shows strong incentive to centralize spending, but it looks unrealistic as some public good can only be provided by the local authority. I am going to address this issue in the next section. Thus, the principal wants to induce a mixed strategy equilibrium and centralize public spending, and still faces a strong loyalty-competence tradeoff. Under a concentration of powers, the central authority has full incentive to implement political mediocrity and expenditure centralization.

Public good as Co-optation against Mass Revolution

In the prior analysis, public good does not have intrinsic value to the principal. The principal strategically chooses locality's discretion over public good to minimize collective action risk. That is why $k^* = 0$, but it looks unrealistic as some public good has to be provided locally. In this section, I assume that the population can engage in revolt without the leadership of local politicians. I call such revolt "mass revolution". In this case, the principal has to grant some decision discretion to local politicians even under a concentration of powers. Only in this case can local politicians co-opt the population and induce them from abstaining from revolution. We will see that a regime with a concentration of powers features a high risk of collective actions from both the population and the secretary. The reason is that the leadership only grants some spending discretion to the local secretary. Otherwise, the local secretary's signaling device will be too powerful, but this means that it is more difficult to induce the local population to abstain from mass revolution. At the same time, even partial spending discretion strengthens the informativeness of the signal, which increases the risk of collective action from the secretary.

The model only changes slightly. Before the signaling stage, with probability ν , the population solves their collective action problem, and they can engage in a mass revolution. If the population succeeds in mass revolution, they will deprive all benefits enjoyed by both the principal and the secretary. The probability that the population will accept the public good from the local authority and abstain from mass revolution is $\sigma(k)$, and $\sigma(\cdot)$ has $\sigma'(\cdot) > 0$, $\sigma''(\cdot) < 0$ and $\lim_{k \rightarrow 0} \sigma'(k) \rightarrow \infty$. Everything else is the same. $\sigma'(\cdot) > 0$ and $\sigma''(\cdot) < 0$ capture the intuition that the population is more likely to abstain from mass revolution with more lavish public good.

Suppose the principal wants to implement the mixed strategy equilibrium under a concentration of powers. The principal's problem:

$$\begin{aligned} & \max_{W,k} (1 - \nu) \left\{ 1 - \mu \frac{(1 - \lambda)W}{W + S} \frac{\pi \lambda W + k}{c} \right\} [(1 - \lambda)W + S] + \nu \sigma(k) [(1 - \lambda)W + S] \\ & = \max_{W,k} \left\{ (1 - \nu) \left\{ 1 - \mu \frac{(1 - \lambda)W}{W + S} \frac{\pi \lambda W + k}{c} \right\} + \nu \sigma(k) \right\} [(1 - \lambda)W + S] \end{aligned}$$

We can prove that:

Proposition 4

With concentration of powers, the optimal spending discretion is:

$$\sigma'(k) = \frac{1 - \nu \mu}{\nu c}$$

Specifically, $k > 0$, and it increases with ν and c and decreases with μ . The optimal competence is an interior solution.

We can see that the tension between competence and loyalty is magnified if public good can be used to preempt a mass revolution. Specifically, with a positive k , it is easier for the secretary to signal his benevolence; because of this concern, the central authority cannot grant full discretion over public spending to the secretary, but this makes it difficult for the secretary to induce the population to abstain from collective action. Thus, under a concentration of powers, the principal faces risk from both the secretary and the population.

2.5 Delegation, Decentralization, and Political Meritocracy

In this section, I adopt the signaling model I analyzed extensively in section 4. I add a fourth player, the governor, who is in charge of fiscal decisions. We will see that the addition of the new player dramatically changes the equilibrium outcome. I will obtain a very sharp prediction that the central authority unambiguously pushes for decentralization and political meritocracy simultaneously.

Setup

We have four players. The principal, the provincial secretary, the population, and the governor. The timing of the game is very similar:

1. Appointment stage: The central authority chooses the competence of the secretary and the governor, who jointly produce a surplus of W .

The central authority also chooses the degree of decentralization k , which measures how much public good can be provided by the local authority.

Nature determines whether the secretary and the governor are normal or benevolent (both with probability μ and independently distributed). If the secretary does not launch a collective action, status-quo payoffs to a secretary and a governor are both $Q \equiv \lambda W$, $0 < \lambda < 1$. The payoff to the principal is $R + S \equiv (1 - 2\lambda)W + S$, and the payoff to the population is normalized to 0. S is the exogenous rent of being the autocrat.

2. Collusion stage: The secretary can make a credible promise to transfer a fraction of the collective action benefit ηR to the governor in the case that collective action succeeds. The transfer is compensation for the governor's effort to provide public good in the signaling stage. The side contract is offered by the secretary. I assume that when the secretary proposes ηR , it will always be awarded to the governor regardless of the governor's type in the event of a successful collective action. As we will see, the governor's only choice variable is whether to provide the public good; he cannot reject the proposal made by the secretary. Of course, this does not mean that the governor has no bargaining power: if the secretary does not provide sufficient compensation, the governor may just refuse to provide public good. Finally, the collusion behavior is not observable to the population. This is not a critical assumption, but it fits reality nicely.

If I do not allow collusion opportunities for local politicians, then the main proposition of the paper is trivial to prove. However, even with fully credible collusion, we will see that collective action risk still decreases dramatically.

3. Signaling stage: Instead of the secretary, it is the governor who decides whether to provide public good $e(k)$, which costs k ; the population values the public good as $e(k)$. Now the governor bears the entire cost of public good provision k . As Shirk (1992 & 1993) and Guo (2009) note, the governor is in charge of economic policy and everyday management of the administration.

4. Mobilization stage: Nature decides whether there is a chance for collective action with probability π . If there is no chance, the normal secretary's payoff is Q , the population obtains $e(k)$, and the governor obtains $Q - k$. For a benevolent secretary, the payoff is $e(k) + Q$, and for a benevolent governor, it is $e(k) - k + Q$.

The secretary decides whether to lead a collective action; the population decides whether to join with a cost of c .

If the secretary launches a collective action and the population does not join, the normal secretary will obtain 0 and the normal governor will obtain $Q - k$. The benevolent secretary will obtain $e(k)$, and the benevolent governor will obtain $Q + e(k) - k$. Note that status-quo payoffs for the secretary are lost because the principal sacks the secretary; but the principal will not sack the governor if the secretary launches a failed collective action. The assumption tracks reality very well. In the three major episodes when provincial secretaries tried to challenge the central authority (Chen Xitong, Chen Liangyu, Bo Xilai), the secretaries were severely punished, but all three governors retained their positions.

In line with Lemma 1, I am going to assume that {launch} or {not launch} must be pure strategies. The assumption is not necessary for my results. An appendix is going to address the situation where the secretary can mix between {launch} and {not launch}, but such

mixed strategy equilibrium greatly complicates my analysis and needs much discussion to clarify my results.

5. Divide the pie: Whoever leads the collective action will distribute its benefit, so the secretary will decide whether to award the benefit $(1 - \eta)R$ to the population. The normal secretary obtains $(1 - \eta)R + Q + S$ if he captures all benefits and obtains $Q + S$ if he awards $(1 - \eta)R$ to the population. The benevolent secretary obtains $(1 - \eta)R + Q + S + e(k)$ if he captures all benefits and obtains $Q + \gamma(1 - \eta)R + S + e(k)$ if he awards R to the population.

From the setup, we can see the key tension brought by delegation. The population wants to join the collective action launched by a benevolent secretary but not a normal one. However, the signaling device is in the hands of the governor rather than the secretary. Thus, public good provision is not very informative regarding the benevolence of the secretary. We will see that in equilibrium, the population learns nothing from public good provision about the secretary. This completely forestalls collective action and enables the central authority to decentralize spending and appoint the most competent local politicians.

Characterizing Behaviors of the Population and Local Politicians

This section is in parallel with section 4.2. We will see that there is a cutoff of k , above which we have pure strategy equilibrium, and below another cutoff, we have mixed strategy equilibrium. We will see that for any value of k , delegation always reduces collective action risk vis-a-vis concentration of powers, which improves political meritocracy.

Lemma 5 (Commitment of Benevolent Governor)

In any PBE, the benevolent secretary always provides public good.

Proposition 5 (No Collective Action Risk in Pure Strategy Equilibrium)

If $k \geq \pi R$, then in any pure strategy equilibrium, the benevolent governor always provides public good, while the normal governor does not. The secretary never initiates a collective action, and the central authority appoints the most competent secretary and governor.

Proposition 5 is the key result of this paper. It says that with sufficient high decentralization, delegation will completely forestall collective action. The intuition is that the normal governor finds it too costly to provide public good for signaling purposes with a high degree of decentralization. This leaves only the benevolent governor to provide public good, which reveals full information about the benevolence of the governor. However, it tells the population nothing about the benevolence of the secretary, yet the population wants to learn about the secretary rather than the governor, because it will be the secretary who leads the

collective action and distributes associated benefits. Consequently, the population refuses to join any collective action. As collective action is completely forestalled, the central authority will decentralize spending and appoint the most competent secretary and governor. This is the full force of economic delegation at work. The secretary has the means to mobilize and organize the population through extensive party organizations and propaganda apparatuses, but he cannot credibly communicate with the population because economic power lies with the governor.

Lemma 6

1. If $k \leq \frac{\pi Q[R-(2-\mu)c]}{Q+S+(2-\mu)c}$, then there are two equilibrium outcomes.

1.1 A mixed equilibrium exists where the secretary mixes between whether to propose public good, and the population mixes between whether to join the collective action. The normal governor provides public good only if the secretary promises a transfer, and the benevolent governor always provides public good.

Denote that the normal secretary provides public good with probability x ; the population chooses {join} with probability y . We have:

$$\hat{x} = \frac{\mu}{(1-\mu)^2} \left(\frac{(1-\eta^*)R}{c} - 2 + \mu \right) = \frac{\mu}{1-\mu} \left\{ \frac{(1-\lambda)(1-\eta^*)W(a)}{c} - 1 \right\}$$

$$\hat{y} = \frac{Q}{Q + (1-\eta^*)R + S} = \frac{\lambda W(a)}{[\lambda + (1-\hat{\eta})(1-\lambda)]W(a) + S}$$

$\hat{\eta}$ satisfies that:

$$\hat{\eta}R = \frac{k(Q + R + S)}{\pi Q + k}$$

Specifically, $\hat{\eta}R$ increases with k .

In addition, the secretary chooses {launch} after the governor provides public good.

1.2 A pure strategy equilibrium exists, where both normal and benevolent secretaries do not propose, only the benevolent governor provides public good, and the population does not join collective action.

2. With $\frac{\pi Q[R-(2-\mu)c]}{Q+S+(2-\mu)c} < k < \pi R$, only the above pure strategy equilibrium in Proposition 5 exists.

I am going to focus on the risky equilibrium. From Lemma 6, we can prove Proposition 6:

Proposition 6 (Reduced Collective Action Risk in Mixed Strategy Equilibrium)

If $k \leq \frac{\pi Q[R-(2-\mu)c]}{Q+S+(2-\mu)c}$, the probability of collective action is:

$$Pr(\text{collective action}) = \mu \frac{(1-\hat{\eta})R}{c} \frac{\pi Q}{Q+(1-\hat{\eta})R+S} = \mu \frac{(1-\hat{\eta})(1-2\lambda)W}{(\lambda+(1-\hat{\eta})(1-2\lambda))W+S} \frac{\pi\lambda W}{c}$$

Specifically, with the same level of competence, the probability of collective action under delegation is lower than that under concentration of powers:

$$\begin{aligned} Pr(\text{collective action}|\text{delegation}) &= \mu \frac{(1-\hat{\eta})(1-2\lambda)W}{(\lambda+(1-\hat{\eta})(1-2\lambda))W+S} \frac{\pi\lambda W}{c} \\ &< \mu \frac{(1-\lambda)W}{W+S} \frac{\pi\lambda W+k}{c} = Pr(\text{collective action}|\text{concentration}) \end{aligned}$$

In addition, $Pr(\text{collective action})$ is a decreasing function of k .

Thus, the result is robust even if the principal does not have the flexibility to choose k . There are two channels. First, the governor has to be compensated by a fraction of the benefit, which lowers the return from collective action for the secretary and the population.

Second, k “disappears” in $Pr(\text{collective action}|\text{delegation})$ (of course, it does not really disappear as the optimal threshold $\hat{\eta}$ still depends on k). Specifically, it appears in the term $\frac{\pi\lambda W+k}{c}$ without delegation, but the term becomes $\frac{\pi\lambda W}{c}$ under delegation. This is where delegation is at work again. Public good provision is an informative signal without delegation because it is costly for the secretary to provide. Under delegation, the governor controls public good and henceforth bears the full cost of public good provision. This makes public good very cheap for the secretary, as he does not generate any efforts to provide it. However, we know that a cheap signal is not informative, which makes the population much less convinced about the benevolence of the secretary. Consequently, the population is less willing to join collective action.

Principal’s Problem under Delegation

The principal’s problem is:

$$\max_{W,k} 1\{k \leq \frac{\pi Q[R-(2-\mu)c]}{Q+S+(2-\mu)c}\} \max\left\{1 - \mu \frac{(1-\lambda)(1-\eta^*)W}{(\lambda+(1-\eta^*)(1-2\lambda))W+S} \frac{\pi\lambda W}{c}\right\}$$

$$[(1 - 2\lambda)W(a) + S], c\} + 1\{k \geq \frac{\pi Q[R - (2 - \mu)c]}{Q + S + (2 - \mu)c}\}([1 - 2\lambda)W + S]$$

The solution is quite straightforward. The principal cannot do better than choosing $k^* \geq \frac{\pi Q[R - (2 - \mu)c]}{Q + S + (2 - \mu)c}$, and $W = \bar{W}$. Thus, he will always implement the pure-strategy equilibrium. The normal governor finds it too costly to provide public good, so only the benevolent governor provides public good. Consequently, public good provision tells the population nothing about the secretary, and all collective action will fail. In this environment, the central authority feels very confident in appointing the most competent local politicians. This is the key result of the paper. Strong decentralization catalyzes the effect of delegation, which enables the central authority to promote full political meritocracy.

Note that the meritocracy still prevails even if there is not too much flexibility on k . Suppose that k is restricted to a specific range or even a singleton, which may be the case if the benevolent secretary has a bliss point of k . Then, by Proposition 6, delegation still strongly supports political meritocracy *vis-a-vis* no delegation. In addition, note that decentralization has a different effect on meritocracy under different political regimes: under a concentration of powers, it strengthens the informativeness of signal and undermines political meritocracy. Under delegation, decentralization helps prevent collusion between the governor and the secretary and thus promotes meritocracy.

2.6 Weak Delegation: The Secretary's Complete Dominance over the Governor

In this section, I assume that the secretary can issue orders to the governor without any compensation, and the governor has to follow whatever the secretary directs. This is the extreme form of the local party-state relationship in which the governor has the weakest possible strength *vis-a-vis* the secretary. Even with such a weak governor, collective action risk will be reduced substantially, and central authority will push for meritocracy. As the cost of public good provision is still borne by the governor, the signal remains uninformative. It forestalls effective communication between the secretary and the population.

The setup is very similar to that in section 5.1. The difference here is that the governor has to do whatever the secretary directs. The secretary does not need to compensate the governor for his effort. Thus, the collusion stage disappears, as the secretary and governor behave almost as a single agent. The difference from the single agent case is that the governor still bears the cost of public good provision.

Lemma 7

Mixed equilibrium exists where the secretary mixes between whether to propose public good and the population mixes between whether to join.

Denote that the normal secretary chooses {provide public good} with probability \hat{x}^s ; the population chooses {join} with probability \hat{y} . We have:

$$\hat{x}^s = \frac{\mu}{1-\mu} \left\{ \frac{R}{c} - 1 \right\} = \frac{\mu}{1-\mu} \left\{ \frac{(1-2\lambda)W}{c} - 1 \right\}$$

$$\hat{y} = \frac{Q}{Q+R+S} = \frac{\lambda W}{(1-\lambda)W+S}$$

In addition, the secretary chooses {launch} after the governor provides public good.

The proof is almost the same as Lemma 2 and thus is omitted.

Proposition 7 (Reduced Risk in Mixed Strategy Equilibrium)

The probability of collective action is:

$$Pr(\text{collective action}) = \mu \frac{R}{c} \frac{\pi Q}{Q+R+S} = \mu \frac{(1-2\lambda)W}{(1-\lambda)W+S} \frac{\pi \lambda W}{c}$$

Specifically, with the same level of competence, the probability of collective action under weak delegation is lower than that under no delegation:

$$\begin{aligned} Pr(\text{collective action}|\text{weak delegation}) &= \mu \frac{(1-2\lambda)W}{(1-\lambda)W(a)+S} \frac{\pi \lambda W}{c} \\ &< \mu \frac{(1-\lambda)W(a)}{W(a)+S} \frac{\pi \lambda W(a)+k}{c} = Pr(\text{collective action}|\text{no delegation}) \end{aligned}$$

The mechanical channel is that the governor also captures a fraction of the economic pie, so the collective action benefit is reduced. The more important channel is that k does disappear in the formula under weak delegation. Thus, our original intuition survives even with such a weak governor: the cost of public good provision falls on the governor, which means that the signal is very uninformative for the secretary.

Different from prior cases, there is another equilibrium featuring no collective action:

Proposition 8 (Public Good Provision but No Collective Action)

There is a (family of) PBE in which the benevolent secretary provides public good and the normal secretary provides with probability $\hat{x}^s > \frac{\mu}{1-\mu}(\frac{R}{c} - 1)$. Neither of them initiates any collective action. The population does not join any collective action.

We have another equilibrium where both normal and benevolent secretaries provide public good, and collective action is completely forestalled. Thus, in this section with the weak governor, we encounter multiple equilibria: the mixed strategy equilibrium with (reduced) collective action risk and the pure strategy equilibrium without any risk. However, whatever equilibrium arises, the regime is more stable, meritocratic, and decentralized than the case with no delegation.

2.7 Dynamic Concern and Ambiguity about the Party-Government Relationship

Central authority in China exhibits a reluctance to clarify the party-government relationship. Although a Chinese governor is unambiguously much stronger than a Soviet one, it is unclear whether the Chinese secretary dominates over the governor. Judging from our analysis before, it is never optimal to allow the secretary's dominance over the governor. There are two possible equilibrium outcomes, with one of them featuring a positive risk of collective action. By comparison, a power balance between the secretary and the governor has a single equilibrium with no collective action at all, so why does the central authority maintain such ambiguity? Why does the central authority not clearly separate economic and political powers?

The reason is quite clear if we take a dynamic view. Until now, my model is essentially static. There is only one period; after serving one term, both politicians retire. However, if they serve more than one term, then there is an obvious problem with the power balance between the secretary and the governor. Only the benevolent governor provides public good in equilibrium, which fully reveals her identity. If the governor becomes the new secretary in the next period, he can easily persuade the population and challenge the central authority.

³

However, if the secretary completely dominates over the governor, then the population learns nothing about the governor in equilibrium. Thus, the concern about future collective action fully vanishes. Note that with complete dominance, the central authority may face a positive probability of collective action today (Proposition 7).

³One may suggest rotation as a solution. Rotation may partially alleviate the problem. However, also note that the reputation of local politicians can easily travel across provinces.

The central authority can do better by creating uncertainty toward the party-government relationship. Formally, suppose we have two periods. The governor in the first period will be promoted as secretary in the second period and plays the same game again. The governor in the second period is a new player. Then the game ends. Suppose in both periods that the secretary dominates over the governor with probability ξ , and there is a power balance with probability $1 - \xi$. This is the new policy instrument available to the central authority to achieve a better outcome.

We can assume that the principal imposes a very small punishment $\epsilon > 0$ on a secretary who does not provide public good. In this case, a strong secretary will always provide public good with probability 1, regardless of his benevolence. Then, when the population observes a governor providing public good, they reason that he is benevolent with probability:

$$\frac{(1 - \xi)\mu + \xi\mu}{(1 - \xi)\mu + \xi} = \frac{\mu}{\xi + (1 - \xi)\mu}$$

For $\xi > \frac{\mu}{1-\mu}(\frac{R}{c} - 1)$, we will have:

$$U(\text{population}|\text{join}) = \frac{\mu}{\xi + (1 - \xi)\mu}[R+e-c] + (1 - \frac{\mu}{\xi + (1 - \xi)\mu})[e-c] < e = U(\text{population}|\text{not})$$

Thus, the population will refuse to join collective action.

For people who are familiar with the Bo Xilai Incident (articulated later), the model says that the population will not join any collective action launched by Huang Qifan, the governor of Chongqing, even if he will be promoted as the provincial secretary. The rationale is quite simple: Huang did provide lavish public good during Bo Xilai's reign, but it probably has nothing to do with his own benevolence. More likely, he was forced by Bo to provide it. This means that the public good provision contains no information. Thus, even in a dynamic setting, the central authority feels safe. The analysis provides a simple yet powerful reasoning behind the seemingly bizarre design to allow the secretary to dominate over the governor. By doing so, not only is potential collective action launched by the secretary forestalled but also one launched by the governor in the future when he becomes the secretary.

However, ξ cannot be too large if the central authority worries about the equilibrium with a positive risk of collective action when the secretary dominates over the governor. I am going to characterize the value of ξ that forestalls collective action when the secretary dominates over the governor.

Proposition 9

If $\xi < \frac{c-\mu R}{R(1-\mu)}$, then the equilibrium featuring a positive risk of collective action when the secretary dominates over the governor can be eliminated.

Corollary 1

If $c > \sqrt{\mu}R$, then for

$$\xi^* \in \left(\frac{\mu}{1-\mu} \left(\frac{R}{c} - 1 \right), \frac{c - \mu R}{R(1-\mu)} \right)$$

the population will refuse to join the collective action of either the current secretary or future secretary who is currently governor. Given this, only the benevolent and strong secretary and the benevolent and strong governor will provide public good, and all secretaries will abstain from collective action.

$c > \sqrt{\mu}R$ is slightly stronger than Assumption 1, which states that $c > \mu R$. But if $c > \sqrt{\mu}R$ is true, then the central authority can manipulate the belief of the population regarding the strength of the governor, and the result is that both current and future collective actions will be forestalled. Thus, the first one achieves best for the central authority when the secretary dominates over the governor with probability $\xi^* \in \left(\frac{\mu}{1-\mu} \left(\frac{R}{c} - 1 \right), \frac{c - \mu R}{R(1-\mu)} \right)$. Thus, the central authority constantly emphasizes the leadership of the party over the government; but at the same time, it also stresses that the power of “first hand” should be divided and constrained (e.g., People’s Daily, 2014).

The analysis echoes research on strategic media censorship in China (Lorentzen, 2014). When the economy is doing badly, the propaganda department allows little media freedom. When the economy is doing fine, censorship is loosened and media reporting is relatively free. By doing this, the population always receives a constant signal about the state of the world, so they cannot infer whether governance is bad and thus whether it is a good opportunity to revolt. My analysis shares a similar flavor. By creating uncertainty regarding the strength of the governor, the central authority “confuses” the population and induces the population to abstain from collective actions launched by either the current secretary or future secretary.

Such stochastic dominance of the secretary over the governor can function very well, but it still has a huge cost. The dual leadership is actually a good framework to design checks and balances that will help fight corruption and rent-seeking (Persson, Tabellini, Roland, 2000). The CCP itself recognizes the point. However, when the secretary gains dominance over the governor, such power duality has obviously no bite in fighting corruption.

2.8 Case Studies of Modern Chinese Politics

At the heart of my theory is that the population shows strong loyalty and support for local politicians who engage with aggressive redistribution or public good provision. This is

actually one of the most important themes of Chinese politics. Ever since Imperial China, leaders of many rebellions redistributed land to peasants, which has helped leaders gain strong loyalty from peasants. Many dynasties declined or fell as a result of such rebellions. The CCP is the modern master of such tactics. The CCP systematically enforced “land reform” in its revolutionary bases during the second and third Chinese Civil Wars (1927-1937; 1945-1949), winning strong support from peasants, which sealed the victory of the Chinese Communists (Pepper, 1999). In Mao (1945), Mao Zedong himself stated that “*Our party must bring tangible benefits to the people. Only then will the masses support us and oppose the Kuomintang (the party that ruled the Republic of China from 1927 to 1949) attacks. Otherwise, the masses will be unable to see clearly which of the two parties, the Kuomintang or the Communist Party, is good and which is bad*” (translated and cited by Pepper, 1999). Here, Mao explicitly identified redistribution and public good provision as a signaling device.

Quite recently, the former Chongqing secretary Bo Xilai reincarnated the old tactic. An excellent survey article is Zhao (2012). During Bo’s reign of the province (2007-2012), he advocated the “Chongqing Model”, which emphasized social and economic equality. First, the administration implemented phenomenal public good provision. “Chongqing spent more than half of all government expenditures on improving public welfare, particularly the livelihood of workers and farmers (Zhao, 2012)”. Cheap public housing was extensively built, a large number of city “Hukou” were granted to farmers, and the whole bureaucratic system was mobilized to meet the needs of grass-root residents. Moreover, the intense propaganda of Maoist and socialist values complemented public good provision. As a consequence, the Bo Xilai administration enjoyed enormous popularity, in particular among low-income households. The Bo Xilai Saga still had many details that remained unclear, such as whether Bo did intend to challenge the central leadership, but such rumors are widespread, and it is accepted that Bo at least aspired to rely on his remarkable achievement in Chongqing to enter the Standing Committee of the CCP Politburo. This is the supreme decision-making group in China. However, during the months when the police chief of Chongqing defected and the Bo Xilai Incident erupted, the popularity of Bo’s policy did not transform into real support for Bo from the population. There was no public rally or protest to support Bo, who was eventually taken into custody and sent to prison. The saga revealed that disloyalty of provincial politicians is still a major concern for central leaders in China, but at the same time, even an extremely artful politician such as Bo was constrained by the system to transform popularity to mass support. This is quite consistent with my theory in section 7. It is well known that Bo Xilai was an unusually powerful and ambitious politician with a princeling background, and the governor of Chongqing might have been completely dominated over by Bo, but the strength of the Chinese party-state structure is that even with such a weak governor, there is hope to fully forestall collective action for the central authority. In the end, the Bo Xilai administration provided public good aggressively, but the population remained immobilized in Bo’s endeavor to challenge central authority.

2.9 Case Studies from Imperial China

An Overview

Separation of powers is not a new feature of Chinese politics. Chinese rulers long recognized the key role of “divide and rule” in stabilizing their regimes. The early maturity of the Chinese polity means that we have rich empirical variation over many centuries to check the validity of the theory that links meritocracy and dual leadership. I am going to give a very brief historical review of stability, meritocracy, and separation of powers here. Most of the narratives here are synthesized from Bai (1996) and Yang (1996).

The Chinese Empire was always very decentralized. It was extremely difficult for the imperial court to communicate with local jurisdictions, so the majority of decisions had to be made by local politicians. In the language of my model, k is constrained to be a large number. The model predicts low meritocracy and stability under a concentration of powers and high meritocracy and stability under a separation of powers. This was indeed the case. The Han Dynasty (202 BCE - 220 CE) experimented with a local separation of powers by appointing censors to monitor provinces. The separation of powers was very primitive: there was no division between political and economic powers, and censors had to monitor several provinces simultaneously. It was able to support only a very weak form of meritocracy (“*Chaju*”), where the recommendation from local nobility and politicians was much more important than performance in the exam. The system collapsed after the Huangjin Rebellion (184 CE), when the emperor awarded all important powers to censors so that they had the capacity to suppress the rebellion, but the decision destroyed the local separation of powers, and China entered into almost four centuries of fragmentation.

During the fragmentation period (184 CE - 589 CE), the emperors did not have enough power to enforce local separation of powers. The emperors did try to establish and enforce separation of powers in the imperial court, but the system did not support any form of meritocracy. Indeed, the fragmentation period was the golden age of the aristocracy in China, when the family of a candidate determined the official position assigned to him. The Sui and Tang Dynasties (589 CE - 907 CE) firmly established separation of powers in the imperial court and at the same time, began to experiment with meritocracy in the form of the civil service exam. However, family origins still played a far more important role in political selection than the exam.

After a brief period of extreme chaos and fragmentation, the Song Dynasty (960 CE - 1276 CE) arose and enforced an elaborate system of checks and balances. In every prefecture, one or two monitoring officials were assigned to work with the governor. Every official document needed the joint signatures of the governor and monitoring officials to be effective. In the provinces, there were governors in charge of fiscal resources, judicial power, and military power. This is the first form of separation of political and economic powers, and the separation was deliberately established to constrain the fiscal governor. Consequently,

local politicians were extremely loyal to the emperor in the Song Dynasty, and the emperor in turn selected almost all politicians based on their merits. The civil service exam entered its golden age, when hundreds of degrees were awarded for every exam. The aristocracy completely vanished from politics, and all important positions were held by politicians with a degree from the civil service exam.

The Yuan Dynasty (1271 CE - 1368 CE) was established by Mongols. The dynasty defied Chinese institutions, and both the separation of powers and meritocracy were very weak. The Ming Dynasty (1368AD-1644AD) re-established the local separation of powers. In the provinces, three governors were in charge of the economy, the judicial court, and the military. The Ming Dynasty also revived the civil service exam as the main vehicle of political selection. The Qing Dynasty (1644 CE-1911 CE) inherited the institutions of the Ming Dynasty and enforced an even more complex network of provincial-level governors with a very high level of competence (Guy, 2010).

Data

To systematically document the empirical relationship between meritocracy and dual leadership in Imperial China, I construct time-series proxies for them from historical records. The proxy of dual leadership is constructed from “*The Twenty-five Histories*”, the official historical record of Imperial China and the single most important source to study Chinese history. I first identify keywords indicating dual leadership. These keywords are non-military political positions with a strong check against the main executive. However, they rank lower than the main executive. They also need to have permanent staff, offices, and residence in the jurisdiction.

I formulate two indices of dual leadership to cross-check each other. The first index includes official positions that satisfy all of the above criteria. They are “*Tongpan*” (surveillant against prefecture mayor), “*Buzheng*” (lieutenant governor or his office), and “*Ancha*” (official in charge of monitoring and judicial affairs in a province or his office). In the second index, I include all official positions that have some flavor of power fragmentation. In addition to the above positions, I add “*Tidian-xingyu*”, “*Tiju-changping*”, “*Xunfu*”, and “*Anfu*”. I construct the second index as a robustness check. The time coverage is from 610 CE to 1910 CE.

Then, I wrote a computer program that matches searched keywords to a year in three steps. The program first locates the keyword in the text. Starting from the keyword, the program searches forward for texts that contain year information. Finally, the program matches the keyword to the year closest to the keyword.

The index of dual leadership is defined as:

$$Dual\ leadership_t = \frac{\#of\ keywords\ in\ decade\ t\ mentioned\ in\ "25\ Histories"}{\#of\ years\ of\ decade\ t\ mentioned\ in\ "25\ Histories"} * 1000$$

Table 2.1: An Example of Normalization

	# duality keywords	# that the decade is mentioned	duality index
decade 1613-1622	8	253	31.6
decade 1636-1645	100	3997	25.0

Basically, I record the frequency of keywords for each decade t . For the first index, I look at the frequency of “*Tongpan*”, “*Buzheng*”, and “*Ancha*”. For the second index, I add “*Tidian-xingyu*”, “*Tiju-changping*”, “*Xunfu*”, and “*Anfu*”. Both indices are normalized by how frequently decade t is mentioned in historical records. Some historical periods are extensively recorded, while others are sparsely mentioned. If I look at the raw frequency of keywords, I overestimate the importance of dual leadership in the historical period covered widely by historical records.

Table 1 is an example to illustrate the normalization. I list two decades and the corresponding statistics. For 1613 CE - 1622 CE, I find only 8 keywords that can serve as a proxy for dual leadership, while I find 100 for 1636 CE - 1645 CE. If I compare the two numbers naively, dual leadership is very strong for the second decade. However, 1636 CE -1645 CE is a pivotal decade in Chinese history with a major dynastic turnover. It obtained far more extensive coverage from historical records than 1613 CE - 1622 CE. As the second column of the table shows, historical records mention the first decade 253 times, but for the second decade, it is 3,997 times. After normalization, the third column shows that the intensity of dual leadership is similar for the two decades. This is consistent with qualitative historical studies that document no structural break between the two decades for dual leadership.

I employ the same methods to construct an index of rebellion. For each decade, the computer program finds the number of “*pan*” and “*luan*” mentioned in historical records (the two words in Chinese that can mean rebellion). I also obtain time-series data from Bai, Kung (2011) about Sino-nomadic conflicts and climate change in Chinese history.

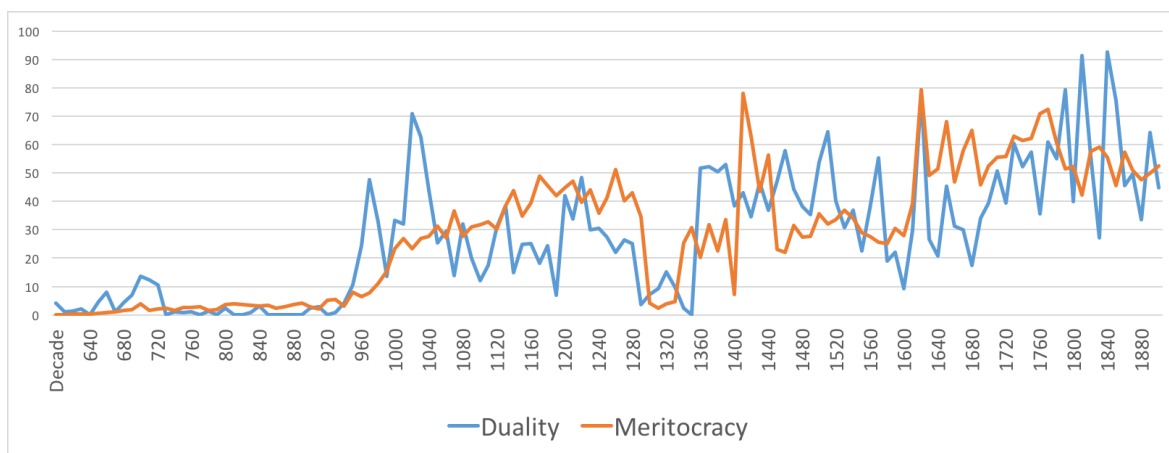
I construct a meritocracy index from another source, the eminent China Biographical Database. The database includes a vast number of biographies for historical figures. I can identify whether a particular historical figure obtained a degree from the civil service exam. This is a good indicator of competence for senior officials since a significant fraction of them hold civil exam degrees. Specifically, I define the index as:

$$meritocracy_t = \frac{\# \text{ of persons with degrees from civil exam in decade } t}{\# \text{ persons in China Biographical Database in decade } t} * 100$$

Graphic Evidence and Statistical Analysis

Figures 1 and 2 visualize the relationship of meritocracy and dual leadership. Figures 1 and 2 depicts the first and second index of dual leadership, respectively. We can see a

Figure 2.1: Meritocracy and Dual Local Leadership



strong correlation of meritocracy and dual leadership over 1,300 years of Chinese history. In addition, note that dual leadership rose and fell before meritocracy, implying Granger causality between the two time-series variables. The major events documented by the time series proxies are consistent with qualitative studies from historians.

Figures 3, 4, and 5 summarize how meritocracy, dual leadership, and rebellions interact with each other. We can observe a prominent negative correlation between rebellions and dual leadership.

To summarize the data quantitatively, I implement some basic statistical exercises. I endeavor no attempt to establish causality. The statistical analysis aims at a succinct summary of the empirical patterns in the time-series data. Another aspiration is to show the first-order importance of the political institutions I have analyzed in my theoretical model.

Specifically, I want to document a long-run correlation between these time-series proxies for meritocracy and dual leadership, if any. The econometric tool for long-run correlation is cointegration analysis (Engle, Granger, 1987), which proceeds in three steps. The first step tests whether time-series variables have unit roots through the Dickey-Fuller test. Intuitively, the unit root test detects whether a time-series variable is highly persistent. Only variables with a unit root can possibly have a long-run correlation with other variables.

I find that proxies for meritocracy and dual leadership have a unit root, but not so for the rebellion proxy. This is consistent with the intuition that meritocracy and dual leadership should be highly persistent as political institutions. The frequency of rebellions, however, is not a “stock” variable and should not be persistent.

Figure 2.2: with Alternative Proxy for Dual Leadership

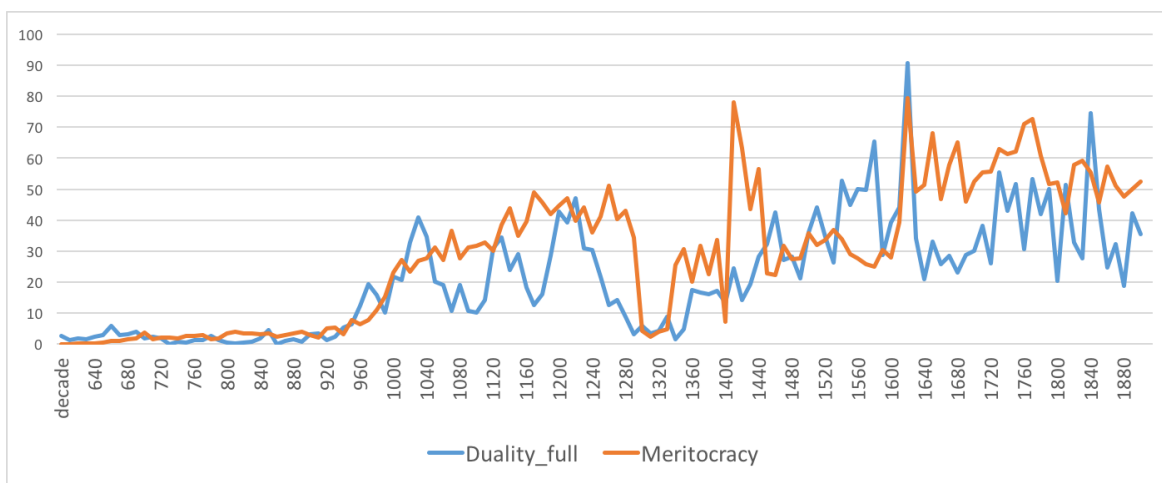


Figure 2.3: Rebellion, Dual Leadership, and Meritocracy

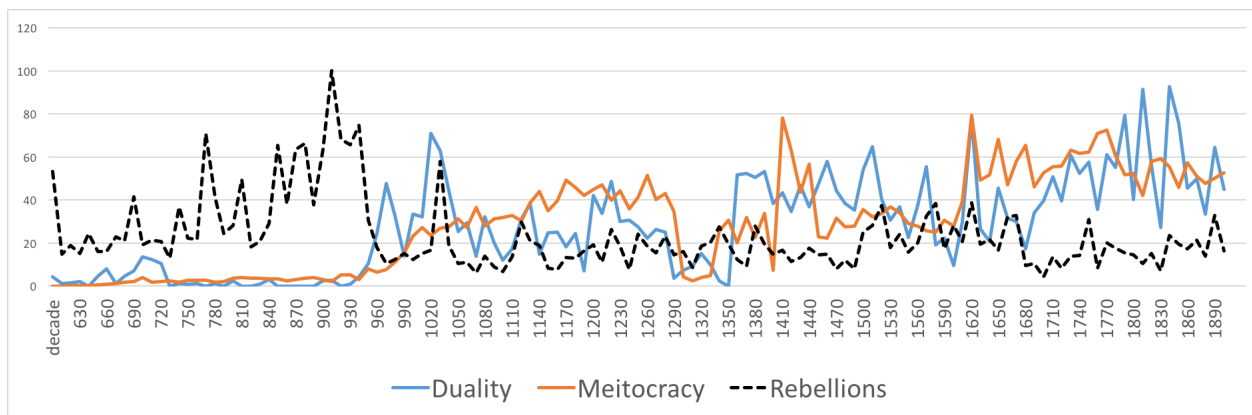


Table 2.2: The Dickey-Fuller Test of Cointegration

Dickey-Fuller Statistic	lag 0	lag 1	lag 2	lag 3
first index of dual leadership	-7.583	-5.758	-4.639	-4.174
second index of dual leadership	-5.861	-4.196	-3.568	-3.567

Critical value of the adjusted Dickey-Fuller test 10%: -3.03, 5%: -3.37, 1%: -4.07

The next step is the Dickey-Fuller Test of Cointegration between meritocracy and dual leadership. Cointegration analysis detects the long-run correlation between two time-series variables with unit roots. Denote Y_t as a proxy for meritocracy and X_t as a proxy for dual leadership. First, I estimate $Y_t = \alpha + \theta X_t + z_t$. Then, I run an adjusted Dickey-Fuller test

Figure 2.4: Rebellion and Dual Leadership

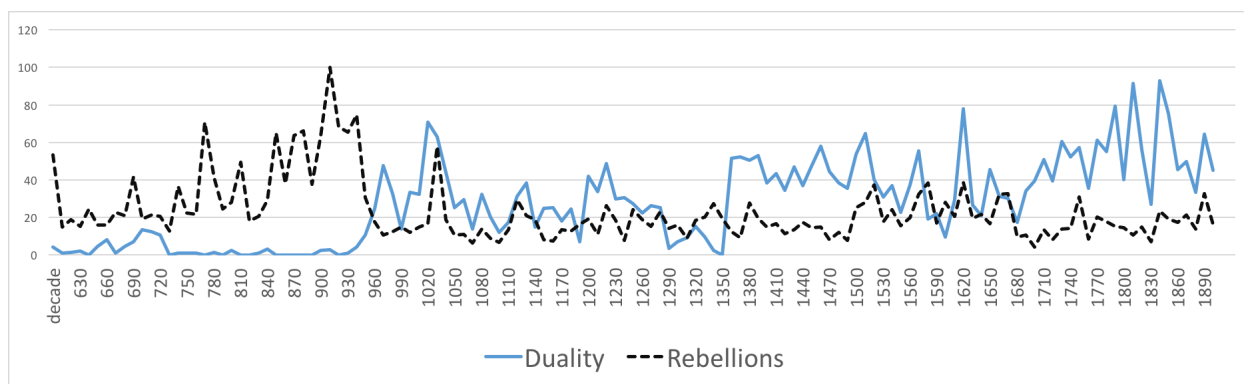
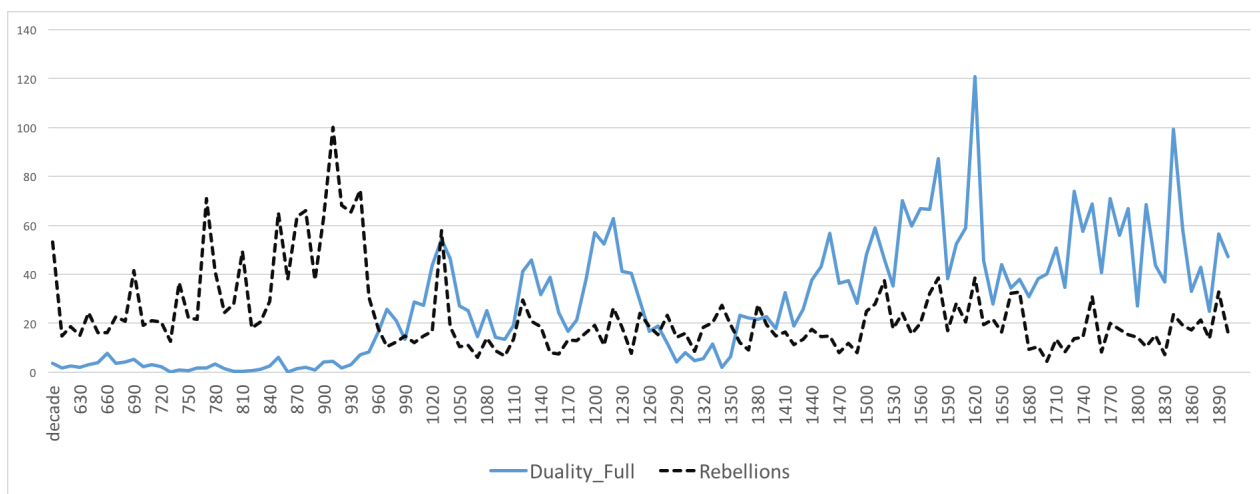


Figure 2.5: Rebellion and Dual Leadership with Alternative Measure



on \hat{z}_t , the residual from fitting $Y_t = \alpha + \theta X_t + z_t$. If \hat{z}_t does not have a unit root, it implies cointegration between Y_t and X_t .

Table 2 lists the Dickey-Fuller statistic for the two indices of dual leadership. I also consider different lags of \hat{z}_t for the Dickey-Fuller Test. The null hypothesis is that z_t is a random walk (so it has a unit root). The critical value of the adjusted Dickey-Fuller test is also listed. The Dickey-Fuller test is one-sided, so if the statistic is smaller than the critical value, the null hypothesis is rejected. We can see that in most cases, the null hypothesis is rejected. This implies that z_t is stationary, so the correlation between Y_t and X_t is not spurious. Statistical evidence supports the conjecture that meritocracy and dual leadership are correlated in the long run.

In the last step, we try to understand Granger causality in cointegration. Suppose that a shock to Y_t is followed by a change in X_{t+1} . Then, Y Granger causes X . In other

words, stronger dual leadership means that political selection in the future becomes more meritocratic. It is important to note that Granger causality is distinct from causality in the Rubin causal model. Granger causality is a purely statistical exercise that documents the dynamic pattern of two cointegrated variables. Nevertheless, Granger causality does suggest the possible direction of causality between cointegrated variables.

The Vector Error Correction Model (VECM) is the statistical tool to document Granger causality for cointegrated variables. It is an extension of the Vector Autoregression Model (VAR), which applies to time-series variables with unit roots. I run the following specification:

$$\Delta Y_t = \beta_{10} + \beta_{11}\Delta Y_{t-1} + \dots + \beta_{1p}\Delta Y_{t-p} + \gamma_{11}\Delta X_{t-1} + \dots + \gamma_{1p}\Delta X_{t-p} + \alpha_1 z_{t-1}^{\wedge} + u_{1t}$$

$$\Delta X_t = \beta_{20} + \beta_{21}\Delta Y_{t-1} + \dots + \beta_{2p}\Delta Y_{t-p} + \gamma_{21}\Delta X_{t-1} + \dots + \gamma_{2p}\Delta X_{t-p} + \alpha_2 z_{t-1}^{\wedge} + u_{2t}$$

Thus, for both Y_t and X_t , we take the first difference and include up to p lags of both variables in the specification. Moreover, the lagged residual from the last period z_{t-1}^{\wedge} also appears on the right-hand side of the specification.

α_1 and α_2 are the key coefficients in the VECM model. They quantify the direction of Granger causality. Assume that in the data-generating process, $\alpha_1 < 0$ and $\alpha_2 = 0$. This implies that X Granger causes Y . To see why, suppose we start from an equilibrium relationship in the last period so that $z_{t-1} = 0$. Now a shock hits the system so that X_t increases. As $z_t = Y_t - \theta X_t < 0$, the system deviates from equilibrium. As $\alpha_1 < 0$, $\Delta Y_{t+1} \approx \alpha_1 z_t > 0$. Thus, Y_{t+1} increases, but because $\alpha_2 = 0$, $\Delta X_{t+1} \approx \alpha_2 z_t = 0$. In a word, an increase in X_t is accompanied by an increase in Y_{t+1} but no change in X_{t+1} .

Exactly the same argument goes if Y_t decreases. In this case, $z_t = Y_t - \theta X_t < 0$ so that Y_t increases but there is no change in X_{t+1} . Intuitively, Y moves to restore the equilibrium relationship when there is a shock to the system, suggesting (but far from proving) a causal chain from X to Y .

If $\alpha_1 = 0$ and $\alpha_2 < 0$, similar argument concludes that Y_t Granger causes X_t .

The regression results are listed in Tables 3 and 4 for the two indices of dual leadership. The coefficients listed are α . α_1 is the coefficient on lag z for the equation where the first difference in meritocracy (D.meritocracy in the tables) is the dependent variable. α_2 is the coefficient on lag z for the equation where the first difference in dual leadership (D.duality in the tables) is the dependent variable. Stability denotes whether the system converges

Table 2.3: Meritocracy and the First Measure of Dual Leadership

	(1)	(2)	(3)
	Johansen's VECM	VECM	VECM
D. meritocracy			
Lag z	-0.0962 (0.04677)	-0.256** (0.0719)	-0.287** (0.0810)
D. duality			
Lag z	0.283*** (0.06787)	0.216 (0.114)	0.164 (0.111)
<i>Controls</i>	No	No	Yes
<i>AIC</i>	2021.9	1989.0	1866.3
<i>Lags</i>	3	4	4
<i>stability</i>	stable	stable	stable
<i>N</i>	128	128	118

Table 2.4: Meritocracy and the Second Measure of Dual Leadership

	(1)	(2)	(3)
	Johansen's VECM	VECM	VECM
D. meritocracy			
Lag z	-0.350*** (0.0846)	-0.329*** (0.0792)	-0.348*** (0.0810)
D. duality			
Lag z	0.0323 (0.0338)	0.0395 (0.0314)	0.0135 (0.0269)
<i>Controls</i>	No	No	Yes
<i>AIC</i>	1653.8	1653.8	1535.8
<i>Lags</i>	3	4	4
<i>stability</i>	stable	stable	stable
<i>N</i>	128	128	118

In both Tables 2.3 and 2.4, standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ (compared to the adjusted Dickey-Fuller critical values). Lags are selected based on AIC. All specifications include time trends.

to an exogenous shock, given the estimated coefficients. Column (1) applies Johansen's VECM procedure (Johansen, 1991), a standard package in VECM analysis. A drawback is that Johansen's procedure does not allow for inclusion of control variables, so I run VECM manually in Column (2) and in Column (3), add control variables such as climate, Sino-nomadic conflicts, and monitoring officials without residence in the local jurisdiction. The lags are chosen by Akaike's Information Criterion (AIC). All columns actually have the same number of lags because Column (1) lists the number of lags at level, while Columns (2) and (3) list the number of lags in first difference.

We can see that the regression results strongly favor that dual leadership Granger causes meritocracy while meritocracy does not Granger cause dual leadership. This is a statistical summary of what we can intuitively detect from Figures 1 and 2 that dual leadership leads the movement of meritocracy. For many centuries, the Chinese polity established strong meritocracy only after the consolidation of dual leadership.

2.10 Conclusion

We have seen that the power duality between the secretary and the governor serves as the institutional foundation of China's stability, decentralization, and political meritocracy. The unique party-state structure is a modern incarnation of Imperial China's institution mixed with Leninism. After Maoist China experimented with the strict form of Leninism in which the party directly controlled economic power, China has delegated economic power to the governor, which achieves a subtle power balance in local politics. This is in line with the literature from political science (e.g., Tsai, (2007)), which argues that China usually implements remarkable reforms within an existing institutional framework.

The subtle duality of party-government minimizes the likelihood of collective action, either organized by the population or led by officials. However, the design works well only if the population can be co-opted with material benefits. If the population strongly demands both economic and political rights, the party-government duality can be a key friction in dealing with mass revolt. Thus, the current Chinese institution may look extremely robust, but it can be fragile to large shocks that require the coordination of political and economic powers from the state.

Recent research of authoritarian and totalitarian regimes tries to understand how propaganda and public good provision consolidate autocrats' power. Adena et al. (2014) find that radio coverage helped the Nazis win popular support. Voigtlaeder and Voth (2015) noted that highway construction also significantly improved the support for the Nazi regime. Importantly, the effect was particularly strong when it was complemented by radio coverage. Thus, in a consolidated autocracy, propaganda and economic powers complement each other. As a local politician in an authoritarian regime is a "petty dictator" himself (Lieberthal, 2005), the concentration of local economic and political power is detrimental to

the central authority in any autocratic regime. They should always try to separate political and economic power, and my theory should extend beyond China.

Indeed, Vietnam has a history of a party-state relationship that parallels the Chinese experience. An excellent survey is Tran (2004). Following the establishment of central planning, the Vietnamese Communist Party (VCP) took comprehensive control of economic management, and the government's power was substantially undermined. The economic performance of Vietnam was very poor during the central planning decades, and consequently, Vietnam began her reform "*Doi Moi*" in 1986. The government gained significant power and autonomy in economic policymaking, and the regime also began to grant more discretion to local politicians. Political meritocracy started to emerge, and it reached a landmark in the Tenth National Congress of the VCP in 2006, when almost all revolutionary cadres stepped down from leadership and a new generation of politicians took control. The experience of Vietnam provides another example of how dual leadership provides a foundation of decentralization and positive political selection.

A similar arrangement also prevails in many historical regimes. Finer's *A History of Government* states that controlling local officials is a key problem for any autocracy. A careful reading of Finer (1997) reveals that the Ottoman Empire, Russian Empire, and Spanish Colonial Empire all established institutions to constrain the power of appointed local leaders that followed the principle of "checks between higher ranked and lower ranked". Finer (1997) asks: "what care (do) the Turks take to preserve the body of their Empires free of faction and rebellion?" The first answer proposed by Finer is that "governors...shared some of their authority with the *defterdars* (fiscal governor), the chief *kadi*, and the Janissary commanders. In addition, the *timars* (provinces) were now allocated by the palace, so governors could not build a local power-base" (page 1194, Finer, 1997). Under Catherine the Great, each Russian *guberniya* (province) had "a governor, plus a deputy governor in charge of finance. The governor did not himself issue orders. In the fashion of the day, he presided over a collegiate board which did this – the governor, his deputy, and two appointed councillors" (page 1420, Finer, 1997). As for the Spanish Colonial Empire, "(Viceroyalties) were ... immense, and tiers of intermediate officers were necessarily interposed between the viceroys and the cabildos at the base. Such were the presidents, and the captains-general, who enjoyed very great discretion. They did not take their orders from the viceroy as one would expect, but directly from the Crown which appointed and removed them, and it was to the Crown they reported; so that they often acted in disregard of the viceroy" (page 1388, Finer, 1997). Although details vary, a general pattern does emerge: different from liberal democracy, "separation of power" in autocracy does not emphasize too much on strong checks and balances. Instead, the regime usually relies on a lower ranked official who controls everyday management and thus real power. The comparison I am making here is simple. We need to pay much more attention to this type of institution to deepen our understanding of authoritarianism.

Chapter 3

Aristocracy, Meritocracy, and Property Rights

3.1 Introduction

In a Weberian bureaucracy, the recruitment process should be impersonal. So a Weberian bureaucracy should be meritocratic rather than aristocratic. In practice, political organizations can be captured by “political dynasties”. Dal Bo et al. (2009) shows that even in the US Congress, “legislators who hold power for longer become more likely to have relatives entering Congress in the future”. Given that political dynasties do not merely reflect the inter-generational transmission of ability (Dal Bo et al. 2009), “patrimonialization of state” (Fukuyama, 2011) will weaken state capacity.

What causes a meritocratic bureaucracy in some states and historical periods but not others? This paper starts with the intuition that with weak property rights, *a meritocracy and its private economy reinforce each other*. On the one hand, a thriving private economy produces many candidates to compete in political recruitment. The intensified political competition prevents any family from holding dynastic political power, helping the formation of a meritocratic rather than an aristocratic government. On the other hand, a meritocratic government promotes a thriving private economy. The straightforward mechanism is that the meritocratic government strengthens various dimensions of state capacity. More importantly, the government constituted mainly by novice politicians bestows large *de facto* bargaining power to private entrepreneurs rather than the government. This is especially important in regimes without good property rights.

The reinforcement between a good government and a productive economy can be either backward-looking or forward-looking. The backward-looking reinforcement induces a path dependence. An aristocrat-controlled government has high bargaining power against commoner households. So commoner households can be too impoverished to make the necessary

education investment for their sons and daughters to enter the government. As a consequence, the government in the future will also be controlled by aristocrats. By contrast, a government controlled by novice politicians means that commoner households can get large economic surpluses. This allows commoner participation in political recruitment so that many first-generation politicians will work in the future government.

The backward-looking reinforcement primarily relies on credit constraints. In an economy without any credit constraint, a self-enforcing circle can still arise due to a self-fulfilling expectation. One possible interpretation is through Greif (2006): the commoner's belief on the return to the political career is codified as a culture/belief which can encourage or discourage investment that promotes political mobility. The backward-looking reinforcement is straightforward to formalize, while forward-looking reinforcement is more subtle. The later is the focus of my model and is elaborated as follows.

Overview of the Model

In many societies, it takes two steps for a commoner to become a politician. The commoner first needs to make an investment to enter an elite “social club”. From there the novice elite has a chance to ascend further and become a politician. The commoner household wants to obtain the elite status if the household expects a high return from it. Among its many benefits, the elite status confers a platform to develop a personal relationship with politicians¹. In other words, the elite status provides access to political patronage.

Also, suppose that the commoner has two production technologies, one traditional and one new. To adopt the new technology, the commoner needs to obtain the elite status to access political patronage². This is because the new technology requires a large *ex ante* investment that can be easily extracted or exploited, and political patronage confers the

¹There are many examples of political selection like this. For instance, the civil service exam in Imperial China has several rounds. The first round confers a gentry status that allows the degree holders to interact with magistrates as equals. The gentry status can also be bought. Only success in the next few rounds allows the degree holders to serve as politicians (see Ho, 1962; Elman, 2000; Chen, Kung, Ma, 2016). In many Continental European nations, sovereigns recruited candidates for prestigious government positions from nobilities. Rich merchants, however, can buy the nobility status. A nobility status allows a merchant to enter the high society with many social benefits (Doyle, 2009). Indeed, studies on European aristocracy explicitly recognizes the crucial importance of numerous categories of investment to sustaining the aristocratic status. Even in a modern democracy, entering electoral competition requires large *ex ante* investment to improve a candidate's name recognition and evaluation among the public (Dal Bo et al. 2009), and canonical models of electoral competition predict that electoral result is a lottery (Gelbach, 2013). To simplify the analysis, the second round of the political competition is assumed to be a lottery in my model.

²For example, households in medieval south China can choose between rice production and other traditional grains, such as wheat or millet (Glahn, 2016). Rice production yields much higher output, but also requires a substantial *ex ante* investment. As the *ex ante* investment for rice production is large, a rice economy has a much higher demand for property rights, which political patronage can readily supply. Glahn (2016) shows that the initial expansion of rice production in south China requires a substantial guarantee of property rights.

protection of such investment. So the commoner's (marginal) return from an elite status has two parts. First, access to political patronage allows the adoption of the modern technology. This part of the marginal return increases with more commoners in the bureaucracy because they reduce the bargaining power of the bureaucracy. Second, the novice elite may further become a politician who can extract economic rents from political patronage. This part of the marginal return decreases with more commoners in the bureaucracy, as a result of more intense political competition and the reduced bargaining power of politicians. *We will see that the first part dominates the second part under plausible scenarios, so the marginal net return to elite status is an increasing function of other commoners' participation.*

Thus, multiple stable equilibria can arise. In the aristocracy equilibrium, a commoner expects a low return from the elite status. So no commoner obtains the elite status, and political positions are highly dynastic. This justifies the commoner's belief: a dynastic government means low protection of property rights and consequently a low return to the elite status. In the end, all commoners adopt the traditional technology, which also leaves few rents to politicians that further reinforces the commoner's belief.

In a meritocracy, commoners believe that return to the elite status is high, so they actively enter the elite "social club"³. Consequently, political competition is intense, and many politicians are from commoner households. Their low bargaining power guarantees that novice elites who seek political patronage can reap a high return. This justifies the commoner's belief of the high return to the elite status, and a substantial fraction of households adopt the new technology. In equilibrium, the economy is highly productive, creating a huge demand for property rights that propels commoners to obtain elite status.

Why does a government constituted by novice politicians bestow high bargaining power to households? There are two possible mechanisms. Firstly, a dynastic politician has higher bargaining power because his father's political career accumulates substantial social capital transmissible to the son (Dal Bo et al. 2009). By contrast, a novice politician was not endowed with such social capital, and his bargaining power is naturally lower.

Secondly, a government constituted mostly by novice politicians necessities stronger personnel control. This mechanism is more applicable to non-democratic regimes, where the autocrat is especially anxious to keep politicians loyal and compliant. A government constituted by dynastic politicians reduces the benefit of personnel control because social control such as marriage network is already very effective. However, marriage network is not powerful in a meritocracy, as the autocrat cannot automatically control politicians in the next generation through concurrent marriage. Thus, the principal has to establish strong personnel rules such as frequent rotation, checks and balances, and promotion based on performances. An

³In an ideal meritocracy, a commoner should be able to enter the civil service without an elite status. My paper only focuses on another dimension of meritocracy, that is how many politicians are from commoner backgrounds. These two features can be analyzed separately.

important consequence of personnel control is that the weakened politician also wields less bargaining power against households, which essentially constitutes a credible commitment of secured property rights⁴.

The workhorse model assumes that commoners without elite status get zero protection for their income from the new technology. In the extended model, I add legal property rights that cover every household. I show that sufficiently strong legal property rights induce a net marginal return decreasing with commoner participation in the bureaucratic recruitment, resulting in a unique equilibrium. When legal property rights are stronger, the bureaucracy becomes more dynastic: commoners who get sufficient protection from legal property rights have little incentive to obtain the elite status for political patronage.

Section 3.2 covers the workhorse model without legal property rights. I discuss how to formalize the two mechanisms that link a meritocracy with a low bargaining power of politicians. Historical examples from Imperial China are employed to illustrate the theory. Section 3.3 shows that better legal property rights induce a more dynastic bureaucracy. To show the empirical relevance of the model, meritocracies in Imperial China and Ottoman Empires are compared, as well as state building in common law vs. civil law countries. Section 3.4 concludes and discusses possible future works.

3.2 The Workhorse Model

The Timeline

There are two periods in the model.

- At $t=1$:
 - There are a unit mass of households. μ are politician households (i.e., a family member serves as a politician). $1 - \mu$ are non-politician households. Each household $i \in [0, 1]$ receives an exogenous wealth of $w(i)$.
 - The non-politician households choose an investment level e . With a probability $p(e)$, the household will be awarded the gentry/elite status next period. With a probability of $1 - p(e)$, the household fails to obtain the status and remains a commoner household.

⁴Alternatively, one can argue that personnel control is easy to establish in a meritocratic government than an aristocratic government. The aristocrats accumulate sufficient political capital and dense social network to resist the autocrat's attempt to establish strong personnel control over them. Lacking such political and social capital, politicians from commoner households accept personnel control without much protest.

- Politician households in $t = 1$ are awarded the gentry status in $t = 2$ automatically⁵.
- At $t=2$:
 - Among the gentries, the sovereign runs a lottery to select μ as politicians in $t = 2$.
 - Each household has two production technologies. The traditional technology produces y ; the new technology produces $w > y$. Traditional technology is immune to political extraction. On the contrary, new technology requires political patronage to be functioning. It also only allows a household to retain $\hat{\beta}w$:
 - * If the household resides a jurisdiction ruled by a novice politician (i.e., his father is not a politician), then $\hat{\beta} = 1$ for gentry households (including politician households, who are gentry themselves), and $\hat{\beta} = 0$ for commoner households.
 - * If the household resides in a jurisdiction ruled by a dynastic politician (i.e., his father is a politician), all categories of households get 0⁶.
 - * Households select production technology *after* categorization of households into the classes of gentry and commoner. Politicians arrive in their jurisdictions after all households have selected a production technology.

Meritocracy And *de facto* Property Rights in Political Patronage

By backward induction, at $t = 2$, a commoner household always chooses the traditional technology. A gentry household chooses the new technology if:

$$E[\beta]w \geq y \tag{3.1}$$

$E[\beta]$ is the expected bargaining power of the gentry household. Suppose at $t = 1$, the investment made by all (other) non-politician households is \hat{e} . Then:

$$E[\beta] = \frac{(1 - \mu)p(\hat{e})}{(1 - \mu)p(\hat{e}) + \mu} * 1 + \frac{\mu}{(1 - \mu)p(\hat{e}) + \mu} * 0 = \frac{(1 - \mu)p(\hat{e})}{(1 - \mu)p(\hat{e}) + \mu} \tag{3.2}$$

With probability $\frac{(1-\mu)p(\hat{e})}{(1-\mu)p(\hat{e})+\mu}$, the gentry household’s jurisdiction is ruled by a novice politician. The gentry household can retain everything. With probability $\frac{\mu}{(1-\mu)p(\hat{e})+\mu}$, the gentry

⁵Bestowing the gentry/elite status to politicians’ offsprings is used by many societies to provide dynamic incentives. Alternatively, we can endogenize politician households’ decision to invest, but this complicates the analysis substantially without new insights. The automatic grant of gentry status to politician’s son allows us to only track the decision of non-politician households.

⁶We can assume that gentry household who used to be politician can keep everything, and this will not change our results.

household's jurisdiction is ruled by a dynastic politician. In this case, the gentry household surrenders everything to the politician.

From (3.1) and (3.2), gentry household chooses the new technology if:

$$p(\hat{e}) \geq \frac{\mu}{1-\mu} \frac{y}{w-y} \tag{3.3}$$

Now go back to $t = 1$. The key control variable is the investment e in $t = 1$ made by non-politician households⁷. They solve the problem:

$$\begin{aligned} \max_e \quad & w(i) - kc(e) + \mathbf{1}\{p(\hat{e}) \geq \frac{\mu}{1-\mu} \frac{y}{w-y}\} \{p(e)[E(\beta)w] + (1-p(e))y\} \\ & + \mathbf{1}\{p(\hat{e}) \leq \frac{\mu}{1-\mu} \frac{y}{w-y}\} y \end{aligned} \tag{3.4}$$

As before, $E(\beta) = \frac{(1-\mu)p(\hat{e})}{(1-\mu)p(\hat{e})+\mu}$. The household receives an income of $w(i)$ in $t = 1$ that is irrelevant to the maximization problem. They can generate an investment of e at a cost of $kc(e)$. The return to this investment depends on the strategy chosen by other households.

If other non-politician households at $t = 1$ choose a high enough investment such that $p(\hat{e}) \geq \frac{\mu}{1-\mu} \frac{y}{w-y}$, then my investment has a marginal return > 0 . In this case, I know that I will choose the modern technology if I become a gentry in the future. The gentry status will be obtained with probability $p(e)$, which depends on my investment today.

What about the rents I collect if I am lucky enough to serve as a politician? This is zero in equilibrium. With probability $\frac{\mu}{(1-\mu)p(\hat{e})+\mu}$, I become a politician with $\frac{1-\mu}{\mu}$ families under my jurisdiction. Among them, $(1-\mu)p(\hat{e}) + \mu$ are gentry, and as a novice politician, I can extract nothing from them. $(1-\mu)(1-p(\hat{e}))$ are commoners. But the commoners will always adopt the traditional technology, so in equilibrium, I can extract nothing from commoners either.

But if other households choose \hat{e} such that $p(\hat{e}) \leq \frac{\mu}{1-\mu} \frac{y}{w-y}$, the marginal benefit from the investment is zero: even if I obtain the gentry status, it is likely that I am going to meet a dynastic politician. In this case, the expected return from the new technology is smaller than the old technology, which does not require property rights at all. So my utility at $t = 2$ is $(1-p(\hat{e}))(y) + p(\hat{e})y = y$, with zero marginal benefit to gentry status. Notice that again, even if the non-politician household is lucky to produce a novice politician in $t = 2$, the political rent is zero.

⁷Recall gentry status is automatically granted to politician households in $t = 1$.

What are the equilibria? If $p(\hat{e}) > \frac{\mu}{1-\mu} \frac{y}{w-y}$, we have the Euler equation:

$$kc'(e^*) = p'(e^*) \left[\frac{(1-\mu)p(\hat{e})}{(1-\mu)p(\hat{e}) + \mu} w - y \right]$$

In equilibrium, we need $e^* = \hat{e}$. So the equilibrium satisfies:

$$kc'(e^*) = p'(e^*) \left[\frac{(1-\mu)p(e^*)}{(1-\mu)p(e^*) + \mu} w - y \right] \quad (3.5)$$

Here, the belief about the return to gentry status is sufficiently high. This induces a large investment to gentry “training”, which means that novice politicians dominate the future government. This justifies the belief about the fruitful return of gentry investment. The aggregate output $= [(1-\mu)p(e^*) + \mu]w + (1-\mu)(1-p(e^*))y$. Also, only a small fraction of households can hold political power for both periods $(\frac{\mu^2}{\mu+(1-\mu)p(e^*)})$. Most politicians are either novice or lose political power in $t = 2$.

We have another equilibrium. If $p(\hat{e}) < \frac{\mu}{1-\mu} \frac{y}{w-y}$, the problem is:

$$\max_e z - kc(e) + y$$

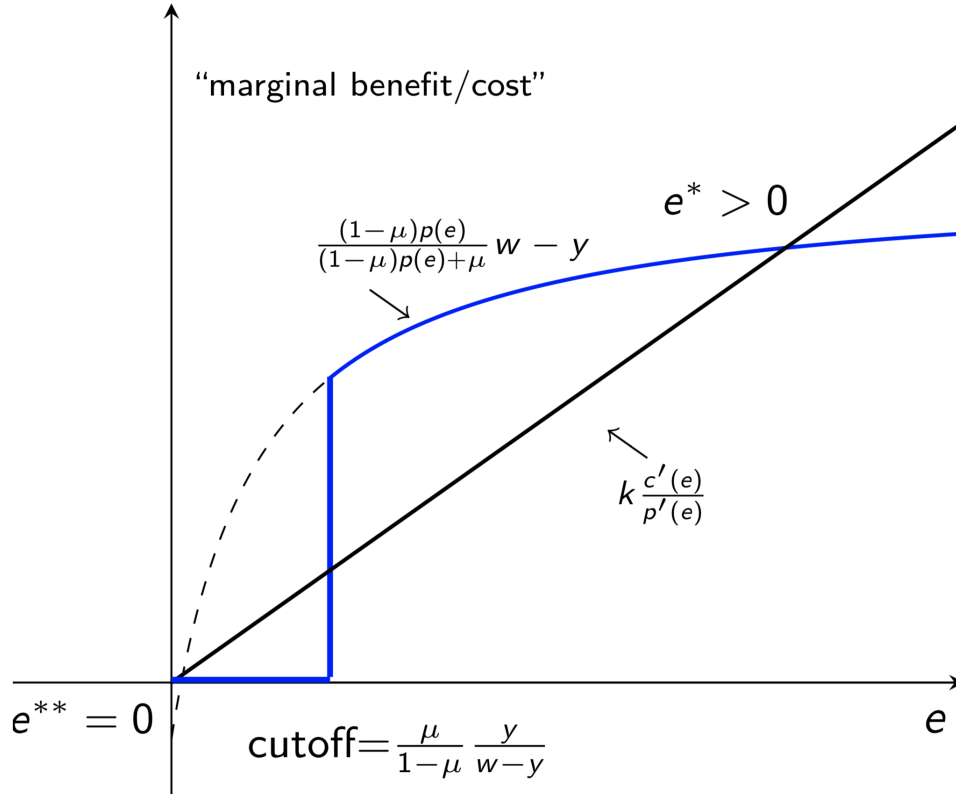
which means that $e^* = 0$. In equilibrium, $\hat{e} = e^* = 0$. as everyone believes that the future return to the gentry status does not justify a serious investment, dynastic politicians dominate the future government, which rationalizes a zero investment on the gentry status. The aggregate output is y . No one uses the new technology in the fear of expropriation. The government is fully dynastic.

To examine the stability of these equilibria, I am looking at a special example. Assume $k = \frac{1}{2}$, $c(e) = we^2$, and $p(e) = e$. Then stability requires that $\frac{(1-\mu)e}{(1-\mu)e+\mu}w - y$ has a slope $< w$ at $e^* = \frac{(1-\mu)e^*}{(1-\mu)e^*+\mu}w - y$. $d[\frac{(1-\mu)e}{(1-\mu)e+\mu}w - y]/de|_{e^*} = \frac{\mu(1-\mu)}{[(1-\mu)e^*+\mu]^2}w < w$, so $e^* > \frac{\sqrt{\mu(1-\mu)}-\mu}{1-\mu}$. $\max_\mu \frac{\sqrt{\mu(1-\mu)}-\mu}{1-\mu} = \frac{1}{4}$. So a sufficient condition is $e^* > \frac{1}{4}$.

A key observation here is that $e^* = 0$ is a stable equilibrium regardless of functional forms. Suppose a measure of ν non-politicians in period $t = 1$ try to break the dismal equilibrium by obtaining the gentry status and (with almost probability 1) entering government. Another way to conceptualize this collective action is that the emperor/sovereign recruits commoners to balance the power of aristocrats. The emperor’s capacity allows him to recruit ν commoners. As long as

$$\nu < \frac{\mu}{1-\mu} \frac{y}{w-y} \quad (3.6)$$

Figure 3.1: Multiple Equilibria: Aristocracy vs. Meritocracy



it is to the self-interest of every coalition member to deviate from the recommended strategy and not obtain the gentry status. For the emperor, although he is strong enough to recruit ν commoners, this does not incentivize commoner's voluntary entrance. In the end, the emperor's heir will not be able to sustain the meritocracy if the heir is weaker than his father.

Look at Equation (3.6); we see that escaping the dismal equilibrium is more likely when ν increases (emperor is more powerful) and when w increases (the new technology becomes more productive). In the historical illustrations from Imperial China, we will see how the productivity boost around 1,000 AD from the rise of rice economy finally allows the powerful emperors to establish a stable meritocracy which also guarantees the sustainability of the new production mode.

These equilibria are illustrated in the graph. It depicts the special case where $k = \frac{1}{2}$, $c(e) = e^2$, and $p(e) = e$. The blue curve is the marginal benefit, and it has three segments: the segment where $e > \frac{\mu}{1-\mu} \frac{y}{w-y}$ is $\frac{(1-\mu)e}{(1-\mu)e+\mu} w - y$, a vertical segment $e = \frac{\mu}{1-\mu} \frac{y}{w-y}$, and a horizontal segment where marginal benefit = 0. The black line is the marginal cost $(\frac{1}{2}e^2)' = e$.

We can see the two stable equilibria $e^{**} = 0$ and $e^* > 0$, along with an unstable equilibrium where $e = \frac{\mu}{1-\mu} \frac{y}{w-y}$.

Stable equilibrium also makes comparative statics results standard. For example, when w increases, e^* increases: a more productive new technology stimulates competition for the gentry status.

These results are summarized as follows:

Proposition 1

Denote $e^* = kc'(e^*) = p'(e^*)[\frac{(1-\mu)p(e^*)}{(1-\mu)p(e^*)+\mu}w - y]$.

1. *Regularity conditions guarantee that e^* is a stable equilibrium with a meritocratic government and the total output = $[\mu + p(e^*)(1 - \mu)]w + (1 - p(e^*))(1 - \mu)y$. Moreover, $w \uparrow e^*$, $y \downarrow e^*$.*

2. *$e^{**} = 0$ is always a stable equilibrium with a fully dynastic government and the total output = y . Moreover, change in w or y has no effect on $e^{**} = 0$.*

Meritocracy and Personnel Control

Above analysis shows one mechanism how a meritocratic government can boost the bargaining power of gentry households: novice politicians are less powerful than dynastic politicians when they interact with local gentries. Another important mechanism is that a government dominated by novice politicians induces the sovereign to exert strong personnel control over it. The enhanced personnel control reduces the bargaining power of politicians against the private economy.

I define personnel control as institutionalized rules imposed on politicians such as rotation, checks and balances, and promotion based on performance. Why the emperor does not exert personnel control in an aristocracy but does so in a meritocracy? One reason is that social control such as marriage networks between aristocrats and the imperial clan already provides a powerful instrument for the emperor to ensure compliance (Tackett, 2014). So the benefit of personnel control only extends to novice politicians. In a government dominated by novice politicians, marriage networks do not work well because politicians' sons were unlikely to become politicians, while the aristocratic government delivers a high continuation value of imperial marriage for the emperor.

With only social control, the emperor can ensure (some) compliance from politicians. However, an unchecked politician with a long tenure in his jurisdiction can easily dominate its households. For example, the politician can patiently collect information about the local economy so that he can calibrate rent extraction based on the realized productivity. This

denies gentry households their information rent. By comparison, a politician expecting frequent rotation will not bother knowing everything about the local economy. Complemented by promotion incentives or checks and balances, rotation can guarantee substantial information rent to gentry households. In other words, gentry households enjoy high bargaining power when politicians face strong personnel control. This is the key insight of Chapter 1 of this dissertation; in this Chapter, I assume its validity.

I am going to employ a reduced form approach to model the link between meritocracy and personnel control. A version with a micro-foundation will be for future work. Assume that the emperor can impose a degree of personnel control of I on politicians with a cost of $C(I)$, $C'(\cdot) > 0$, $C''(\cdot) < 0$. The emperor solves the problem:

$$\frac{(1-\mu)p(e)}{(1-\mu)p(e)+\mu}I - C(I)$$

Among all politicians, $\frac{(1-\mu)p(e)}{(1-\mu)p(e)+\mu}$ are novice. The benefit of personnel control I is $\frac{(1-\mu)p(e)}{(1-\mu)p(e)+\mu}$ because dynastic politicians are already compliant thanks to social control.

The solution to emperor's problem is:

$$I^* = C^{-1}\left(\frac{(1-\mu)p(e)}{(1-\mu)p(e)+\mu}\right)$$

So the intensity of personnel control increases with the fraction of novice politicians. Now suppose that with an intensity of personnel control at I^* , a gentry household's bargaining power is $\beta = G(I^*)$ against either a novice or dynastic politician, $G'(\cdot) > 0$. This assumption of equal bargaining power for novice and dynastic politicians is to isolate the mechanism that a meritocratic government breeds strong personnel control. We have:

$$\beta = G\left\{C^{-1}\left[\frac{(1-\mu)p(e)}{(1-\mu)p(e)+\mu}\right]\right\} \equiv H\left[\frac{(1-\mu)p(e)}{(1-\mu)p(e)+\mu}\right]$$

$H'(\cdot) > 0$. So a household with gentry status chooses the modern technology if:

$$H\left[\frac{(1-\mu)p(\hat{e})}{(1-\mu)p(\hat{e})+\mu}\right]w \geq y$$

$$p(\hat{e}) \geq \frac{\mu}{1-\mu} \frac{H^{-1}\left(\frac{y}{w}\right)}{1-H^{-1}\left(\frac{y}{w}\right)}$$

A non-politician household at $t = 1$ solves the problem:

$$\begin{aligned} \max_e \quad & z - kc(e) + \mathbf{1}\{p(\hat{e}) \geq \frac{\mu}{1-\mu} \frac{H^{-1}(\frac{y}{w})}{1-H^{-1}(\frac{y}{w})}\} \{p(e)[\beta w + (1-\mu)(1-\beta)w] + (1-p(e))y\} \\ & + \mathbf{1}\{p(\hat{e}) \leq \frac{\mu}{1-\mu} \frac{H^{-1}(\frac{y}{w})}{1-H^{-1}(\frac{y}{w})}\} y \end{aligned} \quad (3.7)$$

an expression very similar as the Equation (3.4). The additional term we have is $(1-\mu)(1-\beta)w$, which is the expected rent a gentry household collects as a future politician. Specifically,

$$(1-\mu)(1-\beta)w = \frac{\mu}{(1-\mu)p(\hat{e}) + \mu} \frac{1-\mu}{\mu} [(1-\mu)p(\hat{e}) + \mu](1-\beta)w$$

Conditional on gentry status, with probability $\frac{\mu}{(1-\mu)p(\hat{e}) + \mu}$ the household's son will be selected to serve as the politician in $t = 2$. He rules over $\frac{1-\mu}{\mu}$ households, among them $(1-\mu)p(\hat{e}) + \mu$ are gentries (including politician households). He can extract $(1-\beta)w$ from each gentry household.

Rearrange (3.7). The problem of a non-politician household at $t = 1$ is:

$$\begin{aligned} \max_e \quad & z - kc(e) + \mathbf{1}\{p(\hat{e}) \geq \frac{\mu}{1-\mu} \frac{H^{-1}(\frac{y}{w})}{1-H^{-1}(\frac{y}{w})}\} \{p(e)[\mu H[\frac{(1-\mu)p(\hat{e})}{(1-\mu)p(\hat{e}) + \mu}]w + (1-\mu)w] + (1-p(e))y\} \\ & + \mathbf{1}\{p(\hat{e}) \leq \frac{\mu}{1-\mu} \frac{H^{-1}(\frac{y}{w})}{1-H^{-1}(\frac{y}{w})}\} y \end{aligned} \quad (3.8)$$

We again have an equilibrium with $e^* = 0$.

Assume that $p(\hat{e}) \geq \frac{\mu}{1-\mu} \frac{H^{-1}(\frac{y}{w})}{1-H^{-1}(\frac{y}{w})}$, Euler equation yields:

$$k \frac{c'(e^*)}{p'(e^*)} = \{[\mu H[\frac{(1-\mu)p(\hat{e})}{(1-\mu)p(\hat{e}) + \mu}]w + (1-\mu)w - y\}$$

Equilibrium again requires $e^* = \hat{e}$. Depending on the curvature of $G(\cdot)$, we may have multiple equilibria even for $e^* \geq \frac{\mu}{1-\mu} \frac{H^{-1}(\frac{y}{w})}{1-H^{-1}(\frac{y}{w})}$, since $\{\mu H[\frac{1}{2\kappa} \frac{(1-\mu)p(\hat{e})}{(1-\mu)p(\hat{e}) + \mu}]w + (1-\mu)w - y\}$ is an increasing function of e .

by novice politicians, inducing the emperor to exert strong personnel control. The personnel control boosts the bargaining power of gentry households, validating the commoner’s belief. The belief on the low return to the gentry status can also be self-fulfilling because the induced dynastic government reduces the equilibrium level of personnel control. Unchecked politicians wield a high bargaining power against gentry households, which validates the belief of the low return to the elite status.

Proposition 2

Denote $e^* : k \frac{c'(e^*)}{p'(e^*)} = \{[\mu H[\frac{(1-\mu)p(\hat{e})}{(1-\mu)p(\hat{e})+\mu}]w + (1 - \mu)w - y\}$.

1. *Regularity conditions guarantee that e^* is a stable equilibrium with meritocratic government and total output = $[\mu + p(e^*)(1 - \mu)]w + (1 - p(e^*))(1 - \mu)y$. The emperor exerts strong personnel control over the government.*

2. *$e^{**} = 0$ is always a stable equilibrium with fully dynastic government and total output = y . The emperor exerts no personnel control over the government.*

Historical Illustration: Stability of Meritocracy in Imperial China

The multiple stable equilibria illuminate many historical episodes in political history. To fight against the entrenched nobility, sovereigns usually recruited commoners actively. But in many cases, only very powerful sovereigns succeeded, and their successes were short-lived. In other cases, the meritocratic government reproduced itself smoothly without any active imperial oversight. The most vivid example comes from China, where many powerful emperors recruited commoners and purged aristocrats to consolidate the imperial power. But before 1,000 AD, such efforts could not survive long after the end of the strong emperor’s reign (such as Emperor Wu of Former Han Dynasty and Empress Wu Zetian of Tang/Zhou Dynasty, see *Finer, (1997a, 1997b)*). Aristocratic clans quickly reclaimed control of government afterward. After 1,000 AD, the meritocratic government reproduced itself smoothly without any imperial oversight. As a consequence, the only household that could bequest political power consistently was the imperial family. Even when the whole society was on the verge of collapse (e.g., shrinking fiscal capacity, widespread peasant revolts, threatening international conflicts), one institution always in normal operation was the civil service exam. Indeed, political selection through the exam was stubbornly robust well into the last few years of many dynasties (such as Song (960-1276), Ming (1368-1644), and Qing (1644-1911)).

My theory gives a simple economic explanation. 1,000 AD is both political and economic watersheds in Chinese history. China experienced “Tang-Song Transition” from late Tang Dynasty (618-907) to early Song Dynasty, marking striking “modernization” featured by the rise of a market economy and a meritocratic government. Political historians usually emphasize the transition from an aristocracy to a meritocracy (*Tackett, 2014*), while economic

historians focus on how the market economy, long-distance trade, and monetized taxation experienced unprecedented expansion (Glahn, 2016). A prototype industrial revolution might also occur in the early Song Dynasty (McNeill, 1982). Meanwhile, a central puzzle in Chinese history (as well as modern China studies) is why the booming market economy could persistently work well without the protection of formal or legal property rights. My theory shows that the political and economic transformations were tightly intertwined. A booming market economy is protected by a meritocratic government well disciplined under strict personnel rules. *Different from legal property rights, such de facto property rights are only bestowed to social elites who have access to political patronage. It is precisely the lack of legal property rights that drove commoners to actively participate in the civil service exam, which intensified political competition and prevented the formation of aristocratic families.* The Song meritocratic bureaucracy is boosted by a highly productive rice economy that requires substantial *ex ante* investment and consequently created a huge demand for property protection⁸.

One of the key assumptions of the model to generate multiple equilibria is the existence of dual economy. This is realistic in historical China, as the shift of the economic center from the north to the south opened up the opportunity for a highly productive rice economy that requires large *ex ante* investment. In principle, one can also engage in wheat production in south China. It requires much less investment (thus subject much less to holdup problem) but also much less productive. The grand shift to a flourishing rice economy in Song Dynasty requires good property protection that can incentivize *ex ante* investment, while wheat production is vividly less demanding on property rights.

Also, Song economy experienced an unprecedented expansion of private manufacturing and long-distance trade. Late Ming and Middle Qing Dynasties also enjoyed similar commerce prosperity. Such complex economic activities require even stronger property rights than rice farming. This is due to the easiness of predating commerce and manufacturing, manifested by the surge of commerce tax revenue in Song Dynasty. If property rights were not enforced in a *de facto* way, such flourishing trade and manufacturing were hardly sustainable (Greif, 2006). Merchants have the option to shift back to agriculture production that is much less extractable, creating a dual economy as the model requires.

Also, historical studies confirm that Chinese merchants did enthusiastically invest in their son's education. For example, the famous merchant groups from Huizhou and Shanxi usually had at least one son fully devoted to the preparation for the civil service exam (Ho, 1962;

⁸Such shift to rice as the main grain was a consequence of the repeated Sino-nomadic conflicts that periodically roamed northern China, pushing the population to migrate southward where rice was the more appropriate grain. Glahn (2016) showed that the conflict-driven migration and positive technological shocks (such as new rice species from Vietnam) stimulated the formation of the rice economy in Song China. He also emphasized how important property rights were for a thriving rice economy. The associated productivity boost through the rice economy was also an important contributor to the marketization of Chinese economy (Glahn, 2016).

Glahn, 2016). The enormous success of this strategy is manifested by the resilience of the merchant class in late Imperial China through many political and social upheavals, without any formal protection of property (Glahn, 2016).

Consistent with the model, personnel control was also dramatically expanded in early Song Dynasty along with the recruitment of novice politicians. Tang aristocratic politicians faced feeble personnel control from the emperor, whose half-hearted attempt to establish checks and balances were quickly nullified (Tackett, 2014). Meanwhile, marriages between the imperial clan and aristocrats were extensive (Tackett, 2014). By contrast, Song emperors forcefully implemented an elaborated system of personnel control with routinized rotation, overly complicated checks and balances, and numerous other measures. What is especially striking is that marriages between the imperial clan and politicians also became much rarer (Watson, 1991; Chaffee, 1999).

The question is, does politicians actually hold low bargaining power against gentry class under a meritocracy (and strong personnel control)? This is explicitly documented by historical sources. Chang (1955)'s discussion is so vivid:

“In a handbook for magistrate, a passage on the appropriate treatment of gentry members by magistrates reads as follows:

‘In administrating the affairs of a district, the high families [*shen*] are not to be disturbed. They should be met in a virtuous manner and received courteously and are not to be repressed by power and prestige. Also, the scholars [*shih*] are at the head of the people... When they happen to come because of public affairs, if they are sincere and self-respecting, they should be consulted on problems of whether bandits exist in their villages, what the jobs of the villagers are, whether the customs of their places are praiseworthy...’ ”

The last sentence explicitly acknowledges the information advantage of gentries. In my model, such advantage to gentries is a natural consequence of the strong personnel control over the magistrates, who face frequent rotation. Chung (1955) goes on and further discusses extensively the huge power of local gentries against magistrates.

The ideological shift around 1,000 AD is also very intriguing. The dominance of Buddhism in Tang Dynasty vanished in early Song Dynasty, manifested by a revival of Confucianism unprecedented in breadth and depth. Buddhism discourages any social engagement, while Confucianism emphasizes the value of education, the manners of the educated class, and firm loyalty to the sovereign. Interpreted through my model, the Confucian ideology that values education codifies people's belief of high return to education. *Confucianism serves as a coordination device that propels commoner participation in the political competition, which reinforces and justifies the ideology.* Indeed, recent research finds that Confucianism continue its deep influence on modern Chinese society 100 years after the abolishment of imperial civil service exam, especially its dimension that values education (Chen, Kung, Ma, 2016).

3.3 Patronage, Legal Property Rights, and Meritocratic Government

Above analysis assumes that property will not be protected at all without political patronage. This applies to a few prominent examples like China. But China is a unique civilization without traditional rule of law (Fukuyama, 2011; Ma, Zanden, 2011). In other major civilizations, rule of law is an important feature of the political process, so even commoners enjoyed some property rights. This surely applies to Western Europe, but traditional rule of law was also consolidated in Islamic and Indian civilizations (Fukuyama, 2011; Kuran, 2011). Fukuyama (2011) attributes the origin of rule of law to religion, which creates binding rules that transcend political authority. In this section, I am going to show how legal property rights (legal PRs for short) associated with rule of law affects political dynasties. Legal PRs reduce the commoner’s incentive to enter “elite club” due to smaller return from political patronage: strong PRs already provides decent protection of commoner’s property. This creates an immense difficulty for the sovereign who tries to prevent the patrimonialization of the government. In the end, building a meritocratic government will be costly and challenging in states with strong PRs.

The extended model follows Section 3.2. The only difference is that everyone gets minimum property rights $\underline{\beta}$. Now a commoner facing either dynastic or novice politician can keep $\underline{\beta}w$ if the commoner chooses the new technology. Also, gentry households facing dynastic politicians can keep $\underline{\beta}w$.

When $\underline{\beta}w < y$ or $\underline{\beta} < \frac{y}{w}$, households who turn out to be commoners in $t = 2$ still adopt the traditional technology. Legal PRs are not strong enough to make new technology profitable for the commoner. We have very similar analysis as section 2.1.

When $\underline{\beta}w \geq y$ or $\underline{\beta} \geq \frac{y}{w}$, all households adopt the new technology, including the commoners. In this case, the problem of non-politician households in $t = 1$ is:

$$\max_e w(i) - kc(e) + p(e)\{[E(\beta)w + \frac{(1 - \mu)^2}{(1 - \mu)p(\hat{e}) + \mu}(1 - p(\hat{e}))(1 - \underline{\beta})w] + (1 - p(e))\underline{\beta}w\}$$

If the household invest e today with a cost of $kc(e)$, tomorrow with probability $p(e)$ it earn the gentry status. In that case, the expected bargaining power is $E(\beta)$, which is:

$$E(\beta) = \frac{(1 - \mu)p(\hat{e})}{(1 - \mu)p(\hat{e}) + \mu} * 1 + \frac{\mu}{(1 - \mu)p(\hat{e}) + \mu} * \underline{\beta} = \frac{(1 - \mu)p(\hat{e}) + \mu\underline{\beta}}{(1 - \mu)p(\hat{e}) + \mu}$$

With probability $\frac{(1-\mu)p(\hat{e})}{(1-\mu)p(\hat{e})+\mu}$, the household meets a novice politician. It keeps all output in $t = 2$. With probability $\frac{\mu}{(1-\mu)p(\hat{e})+\mu}$, the household meets a dynastic politician, so I keep $\underline{\beta}$ fraction of total output w .

The second term $\frac{(1-\mu)^2}{(1-\mu)p(\hat{e})+\mu}(1-p(\hat{e}))(1-\underline{\beta})w$ is the expected rent the household collect as a novice politician, conditional on the gentry status:

$$\frac{(1-\mu)^2}{(1-\mu)p(\hat{e})+\mu}(1-p(\hat{e}))(1-\underline{\beta})w = \frac{\mu}{(1-\mu)p(\hat{e})+\mu} \frac{(1-\mu)}{\mu}(1-\mu)(1-p(\hat{e}))(1-\underline{\beta})w$$

With probability $\frac{\mu}{(1-\mu)p(\hat{e})+\mu}$, a member of the household is selected as a (novice) politician in $t = 2$, conditional on the gentry status. In its jurisdiction, there are $\frac{(1-\mu)}{\mu}$ households, in which $(1-\mu)(1-p(\hat{e}))$ are commoners. The politician can collect $(1-\underline{\beta})w$ from each of them. Consistent with the first half of Section 3.2, a novice politician cannot collect rents from gentry households.

With probability $1-p(\hat{e})$, the household fails to obtain the gentry status. In that case, the household keeps $\underline{\beta}w$.

So the problem of non-politician households in $t = 1$ is:

$$\begin{aligned} \max_e \quad & z - kc(e) + p(e) \left\{ [E(\beta)w + \frac{(1-\mu)^2}{(1-\mu)p(\hat{e})+\mu}(1-p(\hat{e}))(1-\underline{\beta})w] + (1-p(e))\underline{\beta}w \right\} \\ = \max_e \quad & z - kc(e) + p(e) \left\{ \left[\frac{(1-\mu)p(\hat{e}) + \mu\underline{\beta}}{(1-\mu)p(\hat{e})+\mu} w + \frac{(1-\mu)^2}{(1-\mu)p(\hat{e})+\mu}(1-p(\hat{e}))(1-\underline{\beta})w \right] + (1-p(e))\underline{\beta}w \right\} \end{aligned}$$

The Euler equation:

$$kc'(e) = p'(e) \left\{ \frac{(1-\mu)p(\hat{e}) + \mu\underline{\beta}}{(1-\mu)p(\hat{e})+\mu} w + \frac{(1-\mu)^2}{(1-\mu)p(\hat{e})+\mu}(1-p(\hat{e}))(1-\underline{\beta})w - \underline{\beta}w \right\}$$

$$kc'(e) = p'(e) \left\{ \frac{(1-\mu)p(\hat{e})}{(1-\mu)p(\hat{e})+\mu}(1-\underline{\beta})w + \frac{(1-\mu)^2}{(1-\mu)p(\hat{e})+\mu}(1-p(\hat{e}))(1-\underline{\beta})w \right\}$$

$$kc'(e) = p'(e) \left\{ \frac{(1-\mu)p(\hat{e})}{(1-\mu)p(\hat{e})+\mu} + \frac{(1-\mu)^2}{(1-\mu)p(\hat{e})+\mu}(1-p(\hat{e})) \right\} (1-\underline{\beta})w$$

In equilibrium, $e^* = \hat{e}$, we have:

$$kc'(e^*) = p'(e^*) \left\{ \frac{(1-\mu)p(e^*)}{(1-\mu)p(e^*) + \mu} + \frac{(1-\mu)^2}{(1-\mu)p(e^*) + \mu} (1-p(e^*)) \right\} (1-\underline{\beta})w$$

From the Euler equation, we can see two effects of property rights. The right hand side is the marginal benefit. The term $(1-\underline{\beta})$ shows that the incentive to obtain gentry status for property protection is reduced when legal PRs is decent. However, there is an additional effect manifested by the term $p'(e^*) \frac{(1-\mu)^2}{(1-\mu)p(e^*) + \mu} (1-p(e^*)) (1-\underline{\beta})w$. This is the marginal expected rent a non-politician household in $t=1$ can collect in $t=2$ from serving as a novice politician. Because of good PRs, even commoner household now adopt new technology that is extractable (subject to legal PRs constraint). This increases the value of political positions and increases the equilibrium investment in e .

Proposition 3

1. If $\underline{\beta} < \frac{y}{w}$, we have similar results as in Proposition 1.
2. If $\underline{\beta} > \frac{y}{w}$, the dismal equilibrium $e^{**} = 0$ is eliminated. In equilibrium, $kc'(e^*) = p'(e^*) \left\{ \frac{(1-\mu)p(e^*)}{(1-\mu)p(e^*) + \mu} + \frac{(1-\mu)^2}{(1-\mu)p(e^*) + \mu} (1-p(e^*)) \right\} (1-\underline{\beta})w$. Total output = w .

We can do standard comparative statics and have the following result:

Proposition 4

Assume $0 < \mu < \frac{1}{2}$ and $\underline{\beta} > \frac{y}{w}$, then

1. Equilibrium is unique.
2. $\frac{de^*}{d\underline{\beta}} < 0$: better legal property rights induces less investment to gentry status and the share of dynastic politicians increases.

Notes to Figure 3: $MB(\beta) = \frac{(1-\mu)p(e)}{(1-\mu)p(e)+\mu} + \frac{(1-\mu)^2}{(1-\mu)p(e)+\mu} (1-p(e)) \} (1-\beta)w$. $MC = k \frac{c'(e)}{p'(e)}$.

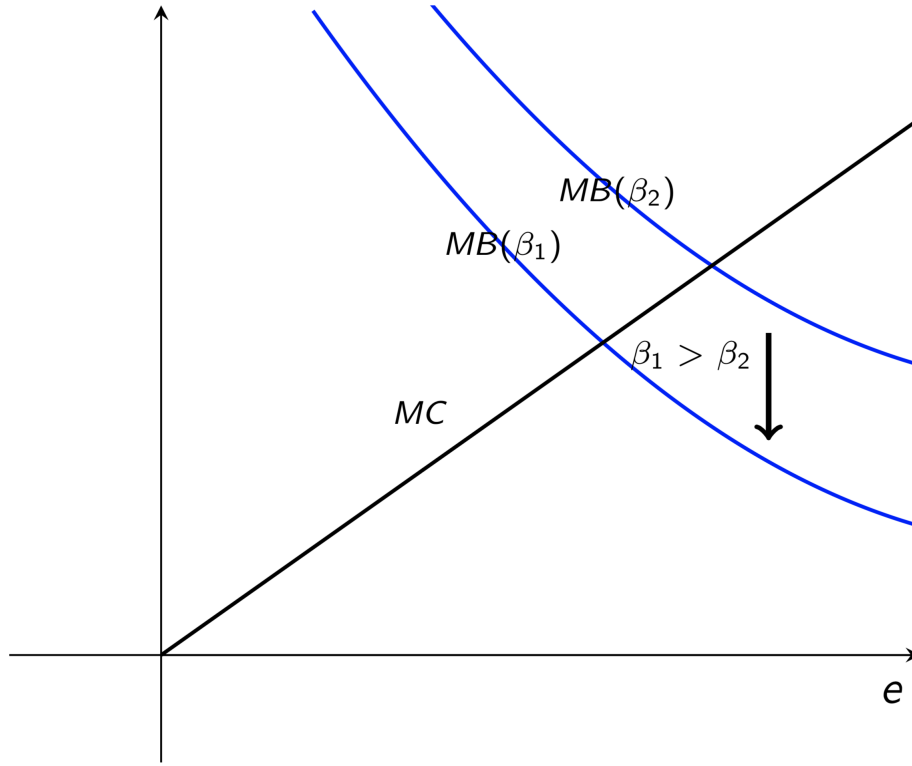
Proof:

Denote $\Delta(e) = \frac{c'(e)}{p'(e)}$. We know $\Delta'(e) > 0$. Denote $\Pi(e) = \frac{(1-\mu)p(e)}{(1-\mu)p(e)+\mu} + \frac{(1-\mu)^2}{(1-\mu)p(e)+\mu} (1-p(e))$. Apply total differentiation to the Euler equation:

$$k\Delta'(e)de = (1-\underline{\beta})w\Pi'(e)de - \Pi(e)w d\underline{\beta}$$

$$\frac{d\underline{\beta}}{de} = \frac{\Pi(e)}{(1-\underline{\beta})w\Pi'(e) - k\Delta'(e)}$$

Figure 3.3: Better Legal Property Rights lead to a More Entrenched Bureaucracy



A sufficient condition for $\frac{de}{d\beta} < 0$ is $\Pi'(e) < 0$.

$$\Pi'(e) = \frac{d\left\{\frac{(1-\mu)p(e)}{(1-\mu)p(e)+\mu} + \frac{(1-\mu)^2}{(1-\mu)p(e)+\mu}(1-p(e))\right\}}{dp} p'(e)$$

$$\frac{(1-\mu)p}{(1-\mu)p+\mu} + \frac{(1-\mu)^2}{(1-\mu)p+\mu}(1-p) = (1-\mu)\frac{1-\mu+\mu p}{(1-\mu)p+\mu}$$

$$d\left\{\frac{1-\mu+\mu p}{(1-\mu)p+\mu}\right\}/dp = \frac{2\mu-1}{[(1-\mu)p+\mu]^2} < 0$$

which reduces to $\mu < \frac{1}{2}$.

Notice $\mu < \frac{1}{2}$ guarantees that the “marginal benefit” $\Pi(e) = \frac{(1-\mu)p(e)}{(1-\mu)p(e)+\mu} + \frac{(1-\mu)^2}{(1-\mu)p(e)+\mu}(1-p(e))$ is a decreasing function of e . So we have unique equilibrium.

Q.E.D.

The results are illustrated in Figure 3.3. As long as $\mu < \frac{1}{2}$, or less than half of the population serve as politicians, then better property rights induce a more dynastic government. When $\underline{\beta}$ increases, the marginal benefit curve shifts downward, and it results in lower e^* .

China and Ottoman Empire: Voluntary v.s. Forced Meritocracies

The theory predicts that regimes with strong legal property rights will have more dynastic government and find it difficult to maintain a meritocracy. The simple point shed light on empirical regularities in different settings.

For instance, apart from Imperial China, another prominent example of a meritocratic government in the pre-industrial world was Ottoman Empire. The cornerstone of the Ottoman meritocracy is the *Devshirme*, through which sultan's scouts coercively recruited boys from Christian commoner families (Coser, 1972; Fukuyama, 2011). These boys were converted to Islam and got excellent civilian or military training. The very best served as top politicians and military commanders, creating an Ottoman bureaucracy controlled by non-Turkish peoples. The sultan's initial motivation to institutionalize *Devshirme* was precisely to weaken the Turkish nobility, and the sultan surely succeeded in it. But *Devshirme* proved to be much less persistent than the civil service exam in China. It collapsed in the 17th century, two hundred years before the demise of Ottoman Empire itself.

The contrast between voluntary and forced meritocracies in China and Ottoman Empire is striking, and my theory attributes it to the better property rights in Ottoman Empire than China⁹. The rule of Islamic law in Ottoman Empire conferred sacred protection of *waqf*, “an unincorporated trust founded under Islamic law by a person for the provision of a designated service in perpetuity” (Kuran, 2004). In practice, *waqf* became the major institution for the rich people to protect their property, as nothing prevents the *mutawalli* (the trustee and the manager) from paying himself a handsome salary from the fund and appointing his offsprings as future *mutawalli*. This gives rich families a strong incentive to have a family member well trained in Islamic law rather than being a politician. The low enthusiasm to enter the government forced the sultan to rely on coercive recruitment of commoners that required strong state capacity beforehand. Bad shocks to state capacity rendered such recruitment impossible to sustain, unlikely Chinese meritocracy whose entrance was voluntary.

⁹Property rights in Ottoman Empire were not well protected by the standard of West Europe, especially England. The key point I want to make is that property rights are better in Ottoman Empire than Imperial China.

Comparative Law and State Building

The comparison between common law and civil law boasts a huge literature (surveyed by La Porta et. al. 2008). It finds that common law provides better property rights than civil law, and the difference breeds numerous economic consequences.

A dimension of the comparison of law systems relatively unexplored is how it affects state building. In the 19th century, state building was active in France and Prussia, the two states that had strong civil law tradition. Many prestigious French *Grande Ecoles* were established during the French Revolution, so they were open to commoners. They recruited students through an extremely competitive process, and students were automatically conferred civil servant status. They formed the backbone of the French technocracy. The Prussian state building in 18 and 19th centuries was equally if not more phenomenal, and Max Weber treated the Prussian bureaucracy as the model of the modern bureaucracy. By comparison, state building in England lagged behind: bureaucratic appointment was determined by patronage or purchase until 1854 (Xu, 2018). Max Weber also argued that the US bureaucracy was hopelessly corrupted. Fukuyama (2014) attributes the divergence in state building to the different sequences in state building and democratization. He argues that the early democratization before the rise of modern state induced elected politicians to buy votes through spoils system, which blocks the building of a meritocratic bureaucracy. My theory provides an alternative explanation based on legal traditions. The better property rights in common law countries reduced the incentive to engage in political patronage. In addition, law career path enjoys a much high prestige as investment in legal education provides better protection of property. On the contrary, the legal institution in civil law countries is intertwined with state apparatus. Together with relatively worse property rights, civil law tradition pushed commoner to actively enter the civil service that eases state building.

The explanation based on my theory also accounts for a few puzzles. For example, although the degree of property rights protection differs significantly between common law and civil law countries, development outcomes such as GDP per capita were quite similar. This contradicts classical economic theories (Hart, Grossman, 1986; Aghion, Holden, 2011) that establish a causal link between property rights and economic output. My theory shows that the adverse effect of worse property rights is alleviated by: 1. the *de facto* bargaining power of entrepreneurs are high because government mostly staffed by novice politicians has a low bargaining power; 2. state capacity in civil law countries is higher because of competitive recruitment. The analysis also gives an alternative explanation of the Anglo-American disdain of government, in contrast with Continental Europe's higher expectation (e.g., Alesina, Angeletos, 2005).

3.4 Conclusion

Mann (1986) asserts that “societies are constituted of multiple overlapping and intersecting sociospatial networks of power”. If we believe that Mann (1986)’s manifesto should be taken seriously, it is important to understand the interaction among political, economic, military, and ideological power networks. In this paper, I show how a meritocratic government and a thriving private economy reinforce each other, especially in societies without serious property rights. Although not explicitly modeled, the theory also touches on the crucial role of (Confucian) ideology that serves as a coordination device to propel a virtuous reinforcement between the market and the meritocracy.

However, these multiple equilibria without property rights mean that violent fluctuations are always a potential problem, as manifested by China’s striking swing between extremely impressive and disastrous economic performances¹⁰. This is especially an ominous concern in modern society, where traditional ideology such as Confucianism that propels the virtuous cycle has been under systematic attack. By contrast, equilibrium uniqueness reveals that decent legal property rights can deliver much more stable development outcomes. This is something to think about carefully, especially for regimes that rely heavily on political patronage to provide protection of property.

¹⁰Indeed, multiple equilibria is also a prominent modeling strategy to think about business cycles (see Farmer, 1999).

Bibliography

- [1] Acemoglu, Daron, Philippe Aghion, and Fabrizio Zilibotti. "Distance to Frontier, Selection, and Economic Growth." *Journal of the European Economic Association* 4, no. 1 (2006): 37-74.
- [2] Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya. "Radio and the Rise of the Nazis in Prewar Germany.", working paper (2014).
- [3] Aghion, Philippe, and Peter Howitt. "A Model of Growth through Creative Destruction". *Econometrica* 60, (1992): 323–351.
- [4] Aghion, Philippe, Ufuk Akcigit, and Peter Howitt. "What do We Learn from Schumpeterian Growth Theory?". No. w18824. National Bureau of Economic Research, 2013.
- [5] Aghion, Philippe, and Richard Holden. "Incomplete Contracts and the Theory of the Firm: What Have We Learned over the Past 25 Years?." *The Journal of Economic Perspectives* 25, no. 2 (2011): 181-197.
- [6] Aghion, Philippe, and Peter Howitt. *Endogenous Growth Theory*. MIT Press, 1998.
- [7] Alesina, Alberto, and George-Marios Angeletos. "Fairness and Redistribution." *The American Economic Review* 95, no. 4 (2005): 960-980.
- [8] Bai, Chong-En, Chang-Tai Hsieh, and Zheng Michael Song. "Crony Capitalism with Chinese Characteristics." The University of Chicago working paper (2014).
- [9] Bai, Gang eds., *Zhongguo Zhengzhi Zhidu Tongshi (A General History of China's Political Institutions)* (in 10 volumes), Shehui Kexue Chubanshe (Press of Social Science Literature), 1996.
- [10] Bai, Ying, and Titi Zhou. "'Mao's Last Revolution': A Loyalty–Competence Tradeoff." working paper (2014).
- [11] Bardhan, Pranab. "Decentralization of governance and development." *Journal of Economic Perspectives* (2002): 185-205.

- [12] Besley, Timothy, and Torsten Persson. *Pillars of Prosperity: The Political Economics of Development Clusters*. Princeton University Press, 2011.
- [13] Berkowitz, Daniel, and Wei Li. "Tax Rights in Transition Economies: A Tragedy of the Commons?." *Journal of Public Economics* 76, no. 3 (2000): 369-397.
- [14] Brandt, Loren, Johannes Van Biesebroeck, and Yifan Zhang. "Creative Accounting or Creative Destruction? Firm-level Productivity Growth in Chinese Manufacturing." *Journal of Development Economics* 97, no. 2 (2012): 339-351.
- [15] Caselli, Francesco, and Massimo Morelli. "Bad Politicians." *Journal of Public Economics* 88, no. 3 (2004): 759-782.
- [16] Chaffee, John W. *Branches of Heaven: A History of the Imperial Clan of Sung China*. Harvard University Press, 1999.
- [17] Chang, Chung-li. *The Chinese Gentry: Studies on Their Role in Nineteenth-century Chinese Society*. University of Washington Press, 1955
- [18] Chassang, Sylvain. "Building Routines: Learning, Cooperation, and the Dynamics of Incomplete Relational Contracts." *The American Economic Review* 100, no. 1 (2010): 448-465.
- [19] Chen, Shuo. "From Governance to Institutionalization: Political Selection from the Perspective of Central-local Relations in China—Past and Present (1368-2010)" Department of Economics, Fudan University Working Paper (2016).
- [20] Chen, Ting, James Kai-Sing Kung, and Chicheng Ma. "Long Live Keju! The Persistent Effects of China's Imperial Examination System." (2016), working paper.
- [21] Coser, Lewis A. "The Alien as a Servant of Power: Court Jews and Christian Renegades." *American Sociological Review* (1972): 574-581.
- [22] Dal Bo, Ernesto, Pedro Dal Bo, Jason Snyder. "Political Dynasties." *The Review of Economic Studies* 76, no. 1 (2009): 115-142.
- [23] Dal Bo, Ernesto, and Martin A. Rossi. "Term Length and the Effort of Politicians." *The Review of Economic Studies* 78, no. 4 (2011): 1237-1263.
- [24] Dragu, Tiberiu, Xiaochen Fan, and James Kuklinski. "Designing Checks and Balances." *Quarterly Journal of Political Science* 9, no. 1 (2014): 45-86.
- [25] Doyle, William. *Aristocracy and its Enemies in the Age of Revolution*. Oxford University Press, 2009.

- [26] Egorov, Georgy, and Konstantin Sonin. "Dictators And Their Viziers: Endogenizing The Loyalty–Competence Trade-Off." *Journal of the European Economic Association* 9, no. 5 (2011): 903-930.
- [27] Elman, Benjamin A. *A Cultural History of Civil Examinations in Late Imperial China*. University of California Press, 2000.
- [28] Engle, Robert F., and Clive WJ Granger. "Co-integration and Error Correction: Representation, Estimation, and Testing." *Econometrica: Journal of the Econometric Society* (1987): 251-276.
- [29] Eriksson, Tor, and Jaime Ortega. "The Adoption of Job Rotation: Testing the Theories." *Industrial and Labor Relations Review* 59, no. 4 (2006): 653-666.
- [30] Evans, Peter, and James E. Rauch. "Bureaucracy and Growth: A Cross-National Analysis of the Effects of 'Weberian' State Structures on Economic Growth." *American Sociological Review* (1999): 748-765.
- [31] Farmer, Roger EA. *The Macroeconomics of Self-fulfilling Prophecies*. MIT Press, 1999.
- [32] Ferguson, Thomas, and Hans-Joachim Voth. "Betting on Hitler—The Value of Political Connections in Nazi Germany." *The Quarterly Journal of Economics* 123, no. 1 (2008): 101-137.
- [33] Finer, Samuel Edward. *The History of Government from the Earliest Times: Ancient Monarchies and Empires*. Vol. 1. Oxford University Press, 1997.
- [34] Finer, Samuel Edward. *The History of Government from the Earliest Times: The Intermediate Ages*. Vol. 2. Oxford University Press, 1997.
- [35] Fukuyama, Francis. *The Origins of Political Order: From Pre-Human Times to the French Revolution*. Farrar, Straus and Giroux, 2011.
- [36] Fukuyama, Francis. *Political Order and Political Decay: From the Industrial Revolution to the Globalization of Democracy*. Macmillan, 2014.
- [37] Gehlbach, Scott. *Formal Models of Domestic Politics*. Cambridge University Press, 2013.
- [38] Gibbons, Robert. "Piece-rate Incentive Schemes." *Journal of Labor Economics* 5, no. 4, Part 1 (1987): 413-429.
- [39] Glahn, Richard von. *An Economic History of China: From Antiquity to the Nineteenth Century*. Cambridge University Press, 2016.
- [40] Glazer, Amihai. "Allies as Rivals: Internal and External Rent Seeking." *Journal of Economic Behavior & Organization* 48, no. 2 (2002): 155-162
- .

- [41] Gordon, Roger, and Wei Li. "Tax Structures in Developing Countries: Many Puzzles and a Possible Explanation." *Journal of Public Economics* 93, no. 7 (2009): 855-866.
- [42] Greif, Avner. *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade*. Cambridge University Press, 2006.
- [43] Grossman, Gregory. *The party as Manager and Entrepreneur*. No. 56. Wilson Center, Kennan Institute for Advanced Russian Studies, 1978.
- [44] Grossman, Sanford J., and Oliver D. Hart. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy* 94, no. 4 (1986): 691-719.
- [45] Guo, Gang. "China's Local Political Budget Cycles." *American Journal of Political Science* 53, no. 3 (2009): 621-632.
- [46] Guy, R. Kent. *Qing Governors and Their Provinces: the Evolution of Territorial Administration in China, 1644-1796*. University of Washington Press, 2010.
- [47] Hart, Oliver. *Firms, Contracts, and Financial Structure*. Clarendon Press, 1995.
- [48] Hartman, Charles. "Sung Government and Politics", in *The Cambridge History of China, Vol. 5: Sung China, 960-1279 AD, Part 2*, edited by John W. Chaffee, Denis Twitchett. Cambridge University Press, 2015.
- [49] Ho, Ping-ti. *The Ladder of Success in Imperial China: Aspects of Social Mobility, 1368-1911*. Vol. 1. New York: Columbia University Press, 1962.
- [50] Holmstrom, Bengt, and Paul Milgrom. "Multitask Principal-agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, and Organization* 7 (1991): 24-52.
- [51] Holmstrom, Bengt, and Jean Tirole. "The Theory of the Firm." *Handbook of Industrial Organization* 1 (1989): 61-133.
- [52] Hough, Jerry F. *The Soviet Prefects: The Local Party Organs in Industrial Decision-making*. Harvard University Press, 1969.
- [53] Jia, Ruixue. "Pollution for promotion." Working paper, (2014).
- [54] Jia, Ruixue, Masayuki Kudamatsu, and David Seim. "Political Selection in China: The Complementary Roles of Connections and Performance." *Journal of the European Economic Association* (2015).
- [55] Jin, Hehui, Yingyi Qian, and Barry R. Weingast. "Regional Decentralization and Fiscal Incentives: Federalism, Chinese Style." *Journal of Public Economics* 89, no. 9 (2005): 1719-1742.

- [56] Jin, Xin, and Michael Waldman. "Lateral Moves, Promotions, and Task-specific Human Capital: Theory and Evidence." (2017).
- [57] Johansen, Soren. "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica: Journal of the Econometric Society* (1991): 1551-1580.
- [58] Khwaja, Asim Ijaz, and Atif Mian. "Tracing the Impact of Bank Liquidity Shocks: Evidence from an Emerging Market." *The American Economic Review* 98, no. 4 (2008): 1413-1442.
- [59] Kung, James Kai-sing, and Shuo Chen. "The tragedy of the Nomenklatura: Career Incentives and Political Radicalism during China's Great Leap Famine." *American Political Science Review* 105, no. 1 (2011): 27-45.
- [60] Kuran, Timur. "Why the Middle East is Economically Underdeveloped: Historical Mechanisms of Institutional Stagnation." *The Journal of Economic Perspectives* 18, no. 3 (2004): 71-90.
- [61] Laffont, Jean-Jacques. *Incentives and Political Economy*. Oxford University Press, 2001.
- [62] La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer. "The Economic Consequences of Legal Origins." *Journal of Economic Literature* 46, no. 2 (2008): 285-332.
- [63] Laffont, Jean-Jacques. *Incentives and Political Economy*. Oxford University Press, 2000.
- [64] Lazear, Edward P., and Sherwin Rosen. "Rank-Order Tournaments as Optimum Labor Contracts." *Journal of Political Economy* 89, no. 5 (1981): 841-864.
- [65] Lazear, Edward P., and Paul Oyer. "Personnel Economics". In Robert Gibbons, and John Roberts. eds *Handbook of Organizational Economics*. (2012): 479-519.
- [66] Levin, Jonathan. "Relational Incentive Contracts." *The American Economic Review* 93, no. 3 (2003): 835-857.
- [67] Lieberthal, Kenneth. *Governing China: From Revolution to Reform (Second Edition)*. W. W. Norton & Company (2003).
- [68] Lorentzen, Peter. "China's Strategic Censorship." *American Journal of Political Science* 58, no. 2 (2014): 402-414.
- [69] Luo, Er'gang, *Luying Bingzhi (Record of Green Standard Army)*, Zhonghua Shuju (Zhonghua Book Company), 1984.
- [70] Ma, Debin and Jan Luiten van Zanden eds. *Law and Long-term Economic Change: a Eurasian Perspective*. Stanford University Press, 2011.

- [71] Magaloni, Beatriz. *Voting for Autocracy: Hegemonic Party survival and Its Demise in Mexico*. Cambridge: Cambridge University Press, 2006.
- [72] Magaloni, Beatriz, and Ruth Kricheli. "Political Order and One-party Rule." *Annual Review of Political Science* 13 (2010): 123-143.
- [73] Mann, Michael. *The Sources of Social Power, Vol. I: A History of Power from the Beginning to 1760 AD*. Cambridge University Press (1986).
- [74] Mao Tse-tung (Mao Zedong), "Building Stable Base Areas in the Northeast," Dec. 28, 1945, *Selected Works*, 4:82-83.
- [75] Maskin, Eric, Yingyi Qian, and Chenggang Xu. "Incentives, Information, and Organizational Form." *The Review of Economic Studies* 67, no. 2 (2000): 359-378.
- [76] McNeill, William H. *The Pursuit of Power: Technology, Armed Force, and Society*. University of Chicago Press, 1982.
- [77] North, Douglass C., and Barry R. Weingast. "Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England." *The Journal of Economic History* 49, no. 4 (1989): 803-832.
- [78] Olson, Mancur. *The Rise and Decline of Nations: Economic Growth, Stagflation and Social Rigidities*. Yale University Press, 1982.
- [79] Olson, Mancur. "Dictatorship, Democracy, and Development." *American Political Science Review* 87.03 (1993): 567-576.
- [80] Ortega, Jaime. "Job Rotation as a Learning Mechanism." *Management Science* 47, no. 10 (2001): 1361-1370.
- [81] Pepper, Suzanne. *Civil war in China: the Political Struggle 1945-1949*. Rowman & Littlefield Publishers, 1999.
- [82] People's Daily (Renmin Ribao), "Avoid Wrong Path in the Reform to Divide the Power of 'First Hand' (Fangzhi yibaoshou fenquan gaige zoupian)" (in Chinese), 2014.
- [83] Persson, Petra, and Ekaterina Zhuravskaya. "The Limits of Career Concerns in Federalism: Evidence from China." *Journal of European Economic Association*, forthcoming.
- [84] Persson, Torsten, Gerard Roland, and Guido Tabellini. "Separation of Powers and Political Accountability." *The Quarterly Journal of Economics* (1997): 1163-1202.
- [85] Persson, Torsten, Gerard Roland, and Guido Tabellini. "Comparative Politics and Public Finance." *Journal of political Economy* 108, no. 6 (2000): 1121-1161.

- [86] Persson, Torsten, and Guido Tabellini. *Political Economics: Explaining Public Policy* (2000). The MIT Press.
- [87] Qian, Yingyi. "Incentives and Loss of Control in an Optimal Hierarchy." *The Review of Economic Studies* 61, no. 3 (1994): 527-544.
- [88] Qian, Yingyi, and Gerard Roland. "Federalism and the Soft Budget Constraint." *American Economic Review* 88, no. 5 (1998).
- [89] Qian, Yingyi, Gerard Roland, and Chenggang Xu. "Coordination and Experimentation in M-form and U-form Organizations." *Journal of Political Economy* 114, no. 2 (2006): 366-402.
- [90] Qian, Yingyi, and Barry R. Weingast. "Federalism as a commitment to preserving market incentives." *The Journal of Economic Perspectives* (1997): 83-92.
- [91] Raith, Michael, and Guido Friebel. "Managers, Training, and Internal Labor Markets." (2014).
- [92] Roland, Gerard. *Transition and Economics: Politics, Markets, and Firms*. MIT press, 2000.
- [93] Rose-Ackerman, Susan, and Bonnie J. Palifka. *Corruption and Government: Causes, Consequences, and Reform*. Cambridge University Press, 2016.
- [94] Sanchez de la Sierra, Raul. "On the Origins of the State: Stationary Bandits and Taxation in Eastern Congo". (2017).
- [95] Shambaugh, David L. *China's Communist Party: Atrophy and Adaptation*. University of California Press, 2008.
- [96] Shih, Victor, Christopher Adolph, and Mingxing Liu. "Getting ahead in the communist party: explaining the advancement of central committee members in China." *American Political Science Review* 106, no. 01 (2012): 166-187.
- [97] Shirk, Susan L. "The Chinese Political System and the Political Strategy of Economic Reform." in *Bureaucracy, politics, and decision making in post-Mao China* , (1992): 60-93, edited by Kenneth G. Lieberthal and David M. Lampton.
- [98] Shirk, Susan L. *The Political Logic of Economic Reform in China*. University of California Press, 1993.
- [99] Svobik, Milan W. *The Politics of Authoritarian Rule*. Cambridge University Press, 2012.
- [100] Tackett, Nicolas. *The Destruction of the Medieval Chinese Aristocracy*. Harvard University Press, 2014.

- [101] Tsai, Kellee S. *Capitalism without Democracy: The Private Sector in Contemporary China*. Cornell University Press, 2007.
- [102] Tran Tri Thu Thong, "Local Politics and Democracy in a Muong Ethnic Community", in Kerkvliet, Benedict J. Tria, and David G. Marr, eds. *Beyond Hanoi: Local Government in Vietnam*. Institute of Southeast Asian Studies, 2004.
- [103] Voigtlaeder, Nico, and Hans-Joachim Voth. Highway to Hitler. No. w20150. National Bureau of Economic Research, 2014.
- [104] Wang, Shaoda. "Fiscal Competition and Coordination: Evidence from China." Working Paper. (2016).
- [105] Watson, Rubie Sharon. *Marriage and Inequality in Chinese Society*. University of California Press, 1991.
- [106] Wedeman, Andrew. "Crossing the River by Feeling the Stones or Carried Across the Current? The Transformation of the Chinese Automotive Sector", in *Beyond the Middle Kingdom. Comparative Perspectives on China's Capitalist Transformation*, edited by Scott Kennedy. Stanford: Stanford University Press (2011): 66-88.
- [107] Weingast, Barry R. "The Economic Role of Political Institutions: Market-preserving Federalism and Economic Development." *Journal of Law, Economics, & Organization* (1995): 1-31.
- [108] Williamson, Oliver E. *The Economic Institutions of Capitalism*. Simon and Schuster, 1985.
- [109] Xi, Tianyang, Yang Yao, and Muyang Zhang. "Bureaucratic Capability and Political Opportunism: An Empirical Investigation of City Officials in China." Working Paper. (2017).
- [110] Xu, Chenggang. "The Fundamental Institutions of China's Reforms and Development." *Journal of Economic Literature* (2011): 1076-1151.
- [111] Xu, Guo. "The Costs of Patronage: Evidence from the British Empire." Job Market Paper (2017).
- [112] Yang, Xuewei eds., *Zhongguo Kaoshi Tongshi (A General History of Examinations in China)* (in 4 volumes), Shoudu Shifan Daxue Chubanshe (Press of Capital Normal University), 1996.
- [113] Young, Alwyn. "The Razor's Edge: Distortions and Incremental Reform in the People's Republic of China." *The Quarterly Journal of Economics* 115, no. 4 (2000): 1091-1135.
- [114] Zhao, Yuezhi. "The struggle for socialism in China: The Bo Xilai saga and beyond." *Monthly Review* 64, no. 5 (2012): 1.

Appendix A

Appendices of *Rotation, Meritocracy, and Property Rights*

A.1 Proofs for Propositions in the Text

Proposition 1 *The entrepreneur will invest under sufficiently strong performance-based rewards.*

Algebraically, denote \hat{R} such that $1 - F(w^(\hat{R}))\{E[y|y \geq w^*(\hat{R})] - w^*(\hat{R})\} = k$. The entrepreneur expects a non-negative profit if $R \geq \hat{R}$.*

Proof: By Assumption 1, $k > [1 - F(w^*(R = 0))]\{E(y|y \geq w^*(R = 0)) - w^*(R = 0)\}$. Recall that $\frac{w^*f(w^*)}{1-F(w^*)} = \frac{1}{1+R}$. As $R \rightarrow \infty$, $\frac{w^*f(w^*)}{1-F(w^*)} \rightarrow 0$. So $\lim_{R \rightarrow \infty} w^*(R) = 0$. $\lim_{R \rightarrow \infty} [1 - F(w^*(R))]\{E(y|y \geq w^*(R)) - w^*(R)\} = [1 - F(0)]\{E[y|y \geq 0] - 0\} = E[y] > k$ by Assumption 1.

Notice that:

$$\begin{aligned} & \frac{\partial}{\partial R} [1 - F(w^*)]\{E(y|y \geq w^*(R)) - w^*\} = \\ & \frac{\partial}{\partial R} \left\{ \int_{w^*}^{\bar{y}} z f(z) dz - [1 - F(w^*(R))]w^* \right\} \\ & = -w^*(R)f(w^*) \frac{dw^*}{dR} - \{[1 - F(w^*(R)) - w^*(R)f(w^*)]\} \frac{dw^*(R)}{dR} \end{aligned}$$

$$= -wf(w^*) \frac{dw^*}{dR} > 0 \text{ as } \frac{dw^*}{dR} < 0$$

So $\exists! \hat{R}$ such that $\forall R \geq \hat{R}, k \leq [1 - F(w^*)]\{E(y|y \geq w^*) - w^*\}$. This covers the sunk cost of investment for the entrepreneur.

Q.E.D.

Proposition 2 1. *Sufficiently strong rotation and performance-based rewards incentivizes the entrepreneur to invest.*

In algebra, denote $\hat{R}(k)$ and $\underline{\pi}(R, c)$ such that $[1 - F(w^*(\hat{R}))]\{E(y|y \geq w^*(\hat{R})) - w^*(\hat{R})\} = k$ and $\underline{\pi}(R, c) = 1 - \frac{c}{(1+R)E[y] - [1 - F(w^*(R))]\{w^* + E[Ry | w^*(R) \leq y \}]}$. If $R > \hat{R}(k)$ and $\pi \geq \underline{\pi}(R, c)$, the entrepreneur will reap non-negative return.

2. $\frac{\partial \underline{\pi}(R, c)}{\partial R} > 0$: the minimum rotation frequency increases with stronger performance-based rewards. If minimum rotation frequency does not change, stronger performance-based rewards incentivize politician in $t = 1$ to pay c , and politician in $t = 2$ will fully predate entrepreneur.

Proof: The first claim: If $\pi < \underline{\pi}$, the politician pays c , and the entrepreneur anticipates a net profit of $-k$ if he invests.

If $\pi \geq \underline{\pi}$, the politician will not pay c . The entrepreneur anticipates a net profit of $[1 - F(w^*(R))]\{E(y|y \geq w^*(R)) - w^*(R)\} - k$.

Proposition 1 already shows that under an uninformed politician, $\exists! \hat{R}$ such that $\forall R \geq \hat{R}, k \leq [1 - F(w^*)]\{E(y|y \geq w^*) - w^*\}$. This covers the sunk cost of investment for the entrepreneur.

The second claim: $\underline{\pi} = 1 - \frac{c}{(1+R)E[y] - [1 - F(w)]\{w + E[Ry | y \geq w \]}}$. So it is sufficient to prove that $\Delta \equiv (1 + R)E[y] - \max_w [1 - F(w)]\{w + E[Ry | y \geq w \]}$ is an increasing function of R . Apply Envelope Theorem,

$$\begin{aligned} \frac{\partial \Delta}{\partial R} &= E[y] - [1 - F(w)]E[y|y \geq w^*] \\ &= \int_{\underline{y}}^{\bar{y}} zf(z)dz - [1 - F(w)] \frac{\int_{w^*}^{\bar{y}} zf(z)dz}{[1 - F(w)]} = \int_{\underline{y}}^{w^*} zf(z)dz > 0 \end{aligned}$$

Q.E.D.

Proposition 3 Denote $\underline{\pi}(R) \equiv 1 - \frac{c}{(1+R)E[y] - [1-F(w^*(R))]\{w^*(R) + E[Ry | y \geq w^*(R)]\}}$,

$\bar{\pi}(R) \equiv 1 - \frac{\eta k}{[1-F(w^*(R))]\{w^*(R) + E[Ry | y \geq w^*(R)]\}}$. For $R \geq 0$, $\exists \pi \in [\underline{\pi}(R), \bar{\pi}(R)]$, such that the local politician neither acquire information nor steal capital.

Moreover, if $R' < R''$, then $\underline{\pi}(R') < \underline{\pi}(R'')$, and $\bar{\pi}(R') < \bar{\pi}(R'')$, In other words, the complementarity between rotation and performance-based rewards persist.

Proof:

To have $\underline{\pi} \leq \bar{\pi}$, we need:

$$1 - \frac{c}{(1+R)E[y] - [1-F(w^*)]\{w^* + E[Ry | y \geq w^*]\}} \leq 1 - \frac{\eta k}{[1-F(w^*)]\{w^* + E[Ry | y \geq w^*]\}}$$

$$D \equiv (c + \eta k)([1-F(w^*)]\{w^* + E[Ry | y \geq w^*]\}) - \eta k(1+R)E[y] \geq 0$$

for $R = 0$, we have $D(R = 0) = (c + \eta k)(1 - F(w^*(0)))w^*(0) - \eta kE[y] > 0$, or $\frac{c}{\eta k} > \frac{E[y]}{[1-F(w^*(0))]w^*(0)} - 1$.

Now we need $\Delta(R)$ to be an increasing function:

$$\Delta'(R) = (c + \eta k)([1-F(w^*)]E[y | y \geq w^*]) - \eta kE[y] > 0$$

$$\frac{c}{\eta k} > \frac{\int_y^{w^*} z f(z) dz}{\int_{w^*}^y z f(z) dz} = \frac{F(w^*)E[y | y \leq w^*]}{[(1-F(w^*))E[y | y \geq w^*]}$$

It is easy to see that $\max_R \frac{\int_y^{w^*} z f(z) dz}{\int_{w^*}^y z f(z) dz} = \frac{\int_y^{w^*(R=0)} z f(z) dz}{\int_{w^*(R=0)}^y z f(z) dz}$. So it is sufficient that:

$$\frac{c}{\eta k} > \frac{F(w^*(R=0))E[y | y \leq w^*(R=0)]}{[(1-F(w^*(R=0)))E[y | y \geq w^*(R=0)]}$$

The second claim can be proved by noting that both $\underline{\pi}(R)$ and $\bar{\pi}(R)$ are increasing function of R . $\underline{\pi}'(R) > 0$ has been shown in Theorem 1. For $\bar{\pi}'(R) > 0$, it is shown in the text.

Q.E.D.

Proposition 4: *Suppose the politician paid c and is thus informed about the old project that has been invested.*

1. *She will endorse the new project with probability $F(\frac{1}{1+R}[1 - G(\tilde{w})]\{\tilde{w} + E_g[Ry|y \geq \tilde{w}]\})$, which increases with R : performance-based rewards encourage adaptation.*

2. $\forall R < \infty$, $F(\frac{1}{1+R}[1 - G(\tilde{w})]\{\tilde{w} + E_g[Ry|y \geq \tilde{w}]\}) < F(E_g(y))$, *the first-best probability of adaptation. An informed politician holds entrenched interests in the old project, not matter how strong performance-based rewards are.*

Proof: The politician will get $(1 + R)y$ from the old project, while get an expected value of $[1 - G(\tilde{w})]\{\tilde{w} + E_g[Ry|y \geq \tilde{w}]\}$ from the new project. He will endorse the new project if $(1 + R)y \leq [1 - G(\tilde{w})]\{\tilde{w} + E_g[Ry|y \geq \tilde{w}]\}$, which happens with probability $F(\frac{1}{1+R}[1 - G(\tilde{w})]\{\tilde{w} + E_g[Ry|y \geq \tilde{w}]\})$.

For the first claim:

$$\begin{aligned} & \partial[F(\frac{1}{1+R}[1 - G(\tilde{w})]\{\tilde{w} + E_g[Ry|y \geq \tilde{w}]\})]/\partial R = \\ & = f\{-\frac{1}{(1+R)^2}[1 - G(\tilde{w})]\{\tilde{w} + E_g[Ry|y \geq \tilde{w}]\} + \frac{1}{1+R} \int_{\tilde{w}}^{\bar{y}} zg(z)dz\} \\ & = f\{\frac{(1+R) \int_{\tilde{w}}^{\bar{y}} zg(z)dz - [1 - G(\tilde{w})]\tilde{w} - R \int_{\tilde{w}}^{\bar{y}} zg(z)dz}{(1+R)^2}\} = f\frac{\int_{\tilde{w}}^{\bar{y}}(z - \tilde{w})g(z)dz}{(1+R)^2} > 0 \end{aligned}$$

So the upper bound of $F(\frac{1}{1+R}[1 - G(\tilde{w})]\{\tilde{w} + E_g[Ry|y \geq \tilde{w}]\})$ is $\lim_{R \rightarrow \infty} F(\frac{1}{1+R}[1 - G(\tilde{w})]\{\tilde{w} + E_g[Ry|y \geq \tilde{w}]\}) = F(\lim_{R \rightarrow \infty} \frac{[1 - G(\tilde{w})]\{\tilde{w} + E_g[Ry|y \geq \tilde{w}]\}}{1+R})$.

Using L'Hospital's Rule, $\lim_{R \rightarrow \infty} \frac{[1 - G(\tilde{w})]\{\tilde{w} + E_g[Ry|y \geq \tilde{w}]\}}{1+R} = \lim_{R \rightarrow \infty} \int_{\tilde{w}(R)}^{\bar{y}} zg(z)dz = \int_{\underline{y}}^{\bar{y}} zg(z)dz = E_g(y)$. Notice that $E_g(y) < \bar{y}$, so $F(E_g(y)) < 1$.

Q.E.D.

Proposition 5 *Suppose the politician didn't pay the cost of learning and a new project arrives with $G(y) \leq F(y)$, $\forall y$. Then the politician will always endorse the new project.*

Proof: By definition,

$$U_2 \equiv \max_w [1 - G(w)] \{w + E_g[Ry | y \geq w]\} = [1 - G(\tilde{w})] \{ \tilde{w} + E_g[Ry | y \geq \tilde{w}]\} \geq$$

$$[1 - G(w^*)] \{w^* + E_g[Ry | y \geq w^*]\} = [1 - G(w^*)] w^* + R \int_{w^*}^{\bar{y}} z g(z) dz$$

Notice that $\int_{w^*}^{\bar{y}} z g(z) dz \geq \int_{w^*}^{\bar{y}} z f(z) dz$ because of F.O.S.D.¹ Also, because $G(y) \leq F(y) \forall y$, $[1 - G(w^*)] w^* \geq [1 - F(w^*)] w^*$. So we have:

$$U_2 \geq [1 - G(w^*)] w^* + R \int_{w^*}^{\bar{y}} z g(z) dz \geq [1 - F(w^*)] w^* + R \int_{w^*}^{\bar{y}} z f(z) dz \equiv U_1$$

So it is optimal for the politician always to endorse the new project.

Q.E.D.

Proposition 6 *1. Sufficiently frequent rotation plus sufficiently strong performance-based rewards will induce a benevolent local government. The old entrepreneur will invest accordingly.*

2. If $p < \frac{\int_y^{w^(0)} z f(z) dz}{\int_y^{w^*(0)} z f(z) dz + \int_y^{\bar{y}} z g(z) dz}$, then complementarity between rotation and performance-based rewards retains. Specifically, if x increases, then $\underline{\pi}$ increases, where:*

$$\underline{\pi} = 1 - \frac{c}{(1-p)\{(1+R)E(y) - U_1\} + p\{(1-\Pi)[E_f[(1+R)y | (1+R)y \geq U_2] - U_2]\}} \quad (\text{A.1})$$

¹This comes from an equivalent definition of first order stochastic dominance. Specifically, “ G f.o.s.d. F ” \Leftrightarrow “ $G(y) \leq F(y) \forall y$ ” \Leftrightarrow “For every weakly increasing utility function u , $\int u(x) dG \geq \int u(x) dF$.”

Proof: Assume the politician doesn't learn the profitability of the first-period project. To induce investment from the first-period entrepreneur with $k > 0$, it must be that $k \leq (1 - p)[1 - F(w^*)]\{E(y|y \geq w^*) - w^*\}$. Proof is similar to in the first part of Proposition 2, though the conditions on \hat{R} is stronger.

From (5), we immediately get $\pi \geq 1 - \frac{c}{(1-p)\{(1+R)E(y)-U_1\}+p\{(1-\Pi)[E_f[(1+R)y|(1+R)y \geq U_2]-U_2]\}} \equiv \underline{\pi}$.

To prove second part, it is sufficient to show that $Q \equiv (1 - p)\{(1 + R)E(y) - U_1\} + p\{(1 - \Pi)[E_f[(1 + R)y|(1 + R)y \geq U_2] - U_2]\}$ is an increasing function of R .

$$\begin{aligned} \frac{\partial Q}{\partial R} &= (1 - p)\{E(y) - \int_{w^*}^{\bar{y}} z f(z) dz\} + p \frac{\partial}{\partial R} \left\{ (1 + R) \int_{\frac{U_2(R)}{1+R}}^{\bar{y}} z f(z) dz - (1 - F\left(\frac{U_2}{1+R}\right)) U_2 \right\} \\ &= (1 - p) \int_{\underline{y}}^{w^*} z f(z) dz + p \frac{\partial}{\partial R} \left\{ (1 + R) \int_{\frac{U_2(R)}{1+R}}^{\bar{y}} z f(z) dz - (1 - F\left(\frac{U_2}{1+R}\right)) U_2 \right\} \end{aligned}$$

Denote $M \equiv \frac{\partial}{\partial R} \left\{ (1 + R) \int_{\frac{U_2(R)}{1+R}}^{\bar{y}} z f(z) dz - (1 - F\left(\frac{U_2}{1+R}\right)) U_2 \right\}$:

$$\begin{aligned} M &= \int_{\frac{U_2(R)}{1+R}}^{\bar{y}} z f(z) dz - (1+R) \frac{U_2(R)}{1+R} f\left[\frac{U_2(R)}{1+R}\right] \frac{\partial}{\partial R} \left[\frac{U_2(R)}{1+R}\right] - \frac{\partial}{\partial R} F\left[\frac{U_2(R)}{1+R}\right] U_2(R) - (1 - F\left[\frac{U_2(R)}{1+R}\right]) \int_{\tilde{w}}^{\bar{y}} z g(z) dz \\ &= \int_{\frac{U_2(R)}{1+R}}^{\bar{y}} z f(z) dz - U_2(R) f\left[\frac{U_2(R)}{1+R}\right] \frac{\partial}{\partial R} \left[\frac{U_2(R)}{1+R}\right] - f\left[\frac{U_2(R)}{1+R}\right] U_2(R) \frac{\partial}{\partial R} \left[\frac{U_2(R)}{1+R}\right] - (1 - F\left[\frac{U_2(R)}{1+R}\right]) \int_{\tilde{w}}^{\bar{y}} z g(z) dz \\ &= \int_{\frac{U_2(R)}{1+R}}^{\bar{y}} z f(z) dz - (1 - F\left[\frac{U_2(R)}{1+R}\right]) \int_{\tilde{w}}^{\bar{y}} z g(z) dz \geq - \int_{\underline{y}}^{\bar{y}} z g(z) dz \end{aligned}$$

So:

$$\frac{\partial Q}{\partial R} \geq (1 - p) \int_{\underline{y}}^{w^*} z f(z) dz - p \int_{\underline{y}}^{\bar{y}} z g(z) dz \geq (1 - p) \int_{\underline{y}}^{w^*(0)} z f(z) dz - p \int_{\underline{y}}^{\bar{y}} z g(z) dz$$

We can immediately see that when $p < \frac{\int_{\underline{y}}^{w^*(0)} z f(z) dz}{\int_{\underline{y}}^{w^*(0)} z f(z) dz + \int_{\underline{y}}^{\bar{y}} z g(z) dz}$, $\frac{\partial Q}{\partial R} \geq 0$.

Q.E.D.

Proposition 7 Denote:

$$\underline{\pi}_1(R) = 1 - \frac{c}{(1+R)E[y] - \max_w [1 - F(w)]\{w + E[Ry | y \geq w]\}}$$

$$\underline{\pi}_2(R) : (1 - \underline{\pi}_2(R)) \max_{w_2} [1 - F(w_2)]\{w_2 + E[Ry | y \geq w_2]\} =$$

$$\max_{w_1} \{-F(w_1)c + (1 - \underline{\pi}_2(R))\{[1 - F(w_1)]\{w_1 + E[Ry | y \geq w_1]\} + F(w_1)(1+R)E[y | y \geq w_1]\}\}$$

$$\hat{R} : k = [1 - F(w^*(\hat{R}))]\{E[y | y \geq w^*(\hat{R})] - w^*(\hat{R})\}, \frac{w^*(\hat{R})f(w^*(\hat{R}))}{1 - F(w^*(\hat{R}))} = \frac{1}{1+R}$$

1. If $\pi \geq \underline{\pi}_1(R)$, $\pi \geq \underline{\pi}_2(R)$, and $R \geq \hat{R}$, the politician does not pay c to learn about y , and the entrepreneur makes the investment k .

2.

$$\frac{\partial \underline{\pi}_1(R)}{\partial R} > 0, \frac{\partial \underline{\pi}_2(R)}{\partial R} > 0$$

So with stronger performance-based rewards, minimum rotation frequency $\underline{\Pi}(R) = \max\{\underline{\pi}_1(R), \underline{\pi}_2(R)\}$ also rises to guarantee the entrepreneur's ex ante investment.

Proof:

Denote:

$$w^* = \operatorname{argmax}_w [1 - F(w)]\{w + E[Ry | y \geq w]\}$$

w^* is the optimal extraction at $t = 2$ for an uninformed politician with no entrepreneur accepting w_1 .

Notice that at $t = 2$ newly appointed politician also honors w'_1 , the contract that the veteran politician made in $t = 1$. Also, if the entrepreneur rejects w'_1 , the information set for a newly appointed politician is exactly the same as the veteran politician. They also have the same utility. So they must offer the same w'_2 .

In $t = 2$, the dominant strategy is for entrepreneur to accept w'_2 if $y \geq w'_2$ and entrepreneur has not accepted in $t = 1$.

Back to $t = 1$. For all $w'_1 \in [y, \bar{y}]$, For all $w'_1 \in [y, \bar{y}]$, I now derive the set of strategies that can satisfy sequential rationality. This enables the politician to predict her payoff if she chooses a specific w'_1 .

1. Fix a $w'_1 \geq w^*$: There are two cases: the politician finds it sequentially rational to pay c or pay 0. The strategy at learning stage is a function of w'_1 . Denote $learn(w'_1) = 1$ if the politician learns after w'_1 and $learn(w'_1) = 0$ otherwise.

1.1 Suppose $learn(w'_1) = 1$: Given that $learn(w'_1) = 1$ is sequentially rational, the entrepreneur accepts any $y - w'_1 \geq 0$ as in $t = 2$ he gets 0 surplus.

Given that $learn(w'_1) = 1$ is sequentially rational, the politician asks to extract $w'_2 = y$ for any entrepreneur who did not accept w'_1 .

The utility of w'_1 if $learn(w'_1) = 1$ is sequentially rational is:

$$U(\{w'_1, learn(w'_1) = 1\}) = -F(w'_1)c + (1-\pi)\{[1-F(w'_1)]\{w'_1 + E[Ry | y \geq w'_1]\} + (1+R)E[y | y \leq w'_1]\} + \pi\hat{U}$$

1.1 Suppose $learn(w'_1) = 0$: If $y - w'_1 > y - w'_2$ or $w'_1 < w'_2$, entrepreneur accepts w'_1 if $y \geq w'_1$. This is impossible as:

$$w'_2 = \operatorname{argmax}_w \operatorname{Prob}(w \leq y | y \leq w'_1) \{w + E[Ry | w \leq y, \text{ given } y \leq w'_1]\} < w'_1$$

If $y - w'_1 \leq y - w'_2$ or $w'_1 \geq w'_2$, entrepreneur rejects w'_1 . In this case:

$$w'_2 = \operatorname{argmax}_w [1 - F(w)] \{w + E[Ry | y \geq w]\} = w^*$$

Indeed $w'_1 \geq w'_2 = w^*$, consistent with the assumption that $w'_1 \geq w^*$.

So given that $learn(w'_1) = 0$ is sequentially rational:

$$U(\{w'_1, learn(w'_1) = 0\}) = (1 - \pi)[1 - F(w^*)] \{w^* + E[Ry | y \geq w^*]\} + \pi\hat{U}$$

So by choosing $w'_1 \geq w^*$, the maximum utility the politician can get is:

$$\max \left\{ -F(w'_1)c + (1 - \pi)\{(1 + R)E[y | y \leq w'_1] + [1 - F(w'_1)]\{w'_1 + E[Ry | y \geq w'_1]\} + \pi\hat{U}\} \right. \\ \left. , (1 - \pi)[1 - F(w^*)] \{w^* + E[Ry | y \geq w^*]\} + \pi\hat{U} \right\}$$

2. Fix a $w'_1 < w^*$:

2.1 Suppose $learn(w'_1) = 1$: The algebra is the same as 1.1. We have:

$$U(\{w'_1, learn(w'_1) = 1\}) = -F(w'_1)c + (1-\pi)\{[1-F(w'_1)]\{w'_1 + E[Ry|y \geq w'_1]\} + (1+R)E[y|y \leq w'_1]\} + \pi\hat{U}$$

1.1 Suppose $learn(w'_1) = 0$: If $y - w'_1 > y - w'_2$ or $w'_1 < w'_2$, entrepreneur accepts w'_1 if $y \geq w'_1$. This is impossible as:

$$w'_2 = \operatorname{argmax}_w \operatorname{Prob}(w \geq w | w \leq w'_1) \{w + E[Ry | w \leq y, \text{given } y \leq w'_1]\} < w'_1$$

If $y - w'_1 < y - w'_2$ or $w'_1 > w'_2$, entrepreneur rejects w'_1 . In this case:

$$w'_2 = \operatorname{argmax}_w [1 - F(w)] \{w + E[Ry | y \geq w]\} = w^*$$

Another contradiction as we assume $w'_1 > w'_2$ and $w^* > w'_1$.

The only possible case is $w'_1 = w'_2$. Given Assumption 4, suppose that the cutoff is \hat{w} such that an entrepreneur with $y \geq \hat{w}$ accepts w'_1 . We have $\hat{w} > w'_1$. Entrepreneur with $w'_1 \leq y < \hat{w}$ accepts w'_2 . To make $w'_2 = w'_1$ sequentially rational, it must be that \hat{w} satisfies:

$$w'_1 = w'_2 = \operatorname{argmax}_w \operatorname{prob}(w \leq y, \text{given } y \leq \hat{w}) \{w + E[Ry | w \leq y, \text{given } y \leq \hat{w}]\}$$

The probability density function of $f(y | y \leq \hat{w}) = \frac{f(y)}{F(\hat{w})} \equiv g(y, \hat{w})$ with support on $[y, \hat{w}]$. We have:

$$\operatorname{prob}(w \leq y, \text{given } y \leq \hat{w}) = \int_w^{\hat{w}} f(z | z \leq \hat{w}) dz = \frac{\int_w^{\hat{w}} f(z)}{1 - F(\hat{w})} dz = \frac{F(\hat{w}) - F(w)}{F(\hat{w})}$$

$$E[y | w \leq y, \text{given } y \leq \hat{w}] = \int_w^{\hat{w}} z g(z, \hat{w} | z \geq w) dz$$

$$g(z, \hat{w} | z \geq w) = \frac{g(z, \hat{w})}{1 - G(w; \hat{w})} = \frac{f(z)/F(\hat{w})}{1 - \int_y^w \frac{f(v)}{F(\hat{w})} dv} = \frac{f(z)/F(\hat{w})}{1 - \frac{F(w)}{F(\hat{w})}} = \frac{f(z)}{F(\hat{w}) - F(w)}$$

Thus we have:

$$E[y|w \leq y, \text{ given } y \leq \hat{w}] = \frac{\int_w^{\hat{w}} z f(z) dz}{F(\hat{w}) - F(w)}$$

So:

$$\text{prob}(w \leq y, \text{ given } y \leq \hat{w}) \{w + RE[y|w \leq y, \text{ given } y \leq \hat{w}]\}$$

$$\begin{aligned} &= \frac{F(\hat{w}) - F(w)}{F(\hat{w})} w + R \frac{F(\hat{w}) - F(w)}{F(\hat{w})} \frac{\int_w^{\hat{w}} z f(z) dz}{F(\hat{w}) - F(w)} \\ &= \frac{[F(\hat{w}) - F(w)]w + R \int_w^{\hat{w}} z f(z) dz}{F(\hat{w})} \end{aligned}$$

As $w'_1 = w'_2 = \text{argmax}_w = \frac{[F(\hat{w}) - F(w)]w + \int_w^{\hat{w}} z f(z) dz}{F(\hat{w})}$, F.O.C. tells us that:

$$\frac{w'_1 f(w'_1)}{F(\hat{w}) - F(w'_1)} = \frac{1}{1 + R}$$

As $\frac{1}{1+R} = \frac{w^* f(w^*)}{1-F(w^*)}$, it is indeed the case that $w'_1 < w^*$: $\frac{1}{1+R} = \frac{w^* f(w^*)}{1-F(w^*)} = \frac{w'_1 f(w'_1)}{F(\hat{w}) - F(w'_1)} > \frac{w'_1 f(w'_1)}{1-F(w'_1)}$.

So if $\text{learn}(w'_1) = 0$ is sequentially rational following a $w'_1 < w^*$, then:

$$U(\{w'_1, w'_1(\text{learn}) = 0\}) = (1 - \pi) \{ [1 - F(\hat{w})] \{w'_1 +$$

²Note that when $w'_1 \geq w^*$, I did not discuss this equilibrium where $w'_1 = w'_2$, $y \geq \hat{w}$ accepts w'_1 , and $w'_2 \leq y < w'_1$ accepts w'_2 . Such equilibrium does not exist, as it requires:

$$w'_1 = w'_2 = \text{argmax}_w \frac{[F(\hat{w}) - F(w)]w + R \int_w^{\hat{w}} z f(z) dz}{F(\hat{w})}$$

We still need $\frac{w^* f(w^*)}{1-F(w^*)} = \frac{1}{1+x} = \frac{w'_1 f(w'_1)}{F(\hat{w}) - F(w'_1)} > \frac{w'_1 f(w'_1)}{1-F(w'_1)}$, so $w'_1 < w^*$. A contradiction to the assumption that $w'_1 \geq w^*$.

$$E[Ry|y \geq \hat{w}] + F(\hat{w}) \frac{[F(\hat{w}) - F(w'_1)]w'_1 + R \int_{w'_1}^{\hat{w}} z f(z) dz}{F(\hat{w})} \} + \pi \tilde{U}$$

$$U(\{w'_1, learn(w'_1) = 0\}) = (1-\pi)\{[1-F(\hat{w})]\{w'_1 + E[Ry|y \geq \hat{w}]\} + [F(\hat{w}) - F(w)]w + R \int_w^{\hat{w}} z f(z) dz\} + \pi \tilde{U}$$

The maximum utility of the politician from a $w'_1 < w^*$ is:

$$max \left\{ -F(w'_1)c + (1-\pi)\{[1-F(w'_1)]\{w'_1 + E[Ry|y \geq w'_1]\} + (1+R)E[y|y \leq w'_1] + \pi \hat{U}\}, \right.$$

$$\left. (1-\pi)\{[1-F(\hat{w})]\{w'_1 + E[Ry|y \geq \hat{w}]\} + [F(\hat{w}) - F(w)]w + R \int_w^{\hat{w}} z f(z) dz\} + \pi \hat{U} \right\}$$

We have derived the function of politician's payoff at $t = 1$ for any w'_1 and learning decision. We can use this function to find the sequential equilibrium strategy.

I want to ensure that $learn(w'_1) = 0$ is the strategy on equilibrium path. That gives a utility of:

$$U(w'_1, learn(w'_1) = 0) =$$

$$\begin{cases} (1-\pi)[1-F(w^*)]\{w^* + E[Ry|y \geq w^*]\} + \pi \hat{U} & w'_1 \geq w^* \\ (1-\pi)\{[1-F(\hat{w})]\{w_1 + E[Ry|y \geq \hat{w}]\} + \{[F(\hat{w}) - F(w)]w + R \int_w^{\hat{w}} z f(z) dz\} + \pi \hat{U} & w'_1 < w^* \end{cases}$$

Notice that:

$$[1-F(\hat{w})]\{w_1 + E[Ry|y \geq \hat{w}]\} + \{[F(\hat{w}) - F(w)]w + R \int_w^{\hat{w}} z f(z) dz\}$$

$$= w'_1 \int_{\hat{w}}^{\bar{y}} f(z) dz + R \int_{\hat{w}}^{\bar{y}} z f(z) dz + w'_1 \int_{w'_1}^{\hat{w}} f(z) dz + R \int_{w'_1}^{\hat{w}} z f(z) dz$$

$$\begin{aligned}
 &= w'_1 \int_{w'_1}^{\bar{y}} f(z)dz + R \int_{w'_1}^{\bar{y}} z f(z)dz = [1 - F(w'_1)]\{w'_1 + E[Ry | y \geq w'_1]\} \\
 &\leq [1 - F(w^*)]\{w^* + E[Ry | y \geq w^*]\}
 \end{aligned}$$

So it is always optimal to choose $w'_1 \geq w^*$.

In that case, I want to make sure that at the stage of learning, it is not optimal to learn:

$$-c + (1 - \pi)(1 + R)E[y] \leq (1 - \pi)[1 - F(w^*)]\{w^* + E[Ry | y \geq w^*]\}$$

And at the stage to decide w'_1 , the utility from not learning is also larger than the highest utility from learning:

$$\begin{aligned}
 &(1 - \pi) \max_{w_2} [1 - F(w_2)]\{w_2 + E[Ry | y \geq w_2]\} \geq \\
 &\max_{w_1} \{-F(w_1)c + (1 - \pi)\{[1 - F(w_1)]\{w_1 + E[Ry | y \geq w_1]\} + F(w_1)(1 + R)E[y | y \geq w_1]\}\}
 \end{aligned}$$

$\underline{\pi}(R)$ is defined as:

$$\begin{aligned}
 &(1 - \underline{\pi}(R)) \max_{w_2} [1 - F(w_2)]\{w_2 + E[Ry | y \geq w_2]\} = \\
 &\max_{w_1} \{-F(w_1)c + (1 - \underline{\pi}(R))\{[1 - F(w_1)]\{w_1 + E[Ry | y \geq w_1]\} + F(w_1)(1 + R)E[y | y \geq w_1]\}\}
 \end{aligned}$$

By Envelope Theorem and Implicit Function Theorem:

$$(1 - \underline{\pi}(R))[1 - F(w^*)][E[y | y \geq w^*]] - \underline{\pi}'(R)U^* = (1 - \underline{\pi}(R))\{[1 - F(w'_1)]E[y | y \geq w_1] + F(w'_1)E[y | y \leq w'_1]\} - \underline{\pi}'(R)U^*$$

where $U^* = [1 - F(w^*)]\{w^* + E[Ry | y \geq w^*]\}$, $\tilde{U}^* = [1 - F(w'_1)]\{w'_1 + E[Ry | y \geq w'_1]\} + F(w'_1)(1 + R)E[y | y \geq w'_1]$, we need $\tilde{U}^* - U^* > 0$ for the analysis to be meaningful. So:

$$\begin{aligned}
 \underline{\pi}'(R) &= \frac{1 - \underline{\pi}(R)}{\tilde{U}^* - U^*} \{[1 - F(w'_1)]E[y | y \geq w_1] + F(w'_1)E[y | y \leq w'_1] - [1 - F(w^*)][E[y | y \geq w^*]]\} \\
 &= \frac{1 - \underline{\pi}(R)}{\tilde{U}^* - U^*} \left[\int_{\underline{y}}^{\bar{y}} z f(z)dz - \int_{\underline{y}}^{w^*} z f(z)dz \right] = \frac{1 - \underline{\pi}(R)}{\tilde{U}^* - U^*} \int_{\underline{y}}^{w^*} z f(z)dz > 0
 \end{aligned}$$

$\underline{\Pi}(R) = \max\{\underline{\pi}_1(R), \underline{\pi}_2(R)\}$ is an increasing function of R because $\underline{\pi}_1'(R) > 0$ and $\underline{\pi}_2'(R) > 0$.

Q.E.D.

A.2 The Model when the New Politician is never Informed

In this appendix, I validate my claim that it is not necessary for the newly appointed politician to know everything about the project and we can still uncover the complementarity, though under more restrictive conditions. Suppose that the cost of learning is a random variable $c \sim H(c)$. To illustrate the extreme case, assume that the newly appointed politician knows nothing about the project except $F(y)$, even if the old politician invests c . The old politician will invest c if:

$$c < (1 - \pi)\{(1 + R)E[y] - [1 - F(w^*(R))]\{w^* + E[ry|w^*(R) \leq y]\}\} = (1 - \pi)\Delta(R).$$

Hence, the expected surplus of the entrepreneur should satisfy:

$$\{\pi + (1 - \pi)\{1 - H[(1 - \pi)\Delta(R)]\}\} \overbrace{[1 - F(w^*(R))]\{E(y|y \geq w^*(R)) - w^*(R)\}}^{\equiv S(R)} > k.$$

With probability π , the politician is replaced with a new one who is uninformed, and the entrepreneur reaps a surplus of $S(R)$. With probability $1 - \pi$, the same politician governs in $t = 1$ and $t = 2$, and she does not pay c with probability $1 - H[(1 - \pi)\Delta(R)]$. Thus, the expected surplus for the entrepreneur is $\{\pi + (1 - \pi)\{1 - H[(1 - \pi)\Delta(x)]\}\}S(R) = \{1 - (1 - \pi)\{1 - H[(1 - \pi)\Delta(R)]\}\}S(R)$.

The left-hand side is monotonically increasing in π . Define $\underline{\pi}$ such that:

$$\underbrace{\{1 - (1 - \underline{\pi})H[(1 - \underline{\pi})\Delta(R)]\}}_{M(\underline{\pi}, R)} [1 - F(w^*(R))]\{E(y|y \geq w^*(R)) - w^*(R)\} = k.$$

We have:

$$HS(R)d\underline{\pi} + (1 - \underline{\pi})h\Delta(R)S(R)d\underline{\pi} - (1 - \underline{\pi})^2h\Delta'(R)S(R)dx + M(\underline{\pi}, R)S'(R)dR = 0$$

$$\frac{d\underline{\pi}}{dR} = \frac{(1 - \underline{\pi})^2h\Delta'(R)S(R) - M(\underline{\pi}, R)S'(R)}{HS(R) + (1 - \underline{\pi})h\Delta(R)}.$$

With $\Delta'(R) > 0$, $S'(R) > 0$, we have two competing effects. $S'(R) > 0$ formalizes the notion that performance-based rewards induce an uninformed politician to extract less.

However, stronger performance-based rewards increase the temptation to become informed in the first place, as $\Delta'(R) > 0$ shows. The sign of $\frac{d\pi}{dR}$ depends on the relative strength of the two mechanisms. Nonetheless, the main intuition is robust: the politician will be more likely to invest in information acquisition with stronger performance rewards, and the complementarity between rotation and performance rewards is preserved if the information acquisition effect is sufficiently strong.

For Section 1.4, we have similar results. Again, assume that the old politician's cost of information acquisition $c \sim H(c)$. She will pay c with probability $H((1-\pi)\{(1-p)(1+R)E(y) + p\{\Pi U_2 + (1-\Pi)E[(1+R)y|(1+R)y \geq U_2]\} - \{(1-p)U_1 + pU_2\}\}) \equiv H((1-\pi)\tilde{\Delta}(R))$. The second claim of Proposition 6 shows that $\Delta'(R) > 0$ (as $\pi = 1 - \frac{c}{\tilde{\Delta}(R)}$ and Proposition 6 shows that $d\pi/dR > 0$). The minimum rotation frequency to induce investment from the entrepreneur is:

$$k = (1-p) \underbrace{\{[1 - (1-\pi)H[(1-\pi)\tilde{\Delta}(R)]]\}}_{\equiv \tilde{M}(\pi,R)} \overbrace{\{[1 - F(w^*(R))\{E_f[y|y \geq w^*(R)] - w^*(R)\}]}^{=S(R)}.$$

The entrepreneur with project $y \sim F(y)$ gets support from the uninformed politician only if the new project does not arrive. If a new project arrives, an uninformed politician always supports the new project, while an informed politician awards zero rent to the entrepreneur with project $y \sim F(y)$ regardless of which project she chooses to endorse. However, when the new project does not arrive, we have the same situation as in Section 1.3. Because $\tilde{\Delta}'(R) > 0$, $S'(R) > 0$, the two competing effects determine whether rotation complements performance or not.

A.3 General Functional Forms for Performance-based Rewards

We can construct a more elaborated model for the reduced-form formulation of the key parameter x to facilitate its interpretation. Suppose now that the game is infinitely repeated, with a discount factor $\beta \in (0, 1)$. In each period, the politician has a chance of promotion with probability $V(y, \alpha)$, where $\frac{\partial V}{\partial y} > 0$, $\frac{\partial V}{\partial \alpha} > 0$, and $V \in (0, 1)$: higher output means higher likelihood of promotion, and α is an exogenous shifter that increase the probability of promotion for any output level. Also, $V(0, \alpha) = 0$: there is no promotion opportunity for the politician without any achievements. Suppose in equilibrium the utility of the politician is U , and a promotion event adds a utility of R to that. Suppose that the rotation frequency $\pi = 0$. By definition:

$$\begin{aligned}
 U &= \max_w [1 - F(w)]w + \beta \left\{ \int_{w^*}^{\bar{y}} \{V(z, \alpha)[U + R] + (1 - V(z, \alpha))U\} f(z) dz + \int_{\underline{y}}^{w^*} U f(z) dz \right\} \\
 &= [1 - F(w^*)]w^* + \beta U + \beta \int_{w^*}^{\bar{y}} V(z, \alpha) f(z) dz \\
 U &= \frac{1}{1 - \beta} [1 - F(w^*)]w^* + \frac{\beta}{1 - \beta} R \int_{w^*}^{\bar{y}} V(z, \alpha) f(z) dz
 \end{aligned}$$

The politician derives a utility from a stream of future rents, which is $\frac{1}{1-\beta}[1-F(w^*)]w^*$; she also values the promotion opportunity, which gives her an additional utility of $\frac{\beta}{1-\beta}R \int_{w^*}^{\bar{y}} V(z, \alpha) f(z) dz$.

The F.O.C. yields:

$$1 - F(w^*) - f(w^*)w^* - \beta R V(w^*, \alpha) f(w^*) = 0$$

$$w^* + \beta R V(w^*, \alpha) = \frac{1 - F(w^*)}{f(w^*)} \equiv s(w^*)$$

$s' < 0$ by Monotone Hazard Rate Property. We can do standard comparative statics of w^* with respect to β :

$$dw^* + R V d\beta + \beta R \frac{\partial}{\partial w^*} V(w^*, \alpha) = s' dw^*$$

$$\frac{\partial w^*}{\partial \beta} = - \frac{R V(w^*, \alpha)}{1 - s'(w^*) + \beta R \frac{\partial}{\partial w^*} V(w^*, \alpha)} < 0$$

Similarly, we can derive:

$$\frac{\partial w^*}{\partial R} = - \frac{\beta V(w^*, \alpha)}{1 - s'(w^*) + \beta R \frac{\partial}{\partial w^*} V(w^*, \alpha)} < 0$$

$$\frac{\partial w^*}{\partial \alpha} = - \frac{\beta \frac{\partial}{\partial \alpha} V(w^*, \alpha)}{1 - s'(w^*) + \beta R \frac{\partial}{\partial w^*} V(w^*, \alpha)} < 0$$

So an uninformed politician extracts fewer rents if: promotion opportunity is valuable; the politician is patient; the probability of promotion increases (because, for example, more higher positions are available). These are all very intuitive results.

The Complementarity between Rotation and Performance-based Rewards

This section proves similar results as Proposition 2. If the politician choose to pay c , her payoff will be:

$$\begin{aligned} U' &= E[y] + \beta \int_{\underline{y}}^{\bar{y}} \{V(z, \alpha)[U + R] + (1 - V(z, \alpha))U\} f(z) dz \\ &= E[y] + \beta U + \beta \int_{\underline{y}}^{\bar{y}} V(z, \alpha) R f(z) dz \end{aligned}$$

$$\Delta = U' - U = E[y] + \beta R \int_{\underline{y}}^{\bar{y}} V(z, \alpha) f(z) dz - \{[1 - F(w^*)]w^* + \beta U + \beta R \int_{w^*}^{\bar{y}} V(z, \alpha) f(z) dz\}$$

We find that:

$$\frac{\partial U}{\partial R} = \beta \int_{\underline{y}}^{w^*} V(z, \alpha) f(z) dz > 0, \quad \frac{\partial U}{\partial \beta} = R \int_{\underline{y}}^{w^*} V(z, \alpha) f(z) dz > 0, \quad \frac{\partial U}{\partial \alpha} = \beta R \int_{\underline{y}}^{w^*} \frac{\partial V(z, \alpha)}{\partial \alpha} f(z) dz > 0$$

So the politician wants to accumulate local knowledge if: promotion opportunity is valuable; the politician is patient; the probability of promotion increases. The traditional parameters that should improve welfare all increase the benefit of local knowledge and exacerbate the temptation to learn.

With probability π of rotation, the politician does not pay c if:

$$\begin{aligned} (1 - \pi) \{ [1 - F(w^*)]w^* + \beta U + \beta \int_{w^*}^{\bar{y}} V(z, \alpha) f(z) dz \} + \pi \tilde{U} &\geq \\ (1 - \pi) \{ E[y] + \beta U + \beta \int_{\underline{y}}^{\bar{y}} V(z, \alpha) R f(z) dz \} + \pi \tilde{U} - c & \end{aligned}$$

$$\pi \geq 1 - \frac{c}{E[y] + \beta R \int_{\underline{y}}^{\bar{y}} V(z, \alpha) f(z) dz - \{ [1 - F(w^*)]w^* + \beta U + \beta R \int_{w^*}^{\bar{y}} V(z, \alpha) f(z) dz \}} = 1 - \frac{c}{\Delta} \equiv \pi$$

$$\frac{\partial \pi}{\partial R} > 0, \frac{\partial \pi}{\partial \beta} > 0, \frac{\partial \pi}{\partial \alpha} > 0$$

By taking a reduced-form approach in Section 1.3, we can neatly summarize the key finding that $\frac{\partial \pi}{\partial R} > 0$, $\frac{\partial \pi}{\partial \beta} > 0$, and $\frac{\partial \pi}{\partial \alpha} > 0$ in one equation. We can do so because the mechanisms that drives $\frac{\partial \pi}{\partial R} > 0$, $\frac{\partial \pi}{\partial \beta} > 0$, and $\frac{\partial \pi}{\partial \alpha} > 0$ are similar. Higher values of R , β , or α all raise the benefit of local knowledge that necessitates intense rotation.

The same with Section 1.3, $\pi \geq \underline{\pi}$ is a necessary condition for the entrepreneur to invest. We also need that the entrepreneur expects more surplus than the cost of investment: $S \equiv [1 - F(w^*)]\{E[y|y \geq w^*] - w^*\} \geq k$. We can show that $\frac{\partial S}{\partial R} > 0$, $\frac{\partial S}{\partial \beta} > 0$, $\frac{\partial S}{\partial \alpha} > 0$:

$$S = [1 - F(w^*)]\{E[y|y \geq w^*] - w^*\} = \int_{w^*}^{\bar{y}} z f(z) dz - w^* \int_{w^*}^{\bar{y}} f(z) dz$$

$$\frac{\partial S}{\partial R} = - \int_{w^*}^{\bar{y}} f(z) dz \frac{\partial w^*}{\partial R} > 0$$

Proofs for $\frac{\partial S}{\partial \beta} > 0$, $\frac{\partial S}{\partial \alpha} > 0$ are similar. We can summarize this section by (without loss of generality, set $\beta = 1$):

Proposition A2: 1. *Sufficiently strong rotation and performance-based rewards incentivize the entrepreneur to invest.*

In algebra, denote $\hat{R}(k, \alpha)$ and $\underline{\pi}(R, \alpha, c)$ such that $[1 - F(w^(\hat{R}))]\{E[y|y \geq w^*(\hat{R})] - w^*(\hat{R})\} = k$ and $\underline{\pi}(R, \alpha, c) = 1 - \frac{c}{E[y+V(y, \alpha)R] - [1 - F(w^*)]\{w^* + E[V(y, \alpha)R|w^* \leq y]\}}$. If $R > \hat{R}(k, \alpha)$ and $\pi \geq \underline{\pi}(R, \alpha, c)$, the politician does not pay c , and the entrepreneur will reap a non-negative return.*

2. $\frac{\partial \pi(R, \alpha, c)}{\partial R} > 0$ and $\frac{\partial \pi(R, \alpha, c)}{\partial \alpha} > 0$: *the minimum rotation frequency increases when performance-based rewards are stronger. If the minimum rotation frequency does not change, more intense performance-based rewards incentivize the politician in $t = 1$ to pay c , and the politician in $t = 2$ will fully predate the entrepreneur.*

Notice that these key results impose minimal assumption on $V(y, \alpha)$: we only require that $V(y, \alpha) \in [0, 1]$ is an increasing function on y and α . This is true for all the following sections (except Section A3.4). Specifically, we don't need to make any assumptions on the curvatures of V or the cross derivatives. This demonstrates that the complementarity between rotation and performance-based rewards is highly generic.

Figure A.1: General Functional Forms – fixed α

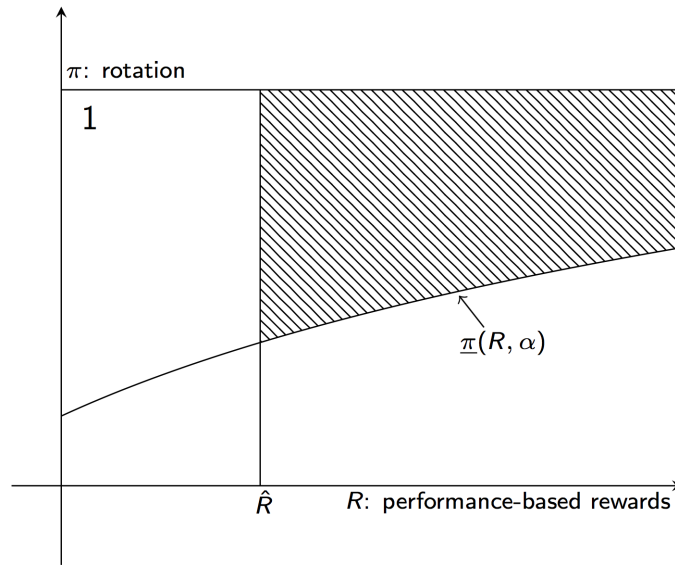
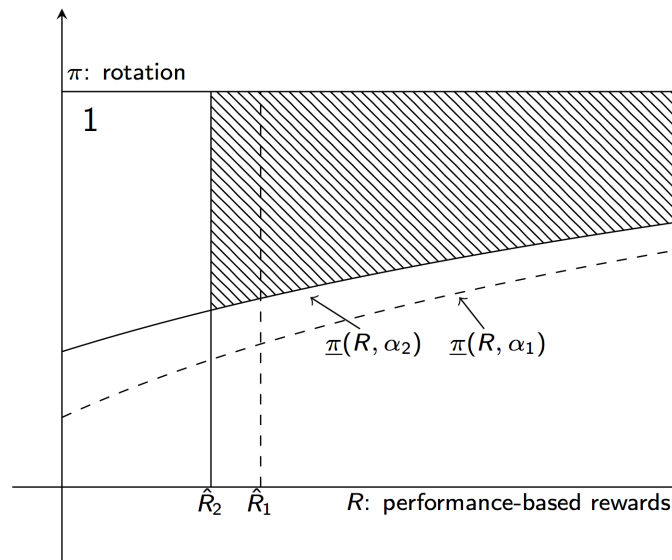


Figure A.2: General Functional Forms – $\alpha_1 > \alpha_2$



Roving bandits and the Complementarity between Rotation and Performance-based Rewards

Without loss of generality, let us re-focus on the two-period model and assumes $\beta = 1$. At the end of her term, the politician either gets a promotion (with payoff R) or enters retirement (with payoff 0). The politician will not steal capital if:

$$\eta k + \pi * \tilde{U} + (1 - \pi) * 0 \leq \pi * \tilde{U} + (1 - \pi) * \{[1 - F(w^*)]w^* + R \int_{w^*}^{\bar{y}} V(z, \alpha) f(z) dz\}$$

$$\pi \leq 1 - \frac{\eta k}{[1 - F(w^*)]w^* + R \int_{w^*}^{\bar{y}} V(z, \alpha) f(z) dz} \equiv \bar{\pi}(R, \alpha)$$

Notice that:

$$\frac{\partial \bar{\pi}}{\partial R} = \frac{\eta k}{\{[1 - F(w^*)]w^* + R \int_{w^*}^{\bar{y}} V(z, \alpha) f(z) dz\}^2} \int_{w^*}^{\bar{y}} V(z, \alpha) f(z) dz > 0$$

Also:

$$\frac{\partial \bar{\pi}}{\partial \alpha} = \frac{\eta k}{\{[1 - F(w^*)]w^* + R \int_{w^*}^{\bar{y}} V(z, \alpha) f(z) dz\}^2} R \int_{w^*}^{\bar{y}} \frac{\partial V(z, \alpha)}{\partial \alpha} f(z) dz > 0$$

So a rise in R , or α raises the stake of promotion opportunity, making the politician less tempted to steal private capital. With similar regularity assumption (as Assumption 3), we have:

Proposition A3: Denote $\underline{\pi} \equiv 1 - \frac{c}{E[y+V(y,\alpha)R] - [1-F(w^*)]\{w^*+E[V(y,\alpha)R|y \geq w^*]\}}$,

$\bar{\pi} \equiv 1 - \frac{\eta k}{[1-F(w^*)]\{w^*+E[V(y,\alpha)R|y \geq w^*]\}}$. $\forall R \geq 0, \exists \pi \in [\underline{\pi}, \bar{\pi}]$, such that the local politician neither acquires information nor steals capital.

Moreover, if $R' < R''$, then $\underline{\pi}(R', \alpha) < \underline{\pi}(R'', \alpha)$, and $\bar{\pi}(R', \alpha) < \bar{\pi}(R'', \alpha)$, Stronger performance-based rewards complement and are complemented by more frequent rotation.

Adaptation: the Informed Politician

This section proves results analogous to Proposition 4. This is the only section where the second derivative matters. Specifically, I need to assume that V is convex in y . A new project arrives with probability p . The productivity of the project follows $G(\cdot)$ that *f.o.s.d.* $F(\cdot)$. The results are for R . They also (trivially) applies to α if α enters $V(y, \alpha)$ linearly.

Proposition A4: *Suppose $V(y, \alpha)$ is convex in y . If the politician paid c and is thus informed about the old project:*

She will endorse the new project with probability $F(\hat{y})$, where \hat{y} satisfies $\hat{y} + V(\hat{y}, \alpha)R = [1 - G(\tilde{w})]\{\tilde{w} + E_g[V(y, \alpha)R|y \geq \tilde{w}]\} \equiv U_2$. $F(\hat{y})$ increases with R : performance-based rewards encourages adaptation.

Proof: Suppose that the old entrepreneur in $t = 1$ has invested. The informed politician supports the new project if:

$$y' + RV(y', \alpha) \leq [1 - G(\tilde{w})]\tilde{w} + R \int_{\tilde{w}}^{\bar{y}} V(z, \alpha)g(z)dz.$$

\tilde{w} maximizes the right hand side of the inequality. y' is a draw from $F(\cdot)$. As the left hand side is monotone in y' , $\exists!$ \hat{y} such that $\hat{y} + RV(\hat{y}, \alpha) = [1 - G(\tilde{w})]\tilde{w} + R \int_{\tilde{w}}^{\bar{y}} V(z, \alpha)g(z)dz$.

The probability that the informed politician supports the new project is thus $F(\hat{y}(R, \alpha))$. We are mainly concerned whether the increase in the value of promotion opportunity R encourages adaptation. As $\frac{dF(\hat{y}(R, \alpha))}{dR} = f \frac{d\hat{y}(R, \alpha)}{dR}$, the sign of $\frac{dF(\hat{y}(R, \alpha))}{dR}$ depends on $\frac{d\hat{y}(R, \alpha)}{dR}$. We have:

$$d\hat{y} + R \frac{\partial V}{\partial \hat{y}} d\hat{y} + V dR = \int_{\tilde{w}}^{\bar{y}} V(z, \alpha)g(z)dz dR$$

$$\frac{d\hat{y}}{dR} = \frac{\int_{\tilde{w}}^{\bar{y}} V(z, \alpha)g(z)dz - V(\hat{y}, \alpha)}{1 + R \frac{\partial V}{\partial \hat{y}}}$$

The sign depends on $\int_{\tilde{w}}^{\bar{y}} V(z, \alpha)g(z)dz - V(\hat{y}, \alpha)$.

By contradiction, suppose that $\int_{\tilde{w}}^{\bar{y}} V(z, \alpha)g(z)dz - V(\hat{y}, \alpha) < 0$. As by definition $\hat{y} + RV(\hat{y}, \alpha) = [1 - G(\tilde{w})]\tilde{w} + R \int_{\tilde{w}}^{\bar{y}} V(z, \alpha)g(z)dz$, we have $\hat{y} - [1 - G(\tilde{w})]\tilde{w} = \int_{\tilde{w}}^{\bar{y}} V(z, \alpha)g(z)dz - V(\hat{y}, \alpha) < 0$, or $\hat{y} < [1 - G(\tilde{w})]\tilde{w}$.

Notice that $\int_{\tilde{w}}^{\bar{y}} V(z, \alpha)g(z)dz = [1 - G(\tilde{w})]E_g[V(y, \alpha)|y \geq \tilde{w}]$. By conditional Jensen's Inequality, as $V(y, \alpha)$ is convex in y , we have:

$$[1 - G(\tilde{w})]V(E[y|y \geq \tilde{w}], \alpha) \leq [1 - G(\tilde{w})]E_g[V(y, \alpha)|y \geq \tilde{w}]$$

Together with $[1 - G(\tilde{w})]E_g[V(y, \alpha)|y \geq \tilde{w}] < V(\hat{y}, \alpha)$ and $\hat{y} < 1 - G(\tilde{w})\tilde{w}$:

$$[1 - G(\tilde{w})]V(E[y|y \geq \tilde{w}], \alpha) < [1 - G(\tilde{w})]E_g[V(y, \alpha)|y \geq \tilde{w}] < V(\hat{y}, \alpha) < V([1 - G(\tilde{w})]\tilde{w}, \alpha).$$

Thus, $[1 - G(\tilde{w})]V(E[y|y \geq \tilde{w}], \alpha) < V([1 - G(\tilde{w})]\tilde{w}, \alpha)$.

As $V(y, \alpha)$ is convex in y , by definition $V([1 - G(\tilde{w})]\tilde{w}, \alpha) = V([1 - G(\tilde{w})]\tilde{w} + G(\tilde{w}) * 0, \alpha) < [1 - G(\tilde{w})]V(\tilde{w}, \alpha) + G(\tilde{w})V(0, \alpha) = [1 - G(\tilde{w})]V(\tilde{w}, \alpha)$.

So we have $[1 - G(\tilde{w})]V(E[y|y \geq \tilde{w}], \alpha) < [1 - G(\tilde{w})]V(\tilde{w}, \alpha)$, or $E[y|y \geq \tilde{w}] < \tilde{w}$, this is a contradiction. So it must be that $\int_{\tilde{w}}^{\bar{y}} V(z, \alpha)g(z)dz - V(\hat{y}, \alpha) > 0$.

Q.E.D.

As before, the first best from the perspective of the principal is to support the old project if and only if $y' \geq E_g(y)$.

The asymptotic result as $R \rightarrow \infty$ becomes: the politician supports the old project if and only if $V(y', \alpha) \geq \int_{\underline{y}}^{\bar{y}} V(z, \alpha)g(z)dz = E_g[V(y, \alpha)]$. So in general, there is still distortion even as $R \rightarrow \infty$, as the politician takes the curvature of promotion function $V(y, \alpha)$ into consideration.

Specifically, as $V(y, \alpha)$ is convex in y , $E_g(V(y, \alpha)) > V(E_g(y), \alpha)$. So if the politician supports the old project (i.e., $V(y', \alpha) \geq E_g[V(y, \alpha)]$), we have $V(y', \alpha) > V(E_g(y), \alpha)$ or $y' > E_g(y)$: the principal also prefers the old project. Thus, there are cases where the principal prefers the old project, but the politician supports the new project: the informed politician excessively adapts when $R \rightarrow \infty$.

Adaptation: the Uninformed Politician

This section is analogous to Proposition 5. Again, she always supports the new project:

$$U_2 \equiv \max_w [1 - G(w)]\{w + \beta RE_g[V(y, \alpha)|y \geq w]\} = [1 - G(\tilde{w})]\{\tilde{w} + \beta RE_g[y|V(y, \alpha) \geq \tilde{w}]\} \geq$$

$$[1 - G(w^*)]\{w^* + \beta RE_g[V(y, \alpha)|y \geq w^*]\} = [1 - G(w^*)]w^* + \beta R \int_{w^*}^{\bar{y}} V(z, \alpha)g(z)dz$$

Notice that $\int_{w^*}^{\bar{y}} V(z, \alpha)g(z)dz \geq \int_{w^*}^{\bar{y}} V(z, \alpha)f(z)dz$ because of F.O.S.D.³ Also, because $G(y) \leq F(y) \forall y$, $[1 - G(w^*)]w^* \geq [1 - F(w^*)]w^*$. So we have:

$$U_2 \geq [1 - G(w^*)]w^* + \beta R \int_{w^*}^{\bar{y}} V(z, \alpha)g(z)dz \geq [1 - F(w^*)]w^* + \beta R \int_{w^*}^{\bar{y}} V(z, \alpha)f(z)dz \equiv U_1$$

So it is optimal for the politician always to endorse the new project. Notice that an uninformed politician is still more much adaptive than an informed politician even as $R \rightarrow \infty$. Proposition A5 is exactly the same as Proposition 5:

Proposition A5: *Suppose the politician didn't pay the cost of learning and a new project arrives with $G(y) \leq F(y)$, $\forall y$. The politician will always endorse the new project.*

Adaptation: the Complementarity between Rotation and Performance Rewards

This section is analogous to Proposition 6. The politician does not pay c if:

$$(1 - \pi)\{(1 - p)U_1 + pU_2\} + \pi\tilde{U} \geq$$

$$(1 - \pi)\left\{(1 - p)\{E_f[y + RV(y, \alpha)]\} + p\Pi U_2 + (1 - \Pi)\{E_f[y + RV(y, \alpha) | y + RV(y, \alpha) \geq U_2]\}\right\} + \pi\tilde{U} - c$$

where $\Pi = F(\hat{y})$, $\hat{y} + RV(\hat{y}, \alpha) = U_2$.

Proposition A6: *1. Sufficiently intense rotation and performance-based rewards induce an adaptive and benevolent local politician. The old and the new entrepreneurs invest accordingly.*

In algebra, denote $\hat{R}(k, \alpha)$ such that $\hat{R}(k, \alpha) = \max\{R_1, R_2\}$, where R_1 satisfies $(1 - p)[1 - F(w^*(R_1, \alpha))]\{E_f(y | y \geq w^*(R_1, \alpha)) - w^*(R_1, \alpha)\} = k$ and R_2 satisfies $[1 - G(\tilde{w}(R_2, \alpha))][E_g(y | y \geq \tilde{w}(R_2, \alpha)) - \tilde{w}(R_2, \alpha)] = k$. Also, $\underline{\pi} = 1 - \frac{c}{(1 - p)\{E(y + V(y, \alpha)R) - U_1\} + p\{(1 - \Pi)[E_f[y + V(y, \alpha)R | y + V(y, \alpha)R \geq U_2] - U_2]\}}$.

If $R > \hat{R}$ and $\pi \geq \underline{\pi}$, the politician does not pay c , and both the old and the new entrepreneurs reap non-negative returns.

³Again, “ G f.o.s.d. F ” \Leftrightarrow “ $G(y) \leq F(y) \forall y$ ” \Leftrightarrow “For every weakly increasing utility function u , $\int u(x)dG \geq \int u(x)dF$.”

2. Denote $w^*(0) = \operatorname{argmax}_w [1 - F(w)]w$. If $p < \frac{\int_{\underline{y}}^{w^*(0)} V(z, \alpha) f(z) dz}{\int_{\underline{y}}^{w^*(0)} V(z, \alpha) f(z) dz + \int_{\underline{y}}^{\bar{y}} V(z, \alpha) g(z) dz}$

and $p < \frac{\int_{\underline{y}}^{w^*(0)} \frac{\partial V(z, \alpha)}{\partial \alpha} f(z) dz}{\int_{\underline{y}}^{w^*(0)} \frac{\partial V(z, \alpha)}{\partial \alpha} f(z) dz + \int_{\underline{y}}^{\bar{y}} \frac{\partial V(z, \alpha)}{\partial \alpha} g(z) dz}$, then $\frac{\partial \pi}{\partial R} > 0$ and $\frac{\partial \pi}{\partial \alpha} > 0$.

Proof: We need:

$$\begin{aligned} \pi &\geq 1 - \frac{c}{\{(1-p)E_f[y + RV(y, \alpha)]\} - U_1 + p\{(1-\Pi)\{E_f[y + RV(y, \alpha)|y + RV(y, \alpha) \geq U_2] - U_2\}} \\ &\equiv 1 - \frac{c}{\Delta}. \end{aligned}$$

Note that $\Delta = (1-p)\Delta_1 + \pi\Delta_2$. $\Delta_1 = E_f[y + \beta RV(y, \alpha)] - U_1$, so:

$$\frac{\partial \Delta_1}{\partial R} = E_f[V(y, \alpha)] - \int_{w^*}^{\bar{y}} V(z, \alpha) f(z) dz = \int_{\underline{y}}^{w^*} V(z, \alpha) f(z) dz.$$

As before, denote \hat{y} such that $\hat{y} + RV(\hat{y}, \alpha) = [1 - G(\tilde{w})]\tilde{w} + R \int_{\tilde{w}}^{\bar{y}} V(z, \alpha) g(z) dz \equiv U_2$. $\Delta_2 = (1-\Pi)\{E_f[y + RV(y, \alpha)|y + RV(y, \alpha) \geq U_2] - U_2\} = \int_{\hat{y}}^{\bar{y}} (z + RV(z, \alpha)) f(z) dz - [1 - F(\hat{y})]U_2$. So:

$$\frac{\partial \Delta_2}{\partial R} = \int_{\hat{y}}^{\bar{y}} V(z, \alpha) f(z) dz - [\hat{y} + RV(\hat{y}, \alpha)] f(\hat{y}) \frac{\partial \hat{y}}{\partial R} + f(\hat{y}) \frac{\partial \hat{y}}{\partial R} U_2 - [1 - F(\hat{y})] \int_{\tilde{w}}^{\bar{y}} V(z, \alpha) g(z) dz$$

As $\hat{y} + RV(\hat{y}, \alpha) = U_2$, $[\hat{y} + RV(\hat{y}, \alpha)] f(\hat{y}) \frac{\partial \hat{y}}{\partial R} = f(\hat{y}) \frac{\partial \hat{y}}{\partial R} U_2$, so:

$$\frac{\partial \Delta_2}{\partial R} = \int_{\hat{y}}^{\bar{y}} V(z, \alpha) f(z) dz - [1 - F(\hat{y})] \int_{\tilde{w}}^{\bar{y}} V(z, \alpha) g(z) dz \geq - \int_{\underline{y}}^{\bar{y}} V(z, \alpha) g(z) dz$$

To have $\frac{\partial \Delta}{\partial R} > 0$, a sufficient condition can be derived:

$$\frac{\partial \Delta}{\partial R} = (1-p) \frac{\partial \Delta_1}{\partial R} + p \frac{\partial \Delta_2}{\partial R} \geq (1-p) \int_{\underline{y}}^{w^*} V(z, \alpha) f(z) dz - p \int_{\underline{y}}^{\bar{y}} V(z, \alpha) g(z) dz$$

$$\geq (1-p) \int_{\underline{y}}^{w^*(0)} V(z, \alpha) f(z) dz - p \int_{\underline{y}}^{\bar{y}} V(z, \alpha) g(z) dz \geq 0,$$

where $w^*(0) = \operatorname{argmax}_w [1 - F(w)]w$. So if $p < \frac{\int_{\underline{y}}^{w^*(0)} V(z, \alpha) f(z) dz}{\int_{\underline{y}}^{w^*(0)} V(z, \alpha) f(z) dz + \int_{\underline{y}}^{\bar{y}} V(z, \alpha) g(z) dz}$, we have $\frac{\partial \Delta}{\partial R} \geq 0$.

Proof for $\frac{\partial \Delta}{\partial \alpha} > 0$ is similar:

$$\frac{\partial \Delta_1}{\partial \alpha} = R \int_{\underline{y}}^{w^*} \frac{\partial V(z, \alpha)}{\partial \alpha} f(z) dz.$$

$$\begin{aligned} \frac{\partial \Delta_2}{\partial \alpha} &= R \int_{\hat{y}}^{\bar{y}} \frac{\partial V(z, \alpha)}{\partial \alpha} f(z) dz - [\hat{y} + RV(\hat{y}, \alpha)] f(\hat{y}) \frac{\partial \hat{y}}{\partial \alpha} + f(\hat{y}) \frac{\partial \hat{y}}{\partial \alpha} U_2 - [1 - F(\hat{y})] R \int_{\bar{w}}^{\bar{y}} \frac{\partial V(z, \alpha)}{\partial \alpha} g(z) dz \\ &= R \int_{\hat{y}}^{\bar{y}} \frac{\partial V(z, \alpha)}{\partial \alpha} f(z) dz - R[1 - F(\hat{y})] \int_{\bar{w}}^{\bar{y}} \frac{\partial V(z, \alpha)}{\partial \alpha} g(z) dz > -R \int_{\underline{y}}^{\bar{y}} \frac{\partial V(z, \alpha)}{\partial \alpha} g(z) dz \end{aligned}$$

A sufficient condition for $\frac{\partial \Delta}{\partial \alpha} > 0$:

$$\begin{aligned} \frac{\partial \Delta}{\partial \alpha} &= (1-p) \frac{\partial \Delta_1}{\partial \alpha} + p \frac{\partial \Delta_2}{\partial \alpha} \\ &> (1-p) R \int_{\underline{y}}^{w^*} \frac{\partial V(z, \alpha)}{\partial \alpha} f(z) dz - p R \int_{\underline{y}}^{\bar{y}} \frac{\partial V(z, \alpha)}{\partial \alpha} g(z) dz \\ &> (1-p) R \int_{\underline{y}}^{w^*(0)} \frac{\partial V(z, \alpha)}{\partial \alpha} f(z) dz - p R \int_{\underline{y}}^{\bar{y}} \frac{\partial V(z, \alpha)}{\partial \alpha} g(z) dz > 0 \end{aligned}$$

$$p < \frac{\int_{\underline{y}}^{w^*(0)} \frac{\partial V(z, \alpha)}{\partial \alpha} f(z) dz}{\int_{\underline{y}}^{w^*(0)} \frac{\partial V(z, \alpha)}{\partial \alpha} f(z) dz + \int_{\underline{y}}^{\bar{y}} \frac{\partial V(z, \alpha)}{\partial \alpha} g(z) dz}.$$

Q.E.D.

Two Rounds of Bargaining

This section is in analogous to Proposition 7:

Proposition A7: Denote:

$$\underline{\pi}_1 = 1 - \frac{c}{E[y + V(y, \alpha)R] - \max_w [1 - F(w)]\{w + E[V(y, \alpha)R|y \geq w]\}}.$$

$$\underline{\pi}_2 : (1 - \underline{\pi}_2)\max_{w_2} [1 - F(w_2)]\{w_2 + E[V(y, \alpha)R|y \geq w_2]\} =$$

$$\max_{w_1} \{-F(w_1)c + (1 - \underline{\pi}_2)\{[1 - F(w_1)]\{w_1 + E[V(y, \alpha)R|y \geq w_1]\} + F(w_1)E[y + V(y, \alpha)R|y \geq w_1]\}\}.$$

$$\hat{R} : k = [1 - F(w^*(R, \alpha))\{E[y|y \geq w^*(R, \alpha)] - w^*(R, \alpha)\}, w^* + V(w^*, \alpha)R] = \frac{1 - F(w^*)}{w^* f(w^*)}.$$

1. If $\pi \geq \underline{\pi}_1$, $\pi \geq \underline{\pi}_2$, and $R > \hat{R}$, the politician does not pay c to learn about y , and the entrepreneur makes the investment k .

2.

$$\frac{\partial \underline{\pi}_1}{\partial R} > 0, \frac{\partial \underline{\pi}_2}{\partial R} > 0, \frac{\partial \underline{\pi}_1}{\partial \alpha} > 0, \frac{\partial \underline{\pi}_2}{\partial \alpha} > 0$$

So with stronger performance rewards, the minimum rotation frequency $\underline{\Pi}(x) = \max\{\underline{\pi}_1, \underline{\pi}_2\}$ also rises to guarantee the entrepreneur's ex ante investment.

Proof: Denote that

$$w^* = \operatorname{argmax}_w [1 - F(w)]\{w + E[V(y, \alpha)R|y \geq w]\}.$$

w^* is the optimal extraction at $t = 2$ for an uninformed politician with no entrepreneur accepting w_1 .

Notice that at $t = 2$ newly appointed politician also honors w'_1 , the contract that the veteran politician made in $t = 1$. Also, if the entrepreneur rejects w'_1 , the information set for a newly appointed politician is exactly the same as the veteran politician. They also have the same utility. So they must offer the same w'_2 .

In $t = 2$, the dominant strategy is for entrepreneur to accept w'_2 if $y \geq w'_2$ and entrepreneur has not accepted in $t = 1$.

Back to $t = 1$. For all $w'_1 \in [y, \bar{y}]$, For all $w'_1 \in [y, \bar{y}]$, I now derive the set of strategies that can satisfy sequential rationality. This enables the politician to predict her payoff if she chooses a specific w'_1 .

1. Fix a $w'_1 \geq w^*$: There are two cases: the politician finds it sequentially rational to pay c or pay 0. The strategy at learning stage is a function of w'_1 . Denote $learn(w'_1) = 1$ if the politician learns after w'_1 and $learn(w'_1) = 0$ otherwise.

1.1 Suppose $learn(w'_1) = 1$: Given that $learn(w'_1) = 1$ is sequentially rational, the entrepreneur accepts any $y - w'_1 \geq 0$ as in $t = 2$ he gets 0 surplus.

Given that $learn(w'_1) = 1$ is sequentially rational, the politician asks to extract $w'_2 = y$ for any entrepreneur who did not accept w'_1 .

The utility of w'_1 if $learn(w'_1) = 1$ is sequentially rational is:

$$U(\{w'_1, learn(w'_1) = 1\}) = -F(w'_1)c + (1 - \pi)\{[1 - F(w'_1)]\{w'_1 + E[V(y, \alpha)R|y \geq w'_1]\} + E[y + V(y, \alpha)R|y \leq w'_1]\} + \pi\hat{U}$$

1.1 Suppose $learn(w'_1) = 0$: If $y - w'_1 > y - w'_2$ or $w'_1 < w'_2$, entrepreneur accepts w'_1 if $y \geq w'_1$. This is impossible as:

$$w'_2 = argmax_w Prob(w \leq y|y \leq w'_1)\{w + E[V(y, \alpha)R|w \leq y, given y \leq w'_1]\} < w'_1$$

If $y - w'_1 \leq y - w'_2$ or $w'_1 \geq w'_2$, entrepreneur rejects w'_1 . In this case:

$$w'_2 = argmax_w [1 - F(w)]\{w + E[V(y, \alpha)R|y \geq w]\} = w^*$$

Indeed $w'_1 \geq w'_2 = w^*$, consistent with the assumption that $w'_1 \geq w^*$

So given that $learn(w'_1) = 0$ is sequentially rational:

$$U(\{w'_1, learn(w'_1) = 0\}) = (1 - \pi)[1 - F(w^*)]\{w^* + E[V(y, \alpha)R|y \geq w^*]\} + \pi\hat{U}$$

So by choosing $w'_1 \geq w^*$, the maximum utility the politician can get is:

$$\begin{aligned} \max \{ & -F(w'_1)c + (1 - \pi)\{E[y + V(y, \alpha)R|y \leq w'_1] + [1 - F(w'_1)]\{w'_1 + E[V(y, \alpha)R|y \geq w'_1]\} + \pi\hat{U}\} \\ & , (1 - \pi)[1 - F(w^*)]\{w^* + E[V(y, \alpha)R|y \geq w^*]\} + \pi\hat{U} \} \end{aligned}$$

2. Fix a $w'_1 < w^*$:

2.1 Suppose $learn(w'_1) = 1$: The algebra is the same as 1.1. We have:

$$\begin{aligned} U(\{w'_1, learn(w'_1) = 1\}) = & -F(w'_1)c + \\ & (1 - \pi)\{E[y + V(y, \alpha)R|y \leq w'_1] + [1 - F(w'_1)]\{w'_1 + E[V(y, \alpha)R|y \geq w'_1]\} + \pi\hat{U} \end{aligned}$$

2.2 Suppose $learn(w'_1) = 0$: If $y - w'_1 > y - w'_2$ or $w'_1 < w'_2$, entrepreneur accepts w'_1 if $y \geq w'_1$. This is impossible as:

$$w'_2 = \operatorname{argmax}_w \operatorname{Prob}(w \geq w | w \leq w'_1) \{w + E[V(y, \alpha)R|w \leq y, \text{ given } y \leq w'_1]\} < w'_1$$

If $y - w'_1 < y - w'_2$ or $w'_1 > w'_2$, entrepreneur rejects w'_1 . In this case:

$$w'_2 = \operatorname{argmax}_w [1 - F(w)] \{w + E[V(y, \alpha)R|y \geq w]\} = w^*$$

Another contradiction as we assume $w'_1 > w'_2$ and $w^* > w'_1$.

The only possible case is $w'_1 = w'_2$. Given Assumption 4, suppose that the cutoff is \hat{w} such that an entrepreneur with $y \geq \hat{w}$ accepts w'_1 . We have $\hat{w} > w'_1$. Entrepreneur with $w'_1 \leq y < \hat{w}$ accepts w'_2 . To make $w'_2 = w'_1$ sequentially rational, it must be that \hat{w} satisfies:

$$w'_1 = w'_2 = \operatorname{argmax}_w \operatorname{prob}(w \leq y, \text{ given } y \leq \hat{w}) \{w + \beta E[V(y, \alpha)R|w \leq y, \text{ given } y \leq \hat{w}]\}$$

The probability density function of $f(y|y \leq \hat{w}) = \frac{f(y)}{F(\hat{w})} \equiv g(y, \hat{w})$ with support on $[y, \hat{w}]$. We have:

$$\text{prob}(w \leq y, \text{ given } y \leq \hat{w}) = \int_w^{\hat{w}} f(z|z \leq \hat{w})dz = \frac{\int_w^{\hat{w}} f(z)}{1 - F(\hat{w})}dz = \frac{F(\hat{w}) - F(w)}{F(\hat{w})}$$

$$E[V(y, \alpha)|w \leq y, \text{ given } y \leq \hat{w}] = \int_w^{\hat{w}} V(z, \alpha)g(z, \hat{w}|z \geq w)dz$$

$$g(z, \hat{w}|z \geq w) = \frac{g(z, \hat{w})}{1 - G(w; \hat{w})} = \frac{f(z)/F(\hat{w})}{1 - \int_y^w \frac{f(v)}{F(\hat{w})}dv} = \frac{f(z)/F(\hat{w})}{1 - \frac{F(w)}{F(\hat{w})}} = \frac{f(z)}{F(\hat{w}) - F(w)}$$

Thus we have:

$$E[V(y, \alpha)|w \leq y, \text{ given } y \leq \hat{w}] = \frac{\int_w^{\hat{w}} V(z, \alpha)f(z)dz}{F(\hat{w}) - F(w)}$$

So:

$$\text{prob}(w \leq y, \text{ given } y \leq \hat{w})\{w + E[V(y, \alpha)|w \leq y, \text{ given } y \leq \hat{w}]\}$$

$$= \frac{F(\hat{w}) - F(w)}{F(\hat{w})}w + \frac{F(\hat{w}) - F(w)}{F(\hat{w})} \frac{\int_w^{\hat{w}} V(z, \alpha)Rf(z)dz}{F(\hat{w}) - F(w)}$$

$$= \frac{[F(\hat{w}) - F(w)]w + \int_w^{\hat{w}} V(z, \alpha)Rf(z)dz}{F(\hat{w})}$$

As $w'_1 = w'_2 = \text{argmax}_w \frac{[F(\hat{w}) - F(w)]w + \int_w^{\hat{w}} V(z, \alpha)Rf(z)dz}{F(\hat{w})}$, F.O.C. tells us that:

$$w'_1 + RV(w'_1, \alpha) = \frac{F(\hat{w}) - F(w'_1)}{F(w'_1)}$$

Now let us verify that $w'_1 < w^*$. By contradiction, suppose that $w'_1 \geq w^*$. $w^* = \operatorname{argmax}_w \{[1 - F(w)]w + R \int_y^{w^*} V(z, \alpha) f(z) dz\}$. So w^* satisfies:

$$\frac{1 - F(w^*)}{f(w^*)} = w^* + V(w^*, \alpha)R$$

As we assume that $w'_1 \geq w^*$, and that $\frac{1-F(w)}{f(w)}$ is a decreasing function, we have:

$$w'_1 + RV(w'_1, \alpha) = \frac{F(\hat{w}) - F(w'_1)}{F(w'_1)} < \frac{1 - F(w'_1)}{F(w'_1)} \leq \frac{1 - F(w^*)}{F(w^*)} = w^* + RV(w^*, \alpha)$$

As $\{w + RV(w, \alpha)\}$ is an increasing function in w , $w'_1 < w^*$, we obtain the contradiction. So we must have $w'_1 < w^*$.

Thus, if $\operatorname{learn}(w'_1) = 0$ is sequentially rational following a $w'_1 < w^*$:

$$U(\{w'_1, w'_1(\operatorname{learn}) = 0\}) = (1 - \pi) \{ [1 - F(\hat{w})] \{ w'_1 + E[V(y, \alpha)R | y \geq \hat{w}] \} +$$

$$F(\hat{w}) \frac{[F(\hat{w}) - F(w'_1)]w'_1 + \int_{w'_1}^{\hat{w}} V(z, \alpha)Rf(z)dz}{F(\hat{w})} \} + \pi \tilde{U}$$

$$U(\{w'_1, \operatorname{learn}(w'_1) = 0\}) = (1 - \pi) \{ [1 - F(\hat{w})] \{ w'_1 + E[V(y, \alpha)R | y \geq \hat{w}] \} +$$

$$[F(\hat{w}) - F(w)]w + \int_w^{\hat{w}} V(z, \alpha)Rf(z)dz \} + \pi \tilde{U}$$

The maximum utility of the politician from a $w'_1 < w^*$ is:

$$\max \{ -F(w'_1)c + (1 - \pi) \{ [1 - F(w'_1)] \{ w'_1 + E[V(y, \alpha)R | y \geq w'_1] \} + F(w'_1)E[y + V(y, \alpha)R | y \leq w'_1] + \pi \hat{U} \},$$

$$(1 - \pi) \{ [1 - F(\hat{w})] \{ w'_1 + E[V(y, \alpha)R | y \geq \hat{w}] \} + [F(\hat{w}) - F(w)]w + \int_w^{\hat{w}} V(z, \alpha)Rf(z)dz \} + \pi \tilde{U} \}$$

We have derived the function of politician's payoff at $t = 1$ for any w'_1 and learning decision. We can use this function to find the sequential equilibrium strategy.

I want to ensure that $learn(w'_1) = 0$ is the strategy on equilibrium path. That gives a utility of:

$$U(w'_1, learn(w'_1) = 0) = (1 - \pi)[1 - F(w^*)]\{w^* + E[V(y, \alpha)R|y \geq w^*]\} + \pi\hat{U}$$

if $w'_1 \geq w^*$, and

$$\begin{aligned} U(w'_1, learn(w'_1) = 0) &= (1 - \pi)\{[1 - F(\hat{w})]\{w_1 + E[V(y, \alpha)R|y \geq \hat{w}]\} \\ &\quad + \{[F(\hat{w}) - F(w)]w + \int_w^{\hat{w}} V(z, \alpha)Rf(z)dz\} + \pi\hat{U} \end{aligned}$$

if $w'_1 < w^*$.

Notice that:

$$\begin{aligned} &[1 - F(\hat{w})]\{w_1 + E[V(y, \alpha)R|y \geq \hat{w}]\} + \{[F(\hat{w}) - F(w)]w + \int_w^{\hat{w}} V(z, \alpha)Rf(z)dz\} \\ &= w'_1 \int_{\hat{w}}^{\bar{y}} f(z)dz + \int_{\hat{w}}^{\bar{y}} V(z, \alpha)Rf(z)dz + w'_1 \int_{w'_1}^{\hat{w}} f(z)dz + \int_{w'_1}^{\hat{w}} V(z, \alpha)Rf(z)dz \\ &= w'_1 \int_{w'_1}^{\bar{y}} f(z)dz + \int_{w'_1}^{\bar{y}} V(z, \alpha)Rf(z)dz = [1 - F(w'_1)]\{w'_1 + E[V(y, \alpha)R|y \geq w'_1]\} \\ &\leq [1 - F(w^*)]\{w^* + E[V(y, \alpha)R|y \geq w^*]\} \end{aligned}$$

So it is always optimal to choose $w'_1 \geq w^*$.

In that case, I want to make sure that at the stage of learning, it is not optimal to learn:

$$-c + (1 - \pi)E[y + \beta V(y, \alpha)R] \leq (1 - \pi)[1 - F(w^*)]\{w^* + E[V(y, \alpha)R|y \geq w^*]\}$$

And at the stage to decide w'_1 , the utility from not learning \geq the highest utility from learning:

$$(1 - \pi) \max_{w_2} [1 - F(w_2)] \{w_2 + E[V(y, \alpha)R | y \geq w_2]\} \geq \max_{w_1} \{-F(w_1)c + (1 - \pi) \{ [1 - F(w_1)] \{w_1 + E[V(y, \alpha)R | y \geq w_1]\} + F(w_1)E[y + V(y, \alpha)R | y \geq w_1]\} \}$$

$$(1 - \underline{\pi}) \max_{w_2} [1 - F(w_2)] \{w_2 + E[V(y, \alpha)R | y \geq w_2]\} = \max_{w_1} \{-F(w_1)c + (1 - \underline{\pi}) \{ [1 - F(w_1)] \{w_1 + E[V(y, \alpha)R | y \geq w_1]\} + F(w_1)E[y + V(y, \alpha)R | y \geq w_1]\} \}$$

defines $\underline{\pi}$. By Envelope Theorem and Implicit Function Theorem:

$$(1 - \underline{\pi}) [1 - F(w^*)] E[V(y, \alpha) | y \geq w^*] dR - U^* d\underline{\pi} = (1 - \underline{\pi}) \{ [1 - F(w_1^*)] E[V(y, \alpha) | y \geq w_1] + F(w_1^*) E[V(y, \alpha) | y \leq w_1^*] \} dR - \tilde{U}^* d\underline{\pi}$$

where $U^* = [1 - F(w^*)] \{w^* + E[V(y, \alpha)R | y \geq w^*]\}$, $\tilde{U}^* = [1 - F(w_1^*)] \{w_1^* + E[V(y, \alpha)R | y \geq w_1^*]\} + F(w_1^*) E[y + V(y, \alpha)R | y \geq w_1^*]$, we need $\tilde{U}^* - U^* > 0$ for the analysis to be meaningful. So:

$$\begin{aligned} \frac{\partial \underline{\pi}}{\partial R} &= \frac{1 - \underline{\pi}}{\tilde{U}^* - U^*} \{ [1 - F(w_1^*)] E[V(y, \alpha) | y \geq w_1^*] + F(w_1^*) E[V(y, \alpha) | y \leq w_1^*] - [1 - F(w^*)] E[V(y, \alpha) | y \geq w^*] \} \\ &= \frac{1 - \underline{\pi}}{\tilde{U}^* - U^*} \left[\int_{\underline{y}}^{\bar{y}} V(z, \alpha) f(z) dz - \int_{w^*}^{\bar{y}} V(z, \alpha) f(z) dz \right] = \frac{1 - \underline{\pi}}{\tilde{U}^* - U^*} \int_{\underline{y}}^{w^*} V(z, \alpha) f(z) dz > 0 \end{aligned}$$

$\underline{\Pi} = \max\{\underline{\pi}_1, \underline{\pi}_2\}$ is an increasing function of R because $\frac{\partial \underline{\pi}_1}{\partial R} > 0$ and $\frac{\partial \underline{\pi}_2}{\partial R} > 0$.

Similarly, we can show that $\underline{\Pi}$ is an increasing function of α , as $\underline{\pi}_1$ increases with α , and:

$$\frac{\partial \underline{\pi}_2}{\partial \alpha} = \frac{1 - \underline{\pi}}{\tilde{U}^* - U^*} R \int_{\underline{y}}^{w^*} \frac{\partial V(z, \alpha)}{\partial \alpha} f(z) dz > 0$$

Q.E.D.

Appendix B

Appendices for *Meritocracy, Decentralization, and Local Dual Leadership*

B.1 Proofs for Propositions in the Text

Lemma 1

There is no PBE in which either normal or benevolent secretaries mix between {lead a collective action} and {not lead a collective action}.

Proof:

We first derive best response of all players for any history.

1. Normal secretary always capture all collective action benefits (denoted as {capture}); benevolent secretary always awards all collective action to the population (denoted as {award}).

2. Denote $\hat{\mu}$ as population's belief that the population is facing a benevolent secretary after the secretary leads a collective action.

The population join the collective action (denoted as {join}) if $\hat{\mu}(R+e-c)+(1-\hat{\mu})(e-c) > e$, or $\hat{\mu}R > c$; will mix between {join} and {not join} if $\hat{\mu}R = c$; will choose {not join} if $\hat{\mu}R < c$. Denotes the probability that the population choose {join} as \hat{y} .

3.1 Normal secretary will choose {launch a collective action} if $(1-\hat{y})(-k)+\hat{y}(Q+R-k) = \hat{y}(Q+R+S) - k > Q - k$, or $\hat{y} > \frac{Q}{Q+R+S}$; will mix between {launch} and {not launch} if $\hat{y} = \frac{Q}{Q+R+S}$; will choose {not launch} if $\hat{y} < \frac{Q}{Q+R+S}$. Denote the probability that the normal secretary chooses {launch} as \hat{z}_1 .

3.2 Benevolent secretary will choose {launch a collective action} if $(1 - \hat{y})(\gamma e - k) + \hat{y}(Q + \gamma R + S + \gamma e - k) = \hat{y}(Q + \gamma R + S) + \gamma e - k > Q + \gamma e - k$, or $\hat{y} > \frac{Q}{Q + \gamma R + S}$; will mix between {launch} and {not launch} if $\hat{y} = \frac{Q}{Q + \gamma R + S}$; will choose {not launch} if $\hat{y} < \frac{Q}{Q + \gamma R + S}$. Denote the probability that the normal secretary chooses {launch} as \hat{z}_2 .

4.1 Normal secretary will choose {provide public good} if $(1 - \pi)(Q - k) + \pi\{(1 - \hat{z}_1)(Q - k) + \hat{z}_1[(1 - \hat{y})(-k) + \hat{y}(Q + R + S - k)]\} > Q$; will mix between {provide} and {not provide} if $(1 - \pi)(Q - k) + \pi\{(1 - \hat{z}_1)(Q - k) + \hat{z}_1[(1 - \hat{y})(-k) + \hat{y}(Q + R + S - k)]\} = Q$; will choose {not provide} if $(1 - \pi)(Q - k) + \pi\{(1 - \hat{z}_1)(Q - k) + \hat{z}_1[(1 - \hat{y})(-k) + \hat{y}(Q + R + S - k)]\} < Q$.

4.2 Denote $b = \gamma e - k$. Benevolent secretary will choose {provide public good} if $(1 - \pi)(Q + b) + \pi\{(1 - \hat{z}_2)(Q + b) + \hat{z}_2[(1 - \hat{y})b + \hat{y}(Q + \gamma R + S + b)]\} > Q$; will mix between {provide} and {not provide} if $(1 - \pi)(Q + b) + \pi\{(1 - \hat{z}_2)(Q + b) + \hat{z}_2[(1 - \hat{y})b + \hat{y}(Q + \gamma R + S + b)]\} = Q$; will choose {not provide} if $(1 - \pi)(Q + b) + \pi\{(1 - \hat{z}_2)(Q + b) + \hat{z}_2[(1 - \hat{y})b + \hat{y}(Q + \gamma R + S + b)]\} < Q$.

By contradiction, suppose $z_1 \in (0, 1)$. It must be the case that $\hat{y} = \frac{Q}{Q + R + S}$.

In this case, because $\frac{Q}{Q + \gamma R + S} > \frac{Q}{Q + R + S}$, benevolent secretary chooses $z_2 = 1$, or {launch a collective action}. The benevolent secretary will choose {provide public good} as $(1 - \pi)(Q + b) + \pi\{[(1 - \hat{y})b + \hat{y}(Q + R + S + b)]\} = b + (1 - \pi)Q + \pi \frac{Q + R + S}{Q + \gamma R + S} Q > b + (1 - \pi)Q + \pi Q = b + Q > Q$.

Thus, to make the population indifferent between {join} and {not join}, x_1 and z_1 must satisfy $\frac{\mu}{\mu + (1 - \mu)x_1 z_1}(R + e - c) + \frac{(1 - \mu)x_1 z_1}{\mu + (1 - \mu)x_1 z_1}(e - c) = e$, or $\hat{x}_1 \hat{z}_1 = \frac{\mu}{1 - \mu}(\frac{R}{c} - 1)$. So $\hat{x}_1 > 0$. It must be the case that: $(1 - \pi)(Q - k) + \pi\{(1 - \hat{z}_1)(Q - k) + \hat{z}_1[(1 - \hat{y})(-k) + \hat{y}(Q + R + S - k)]\} \geq Q$, or $\pi \hat{z}_1 \hat{y}(Q + R + S) \geq \pi \hat{z}_1 Q + k$. Given that $\hat{y} = \frac{Q}{Q + R + S}$, $\pi \hat{z}_1 \hat{y}(Q + R + S) \geq \pi \hat{z}_1 Q + k$ reduces to $0 \geq k$, which is a contradiction. We conclude that it is impossible for normal secretary to mix between {launch a collective action} and {not launch}.

By contradiction, suppose $z_2 \in (0, 1)$. It must be the case that $\hat{y} = \frac{Q}{Q + \gamma R + S}$.

Because $\frac{Q}{Q + \gamma R + S} < \frac{Q}{Q + R + S}$, the normal secretary always choose {not launch}; consequently, he always choose {not provide public good}. So $x_1 = z_1 = 0$. Here the benevolent secretary has $(1 - \pi)(Q + b) + \pi\{(1 - \hat{z}_2)(Q + b) + \hat{z}_2[(1 - \hat{y})b + \hat{y}(Q + R + S + b)]\} = b + (1 - \pi \hat{z}_2)Q + \pi \hat{z}_2 \hat{y}(Q + R + S) = b + (1 - \pi \hat{z}_2)Q + \pi \hat{z}_2 Q = b + Q > Q$. So benevolent secretary always chooses {provide public good}. But then it is not optimal for the population to choose $\hat{y} = \frac{Q}{Q + \gamma R + S}$, given that only benevolent secretary provides public good and lead a collective action. They should choose $\hat{y} = 1$. We conclude that it is also impossible for benevolent secretary to mix between {launch} and {not launch}.

Q.E.D.

Lemma 2 (Commitment of Benevolent Secretary)

In any PBE, benevolent secretary always provides public good.

Proof:

By contradiction, benevolent secretary will not always provide public good if in a PBE, we have $Q \leq (1 - \pi)(Q + b) + \pi[(1 - \hat{y})b + \hat{y}(Q + \gamma R + S + b)]$.

1. Mixed strategy equilibrium where benevolent secretary mixes between {provide} and {not provide} is impossible.

By contradiction, it must be the case that $Q = (1 - \pi)(Q + b) + \pi[(1 - \hat{y})b + \hat{y}(Q + \gamma R + S + b)]$, or $\hat{y} = \frac{\pi Q - b}{Q + \gamma R + S}$. Note that normal secretary will not provide public good as:

$$(1 - \pi)(Q - k) + \pi[(1 - \hat{y})(-k) + \hat{y}(Q + R + S - k)] = (1 - \pi)Q + \pi \frac{\pi Q - b}{Q + \gamma R + S} (Q + R + S) - k$$

$$< (1 - \pi)Q + \pi Q - b - k = Q - b - k < Q$$

But given this, only benevolent secretary chooses {provide} with positive probability. This means that $\hat{\mu} = 1$, so the population should always join the collective action if the secretary launches one, or $\hat{y} = 1$. This is a contradiction.

2. Pure strategy equilibrium where benevolent secretary chooses {not provide} is not possible. So $Q < (1 - \pi)(Q + b) + \pi[(1 - \hat{y})b + \hat{y}(Q + \gamma R + S + b)]$. The only case such inequality holds is $\hat{y} = 0$. So $Q < (1 - \pi)Q + b$.

2.1 Normal secretary provides public good with positive probability such that $\hat{\mu} < \frac{c}{R}$. Note that if $\hat{\mu} > \frac{c}{R}$, the population will deviate to {join the collective action}.

In this case, as $\hat{\mu} < \frac{c}{R}$, the normal secretary wants to deviate to $x_1 = 0$, or never provides public good. So it is not a possible PBE outcome.

2.2 Normal secretary does not provide public good.

In this case, benevolent secretary also does not provide public good, and he chooses to not lead a collective action if he provides public good (otherwise it is not subgame optimal), and by assumption the population will not join if they observe public good provision. The benevolent secretary can get a higher payoff by deviating to provide public good.

Q.E.D.

Lemma 3 (Impossible Strategies)

The following strategy will not appear in any PBE: 1. {provide} and {not lead} for normal secretary; 2. population chooses {not join} while secretary chooses {launch}; 3. population chooses {join} while secretary chooses {not launch}

Proof:

If the secretary chooses {provide} and {not lead}, given that benevolent secretary always provides public good, the best response of benevolent secretary is to choose {lead}. Given this, $\hat{\mu} = 1$, and the population must join. The governor can get a higher payoff by deviating to {lead}.

“Population chooses {not join} while secretary chooses {launch}” and “population chooses {join} while secretary chooses {not launch}” do not satisfy sequential rationality for both benevolent and normal secretaries.

Q.E.D.

Lemma 4

If $k < \pi(R + S)$, then only mixed equilibrium exists and it is unique. The normal secretary mixes between providing public good and not, and the population mixes between joining the collective action and not.

Denote that the normal secretary provides public good with probability \hat{x} ; the population chooses {join} with probability \hat{y} . We have:

$$\hat{x} = \frac{\mu}{1 - \mu} \left(\frac{R}{c} - 1 \right) = \frac{\mu}{1 - \mu} \left\{ \frac{(1 - \lambda)W}{c} - 1 \right\}$$

$$\hat{y} = \frac{Q + \frac{k}{\pi}}{Q + R + S} = \frac{\lambda W + \frac{k}{\pi}}{W + S}$$

And the secretary chooses {launch} after providing public good.

Proof:

1. Use indifference principle to derive \hat{x} and \hat{y} , assuming that the secretary always *{launch}*.

To make the population indifferent between joining and not joining, the normal secretary must choose \hat{x} so that:

$$\frac{\mu}{\mu + (1 - \mu)\hat{x}}(e + R - c) + \frac{(1 - \mu)\hat{x}}{\mu + (1 - \mu)\hat{x}}(e - c) = e$$

Solve the equation, we get:

$$\hat{x} = \frac{\mu}{1 - \mu} \left(\frac{R}{c} - 1 \right) = \frac{\mu}{1 - \mu} \left\{ \frac{(1 - \lambda)W}{c} - 1 \right\}$$

To make the normal secretary indifferent between providing public good and not providing, the population must choose \hat{y} so that:

$$(1 - \pi)(Q - k) + \pi[\hat{y}(Q + R + S - k) + (1 - \hat{y})(-k)] = Q$$

Solve the equation, we get:

$$\hat{y} = \frac{Q + \frac{k}{\pi}}{Q + R + S} = \frac{\lambda W + \frac{k}{\pi}}{W + S}$$

2. Verify that normal secretary wants to lead a collective action given \hat{y} :

$$\hat{y}(Q + R + S - k) + (1 - \hat{y})(-k) = Q + \frac{1 - \pi}{\pi}k \geq Q - k$$

Verify that benevolent secretary wants to lead a collective action given \hat{y} :

$$\hat{y}(\gamma Q + R + S + e - k) + (1 - \hat{y})(e - k) \geq \gamma Q + e - k$$

It reduces to:

$$\frac{Q + \frac{k}{\pi}}{Q + R + S} \geq \frac{Q}{Q + \gamma R + S}$$

which is true for $\gamma \geq 1$.

Note that the benevolent secretary wants to provide public good given \hat{y} ;

$$(1 - \pi)[Q + e - k] + \pi[\hat{y}(Q + \gamma R + S + e - k) + (1 - \hat{y})(e - k)] \geq Q + e - k > \gamma Q$$

Note that the mixed strategy equilibrium trivially satisfies Assumption as it does not involve off-equilibrium belief.

To check uniqueness, note that it is never optimal for both types to not provide public good and then lead a collective action. If the other type also does not provide public good, then by Assumption 1 the population will not join; if the other type provides public good, the population will join the collective action launched by secretary who does not provide public good if only benevolent secretary does not do so and normal secretary provides. The situation is ruled out by Assumption 2.

1.

{provide, not provide, launch, launch, join} is not a PBE. The normal secretary want to deviate to {provide}:

$$Q < (1 - \pi)(Q - k) + \pi(Q + R + S - k)$$

2.

{provide, not provide, not launch, not launch, not join} is a (family of) equilibrium supported by $\hat{\mu} \in [0, \frac{c}{R})$. Note that no one can single deviate and get a strictly higher payoff, *given the off-equilibrium belief that $\hat{\mu} \in [0, \frac{c}{R})$.*

But note that it does not satisfy Assumption 3:

$$\bar{\mu} = \frac{\mu\pi}{\mu\pi} = 1$$

Assumption 3 restricts that $\hat{\mu} = \bar{\mu} = 1$. In this case, above strategies is not a PBE because the population wants to deviate.

3.

{provide, provide, launch, launch, join} is not a PBE by Assumption 1.

All other pure PBE has been ruled out by Lemmas 1-3.

Q.E.D.

Proposition 1 (Loyalty-competence tradeoff in Mixed-strategy Equilibrium)

If $k < \pi(R + S)$, the probability of collective action is:

$$Pr(\text{collective action}) = \mu \frac{R}{Q + R + S} \frac{\pi Q + k}{c} = \mu \frac{(1 - \lambda)W}{W + S} \frac{\pi W + k}{c}$$

Specifically, $Pr(\text{collective action})$ increases with competence W . It also increases with k , μ , π , and decreases with c .

Proof:

$$\begin{aligned} Pr(\text{collective action}) &= (1 - \mu)x^*y^*\pi + \mu y^*\pi \\ &= \{(1 - \mu)\left\{\frac{\mu}{1 - \mu}\left[\frac{(1 - \lambda)W(a)}{c} - 1\right]\right\} + \mu\} \frac{Q + \frac{k}{\pi}}{Q + R + S} \pi \\ &= \frac{R}{Q + R + S} \frac{\pi Q + k}{c} = \mu \frac{(1 - \lambda)W}{W + S} \frac{\pi W + k}{c} \end{aligned}$$

Q.E.D.

Proposition 2 (Loyalty-competence Tradeoff in Pure-strategy Equilibrium)

If $k \geq \pi(R + S)$, then only pure strategy equilibrium exists.

1.1 Collective action risk in equilibrium

If $R > c$, then in the unique PBE, {provide public good, not provide, lead, lead, join} is the strategy profile.

1.2 No collective action risk in equilibrium with low competence

If $R \leq c$, then in the unique PBE, $\{\text{provide, not provide, not lead, not lead, not join}\}$ is the strategy profile.

Proof:

Denote $\{\cdot, \cdot, \cdot, \cdot, \cdot\}$ as the strategy profile of benevolent secretary and normal secretary in choosing public good provision, leading a collective action, and the population in joining the collective action.

Assume that $R > c$.

It is easy to see that $\{\text{provide public good, not provide, launch, launch, join}\}$ is a PBE supported by the belief of the population $\hat{\mu} = 1$. In this case, the population wants to join. The benevolent secretary wants to provide public good as:

$$(1 - \pi)(Q + \gamma e(k) - k) + \pi(Q + \gamma R + S + \gamma e(k) - k) > (1 - \pi)Q + \pi(Q + \gamma R + S)$$

And the normal secretary does not want to provide because:

$$Q \geq (1 - \pi)(Q - k) + \pi(Q + R + S - k)$$

by assumption $k \geq \pi(R + S)$. Given the strategy profiles, it is optimal for both secretaries to $\{\text{launch}\}$ a collective action after providing public good. But because the normal secretary does not provide public good, it is an off-equilibrium strategy for him.

Note that the PBE trivially satisfies Assumption 3 because there is no off-equilibrium belief at $\{\text{launch}\}$ and $\{\text{not launch}\}$.

1.

$\{\text{provide, not provide, not launch, not launch, not join}\}$ is a (family of) equilibrium supported by $\hat{\mu} \in [0, \frac{c}{R})$. Note that no one can single deviate and get a strictly higher payoff, given the off-equilibrium belief that $\hat{\mu} \in [0, \frac{c}{R})$.

But it does not satisfy Assumption 3:

$$\bar{\mu} = \frac{\mu\pi}{\mu\pi} = 1$$

Assumption 3 restricts that $\hat{\mu} = \bar{\mu} = 1$. In this case, above strategies is not a PBE because the population wants to deviate.

2.

{provide, provide, launch, launch, join} is not a PBE: $\hat{\mu} = \mu$, so the population wants to deviate to {not join}.

All other pure PBE are ruled out by Lemmas 1-3.

Rule out mixed strategy equilibrium:

The benevolent secretary is “committed” to provide public good, and it is impossible to mix between {launch} and {not launch}. Thus, the only possible mixed strategy equilibrium involves the normal secretary to mix between providing and not providing, and the population mixes between join and not join the collective action (if they observe public good). Denote the probability of joining the collective action as $y \in (0, 1)$.

It is profitable for the normal secretary to deviate to not provide public good deterministically, so there is no mixed strategy equilibrium:

$$(1 - \pi)(Q - k) + \pi[y(Q + R + S - k) + (1 - y)(-k)] < (1 - \pi)(Q - k) + \pi(Q + R + S - k) < Q$$

Assume that $R \leq c$.

Now the population never find it optimal to join the collective action. Consequently, only the benevolent secretary provides public good and the normal one not. There is no collective action in equilibrium.

Q.E.D.

Proposition 3 (Implementation of Mixed Equilibrium with Loyalty-competence Tradeoff)

The principal wants to implement the mixed strategy equilibrium; $k^ = 0$, and optimal competence is an interior solution.*

Proof:

If $\lambda R \leq c$, then $\text{Prob}(\text{collective action}|\text{mixed equilibrium})=$

$$\mu \frac{Q}{Q+R+S} \frac{\pi R}{c} < \mu \pi \frac{\lambda R}{c} < \mu \pi =$$

$\text{Prob}(\text{collective action}|\text{pure equilibrium})$.

The probability of collective action under mixed strategy equilibrium is lower than that under pure strategy equilibrium, whatever the competence of secretary. Thus, the principal is guaranteed to get a higher payoff by choosing the mixed strategy equilibrium.

The problem then reduces to

$$\max_{W,k} \left\{ 1 - \mu \frac{(1-\lambda)W}{W+S} \frac{\pi \lambda W + k}{c} \right\} [(1-\lambda)W + S]$$

Obviously, the optimal solution to k is $k^* = 0$.

For W , the first order condition is:

$$(1-\lambda) \left[1 - \mu \frac{(1-\lambda)W}{W+S} \frac{\pi \lambda W + k}{c} \right] + \left[-\mu \frac{(1-\lambda)S}{(W+S)^2} \frac{\pi \lambda W + k}{c} - \mu \frac{(1-\lambda)W}{W+S} \frac{\pi \lambda}{c} \right] [(1-\lambda)W + S] = 0$$

To make sure that the solution to first order condition is globally optimal, we need to check second order condition and make sure that:

$$2(1-\lambda) \left[-\mu \frac{(1-\lambda)S}{(W+S)^2} \frac{\pi \lambda W + k}{c} - \mu \frac{(1-\lambda)W}{W+S} \frac{\pi \lambda}{c} \right] + 2 \left[\frac{\mu(1-\lambda)S}{(W+S)^3} \frac{\pi \lambda W + k}{c} - \mu \frac{(1-\lambda)S}{(W+S)^2} \frac{\pi \lambda}{c} \right] < 0$$

which will be true if:

$$\frac{\mu(1-\lambda)S}{(W+S)^3} \frac{\pi \lambda W + k}{c} - \mu \frac{(1-\lambda)S}{(W+S)^2} \frac{\pi \lambda}{c} < 0$$

which reduces to:

$$k \leq \pi \lambda S$$

As $k^* = 0$ and $S > 0$. It is always true.

Manually check for the case $S = 0$. We will see that it is a quadratic problem with positive linear term and negative quadratic term; so an globally optimal solution exists and is also interior.

Q.E.D.

Lemma 5 (Commitment of Benevolent Governor)

In any PBE, benevolent secretary always provides public good.

Proof:

Note that if both secretaries lead collective action:

$$(1 - \pi)(Q + b) + \pi[(1 - \hat{y})(Q + b) + \hat{y}(Q + \eta R + b)] > Q$$

If only benevolent secretary leads collective action:

$$(1 - \pi)(Q + b) + \pi\{(1 - \mu)(Q + b) + \mu[(1 - \hat{y})(Q + b) + \hat{y}(Q + \eta R + b)]\} > Q$$

If only normal secretary leads collective action:

$$(1 - \pi)(Q + b) + \pi\{\mu(Q + b) + (1 - \mu)[(1 - \hat{y})(Q + b) + \hat{y}(Q + \eta R + b)]\} > Q$$

So the benevolent governor always provides public good.

Q.E.D.

Proposition 5 (No Collective Action Risk in Pure Strategy Equilibrium)

If $k \geq \pi R$, then in any pure strategy equilibrium, the benevolent governor always provides public good, while the normal governor does not. The secretary never initiates a collective action, and the central authority appoints the most competent secretary and governor.

Proof:

$\{\cdot, \cdot; \cdot, \cdot; \cdot\}$ denotes the strategy profile of benevolent and normal governors in providing public good, benevolent and normal secretaries in launching a collective action, and the population if they observe public good provision. Note that with $k \geq \pi R$, we have:

$$(1 - \pi)(Q - k) + \pi(\eta R + Q - k) \leq Q$$

So it doesn't matter whether the secretaries propose or not.

1. In the (family of) PBE {provide, not provide, not launch, not launch, not join} with $\hat{\mu} = \bar{\mu} = \frac{\mu^* \mu^* \pi}{\mu^* \pi} = \mu$ (Assumption 3), the benevolent governor has no incentive to deviate:

$$Q + \gamma e(k) - k > Q$$

The normal governor has no incentive to deviate:

$$Q \geq (1 - \pi)(Q - k) + \pi(\eta R + Q - k)$$

Both secretaries have no incentive to deviate as $Q > 0$.

The population has no incentive to deviate given their belief $\hat{\mu} = \mu$.

2.

{provide, not provide, launch, launch, join} is not a PBE. The normal secretaries want to deviate to {provide}:

$$Q < (1 - \pi)(Q - k) + \pi(Q + R + S - k)$$

3

{provide, provide, launch, launch, join} is not a PBE by Assumption 1.

4.

{provide, provide, launch, launch, join} is not a PBE. Normal governor wants to deviate to {not provide}.

All other PBEs are ruled out by Lemmas 3 and 5.

5. There is no mixed strategy equilibrium where the secretary mixes between {propose} (with probability x) and {not propose}, the normal governor provides only if the secretary propose, and the population mixes between {join} (with probability y) and {not join}. Note this is the only possible mixed strategy equilibrium. But the equilibrium is forestalled because the normal governor finds it undesirable to provide public good:

$$(1 - \pi)(Q - k) + \pi[y(\eta R + Q - k) + (1 - y)(Q - k)] < Q - k + \pi\eta R \leq Q$$

for any $y \in [0, 1]$.

Q.E.D.

Lemma 6

1. If $k \leq \frac{\pi Q[R - (2 - \mu)c]}{Q + S + (2 - \mu)c}$, then there are two equilibrium outcomes.

1.1 A mixed equilibrium exists where the secretary mixes between proposing public good and not, and the population mixes between joining the collective action and not. The normal governor provides public good only if secretary promises a transfer, and the benevolent governor always provides public good.

Denote that the normal secretary provides public good with probability x ; the population chooses {join} with probability y . We have:

$$\hat{x} = \frac{\mu}{(1 - \mu)^2} \left(\frac{(1 - \eta^*)R}{c} - 2 + \mu \right) = \frac{\mu}{1 - \mu} \left\{ \frac{(1 - \lambda)(1 - \eta^*)W(a)}{c} - 1 \right\}$$

$$\hat{y} = \frac{Q}{Q + (1 - \eta^*)R + S} = \frac{\lambda W(a)}{[\lambda + (1 - \hat{\eta})(1 - \lambda)]W(a) + S}$$

$\hat{\eta}$ satisfies that:

$$\hat{\eta}R = \frac{k(Q + R + S)}{\pi Q + k}$$

Specifically, $\hat{\eta}R$ increases with k .

And the secretary chooses {launch} after governor provides public good.

1.2 A pure strategy equilibrium exists, where both normal and benevolent secretaries do not propose, only benevolent governor provides public good, and population does not join collective action.

2. With $\frac{\pi Q[R-(2-\mu)c]}{Q+S+(2-\mu)c} < k < \pi R$, only above pure strategy equilibrium exists.

Proof:

The proof for \hat{x} and \hat{y} is similar to Lemma 4. Assume that the benevolent secretary always proposes.

\hat{x} has to satisfy:

$$\frac{\mu^2 + (1 - \mu)\mu}{\mu + (1 - \mu)(\mu + (1 - \mu)\hat{x})}(1 - \eta)R = c$$

And \hat{y} has to satisfy:

$$Q = (1 - \pi)Q + \pi[(1 - \hat{y}) * 0 + \hat{y}(Q + (1 - \hat{\eta})R + S)]$$

Solve for \hat{x} and \hat{y} , we have $\hat{x} = \frac{\mu}{(1-\mu)^2}(\frac{(1-\hat{\eta})R}{c} - 2 + \mu)$ and $\hat{y} = \frac{Q}{Q+(1-\hat{\eta})R+S}$.

Note that benevolent secretary does always propose:

$$(1 - \pi)Q + \pi[\hat{y}(Q + (1 - \hat{\eta})R + S)] > (1 - \pi)Q + \pi Q = Q$$

$\hat{\eta}$ has to make sure that normal governor find it more desirable to provide public good:

$$(1 - \pi)(Q - k) + \pi[\hat{y}(Q + \eta R - k) + (1 - \hat{y})(Q - k)] \geq Q$$

With $\hat{y} = \frac{Q}{Q+(1-\hat{\eta})R+S}$, rearrange the inequality:

$$\hat{\eta}R = \frac{k(Q + R + S)}{\pi Q + k}$$

$\hat{\eta} \leq 1$ if and only if:

$$\frac{k(Q + R + S)}{(\pi Q + k)R} \leq 1$$

$$k \leq \frac{Q}{Q + S} \pi R$$

Note that we also need to make sure that $\hat{x}, \hat{y} \in [0, 1]$. It is trivial for \hat{y} . For $\hat{x} \leq 1$, we need $\frac{\mu}{(1-\mu)^2} \left(\frac{(1-\eta^*)R}{c} - 2 + \mu \right) < 1$, or $\mu(1-\eta^*)R \leq c$, which is true for any η^* with Assumption 1 $\mu R < c$.

For $\hat{x} \geq 0$, we need $\frac{(1-\eta^*)R}{c} - 2 + \mu > 0$, or $(1-\eta^*)R \geq (2-\mu)c$. So $\eta^* \leq 1 - \frac{(2-\mu)c}{R}$. Note that $1 - \frac{(2-\mu)c}{R} \geq 0$, as it requires $R \geq (2-\mu)c$. As $2c - \mu c < 2c - R$, it suffices to show that $R > 2c - R$, which reduces to $R > c$. Note it is trivial that $1 - \frac{(2-\mu)c}{R} < 1$, so between $\hat{\eta} \leq 1$ and $\hat{\eta} \leq 1 - \frac{(2-\mu)c}{R}$, the later is binding.

So $\frac{k(Q+R+S)}{(\pi Q+k)R} \leq 1 - \frac{(2-\mu)c}{R}$, which reduces to $k \leq \frac{\pi Q[R-(2-\mu)c]}{Q+S+(2-\mu)c}$.

Note that the PBE passes Assumption 3 trivially. The proof of uniqueness is similar as Lemma 4.

Now assume that benevolent secretary does not propose. Note that the normal secretary also has no incentive to propose. If he does so, the best outcome he can expect is that the updated belief of population is $\frac{\mu^2}{\mu+(1-\mu)(1-\mu)}$, which is smaller than μ . Assume that normal governor provides with probability \hat{v} .

1. Suppose both normal and benevolent secretaries abstain. Evoke Assumption 3, we have

$$\hat{\mu} = \frac{[\mu + (1-\mu)\hat{v}]\mu\pi}{[\mu + (1-\mu)\hat{v}]\pi} = \mu$$

So population does not join. Given this, indeed normal and benevolent secretaries abstain, and $\hat{v} = 0$.

2. Suppose benevolent secretary lead,s and normal secretary abstains. Then $\hat{\mu} = \frac{[\mu+(1-\mu)\hat{v}]\mu\pi}{[\mu+(1-\mu)\hat{v}]\mu\pi} = 1$. The population wants to join, and the normal secretary wants to deviate and lead. A contradiction.

3. Suppose benevolent secretary abstains, and normal secretary leads. Then $\hat{\mu} = \frac{[\mu+(1-\mu)\hat{v}]\pi*0}{[\mu+(1-\mu)\hat{v}](1-\mu)\pi} = 0$. The population abstain, and the normal secretary wants to deviate and abstain.

4. Suppose both benevolent and normal secretaries lead. Then $\hat{\mu} = \frac{[\mu+(1-\mu)\hat{v}]\mu\pi}{[\mu+(1-\mu)\hat{v}]\pi} = \mu$. The population wants to abstain, and both normal and benevolent secretary want to deviate and abstain.

The claim is established.

For the case $\frac{\pi Q[R-(2-\mu)c]}{Q+S+(2-\mu)c} < k < \pi R$, note that the mixed strategy equilibrium is not feasible. The proof for the pure strategy equilibrium is exactly the same as before.

Q.E.D.

Proposition 6 (Reduced Collective Action Risk in Mixed Strategy Equilibrium)

If $k \leq \frac{\pi Q[R-(2-\mu)c]}{Q+S+(2-\mu)c}$, the probability of collective action is:

$$Pr(\text{collective action}) = \mu \frac{(1-\hat{\eta})R}{c} \frac{\pi Q}{Q+(1-\hat{\eta})R+S} = \mu \frac{(1-\hat{\eta})(1-2\lambda)W}{(\lambda+(1-\hat{\eta})(1-2\lambda))W+S} \frac{\pi\lambda W}{c}$$

Specifically, with the same level of competence, the probability of collective action under delegation is lower than that under concentration of powers:

$$Pr(\text{collective action}|\text{delegation}) = \mu \frac{(1-\hat{\eta})(1-2\lambda)W}{(\lambda+(1-\hat{\eta})(1-2\lambda))W+S} \frac{\pi\lambda W}{c}$$

$$< \mu \frac{(1-\lambda)W}{W+S} \frac{\pi\lambda W + k}{c} = Pr(\text{collective action}|\text{concentration})$$

And $Pr(\text{collective action})$ is a decreasing function of k .

Proof:

The key is that $\frac{(1-\hat{\eta})(1-2\lambda)W}{(\lambda+(1-\hat{\eta})(1-2\lambda))W+S} < \frac{(1-\lambda)W}{W+S}$ for $\hat{\eta} \in [0, 1]$.

Proposition 7 (Reduced Risk in Mixed Strategy Equilibrium)

The probability of collective action is:

$$Pr(\text{collective action}) = \mu \frac{R}{c} \frac{\pi Q}{Q + R + S} = \mu \frac{(1 - 2\lambda)W}{(1 - \lambda)W + S} \frac{\pi \lambda W}{c}$$

Specifically, with the same level of competence, the probability of collective action under weak delegation is lower than that under no delegation:

$$\begin{aligned} Pr(\text{collective action}|\text{weak delegation}) &= \mu \frac{(1 - 2\lambda)W}{(1 - \lambda)W(a) + S} \frac{\pi \lambda W}{c} \\ &< \mu \frac{(1 - \lambda)W(a)}{W(a) + S} \frac{\pi \lambda W(a) + k}{c} = Pr(\text{collective action}|\text{no delegation}) \end{aligned}$$

Proposition 8 (Public Good Provision but No Collective Action)

There is a (family of) PBE, in which benevolent secretary provides public good and normal secretary provides with probability $\hat{x}^s > \frac{\mu}{1-\mu}(\frac{R}{c}-1)$. Both of them do not initiate any collective action. The population does not join any collective action.

Proof:

$\{\cdot, \cdot, \cdot\}$ denotes whether benevolent secretary wants to provide public good, whether normal secretary wants to provide, and whether the population wants to join.

$\{\text{provide, provide with probability } \hat{x}^s > \frac{\mu}{1-\mu}(\frac{R}{c}-1), \text{not}\}$ is a PBE: the benevolent secretary does not deviate; the normal secretary does not deviate as he will get the same utility, because all cost of public good provision is borne by the governor; the population does not deviate by Assumption 1.

Uniqueness of pure PBE is similar as prior propositions and is omitted here.

Q.E.D.

Proposition 9

If $\xi < \frac{c-\mu R}{R(1-\mu)}$, then the equilibrium featuring positive risk of collective action when secretary dominates governor can be eliminated.

Proof:

Denote \hat{x}^s the probability that a strong secretary provides public good, and \hat{x}^w the probability that a weak secretary proposes to provide public good. \hat{y} is the probability that the population join collective action.

If $\hat{y} = \frac{Q}{Q+(1-\eta)R+S}$, there will be positive risk of collective action from weak secretary, and the strong secretary will never provide public good. Such situation can be eliminated by setting $k > \pi R$.

If $\hat{y} = \frac{Q}{Q+R+S}$ and $k > \pi R$, then under weak secretary, only benevolent governor provides public good. The strong secretary is indifferent between {provide} and {not provide}.

But to have $\hat{y} = \frac{Q}{Q+R+S}$, we need $\hat{\mu}R = c$. So it must be:

$$\frac{(1-\xi)\mu^2 + \xi\mu}{(1-\xi)\mu + \xi[\mu + (1-\mu)\hat{x}^s]}R = c$$

Solve for \hat{x}^s :

$$\hat{x}^s = \frac{\mu}{\xi(1-\mu)} \left\{ [(1-\xi)\mu + \xi] \frac{R}{c} - 1 \right\}$$

To eliminate the equilibrium, we need to guarantee that $\hat{x}^s < 0$, or:

$$[(1-\xi)\mu + \xi] \frac{R}{c} - 1 < 0$$

$$\xi < \frac{c - \mu R}{R(1 - \mu)}$$

Note $\frac{c-\mu R}{R(1-\mu)} < 1$ if $R > c$.

Q.E.D.

Corollary 1

If $c > \sqrt{\mu}R$, then for $\xi^* \in (\frac{\mu}{1-\mu}(\frac{R}{c} - 1), \frac{c-\mu R}{R(1-\mu)})$, the population will refuse to join the collective action of either current secretary or future secretary who is currently governor. Given this, only benevolent and strong secretary and benevolent and strong governor will provide public good, and all secretaries will abstain from collective action.

Proof:

The key is that we need to have:

$$\frac{\mu}{1-\mu}(\frac{R}{c} - 1) < \frac{c - \mu R}{R(1-\mu)}$$

which reduces to $c^2 > \mu R^2$.

The strategies are obviously optimal given our analysis before.

The population's belief is off-equilibrium. So it has to be identified by Assumption 3:

$$\hat{\mu} = \bar{\mu} = \frac{(1-\xi)\mu^2\pi + \xi\mu\pi}{(1-\xi)\mu\pi + \xi\mu\pi} = (1-\xi)\mu + \xi$$

It is indeed not optimal for the population to join as for $\xi < \frac{c-\mu R}{R(1-\mu)}$:

$$[(1-\xi)\mu + \xi]\frac{R}{c} - 1 < 0$$

Q.E.D.

B.2 Micro-foundation for Focusing on Equilibria with Pure Strategy for Section 3.5 and Section 3.6

I am going to provide a micro-foundation to justify my focus on equilibria with pure strategy for {launch} and {not launch}. Let us assume that instead the secretary can choose mixed strategy between {launch} and {not launch}. The normal secretary chooses {launch} with probability z_1 , and the benevolent secretary chooses {launch} with probability z_2 .

Note that it is impossible for the benevolent secretary to mix: if that's the case, $\hat{y}_2 = \frac{Q}{Q+(1-\eta)\gamma R+S}$. This means that the normal secretary never launch, because he needs a $\hat{y}_1 = \frac{Q}{Q+(1-\eta)R+S} > \hat{y}_2$. This means that the population knows for sure that anyone leading a collective action is benevolent, so they do not want to choose $\hat{y} = 0$.

If the normal secretary mix, we must have $\hat{y} = \frac{Q}{Q+(1-\eta)R+S}$. Note that normal secretary always have incentive to mix between {propose collusion} and {not}: $(1-\pi)Q + \pi((1-\hat{z}_1)Q + \hat{z}_1[(1-\hat{y}) * 0 + \hat{y}(Q + (1-\eta)R + S)]) = (1-\pi\hat{z}_1)Q + \pi\hat{z}_1Q = Q$.

The secretary will compensate enough so that normal governor will always provide public good. To induce population to mix between {join} and {not join}: $\frac{[\mu^2+(1-\mu)\mu]\pi}{\{\mu[\mu+(1-\mu)\hat{z}_1]+(1-\mu)[\mu+(1-\mu)\hat{x}_1\hat{z}_1]\}\pi} (1-\hat{\eta})R = c$.

The probability of collective action is:

$$\{\mu[\mu + (1 - \mu)\hat{z}_1] + (1 - \mu)[\mu + (1 - \mu)\hat{x}_1\hat{z}_1]\}\pi\hat{y} = \mu \frac{(1 - \eta)R}{(1 - \eta)R + Q + S} \frac{\pi Q}{c}$$

The normal governor requires an compensation ηR that satisfies:

$$(1 - \pi)(Q - k) + \pi\{(1 - \hat{z}_1)(Q - k) + \hat{z}_1[(1 - \hat{y})(Q - k) + \hat{y}(Q + \eta R - k)]\} = Q$$

$$\eta R = \frac{k(Q + R + S)}{(\pi Q + k)\hat{z}_1}$$

The probability of collective action is:

$$\mu \frac{(1 - \eta)R}{(1 - \eta)R + Q + S} \frac{\pi Q}{c}$$

At first glance, the formula is exactly the same as the one in Proposition 6. But actually the two probabilities are different, as $\hat{\eta}$ depends on \hat{z}_1 . Note that ηR is smallest with $\hat{z}_1 = 1$, which means that $\text{Prob}(\text{collective action}) = \mu \frac{(1-\eta)R}{(1-\eta)R+Q+S} \frac{\pi Q}{c}$ is the highest with $\hat{z}_1 = 1$. So by treating $\hat{z} \in \{0, 1\}$, we already consider the riskiest case for the principal, and is consequently without loss of generality.

For Section 6, the argument is even stronger. We can see that in section 6.2, whatever \hat{z}_1 is, the probability of collective action is the same.

With similar argument, when secretary completely dominates the governor, the only possible case involving mixed strategies is that the normal secretary mixes between {provide} and {not provide}, and the normal secretary mixes between {lead} and {not lead}. To induce such mixed strategy, $\hat{y} = \frac{Q}{Q+R+S}$. To induce population to mix, the normal secretary must choose \hat{x}_1 and \hat{z}_1 such that $\frac{\mu}{\mu+(1-\mu)\hat{x}_1\hat{z}_1} R = c$, or $\hat{x}_1\hat{z}_1 = \frac{\mu}{1-\mu}(\frac{R}{c} - 1)$. So $\text{Prob}(\text{collective action}) = (1 - \mu)\hat{x}_1\hat{z}_1\pi\hat{y} + \mu\pi\hat{y} = \mu \frac{R}{R+Q+S} \frac{\pi Q}{c}$, which is independent of \hat{z}_1 that the normal secretary chooses. So our analysis in Section 6 is without loss of generality.

Let us turn to section 5.2, or when $k > \pi R$. In this case, only benevolent governor provides public good. If we allow the secretary to mix between {lead} and {not lead}, there is another equilibrium with positive probability of collective action. To induce population to mix, $\frac{\mu\hat{z}_2}{(1-\mu)\hat{z}_1+\mu\hat{z}_2} R = c$. Again, the only possible mixed strategy equilibrium involves $\hat{z}_1 \in (0, 1)$ and $\hat{z}_2 = 1$. We have $\frac{\mu}{(1-\mu)\hat{z}_1+\mu} R = c$, so $\hat{z}_1 = \frac{\mu}{1-\mu}(\frac{R}{c} - 1)$. To induce normal secretary to mix, $\hat{y} = \frac{Q}{Q+R+S}$. So the probability of collective action is:

$$\mu\{(1 - \mu)\hat{z}_1\pi\hat{y} + \mu\pi\hat{y}\} = \mu^2 \frac{R}{R + Q + S} \frac{\pi Q}{c}$$

Note that μ denotes the fraction of benevolent secretary. μ is probably small, which means that $\mu^2 \frac{R}{R+Q+S} \frac{\pi Q}{c}$ is truly second-order. As usual, k disappears from the formula.

For $\frac{\pi Q[R-(2-\mu)c]}{Q+S+(2-\mu)c} < k < \pi R$, we know that the normal secretary cannot mix between {propose} and {not propose}. Assume that normal governor provides public good with probability \hat{z}_g . And the normal secretary chooses to {lead} with probability \hat{z}_1 . Then:

$$\hat{\mu} = \frac{[\mu + (1 - \mu)\hat{z}_g]\mu\pi}{[\mu + (1 - \mu)\hat{z}_g][\mu + (1 - \mu)\hat{z}_1]\pi} = \frac{\mu}{\mu + (1 - \mu)\hat{z}_1}$$

So again, to have $\hat{\mu}(1 - \hat{\eta})R = c$, we need $\hat{z}_1 = \frac{\mu}{1-\mu}(\frac{(1-\hat{\eta})R}{c} - 1)$. To induce normal secretary to mix, we need $\hat{y} = \frac{Q}{Q+(1-\hat{\eta})R+S}$.

To compensate normal governor, we need:

$$(1 - \pi)(Q - k) + \pi\{(\mu + (1 - \mu)\hat{z}_1)\hat{y}(Q + \eta R - k) + [1 - (\mu + (1 - \mu)\hat{z}_1)](Q - k)\} \geq Q$$

Note that we need $(1 - \pi)(Q - k) + \pi\{(\mu + (1 - \mu)\hat{z}_1)\hat{y}(Q + \hat{\eta}R - k) + [1 - (\mu + (1 - \mu)\hat{z}_1)](Q - k)\} = Q - k + \pi(\mu + (1 - \mu)\hat{z}_1)\hat{y}\hat{\eta}R \geq Q$, or $\pi(\mu + (1 - \mu)\hat{z}_1)\hat{y}\hat{\eta}R = \pi\mu\frac{R}{c}\hat{y}\hat{\eta}R > k$. Note that as $\frac{\pi Q[R - (2 - \mu)c]}{Q + S + (2 - \mu)c} < k$, we already have $(1 - \pi)(Q - k) + \pi(\hat{y}(Q + \eta R - k) + (1 - \hat{y})(Q - k)) < Q$, or $\pi\hat{y}\hat{\eta}R < k$. As $\mu\frac{R}{c} < 1$, so $\pi\mu\frac{R}{c}\hat{y}\hat{\eta}R < \pi\hat{y}\hat{\eta}R < k$. So the normal governor will never mix. So $\hat{z}_1 = 0$, and the probability of collective action is again:

$$\mu\{(1 - \mu)\hat{z}_1\pi\hat{y} + \mu\pi\hat{y}\} = \mu^2\frac{R}{R + Q + S}\frac{\pi Q}{c}$$

Note that with the additional equilibrium, the maximum probability of collective action with $k > \frac{\pi Q[R - (2 - \mu)c]}{Q + S + (2 - \mu)c}$ is still lower than the case $k > \frac{\pi Q[R - (2 - \mu)c]}{Q + S + (2 - \mu)c}$. Specifically, we want to show that:

$$\mu^2\frac{R}{R + Q + S}\frac{\pi Q}{c} < \mu\frac{(1 - \hat{\eta})R}{c}\frac{\pi Q}{Q + (1 - \hat{\eta})R + S}$$

For the right hand side, the lowest probability of collective action is reached when $k = \frac{\pi Q[R - (2 - \mu)c]}{Q + S + (2 - \mu)c}$. In this case, as $(1 - \hat{\eta})R = (2 - \mu)c$, we have: $\mu\frac{(1 - \hat{\eta})R}{c}\frac{\pi Q}{Q + (1 - \hat{\eta})R + S} = \mu(2 - \mu)\frac{\pi Q}{Q + (1 - \hat{\eta})R + S}$. We need:

$$\mu^2\frac{R}{R + Q + S}\frac{\pi Q}{c} < \mu(2 - \mu)\frac{\pi Q}{Q + (1 - \hat{\eta})R + S}$$

$$\mu\frac{R}{c} < (2 - \mu)\frac{R + Q + S}{Q + (1 - \hat{\eta})R + S}$$

Note that $(2 - \mu)\frac{R + Q + S}{Q + (1 - \hat{\eta})R + S} > 2 - \mu > 1$, so we need $\mu\frac{R}{c} < 1$. This is guaranteed by Assumption 1.

So ignoring the mixture between {lead} and {not lead} in the main text does not change our results at all. The central authority still have strong incentive to implement decentralization and corresponding meritocracy.

Notice that ignoring the mixture between {lead} and {not lead} actually strengthens my analysis in section 7: it eliminates risk of collective action under balance of power. But even in this case, section 7 shows that the central authority still maintains a stochastic advantage of secretary because of dynamic concern that governor will work as secretary in the future.

B.3 A Model with Alternative Timeline

Readers may question whether the results are robust to timeline assumptions. Specifically, what will happen if the population launches spontaneous protest first and then secretary organize it and turns it into collective action? In this section, I am going to show that the model is very robust. Specifically, I will change the timeline during the mobilization stage, and everything else will be the same. We will see that all results of meritocracy are still very strong, but we lose predictions on decentralization. This is not a problem as long as the timeline assumed in workhorse model arises with positive probability. The alternative model also has the advantage of being especially simple and clear. Many results before that relies on technical lemmas can be easily proved under the alternative model.

Benchmark case

The general timeline is the same as before: 1. appointment stage; 2. signaling stage; 3. mobilization stage; 4. divide the pie. The only change is about mobilization stage. I am going to treat decentralization k as an exogenous variable for this section.

Mobilization stage: The population decides whether to launch a spontaneous protest with a cost of c . The secretary then decides whether to abstain or organize and turns the protest into a collective action. If the population launches a protest, and the secretary does not organize, the normal secretary will get $Q - k$, the benevolent secretary will get $Q + \gamma e(k) - k$, and the population gets $e - c$. If the secretary organizes, then the collective action succeeds.

Obviously, in the pie-division stage, the normal secretary will capture all benefits and gets $Q + R - k$, and benevolent secretary will award all benefits to the population and gets $Q + \gamma R + \gamma e - k$.

First, analyze the best response of players in the game.

1. Normal secretary always choose {capture all benefits}; benevolent secretary always choose {award all benefits}.

2. Both normal and benevolent secretary choose {organize a collective action} given that the population launches one.

3. Denote the belief that the secretary is benevolent at the beginning of mobilization stage as $\hat{\mu}$. The population launches a protest if $\mu R > c$; mix (with probability \hat{y}) between {launch} and {abstain} if $\mu R = c$; and choose {abstain} if $\mu R < c$.

4. The normal secretary chooses {provide public good} if

$$(1 - \pi)(Q - k) + \pi[(1 - \hat{y})(Q - k) + \hat{y}(Q + R + S - k)] > Q$$

will mix between {provide} and {not provide} if $(1-\pi)(Q-k) + \pi[(1-\hat{y})(Q-k) + \hat{y}(Q+R+S-k)] = Q$, and chooses {not provide} if $(1-\pi)(Q-k) + \pi[(1-\hat{y})(Q-k) + \hat{y}(Q+R+S-k)] < Q$.

The benevolent secretary always chooses {provide}. Denote $b = \gamma e(k) - k > 0$, we have:

$$(1-\pi)(Q+b) + \pi[(1-\hat{y})(Q+b) + \hat{y}(Q+\gamma R+S+b)] = Q+b + \pi\hat{y}(\gamma R+S) > Q$$

So we prove that benevolent secretary is a commitment type. Note how easy it is to prove this result compare to Lemma 2.

To consider a especially simple case, I assume that $k > \pi(Q+S)$. In this case, the normal secretary also becomes a commitment-type. He is committed to not provide public good at all:

$$(1-\pi)(Q-k) + \pi(Q+R+S-k) < Q$$

Proposition B1 (Loyalty-Competence Tradeoff under Alternative Model)

$\{\cdot, \cdot, \cdot, \cdot, \cdot\}$ denotes whether benevolent and normal secretaries provide public good, whether the population launches a protest, and whether the benevolent and normal secretaries organize the collective action.

If $k > \pi(R+S)$,

B1.1 Collective action risk in equilibrium

If $R \geq c$, then in the unique PBE, {provide public good, not provide, protest, organize, organize} is the strategy profile.

B1.2 No collective action risk in equilibrium with low competence

If $R < c$, then in the unique PBE, {provide, not provide, not protest, organize, organize} is the strategy profile.

\bar{R} denotes principal's surplus produced by the most competent secretary, If $(1-\mu\pi)(\bar{R}+S) \geq (c+S)$, the principal implements equilibrium with positive probability of protest and chooses the most competent secretary. If $(1-\mu\pi)(\bar{R}+S) < (c+S)$, the principal implements the collective-action-free equilibrium with low competence $R(a) = c$

Proof:

From our discussion before, we already show {provide, not provide} and {organize, organize} are optimal. The only thing that we need to verify is the strategy of the population.

If $R \geq c$, given that only benevolent secretary provides public good, $\hat{\mu} = 1$. So $\hat{\mu}R = R > c$, it is optimal for the population protest.

If $R < c$, then $\hat{\mu}R = R < c$. So the population does not protest.

Given the behaviors of secretaries and population, the principal's strategy is trivial to prove.

The uniqueness of pure strategy PBEs is trivial to prove.

Q.E.D.

From the proof, we can see that the model allows an especially simply demonstration of the main benchmark result, loyalty-competence trade-off. We have the same intuition that the central authority either face a competent but revolutionary secretary or a loyal but mediocre one.

Delegation Promotes Meritocracy

The setup is the same as section 5.1. The only difference is that mobilization stage is modified and it is the same as section 8.1. Again, let us analyze the best responses first.

1. Normal secretary always choose {capture all benefits}; benevolent secretary always choose {award all benefits}.

2. Both normal and benevolent secretary choose {organize a collective action} given that the population launches one.

3. Denote the belief that the secretary is benevolent at the beginning of mobilization stage as $\hat{\mu}$. The population launches a protest if $\mu(1 - \eta)R > c$; mix (with probability \hat{y}) between {launch} and {abstain} if $\mu(1 - \eta)R = c$; and choose {abstain} if $\mu(1 - \eta)R < c$.

4. The normal governor chooses {provide public good} if

$$(1 - \pi)(Q - k) + \pi[(1 - \hat{y})(Q - k) + \hat{y}(Q + \eta R + S - k)] > Q$$

will mix between {provide} and {not provide} if $(1-\pi)(Q-k) + \pi[(1-\hat{y})(Q-k) + \hat{y}(Q+R+S-k)] = Q$, and chooses {not provide} if $(1-\pi)(Q-k) + \pi[(1-\hat{y})(Q-k) + \hat{y}(Q+\eta R+S-k)] < Q$.

The benevolent governor always chooses {provide}. Denote $b = \gamma e(k) - k > 0$, we have:

$$(1-\pi)(Q+b) + \pi[(1-\hat{y})(Q+b) + \hat{y}(Q+\gamma(1-\eta)R+\eta R+S+b)] = Q+b + \pi\hat{y}(\gamma(1-\eta)R+\eta R+S) > Q$$

So we prove the property of commitment-type for benevolent governor.

5. Secretaries' collusion behavior will be left in the proof of Proposition 10 to discuss.

Proposition B2 (Risk-free PBE and Meritocracy)

$\{\cdot, \cdot, \cdot, \cdot, \cdot\}$ denotes whether benevolent and normal governors provide public good, whether the population protest, and whether the benevolent and normal secretaries organize the collective action.

If $k > \pi(R+S)$, {provide, not provide, not protest, organize, organize} is the strategy profile in any PBE. There is no collective action, so the central authority chooses the most competent secretary and governor.

Proof:

If $k > \pi(R+S)$, the normal governor never accepts any proposal of collusion. This means that only benevolent governor provides public good, and normal governor does not. Population's posterior on secretary's benevolence is unchanged:

$$\hat{\mu} = \frac{\mu * \mu}{\mu} = \mu$$

Given that $\mu R < c$, the population refuses to launch any protest.

Given that there will be no protest, the secretary will be indifferent between propose collusion or not, as the distribution of revolutionary benefit is off-equilibrium and will not materialize.

Note the uniqueness of the strategy profile {provide, not provide, not protest, organize, organize} is trivial to prove.

Q.E.D.

Delegation Promotes Meritocracy with Weak Governor

The setup is the same as section 6.2. The only difference is that mobilization stage is modified and it is the same as section 8.1. Recall that the governor has no autonomy in decision making, yet all cost of public good provision falls on him. Again, let us analyze the best responses first.

1. Normal secretary always choose {capture all benefits}; benevolent secretary always choose {award all benefits}.
2. Both normal and benevolent secretary choose {organize a collective action} given that the population launches one.
3. Denote the belief that the secretary is benevolent at the beginning of mobilization stage as $\hat{\mu}$. The population launches a collective action if $\mu R > c$; mix (with probability \hat{y}) between {launch} and {abstain} if $\mu R = c$; and choose {abstain} if $\mu R < c$.
4. The normal secretary always choose {provide public good} if $\hat{y} > 0$

$$(1 - \pi)(Q) + \pi[(1 - \hat{y})(Q - k) + \hat{y}(Q + \eta R + S - k)] > Q$$

and will mix between {provide} and {not provide} if $\hat{y} = 0$.

The benevolent secretary always chooses {provide}. We have:

$$(1 - \pi)(Q + \gamma e) + \pi[(1 - \hat{y})(Q + \gamma e) + \hat{y}(Q + \gamma R + S + \gamma e)] = Q + \gamma e + \pi\hat{y}(\gamma R + S) > Q$$

Proposition B3 (Risk-free PBE and Meritocracy)

$\{\cdot, \cdot, \cdot, \cdot, \cdot\}$ denotes whether benevolent and normal secretaries provide public good, whether the population protest, and whether the benevolent and normal secretaries organize the protest.

In any PBE, the population chooses {not protest}, and the benevolent and normal secretaries choose {organize, organize}. The benevolent secretary always provides public good, and the normal secretary provides public good with probability $\hat{x} \geq \frac{\mu}{(1-\mu)}(\frac{R}{c} - 1)$.

Proof:

Note that {protest} cannot be a PBE strategy. As the best response is for both secretaries to provide public good, so $\hat{\mu} = \mu$, and it is optimal for the population to deviate to {not protest}.

To support {not protest}, it must be the case that the normal governor provides public good with probability \hat{x} that satisfies:

$$\frac{\mu}{(1 - \mu)x + \mu} R \leq c$$

So $\hat{x} \geq \frac{\mu}{1-\mu}(\frac{R}{c} - 1)$.

Note that the normal governor cannot increase his utility by single-deviation. He always get Q , given that the population does not protest.

There are no other possible equilibria.