Journal of Economic Behavior and Organization 184 (2021) 606–652

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Intergovernmental communication under decentralization[☆]

Shiyu Bo[a], Liuchun Deng[b], Yufeng Sun[c,*], Boqun Wang[d,e]

[a] Institute for Economic and Social Research, Jinan University, China
[b] Social Sciences Division, Yale-NUS College, Singapore
[c] School of Public Economics and Administration, Shanghai University of Finance and Economics, Shanghai 200433, China
[d] School of Banking & Finance, University of International Business and Economics, China
[e] China Financial Policy Research Center, Renmin University of China,China

ARTICLE INFO

Article history:
Received 3 April 2020
Revised 2 February 2021
Accepted 6 February 2021

JEL classification:
H10
H70
O50
P20

Keywords:
Decentralization
Authoritarian regime
Output
Volatility
Communication
China

ABSTRACT

We develop a model of inter-governmental communication to study the impact of decentralization on economic performance under an authoritarian regime. Decentralization shifts the decision power of policy-making from the central government to the local. The local government has the information advantage, but it also has the loyalty concern to follow the policy prescriptions from the central. We show that the loyalty concern impacts the economic outcome of decentralization by distorting both inter-governmental transmission of information and final policy-making. A strict adherence to the central renders decentralization welfare-reducing, causing low output and high volatility. Our model implications shed light on the history of decentralization reforms in the People's Republic of China. A reinterpretation of our analytical framework also extends the core insights to representative democracies.

1. Introduction

Decentralization, being political or economic, has become a catchword in the ongoing discussion about structural reforms in the developing world. Historically, decentralization did not always produce desirable outcomes and sometimes, in fact, led to socio-economic catastrophes. Why would decentralization fail? A large literature examines the role of heterogeneity, competition, and externality across different localities. In this paper, we investigate the vertical linkage between the central and the local governments and formalize the inter-governmental communication, a less explored channel in the

☆ We would like to thank the editor, Mark Koyama, and two anonymous referees, whose comments and suggestions greatly improve this paper. We are grateful to Ying Chen for helpful conversations. We thank Hülya Eraslan, Pravin Krishna, Cong Liu, Jiwei Qian, Zhiyong Yao, Sezer Yasar, Miaojie Yu, our discussants, Yu-Hsiang Lei, Vanesa Pesqué-Cela, and Andrea Schneider, and participants at conferences for helpful comments. We thank Hongxia Ma for excellent research assistance. This paper subsumes an earlier version entitled "A Tale of Two Decentralizations: Volatility and Economic Regimes". This project was supported by National Natural Science Foundation of China (No. 71803064), Guangzhou Philosophy and Social Science Planning Grant (No. 2020GZGJ42), the Fundamental Research Funds for the Central Universities (No. 20JNTZ16), the 111 Project of China (No. B18026) and the Start-up Grant from Yale-NUS College.

* Corresponding author.
E-mail address: sun.yufeng@mail.shufe.edu.cn (Y. Sun).

https://doi.org/10.1016/j.jebo.2021.02.007
0167-2681/

literature but oftentimes hinted in historical anecdotes. Despite the local having information advantage, as forcefully argued by Hayek (1945), our framework demonstrates how loyalty concern, the political incentives of the local government to follow the policy prescriptions from the central, renders decentralization welfare-reducing through a tightly specified communication process, thus highlighting the distortion deeply rooted in political hierarchy.

Our formal analysis is motivated by the two major decentralization reforms in the history of the People's Republic of China (PRC, henceforth). The "reform and opening up" in 1978, decentralizing economic decision making from the central to the local, has long been understood as the trigger of China's growth over the last four decades. In sharp contrast to the huge success of the 1978 reform is the sometimes forgotten story of PRC's first major decentralization reform in the late 1950s. This wave of decentralization, which has been observed to share common essentials with the 1978 reform, however, produced disastrous outcomes including the Great Chinese Famine, claiming millions of lives (Wu and Reynolds, 1988).[1] Against this backdrop, we build a politico-economic model of intergovernmental communication. Our theoretical result sheds light on the contrasting experience following the two decentralization reforms in China and, perhaps more importantly, regional variation in economic downturns during the first wave of decentralization.[2] The theory may also be used broadly to rationalize the mixed outcomes of decentralization in many transition economies under (semi-)authoritarian regimes over the last three decades.

Our framework identifies two information-based channels through which loyalty concern impacts the economic performance of decentralization. First, loyalty concern directly changes the use of information in local governments' decision making. In the extreme case like the reform in the 1950s, the local bureaucrats' own knowledge of the local economy was often irrelevant as pursuit of economic betterment bore great political risks. Equally important but being less appreciated in the previous studies is the second channel: Loyalty concern alters endogenous allocation of efforts between information acquisition and transmission. The information advantage of being local could completely be squandered when the local bureaucrats are strongly motivated to decipher policy documents from the central rather than directly acquire useful information about the economy.[3] The success of China's 1978 reform can then be attributed to the bundling of promotion with economic performance as documented by Li and Zhou (2005):[4] The incentive of the local bureaucrats to signal loyalty at the expense of the economy through either channel has been substantially dampened.

In the model, a central government and a local government engage in policy making subject to uncertainty. Decentralization shifts the decision power of policy-making from the central government to the local which holds information advantage. Under a decentralized regime, the local government's decision problem has two layers, which give rise to the two aforementioned sources of distortion. The local government first decides on how to allocate its resources between directly acquiring information from the economy and indirectly seeking policy advice from the central government. Based on the information obtained, the local government then makes the policy decision. We demonstrate, with minimum parametric assumption, that decentralization improves the economic performance, bringing about higher output and lower volatility, if and only if the loyalty concern of the local bureaucrats is sufficiently weak.

Our model is directly related to the long-standing debate over centralization versus decentralization. In a seminal paper, Tiebout (1956) first points out the efficiency of decentralization hinges on inter-jurisdictional competition and individual's voting by one's feet. Oates (1972) argues that even though centralization can internalize the spillovers across districts, the accompanying uniformity produces inefficiency, since preferences are heterogeneous. Beyond its economic consequences, Weingast (1995) emphasizes how market-preserving federalism secures the political foundations of markets. The trade-off between conflicts of interests under centralization and (informational or non-informational) externality problems under decentralization is further formalized in a political-economic framework (Besley and Coate, 2003) and recently elaborated in an alternative policy experimentation setting (Cheng and Li, 2019). Alternative theoretical arguments suggest that decentralization could avoid the accountability problem (Seabright, 1996), while it may induce a race-to-the-bottom competition between local governments (Keen and Marchand, 1997) and corrode the state capacity by locally shielding firms from central regulations and tax collectors (Cai and Treisman, 2004). This paper contributes to this literature by offering a new perspective of vertical information transmission. Depending on the institutional contexts, there are varieties of decentralization in practice, being fiscal, administrative, and political (Qian and Roland, 1998; Zhuravskaya, 2000; Bardhan, 2002; Jin et al., 2005; Enikolopov and Zhuravskaya, 2007; Suzuki, 2019; Bo, 2020). Abstracting from its specific content, our work goes to the very nature of decentralization, the shift of decision power from the central to the local government.

This paper joins a large literature on the role of institutions, decentralization and centralization in particular, in economic history (North, 1990; Acemoglu et al., 2005). Political fragmentation, as an extreme case of decentralization, has been

[1] Che et al. (2017) also juxtapose these two reforms in China and construct an overlapping generation model to characterizes the two-way relationship between decentralization and career concern. They mainly examine the nexus between political career incentive and decentralization through the lens of public good provision.

[2] To be sure, the disastrous outcome of the Great Leap Forward is due to the central planning regime. Rather than propose an alternative theory of the Great Leap Forward, our model intends to highlight a structural parameter in the authoritarian regime that could give rise to differential decentralization outcomes.

[3] For example, in Anhui, the province that was hardest hit by the great famine, it is unclear whether the provincial leaders really knew their local situation better than the central did.

[4] As a seminal paper, Li and Zhou (2005) initiate a large literature on politico-economic determinants of personnel control in China. Our work complements this literature by focusing exclusively on the vertical dimension of political hierarchy. As explained in Section 4.5.1, competition across localities further strengthens the channels highlighted in our model.

regarded as a primary driving force in the rise of Europe: frequent warfare led to the advancement of technology and enhancement of state capacity (Tilly, 1990; Dincecco, 2011; Karaman and Pamuk, 2013; Hoffman, 2015). On the contrary, the long-standing political centralization in China contributed to maintaining population growth and governance under relatively low tax rates (Wong, 1997; Rosenthal and Wong, 2011; Ko et al., 2018; Ma and Rubin, 2019). Notably, Sng (2014) offers a neat game-theoretic model to clarify how the agency problem that arises from taxation deteriorates with the size of a country. In the concluding section of that paper, the author discusses decentralization as a potential solution to the agency problem and its associated pitfalls. Our work highlights another form of the agency problem in a large country, the communication friction inherent to a gigantic bureaucratic structure. Beyond Europe and China, of which the comparison is the traditional focus of the Great Divergence literature (Pomeranz, 2000), historical evidence from other parts of the world such as Japan and African countries (Michalopoulos and Papaioannou, 2013; Koyama et al., 2018) also reveals the potential value of centralized and unified regimes in long-run economic growth. Using a Hotelling-type model with endogenous investment in state capacity, Koyama et al. (2018) argue that China's decentralization in response to multiple geopolitical threats was responsible for its failure in building a modern state in the late 19th century, whereas Japan, as an island state, during the same period had more incentives to move towards political centralization and later successfully modernized itself. Our work contributes to this strand of the literature by formalizing a novel source of the heterogeneous outcomes of political decentralization: the local bureaucrats' career incentives.

Specifically, the career incentives of local officials played a crucial role in the failure of the decentralization reform in the late 50s in China. Since political loyalty paid off, officials became blind followers rather than critics of the wishful thinking at the very top (Kung and Chen, 2011; Li and Yang, 2005), and therefore, the benefits of local information, which has long been argued as a key reason for decentralization (Hayek, 1945),[5] cannot be fully reaped following decentralization. This line of informal reasoning is intuitive, but it could not explain why the two waves of decentralization in China yielded completely opposite outcomes nor why there was huge regional variation in socio-economic outcomes during the Great Leap Forward. More fundamentally, it does not clarify the nature of how career incentives, loyalty concern in particular,[6] distorts acquisition and use of local information in an authoritarian regime.[7] This paper attempts to fill this void with a formal framework.

The degree of loyalty concern, which serves as the key explanatory variable in our model, has been examined by an active literature on the roots of loyalty in authoritarian regimes. In his influential book, Svolik (2012) carries out a systematic investigation of the problem of authoritarian power-sharing. Motivated by rich empirical evidence, a formal game-theoretic framework is developed to characterize the balance of power between the dictator and the ruling coalition and how it evolves over time. As pointed out by Svolik (2012), the loyalty concern is an incentive structurally inherent to the authoritarian regime and this concern is crucial in the understanding of politicians decisions in key political events in the history of many autocracies. Egorov and Sonin (2011) endogenize the trade-off between loyalty and competence in a principal-agent framework. Their model highlights how the inherent agency problem in autocracies gives rise to long-term polico-economic inefficiencies. The loss of information from the subordinate's side echoes the "yes-man" theory of Prendergast (1993), according to which an incentive contract could endogenously give rise to inefficient conformity of subordinates to the leaders. This paper abstracts from the underlying political distortion arising from a dictatorial environment. We take the degree of loyalty concern as our model primitive while enriching and focusing on the channel of endogenous information acquisition and communication.

Our work also complements a growing literature that links the outcome of decentralization with the state capacity of the local governments (Besley and Persson, 2014; Bardhan, 2016; Bellofatto and Besfamille, 2018). In the context of our historical setting, Lu et al. (2020) exploit the Red Army's Long March as a quasi-natural experiment to document a causal impact of state capacity proxied by the local communist party membership on socio-economic outcomes.[8] During the Maoist era, counties with stronger state capacity had already made more progresses in education attainment, road construction, and agriculture mechanization, while it was until the post-1978 reform era that state capacity manifested itself in output growth. Their findings substantiate empirically, to some extent, one of our key modeling assumptions that the local government (with sufficient state capacity) holds information advantage and are consistent with our theoretical prediction that, once loyal-driven distortions are alleviated, decentralization raises economic output. At a perhaps more fundamental level, the notion of information processing capacity and the inter-governmental communication friction in our rational inattention framework gets to the heart of the matter in Mann's original writing on two forms of the state power: despotic and infrastructural (Mann, 1984). The success or failure of decentralization hinges on the interplay between despotic and infrastructural power.

[5] Local information, knowingly difficult to measure, has been empirically identified as one of the driving forces of decentralization in the reform era of China; see, for example, an influential empirical study by Huang et al. (2017).

[6] It should be emphasized that loyalty concern is conceptualized and modeled in a relative sense. In the 1980s following the second decentralization reform, political loyalty may still be important in promotion, but due to fiscal decentralization, ideological shifts, and various economic motivations, its relative weight became smaller.

[7] One notable exception is the empirical work by Fan et al. (2016), which documents the information distortion in local governments' reports to the central before and during the great famine.

[8] It cannot be completely ruled out that variation in state capacity across localities may also stem from legacies in imperial China. See, for example, Chen et al. (2020) and Xue (2021).

Moreover, our framework provides a tool to analyze the consequences of decentralization beyond the first moment, that is, the level terms which the empirical work predominantly focuses on (Mookherjee, 2015). Partly due to the lack of theoretical underpinnings, it is until very recently that a burgeoning literature starts to tackle the relationship between decentralization and volatility.[9] In his pioneering paper, Nishimura (2006) studies theoretically whether and how fiscal decentralization could lower volatility of economic growth and the key theoretical predication is tested using the state-level data of the United States by Akai et al. (2009). Incorporating policy-making margin into a macroeconomic framework, Cheng et al. (2018) study both theoretically and empirically the relationship between governmental system and volatility under democracy. Our information-based framework adds to this literature as the economic performance in our model can be both measured by the output level and volatility.

Last, we model the tradeoff between direct information acquisition and inter-governmental communication in the fashion of rational inattention as in Sims (2003). Bolton et al. (2012) touch upon the role of rational inattention in information flows within an organization. With a network grounding, Dessein et al. (2016) discuss how to allocate limited attention optimally in organizing production. A technical contribution of this paper is that we provide the closed form solution to a model of rational inattention with two *conditionally correlated* sources of uncertainty.

The rest of this paper is organized as follows. Section 2 discusses the historical setting that motivates our model. Section 3 describes the baseline model. Section 4 presents the main results of the baseline model and connects them with the historical episodes covered in Section 2. Section 5 discusses theoretical extensions. Section 6 concludes.

2. Historical setting

2.1. Decentralization reforms in the history of the People's Republic of China

Ever since its establishment in 1949, the PRC started building its socialist central planning economy in the Soviet style. In 1956, during the period of the first five-year plan, socialist transformation of the agricultural sector, the handicrafts, and capitalist industry and commerce was largely completed (Bowie, 1962), which marked the accomplishment of the transition into a socialist economy. Very soon the Communist Party leaders realized the issue of over-concentration of decision power in this new central planning regime.[10] The discussion and debate at the very top led to the first wave of decentralization reforms from 1956 to 1958. According to Wu and Reynolds (1988), the reform policy package consists of: (1) transferring the control of central ministry enterprises to the local; (2) planning management reformed to be bottom-up balancing; (3) more autonomy for the local to choose investment projects; (4) more decision power for the local to allocate resources; (5) decentralization of the financial and credit systems. The rapid delegation of power to the local (Zhou, 1984; Lin et al., 2006), together with collectivization (Lin, 1990; Li and Yang, 2005), provides "the institutional basis for the Great Leap Forward" (Wu and Reynolds, 1988). It is noted that even though this wave of decentralization involved substantial delegation of decision-making power, promotion of the local bureaucrats was tightly controlled by the central and, more importantly, was determined by political consideration rather than the economic performance. Local bureaucrats were rewarded for closely following instructions from the central government (Kung and Chen, 2011).

Following an extended period of political and economic turmoil,[11] the second wave of decentralization came as a major ingredient of the famous reform in 1978. The reform has been regarded as the most important factor in the recent growth of China (Xu, 2011). To incentivize local bureaucrats and to foster inter-regional competition, this reform emphasized the great importance of economic betterment, which stands in sharp contrast with the earlier reforms which stigmatized pursuit of economic goals (Li and Zhou, 2005).

Fig. 1 plots China's GDP growth rate and its output volatility over the past six decades. Evidently in the figure, economic growth tanked dramatically following the first wave of decentralization, while the economy enjoyed much higher growth in the post 1978 reform era. Less attention has been paid to the output volatility, but the same, contrasting dynamics followed two waves of decentralization:[12] volatility skyrocketed in the late 1950s and it steadily went down after 1978. Besides contrasting experience between the two reforms, there is also substantial heterogeneity across regions. Fig. 2 plots the evolution of GDP growth rate and output volatility using the data from Henan and Zhejiang, two provinces in central and southern China respectively. Despite similar trends, the growth rate declined much more significantly in Henan during the first decentralization reform.

[9] In particular, Wang and Yang (2016) provide systematic empirical evidence concerning the relationship between decentralization and volatility in China. Since they focus mainly on the second wave of decentralization in China, they find an unambiguously negative impact of decentralization on output volatility. Following their approach, we enrich their findings by examining both waves of decentralization, thus suggesting a more nuanced view of decentralization under an authoritarian regime.

[10] In his famous speech on the relationship between the central and local governments, Mao Zedong pointed out, "the local should be empowered. This helps us build a strong socialist country. It seems not a good idea to squeeze the power from the local" ("On Ten Major Relationships", April 25, 1956).

[11] Due to the disappointing outcome of the first wave of decentralization, there were a sequence of re-centralization and decentralization reforms, albeit at smaller scales, during 1960s and 70s. For more detailed discussions, see Lin et al. (2006).

[12] The output volatility is calculated in a very simple manner, but as shown in Wang and Yang (2016), the same pattern persists if we detrend the GDP series using HP filter and control for conventional economic factors that determine volatility such as financial development, openness, inventory management, and monetary policy.

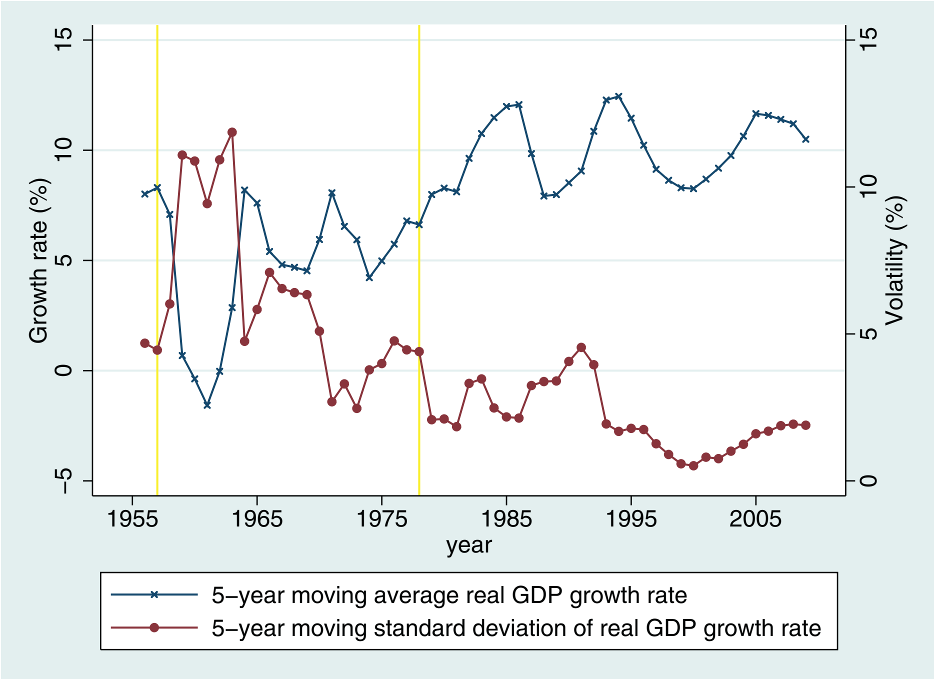


Fig. 1. China's Economic Growth and Volatility.
Notes: (1) The data source is the FRED economic data; (2) Volatility is measured as the standard deviation of the growth rate of real GDP for a five-year moving window; (3) Two vertical yellow lines mark the start year of decentralization reforms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Why did the two decentralization reforms in China produce completely the opposite outcomes? An immediate answer is the regime change: the former is under the central planning economy, while the later took place with the establishment of a new market economy. However, this explanation is silent about why the first decentralization reform had a much larger negative impact on the Henan economy as suggested by Fig. 2. More to the point, it does not clarify the nature of the difference that underlies the two waves of decentralization, which we believe is tied with the incentive of local bureaucrats. To motivate our formal analysis, we present further anecdotal and empirical evidence on how regional or time variation in local officials' career incentives is associated with the outcomes of decentralization reforms.

2.2. Local officials' career incentives: historical anecdotes

The particularly disastrous outcome in Henan during the Great Leap Forward as shown in Fig. 2 is closely tied with the provincial leadership at that time, the Party Secretary Wu Zhipu in particular (Yang, 2009). Since the first wave of the Anti-Rightist Campaign in 1957, a nationwide political purge launched by Mao Zedong, Wu Zhipu enthusiastically embraced, organized, and led the campaign in Henan. While gaining power through the Anti-Rightist Campaign, Wu initiated a series of radical policies in the agricultural sector since the very beginning of the Great Leap Forward. According to Yang (2009), massive hydraulic projects, close planting ("Mi Zhi"), and People's Communes all started in Henan and later extended to the whole country. *The People's Daily*, the Party's official newspaper, shared regularly the success stories and experience from Henan, further boosting the stardom of the province and Wu's leadership during the campaign. Wu Zhipu's single-minded adherence to what was signaled from the very top soon received Mao Zedong's attention. In an official meeting in March 1958, Mao highly praised Wu Zhipu's work by citing inflated numbers on hydraulic construction in Henan. Mao's endorsement incentivized further radicalization of policy-making in Henan. From the summer of 1958, fake reports and misleading statistics became commonplace, and shortly afterwards, Henan was among the provinces hardest hit by the most tragic famine in human history.

In contrast to Henan is Zhejiang's experience. Like many southern provinces, shortly before the establishment of the PRC, Zhejiang started seeing the arrival of the so-called southbound cadres, the party cadres sent by the new central government to take over and consolidate the political power. Soon a power struggle arose between the southbound cadre group and the local guerrilla cadre group.[13] Being backed by the central, the southbound cadre group captured major positions in the new Zhejiang government while the local guerrilla cadre group were largely marginalized during the power transition. However,

[13] For an excellent and insightful account of this historical episode in Zhejiang, see the book by Zhang and Liu (2016).

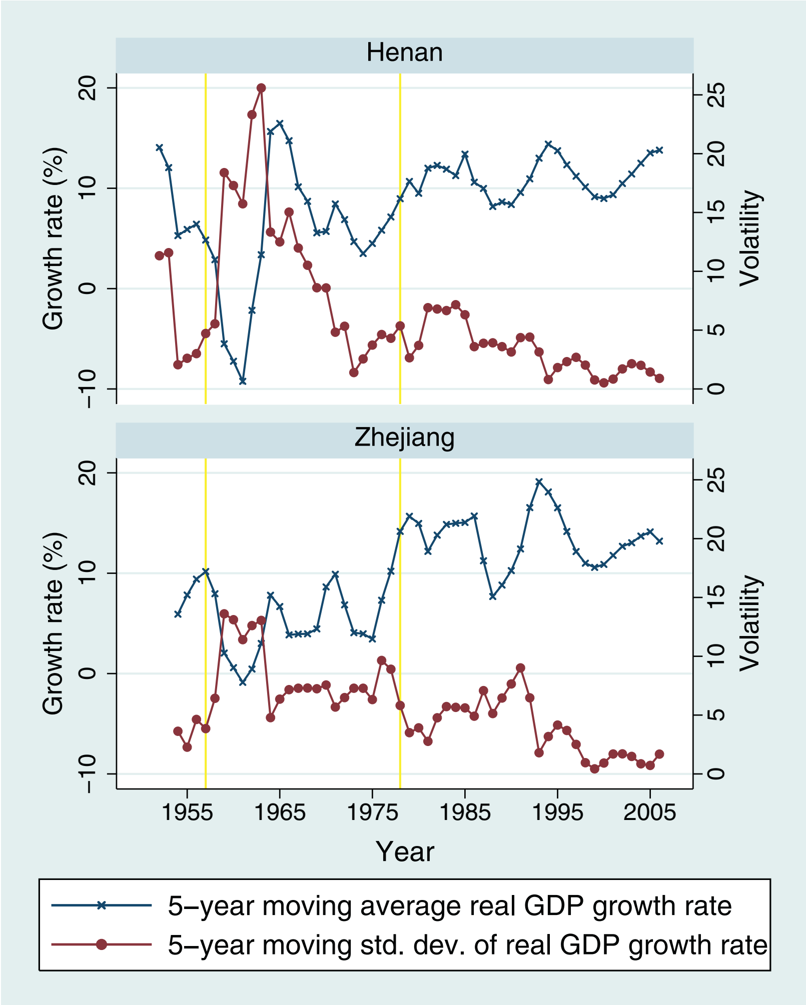


Fig. 2. Economic Growth and Volatility at the Provincial Level.
Notes: (1) The data source is China Compendium of Statistics; (2) Volatility is measured as the standard deviation of the growth rate of real GDP for a five-year moving window; (3) Two vertical yellow lines mark the start year of decentralization reforms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

because the local guerrilla cadres had been working in the region for much longer and thus understood better the local social economy, the new government mainly led by southbound cadres still had to rely on their political competitors in the actual implementation of policies. With little political support from the top, the political survival of local guerrilla cadres hinged on the support from the grassroots. Consequently, the local guerrilla cadre group had much stronger incentives to act in the interests of the masses especially when they sensed the great damage a campaign like the Great Leap Forward may bring about to the local economy (Zhang and Liu, 2016). By deviating from the policy guidance from the central, the local guerrilla cadre group helped mitigate the consequences of the nation-wide policy mistake. Going beyond the Great Leap Forward episode, Zhang et al. (2013) further provide convincing evidence of how this unique power-sharing structure in Zhejiang in the wake of the establishment of the PRC causally impacted local private sector development. The legacy had its influences even during the turbulent period of the Cultural Revolution and manifested itself in the form of a province known for its booming private sector during the post 1978 reform era.

The two contrasting historical anecdotes underscore the vital role played by local officials' career incentives in shaping the regional variation in decentralization reform outcomes. In the next subsection, we provide further empirical evidence to illustrate the structural differences in career incentives between the two decentralization reforms.

2.3. Local officials' career incentives: empirical evidence

Using the historical data, the existing work suggests that the loyalty concern was one of the culprits of the tragedy during the Great Leap Forward (Kung and Chen, 2011). On the other hand, it has been highlighted that economic performance has

Table 1
Correlates with Promotion: 1950–1966 versus 1978–2008.

	Dependent Variable: Promotion from Provincial Party Leadership					
	Maoist Era: 1950–1966			Reform Era: 1978–2008		
	(1)	(2)	(3)	(4)	(5)	(6)
Connection	0.254*		0.290*	−0.024		−0.052
	(0.141)		(0.148)	(0.091)		(0.089)
GDP Growth		0.026	0.058		0.099***	0.102***
		(0.049)	(0.048)		(0.029)	(0.030)
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	60	60	60	188	188	188
R-squared	0.616	0.576	0.627	0.298	0.337	0.339

Notes: (1) The table reports the correlates of promotion for the provincial Party secretaries. The outcome variable is whether a provincial Party secretary was promoted after the tenure. The two independent variables are whether the person had connections to the central leader and the standardized annual provincial GDP growth throughout the tenure. Province fixed effects are controlled. (2) The regression is run separately for two periods: the Maoist era from 1950 to 1966 and the reform era from 1978 to 2008. (3) The data on provincial Party secretaries is compiled from public sources including China Vitae, Xinhuanet, and Wikipedia. GDP data is from the China Compendium of Statistics. (4) Robust standard errors are reported in parentheses. (5) *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

become an important promotion criterion for local bureaucrats in the post 1978 reform era (Li and Zhou, 2005), while how important the loyalty concern remains to be is subject to debate (Jia et al., 2015; Fisman et al., 2020). In this subsection, we demonstrate that *relative* to Maoist China, the loyalty concern plays a less important role in promotion of local bureaucrats in post-reform China.

We follow Jia et al. (2015) to use local officials' connection with the central politicians as the measurement of loyalty, defined by whether they worked at the same time in the same branch of the Party, the government, or the army. We focus on the connection between each provincial Party secretary and the highest central leaders (Mao Zedong, Deng Xiaoping, Jiang Zemin, and Hu Jintao in respective periods during their terms). The cross-sectional regression is specified as follows

$$Promotion_i = \alpha + \beta_1 Connection_i + \beta_2 GDP\_Growth_{ip} + \gamma_p + \epsilon_i, \tag{1}$$

where $Promotion_i$ is an indicator that equals one if provincial Party secretary i got promoted after his/her term as the provincial Party secretary; $Connection_i$ is an indicator that equals one if provincial Party secretary i had a connection with the central leaders as defined above; GDP_Growth_{ip} is the annual GDP growth in province p (to make the interpretation of coefficient easier, standardized by the standard deviation across provinces) in which i served as the Party secretary; γ_p is the province fixed effect. We run the regression separately in two periods, the Maoist era before the Cultural Revolution (1949–1966) and the post-reform era (1978–2008).

The regression results are presented in Table 1. Columns (1)–(3) report the correlates with promotion during the Maoist era. Political connection with the Party leadership enters the regression equation positively and significantly while the role of GDP growth is insignificant. According to the point estimate in Column (3), being connected with the Party leader is associated with an increase in promotion probability by 29.0%. Columns (4)–(6) report the correlates with promotion during the reform era. In contrast to Columns (1)–(3), GDP growth is now estimated significantly positive while political connection loses its significance. According to the point estimate in Column (6), a one standard-deviation increase in GDP growth is associated with a 10.2% increase in promotion probability, compared with the unconditional promotion probability being around 25%. Though this simple regression specification does not aim at establishing causality, the results above are suggestive that given the prominent role of political connections in promotion, local bureaucrats had stronger incentives to signal their loyalty to their leadership during the Maoist era, while the career incentives were relatively more tied with economic growth during the reform era. This structural difference in the incentive structure for the local bureaucrats will be captured by our model as the key explanation of the contrasting outcomes of the two decentralization reforms.

2.4. Historical reforms in other countries

Finally, it is worth emphasizing that even though our discussion revolves around the reform history of China, reforms featuring decentralization of decision making have been taking place across countries under the authoritarian regime since the late 1980s. Among Asian countries, Viet Nam launched its large-scale reform ("Doi Moi") in 1986 which shared many similar characteristics with the China's 1978 reform (World Bank, 1993; St John, 1997). In the same year, Laos initiated a structural reform program called "New Economic Mechanism" ("Chintanakhan Mai") (Stuart-Fox, 2005). A few years later, Cambodia entered a decade-long process of decentralization reform which made its breakthrough in early 2000s (Un and Ledgerwood, 2003; World Bank, 2015). In 2001, known as one of the most radical decentralization reforms, Indonesia started its big bang reform that packages together economic, political, and administrative decentralizations (Kassum et al., 2003;

Yap, 2018). Unlike Viet Nam, Laos, and Cambodia, Indonesia's decentralization reform is accompanied with a prolonged phase of democratization after which the country is no longer under the authoritarian regime. South Korea, to some extent, follows a similar path despite having a more dramatic democratization process and a more gradual process of decentralization. All these reforms share the common ingredient of shifting the economic decision from the central to the local governments. Like the 1978 reform in China, the impact of those reforms is generally positive, albeit less conclusive.[14]

3. The baseline model

We model policy making under uncertainty. There are two players, a central government and a local government. They want to implement an economic policy that hinges on the true state of the economy subject to uncertainty. There are two channels through which the governments can reduce the uncertainty. Each government can directly acquire information of the true state of the economy. It can also acquire information from the other government through the inter-governmental communication which is nevertheless subject to communication friction. In the baseline setting, the communication friction is endogenously determined. We will present a version of the model with exogenous communication friction in the section on extensions, highlighting under what condition the endogenous communication channel is essential to our main results. Each government has a fixed amount of resource, which can be allocated between the two activities: direct information acquisition and indirect information acquisition by reducing the friction in inter-governmental communication.

We consider two economic regimes. Under the centralized regime, the communication is bottom-up. The local government directly acquires information and then sends a noisy signal to the central government. Facing the trade-off between the two information acquisition channels, the central government decides how to allocate its attention resource and chooses the economic policy accordingly. Under the decentralized regime, the communication is top-down. The central government directly acquires information and sends a noisy signal to the local government. The local government allocates its attention resource and then implements its desired economic policy. Fig. 3 illustrates respectively the timeline of the model under the two regimes.

We now proceed to formally specify the information structure, economic regimes, and the decision problem of the governments under each regime.

3.1. The information structure

Denote the true state of the *local* economy by θ. Both the local and central governments hold the same prior about θ, which follows a normal distribution with mean zero and variance σ^2, denoted by $\mathcal{N}(0, \sigma^2)$. Due to information imperfections, governments cannot observe θ perfectly. Instead, they observe θ with a white noise:

$$\theta_c = \theta + z_c, \quad z_c \sim \mathcal{N}(0, \sigma_c^2),$$
$$\theta_\ell = \theta + z_\ell, \quad z_\ell \sim \mathcal{N}(0, \sigma_\ell^2),$$

where θ_c and θ_ℓ are the noisy signals for the central and local governments.[15] The governments can choose to reduce σ_c^2 and σ_ℓ^2 by directly acquiring information of the state of the economy, so both σ_c^2 and σ_ℓ^2 will be endogenously determined. Alternatively, a government can acquire information from the other government through inter-governmental communication subject to friction. This will be specified under two different economic regimes.

3.2. Signaling under the two economic regimes

Under the centralized regime, the local government sends a signal s_ℓ to the central government. Communication is subject to friction. The central government receives a (composite) signal $\mathbf{s}'_\ell$ which has two components, $\mathbf{s}'_\ell = \{s'_{\ell 0}, s'_{\ell 1}\}$. The first component is the signal sent by the local government subject to exogenous communication friction, $s'_{\ell 0} = s_\ell + \epsilon$ with $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$. The second component, $s'_{\ell 1} = s_\ell + \epsilon_c$ with $\epsilon_c \sim \mathcal{N}(0, \tilde{\sigma}_{\epsilon c}^2)$ is the additional information the central government acquires to mitigate the communication friction. We explicitly introduce these two components to clarify the covariance structure between $s'_{\ell 0}$ and $\mathbf{s}'_\ell$, which facilitates the formulation of the information flow constraints to be presented in the next subsection. For the decision problem of the central government, the two components of the signal boil down to a sufficient statistic $s'_\ell = \frac{s'_{\ell 0}/\sigma_\epsilon^2 + s'_{\ell 1}/\tilde{\sigma}_{\epsilon c}^2}{1/\sigma_\epsilon^2 + 1/\tilde{\sigma}_{\epsilon c}^2}$ which follows $\mathcal{N}(s_\ell, \sigma_{\epsilon c}^2)$ with $\sigma_{\epsilon c}^2 = (\sigma_\epsilon^{-2} + \tilde{\sigma}_{\epsilon c}^{-2})^{-1}$. The central government chooses the precision of the composite signal $\sigma_{\epsilon c}^2$ which is bounded above by σ_ϵ^2. Based on the realized information set $\{\theta_c, \mathbf{s}'_\ell\}$, the central government chooses its preferred policy a_c.

Under the decentralized regime, the central government sends a signal s_c. Symmetrically, the local government receives a (composite) signal $\mathbf{s}'_c$ which has two components, $\mathbf{s}'_c = \{s'_{c0}, s'_{c1}\}$. The first component is the raw signal received, $s'_{c0} = s_c + \epsilon$ with $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ and the second component, $s'_{c1} = s_c + \epsilon_\ell$ with $\epsilon_\ell \sim \mathcal{N}(0, \tilde{\sigma}_{\epsilon \ell}^2)$ captures the additional information the local government acquires through the communication channel. For the decision problem of the local government, the two

[14] We plot how output growth and volatility evolve over the reform period for these four Asian countries in the appendix. See Figs. 8 and 9.

[15] Throughout the paper, we will use subscript "c" for variables associated with the central government and subscript "ℓ" for variables associated with the local government.

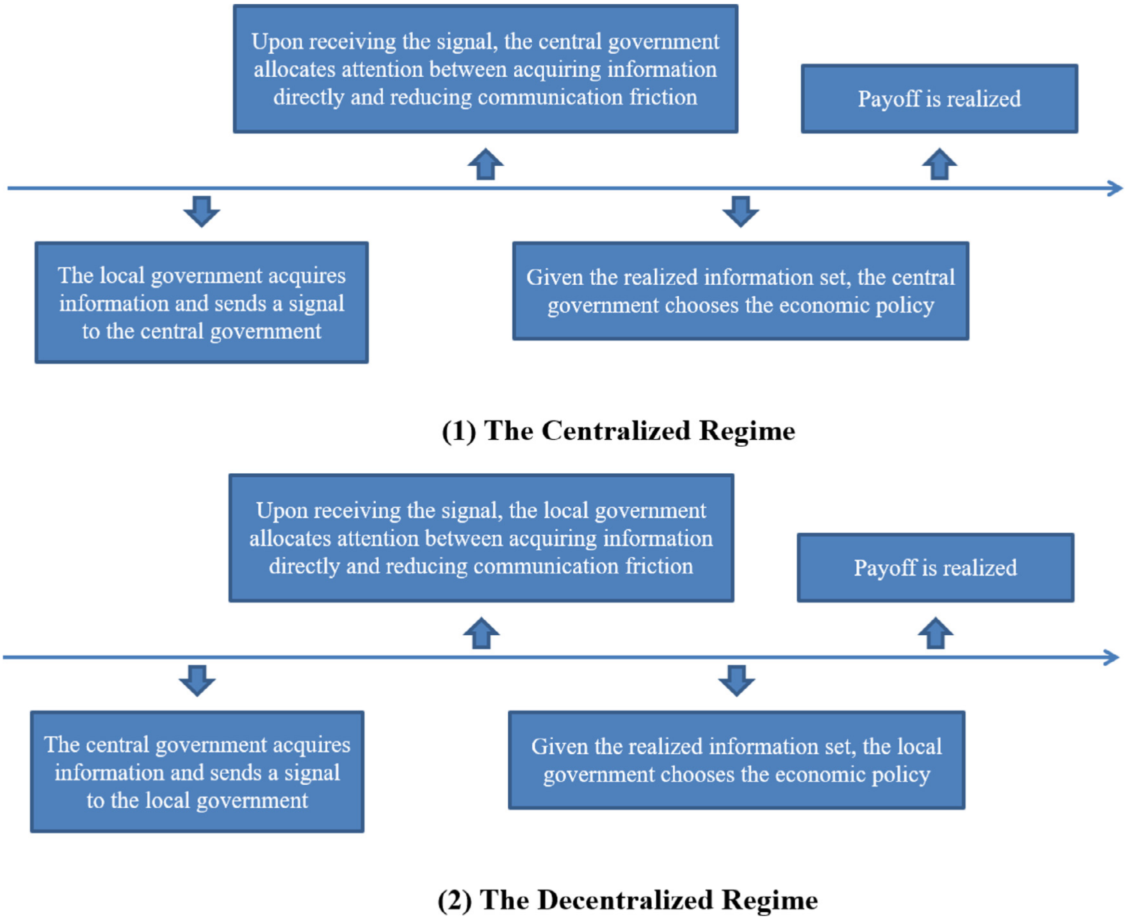


Fig. 3. Timeline.

components boil down to a sufficient statistic $s'_c = \frac{s'_{c0}/\sigma_\epsilon^2 + s'_{c1}/\tilde{\sigma}_{\epsilon\ell}^2}{1/\sigma_\epsilon^2 + 1/\tilde{\sigma}_{\epsilon\ell}^2}$ which follows $\mathcal{N}(s_c, \sigma_{\epsilon\ell}^2)$ with $\sigma_{\epsilon\ell}^2 = (\sigma_\epsilon^{-2} + \tilde{\sigma}_{\epsilon\ell}^{-2})^{-1}$. The local government chooses the precision of the composite signal $\sigma_{\epsilon\ell}^2$ which is bounded above by σ_ϵ^2. Based on the realized information set $\{\theta_\ell, \mathbf{s}'_c\}$, the local government makes its policy decision a_ℓ.

The friction in information transmission is pervasive in any large organization, but it could be particularly severe in the context of an authoritarian government. For the top-down communication, the friction comes from the lack of transparency of discussions and debates at the very top and the tendency of over-simplifying real economic issues in policy documents,[16] not to mention the complication of coupling policy prescription with political propaganda.[17] For the bottom-up communication, it is also essential for the central to read between the lines to better understand the reports from the local. The random or intentional noise accumulates over the long process of reporting from the very bottom of the regime.

From now on, we assume that $s_\ell = \theta_\ell$ and $s_c = \theta_c$. In one of the extensions, we will allow the signal sender to strategically introduce noise into the inter-governmental communication.[18] We assume all the white noises z_c, z_ℓ, ϵ, ϵ_ℓ, and ϵ_c are independent.

[16] In an authoritarian regime, the official policy documents are usually the product of the input from the technocrats, fights and compromises among the few decision makers, and politico-economic needs. Wu (1995) presents an excellent study of the so-called "Documentary Politics" in China. His case studies detail the whole political process of drafting and disseminating official documents, explaining why even specific wording or quotation could have deep connotations.

[17] For example, in the May of 1958, the second meeting of the Eighth National Congress of the Party approved that the "overall strategy" is to "achieve greater, faster, better, and more economical results in building socialism". Due to the political climate in late 1950s, tremendous emphasis was put on quantity and speed with quality and efficiency being effectively unnoticed while communicating this overall strategy to the lower level governments, which contributes to the disastrous Great Leap Forward.

[18] To be sure, misreporting and manipulation are quite common under the authoritarian regime. For the striking example of over-reporting, see the announcements of agricultural output during the Great Leap Forward period in China. In contrast, for fear of the ratchet effect, managers in Soviet Union had great incentive to under-report (Weitzman, 1976). However, since our main focus is on the tradeoff between different channels of information acquisition in relation to the resulting policy choice, we abstract from misreporting in this model.

3.3. The information flow constraint

The decision problem for the government that decides the economic policy, that is, the signal receiver, has two layers. It has to first decide the resource allocation over two channels of information acquisition and then choose the optimal policy based on the information gathered.

We first formalize the resource allocation problem. We assume that each government can only acquire a fixed amount of information following the framework of rational inattention (Sims, 2003; Mackowiak and Wiederholt, 2009). To formalize the notion of information, we define the differential entropy as in the standard information theory, which is a measure of the uncertainty of a continuous random variable.[19]

Definition 1. The differential entropy $H(X)$ of a continuous random variable X with a probability density function $f(x)$ is defined as

$$H(X) = E[-\log_2 f(x)] = -\int f(x)\log_2 f(x)dx.$$

If X follows a multivariate normal distribution with a covariance matrix Σ, it can be shown that the entropy of X is given by

$$H(X) = \frac{n}{2}\log_2(2\pi e) + \frac{1}{2}\log_2|\Sigma|,$$

where n is the dimension of the random variable and $|\Sigma|$ is the determinant of Σ.

Definition 2. The conditional differential entropy $H(X|Y)$ of two continuous random variables X and Y with a joint probability density function $f(x, y)$ is defined as

$$H(X|Y) = -\int f(x,y)\log_2 f(x|y)dxdy.$$

In general, we have

$$H(X|Y) = H(X,Y) - H(Y).$$

Hence, if one is interested in X, the informativeness of an observation Y can be captured by the difference between $H(X) - H(X|Y)$. In other words, the difference between $H(X)$ and $H(X|Y)$ is the reduction of uncertainty with respect to X when Y is observed. In the framework of rational inattention, we assume that each economic agent has limited attention resource, so its information flow constraint is generally given by $H(X) - H(X|Y) \le \kappa$. We now specialize this constraint to our setting.

Under the centralized regime, the information flow constraint for the signal sender, the local government, is given by

$$H(\theta) - H(\theta|\theta_\ell) \le \kappa_\ell \quad \Leftrightarrow \quad \frac{1}{2}\log_2\left(\frac{\mathrm{Var}(\theta)}{\mathrm{Var}(\theta|\theta_\ell)}\right) = \frac{1}{2}\log_2\left(\frac{\sigma^2+\sigma_\ell^2}{\sigma_\ell^2}\right) \le \kappa_\ell, \tag{2}$$

where $\kappa_\ell > 0$ is the capacity of information acquisition of the local government.

For the central government, the reduction of entropy comes from two sources: improved information about both θ and ϵ. The constraint on the entropy reduction is then given by

$$H(\theta,\epsilon|s'_{\ell 0}) - H(\theta,\epsilon|\theta_c,\mathbf{s}'_\ell) \le \kappa_c,$$

where $\kappa_c > 0$ is the capacity of information acquisition of the central government.

Notice that even though θ and ϵ are unconditionally independent, we cannot write the constraint in an additively separable form for θ and ϵ as they might not be independent conditional on the acquired information $(\theta_c, \mathbf{s}'_\ell)$.[20] The following lemma provides a closed form solution to this information flow constraint.[21]

Lemma 1. *The information flow constraint of the central government under the centralized regime is given by*

$$\left(\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_{\epsilon c}^2}\right)\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_c^2}\right) \le \frac{2^{2\kappa_c}(\sigma^2+\sigma_\ell^2+\sigma_\epsilon^2)}{\sigma^2\sigma_\ell^2\sigma_\epsilon^2} + \frac{1}{\sigma_\ell^4} \equiv K_c(\sigma_\ell^2). \tag{3}$$

Eq. 3 may appear a bit complicated, but the choice variables, σ_c^2 and $\sigma_{\epsilon c}^2$ for the central government are multiplicatively separable, which is very important for a sharp characterization of the attention allocation problem. Moreover, we have $K_\ell \ge \left(1/\sigma_\epsilon^2 + 1/\sigma_\ell^2\right)\left(1/\sigma^2 + 1/\sigma_\ell^2\right)$ with the equality if and only if $\kappa_c = 0$, i.e., if the central government has zero information capacity, then it is impossible to acquire any information ($\sigma_{\epsilon c}^2 = \sigma_\epsilon^2$ and $\sigma_c^2 = \infty$). It should be noted that σ_ℓ^2, the

[19] See, for example, chapter 8 in Cover and Thomas (2012) for a standard treatment.

[20] This stands in sharp contrast with the earlier macroeconomic applications of rational inattention such as Mackowiak and Wiederholt (2009). Conditional correlation substantially complicates the analytics of the model.

[21] All the proofs are relegated to the appendix.

signal precision of the local government, enters the above constraint. In what follows, we sometimes write $K_c(\sigma_\ell^2)$ as K_c for simplicity when it would not cause any confusion.

Symmetrically, under the decentralized regime, the information flow constraint for the signal sender, the central government, is given by

$$H(\theta) - H(\theta|\theta_c) \leq \kappa_c \quad \Leftrightarrow \quad \frac{1}{2}\log_2\left(\frac{\mathrm{Var}(\theta)}{\mathrm{Var}(\theta|\theta_c)}\right) = \frac{1}{2}\log_2\left(\frac{\sigma^2+\sigma_c^2}{\sigma_c^2}\right) \leq \kappa_c. \tag{4}$$

For the local government, the constraint on the entropy reduction is given by

$$H(\theta, \epsilon|s'_{c0}) - H(\theta, \epsilon|\theta_\ell, \mathbf{s}'_c) \leq \kappa_\ell.$$

Following the proof of Lemma 1, we can rewrite the information flow constraint for the local government in a multiplicatively separable form of its two choice variables σ_ℓ^2 and $\sigma_{\epsilon\ell}^2$.

Lemma 2. *The information flow constraint of the local government under the decentralized regime is given by*

$$\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma_{\epsilon\ell}^2}\right)\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2}\right) \leq \frac{2^{2\kappa_\ell}(\sigma^2+\sigma_c^2+\sigma_\epsilon^2)}{\sigma^2\sigma_c^2\sigma_\epsilon^2}+\frac{1}{\sigma_c^4} \equiv K_\ell(\sigma_c^2). \tag{5}$$

In what follows, we sometimes simply write $K_\ell(\sigma_c^2)$ as K_ℓ, but again it should noticed that K_ℓ depends on the choice variable of the central government under the decentralized regime.[22] We now impose the key assumption for this baseline setting.

Assumption 1. $\kappa_\ell > \kappa_c$.

In words, we assume that the local government has higher information capacity than the central government. This is a reasonable assumption in our setting for two main reasons. First, since θ is interpreted as the state of the *local* economy, according to Hayek (1945), the local government tends to have the intrinsic advantage of obtaining local information. In the context of China, Huang et al. (2017) have substantiated Hayek's insight by demonstrating a tight link between decentralization of state-owned enterprises and the distance to the oversight government. Second, the central government usually has many more preoccupations, some of which may well be beyond economic considerations, to divert its attention resources. Further, in the presence of multiple localities, which our model abstracts from, the local governments are more likely to be better focused than the central, when it comes to specific issue pertaining to its own locality.

3.4. Output level and volatility

We define the output level Y as a quadratic form

$$Y \equiv Y^* - (a_i - \theta)^2, \quad i = c, \ell,$$

where Y^* is the ideal output level if the policy choice a_c or a_ℓ perfectly matches the true state of the economy θ. In this paper, we are particularly interested in the ex ante expected output level $E(Y)$ and its variance $Var(Y)$.

3.5. The decision problem for each government

Under the decentralized regime, the central government, which is the signal sender, is assumed to be benevolent. Its decision problem is given by

$$\max_{\sigma_c^2} E(Y) = Y^* - E(a_\ell - \theta)^2,$$

subject to Constraint 4.

The local government, which receives the signal from the central, has a two-layer decision problem, which is given by

$$\max_{\sigma_\ell^2, \sigma_{\epsilon\ell}^2} E\left\{\max_{a_\ell} (1-\gamma)\left(Y^* - E[(a_\ell-\theta)^2|\theta_\ell, \mathbf{s}'_c]\right) - \gamma E[(a_\ell - \theta_c)^2|\theta_\ell, \mathbf{s}'_c]\right\},$$

subject to Constraint 5 and $\sigma_{\epsilon\ell}^2 \leq \sigma_\epsilon^2$. Suggested by the objective function, beyond economic output, the local government is also concerned with whether its economic policy follows closely enough the signal sent by the central. The local government has the incentive to stick to the policy prescription of the central, signaling its political loyalty. This additional career concern

[22] As we will show in Propositions 1 and 6, given the same attention budget, the central government can get a more precise signal θ_c under the centralized regime (setting $\sigma_{\epsilon c}^2 = \sigma_\epsilon^2$) than the decentralized regime, while the local government can get a more precise signal θ_ℓ under the decentralized regime (setting $\sigma_{\epsilon\ell}^2 = \sigma_\epsilon^2$) than the centralized regime. In some sense, being a signal receiver softens the information flow constraint. Our main results are not driven by this *de facto* difference in capacity across regimes. This will be clearer when we discuss a variant of the model with σ_ℓ^2 and σ_c^2 being exogenously given.

seems to be a common characteristics in many authoritarian regimes,[23] and it turns out to be the key driving force of our main results.

The parameter $\gamma \in [0, 1]$ captures the importance of signaling political loyalty *relative* to economic consideration in policy making. From now on we simply interpret γ as loyalty concern but our comparative static results should be understood as a change of either political or economic incentives. When $\gamma = 0$, the objective of the local government is perfectly aligned with that of the central government. When $\gamma = 1$, the local government attaches no importance to economic output and only attempts to induce the central government to adopt its policy recommendation. During the reform in the late 1950s, under a more coercive system with greater emphasis on political loyalty, the local government faces a high γ. In the 1978 reform, the change of political climate, as well as more focus placed on economic development, yields a lower γ. It should be noted that γ reflects the institutional constraints faced by the central government and thus is not a choice variable at least in the short run and as far as economic policy-making is concerned.

An important asymmetry arises in the objective function between the local and central government. For simplicity, the central government is assumed to be benevolent. This assumption presumes that the central government can separate political consideration from economic motivation while the local government cannot. Alternatively, we could assume that the central government indeed cares about conformity of the local as well, but as long as its loyalty concern is not as strong as the local's in the domain of economic policy decisions, the intuition of our main results would carry through.

Under the centralized regime, the central government attempts to maximize solely the expected economic output, so its decision problem, consisting of two layers, is given by

$$\max_{\sigma_c^2, \sigma_{\epsilon c}^2} E\Big[\max_{a_c} E(Y|\theta_c, \mathbf{s}'_\ell)\Big] = \max_{\sigma_c^2, \sigma_{\epsilon c}^2} E\Big[\max_{a_c} Y^* - E((a_c - \theta)^2|\theta_c, \mathbf{s}'_\ell)\Big],$$

subject to Constraint 3 and $\sigma_{\epsilon c}^2 \le \sigma_\epsilon^2$.

The local government, which sends the signal to the central government under this regime, is faced with the following decision problem

$$\max_{\sigma_\ell^2} (1-\gamma)\big(Y^* - E(a_c - \theta)^2\big) - \gamma E(\theta_\ell - a_c)^2,$$

subject to Constraint 2. Despite having a similar form, the second term of the payoff function entails a different interpretation. It captures how much the local government cares about whether the central government follows its policy suggestion. This is a reduced-form way to incorporate additional promotion incentive beyond the merit-based rules.[24]

As a technical note, we can dispense with the information flow constraints by directly incorporating a cost item of information acquisition into the objective function as in (Wu, 2019) and Matějka and Tabellini (2020). This alternative modeling strategy would transform a constrained optimization problem into an unconstrained optimization while delivering similar substantive insights. However, because we have to deal with the change in information content of two *conditionally correlated* sources of uncertainty (θ and ϵ), either way of modeling would lead to a technically non-trivial decision problem for the signal receiver.

The equilibrium of this model under each regime is characterized by solving the constrained optimization problem of the signal receiver and sender sequentially. The formal definition of the equilibrium can be found in the appendix. We now turn to the main results of the baseline model.

4. Main results

Under each regime, the government that receives the signal solves its decision problem backwards. The resource allocation of the two channels of information acquisition hinges on the determination of optimal economic policy. For each regime, we first characterize the optimal economic policy for any given resulting information set. We then solve backwards the optimal allocation of the attention resource. The last step is to characterize the optimal decision of the signal sender. After we solve the equilibrium under each regime, we turn to the comparison of economic output and volatility between two regimes.

4.1. Optimal economic policy

Because of the quadratic objective function, the optimal policy of the final policy maker is a linear combination of the signals it receives. Under the centralized regime, the benevolent central government always targets its policy to the expected

[23] Besides the Great Leap Forward in China, Nikita Khrushchev's Corn Campaign is another infamous example. Seeing the increase of corn production as an important part of his agricultural reform, Khrushchev initiated a large scale expansion program of corn production in mid 1950s. With his strong backing, the area of corn cultivation grew exponentially, which was later proven to be quite unproductive and inefficient. However, according to a detailed case study by Hale-Dorrell (2014), even subordinates had recognized the absurdity of the corn campaign much earlier, deception, submission, and fine-tuning were widespread.

[24] Under an authoritarian regime, rather than the actual performance, the promotion of lower-level officers sometimes hinges on whether their policy recommendations are favored and adopted by their superordinates. One of the most dramatic cases is "learning-from-Dazhai-in-Agriculture" movement during the pre-reform era (Meisner, 1978). Thanks to the nation-wide promotion of his model agricultural production, Yonggui Chen, a community-level party secretary rose to become the vice premier of China in less than twenty years.

state of the economy conditional on all the signals it gathers. For the closed-form solution to the central government's policy, see Lemma 12 in the appendix. We can write the distance between the policy and the true state of the economy as

$$E(a_c - \theta)^2 = Var(\theta|\theta_c, \mathbf{s}'_\ell) = \left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2} + \frac{1}{\sigma_\ell^2 + \sigma_{\epsilon c}^2}\right)^{-1}. \tag{6}$$

which suggests that direct information acquisition has a first-order impact on reducing the expected deviation through σ_c^2, while the contribution of inter-governmental communication through $\sigma_{\epsilon c}^2$ is constrained by the quality of the signal sent by the local, σ_ℓ^2.

Under the decentralized regime, the local government's policy choice, formally stated as Lemma 13 in the appendix, is a weighted average of the optimal policy chosen by the local government if it were benevolent (the case of $\gamma = 0$) and the local government's conditional expectation of the signal from the central (the case of $\gamma = 1$). The loyalty concern of the local government introduces distortion directly through the policy-making margin. In the absence of loyalty concern ($\gamma = 0$), the distance between the chosen policy and the true state of the economy is symmetric to what we have obtained under the centralized regime:

$$E(a_\ell - \theta)^2\Big|_{\gamma=0} = Var(\theta|\theta_\ell, \mathbf{s}'_c) = \left(\frac{1}{\sigma^2} + \frac{1}{\sigma_\ell^2} + \frac{1}{\sigma_c^2 + \sigma_{\epsilon\ell}^2}\right)^{-1} \tag{7}$$

On the other hand, when the local government is entirely loyalty driven ($\gamma = 1$), the distance between the policy and its best guess of the central government's signal is of the similar form of harmonic mean:

$$E(a_\ell - \theta_c)^2\Big|_{\gamma=1} = Var(\theta_c|\theta_\ell, \mathbf{s}'_c) = \left(\frac{1}{\sigma_{\epsilon\ell}^2} + \frac{1}{\sigma_c^2 + (1/\sigma^2 + 1/\sigma_\ell^2)^{-1}}\right)^{-1}, \tag{8}$$

which implies that, to minimize the distance between a_ℓ and θ_c, inter-governmental communication tends to be more effective than direct information acquisition.

Since the optimal policy chosen by the signal receiver is always a linear combination of all the signals it obtains, we obtain a tight relationship between the expected output and output volatility.

Lemma 3. *Let a_i be a linear combination of all the signals: θ_ℓ, θ_c, s'_ℓ, and s'_c ($i = c, \ell$). The expected output is given by*

$$E(Y) \equiv Y^* - E(a_i - \theta)^2 \quad \text{and} \quad Var(Y) = 2(Y^* - E(Y))^2.$$

The lemma shows that the output level and volatility move in the opposite direction, and as a result, the comparative statics concerning the output level can easily be re-interpreted in terms of volatility. From now on, we will mainly work with $E(Y)$, or more directly, the deviation of policy from the true state of the economy, $E(a_i - \theta)^2$.

4.2. The equilibrium under the centralized regime

Under the centralized regime, there is a sharp characterization of the equilibrium. First, the central government spends all of its attention budget on direct information acquisition.[25] Since the central government cares only about economic output and direct information acquisition has a first-order impact on economic output, the central has no incentive to divert its resource to inter-governmental communication. As a result, $\sigma_{\epsilon c}^2 = \sigma_\epsilon^2$. Second, the local government also devotes itself to direct information acquisition.[26] On the one hand, the economic motive (the term $(1-\gamma)EY$ in the objective function) incentivizes the local government to increase the precision of the signal it sends to the central. On the other hand, since the central government attaches more importance to the signal sent by the local government if the quality of the signal is higher, the political motive (the term $-\gamma E(a_c - \theta_\ell)^2$) gives additional incentive for the local government to maximize its effort to acquire information. Since all the attention budget of the local is spent on information acquisition, Constraint 2 is binding, leading to $\sigma_\ell^2 = \sigma^2/(2^{2\kappa_\ell} - 1)$. The following proposition formally states the equilibrium characterization under the centralized regime.

Proposition 1. *Under the centralized regime, we have*

$$E(a_c - \theta)^2 = \left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2} + \frac{1}{\sigma_\ell^2 + \sigma_\epsilon^2}\right)^{-1}$$

with $\sigma_\ell^2 = \sigma^2/(2^{2\kappa_\ell} - 1)$, $\sigma_{\epsilon c}^2 = \sigma_\epsilon^2$, and

$$\sigma_c^2 = \left[K_c\left(\frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma_\ell^2}\right)^{-1} - \frac{1}{\sigma^2} - \frac{1}{\sigma_\ell^2}\right]^{-1} = \frac{\sigma^2(\sigma_\epsilon^2 + \sigma_\ell^2)}{(2^{2\kappa_c} - 1)(\sigma^2 + \sigma_\epsilon^2 + \sigma_\ell^2)}.$$

[25] See Lemma 14 in the appendix.
[26] See Lemma 15 in the appendix.

4.3. The equilibrium under the decentralized regime

The equilibrium characterization under the decentralized regime is more involved, which hinges on the degree of loyalty concern γ. To proceed, we first provide the following characterization of how the local government allocates its attention budget for any given γ.

Lemma 4. *Under the decentralized regime, there exist two cutoffs $\underline{\gamma}$ and $\bar{\gamma}$ such that $0 < \underline{\gamma} < \bar{\gamma} < 1$. If $\gamma \le \underline{\gamma}$, the local government specializes in direct information acquisition ($\sigma^2_{\epsilon\ell} = \sigma^2_{\epsilon}$). If $\gamma \ge \bar{\gamma}$, the local government specializes in intergovernmental communication ($\sigma^2_{\ell} = \infty$). If $\gamma \in (\underline{\gamma}, \bar{\gamma})$, the local government allocates its budget to both activities ($\sigma^2_{\ell} < \infty$ and $\sigma^2_{\epsilon\ell} < \sigma^2_{\epsilon}$) with*

$$\frac{1}{\sigma^2_{\ell}} + \frac{1}{\sigma^2} + \frac{1}{\sigma^2_c} = \frac{K_{\ell}^{1/2}(1-\gamma)}{\gamma}, \tag{9}$$

which implies that $\partial \sigma^2_{\ell} / \partial \gamma > 0$.

According to this lemma,[27] the parameter space of γ can be partitioned into three subsets. The local government focuses exclusively on intergovernmental communication if its loyalty concern is sufficiently strong ($\gamma \ge \bar{\gamma}$). Intuitively, if the main concern of the local bureaucrats is to infer the signal sent by the central government, then they can best achieve this goal by directly reducing the inter-governmental communication friction, at the cost of not acquiring any additional information about the true state of the economy.[28] Second, the local government focuses exclusively on direct information acquisition provided that the economic motive is sufficiently strong ($\gamma \le \underline{\gamma}$). This case is symmetric to the equilibrium under the centralized regime. Last, if the loyalty concern is in the intermediate range, then the attention resource will be allocated to both channels with the effort on intergovernmental communication strictly increasing with the intensity of the loyalty concern.

We illustrate the relationship between $1/\sigma^2_{\ell}$ with γ in Fig. 4. Clearly, the precision of the signal acquired by the local government (weakly) decreases with γ. The loyalty concern could heavily influence the information acquisition margin of the local government. On top of the policy-making margin, this is the second margin that the loyalty concern distorts the economy, which turns out to be indispensable for generating differential economic outcomes under decentralization.

We turn to the strategy of the central government under the decentralized regime. A better signal from the central helps the local target the true state of the economy, but it is also possible that a better signal induces the local to spend more effort on inter-governmental communication, leading to a waste of the attention budget from the standpoint of social welfare. In our framework, the first channel dominates, so the central government always maximizes its effort to direct information acquisition. The actual proof is tedious because the strategy of the local government is not differentiable with respect to γ at the two cutoffs and the two cutoffs $\underline{\gamma}$ and $\bar{\gamma}$ themselves are functions of σ^2_c.

Lemma 5. *Under the decentralized regime, for any γ, the central government spends all of its attention resource on information acquisition (Constraint 4 is binding), which leads to*

$$\sigma^2_c = \sigma^2 / (2^{2\kappa_c} - 1).$$

Given the sharp characterization of the strategy of the central government, we can establish the counterpart of Proposition 1 under the decentralized regime.[29] More importantly, we establish the following monotonicity result that is crucial for the comparison of economic performance between the two regimes.

Proposition 2. *Under the decentralized regime, $E(a_{\ell} - \theta)^2$ strictly increases with γ.*

In words, the stronger the loyalty concern is, the further away the economic policy is from the true state of the local economy.

4.4. Comparison between two regimes

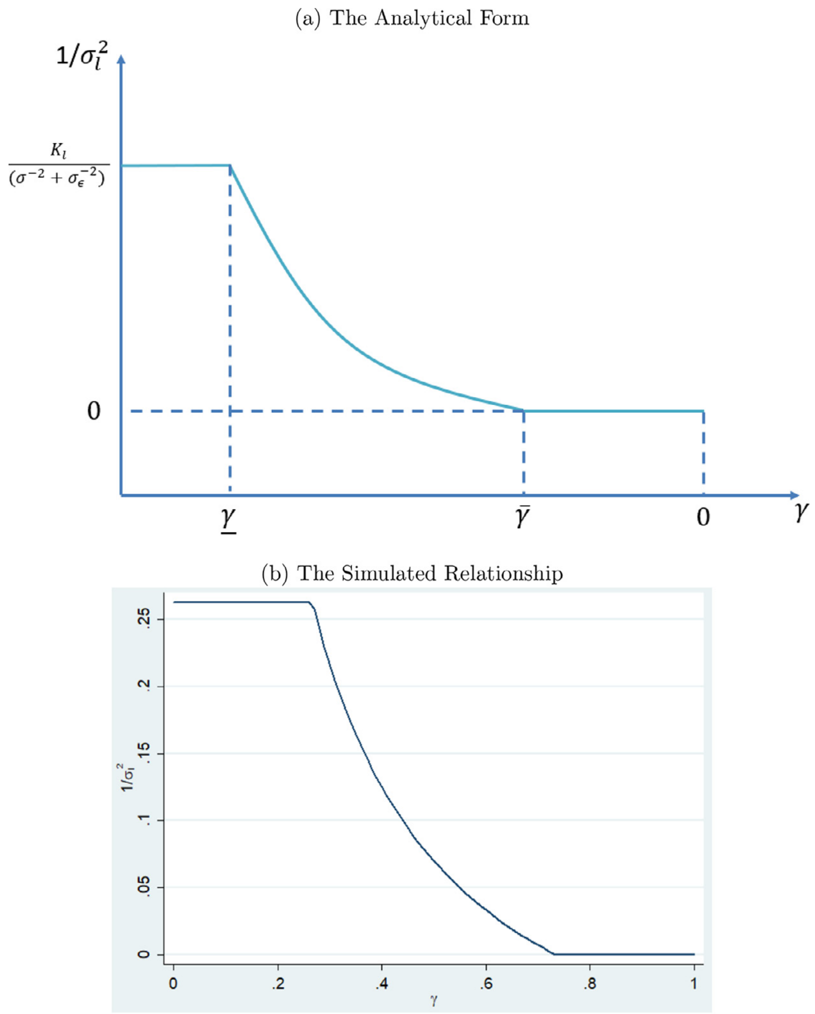
In the absence of loyalty concern, governments under each regime focus exclusively on direct information acquisition. The intergovernmental communication friction has an asymmetric impact on information transmission under the two regimes.

[27] As a technical note, we prove this lemma by first showing that the information flow constraint must be binding. Using the binding constraint, we recast the decision problem as an unconstrained univariate optimization problem. We then solve the optimization problem for γ in different ranges. Moreover, $\underline{\gamma}$ and $\bar{\gamma}$ can be solved in closed form. They are roots in [0,1] to the following two equations respectively:

$$\underline{\gamma}^2 - (1-\underline{\gamma})^2 K_{\ell}^{-1}(1/\sigma^2_{\epsilon} + 1/\sigma^2_c)^2 = 0$$
$$\bar{\gamma}^2 - (1-\bar{\gamma})^2 K_{\ell}(1/\sigma^2 + 1/\sigma^2_c)^{-2} = 0.$$

[28] It is noted that the local government could also infer the central government's signal by acquiring information about θ, but it is indirect and turns out to be less efficient.

[29] See Propositions 6 and 7 for two special cases in the appendix.



(Note: $\sigma^2 = \sigma_\epsilon^2 = 100; \kappa_c = 1; \kappa_\ell = 2; \sigma_c^2 = \sigma^2/(2^{2\kappa_c} - 1)$.)

Fig. 4. The Relationship between $1/\sigma_\ell^2$ ang γ under the Decentralized Regime.

Under the centralized regime, it is the better signal received by the central that becomes noisier, while under the decentralized regime, it is the worse signal received by the local that becomes noisier. To predict the true state of the economy, one high quality signal is better than two mediocre quality signals. Therefore, if $\gamma = 0$, the decentralized regime performs better (higher expected output and lower volatility). The assumption that the local government has higher information capacity (Assumption 1) is essential to this result.[30]

Lemma 6. $E(a_c - \theta)^2 > E(a_\ell - \theta)^2\Big|_{\gamma=0}$.

On the other hand, decentralization with $\gamma = 1$ always worsens economic performance. Distortion in the information acquisition margin leads to strictly less informative signals for the decision maker, the local government. The economic outcome further deteriorates due to distortion in the policy making margin.

Lemma 7. $E(a_c - \theta)^2 < E(a_\ell - \theta)^2\Big|_{\gamma=1}$.

Following Proposition 2 and Lemmas 3, 6, and 7, we now obtain the core result of the paper.

Theorem 1. *There exists a unique $\tilde{\gamma}$ in (0,1) such that $E(a_c - \theta)^2 = E(a_\ell - \theta)^2\Big|_{\gamma=\tilde{\gamma}}$. If $\gamma > \tilde{\gamma}$, decentralization worsens economic performance; if $\gamma < \tilde{\gamma}$, decentralization improves economic performance.*

[30] As made clear in the proof, for $\gamma = 0$, decentralization improves economic performance if and only if $\kappa_\ell > \kappa_c$.

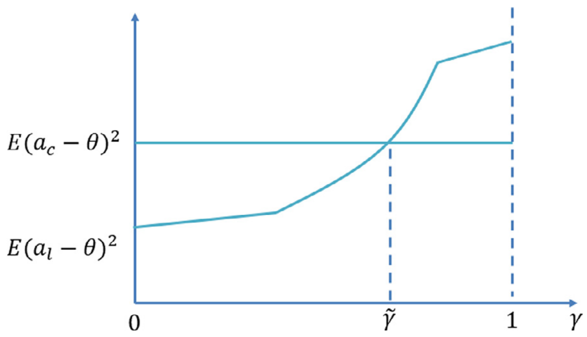


Fig. 5. The Comparison between Two Regimes.

This result highlights the pivotal role played by the loyalty concern in determining the economic outcome of decentralization in an authoritarian regime. Despite the information advantage held by the local, decentralization could be detrimental to the economy if the local bureaucrats have strong incentive to follow the policy suggestions from the central. Fig. 5 illustrates the comparison between two economic regimes in relation to the degree of loyalty concern γ. Notice there are two kinks on the curve of $E(a_\ell - \theta)^2$ at which $\gamma = \underline{\gamma}$ or $\bar{\gamma}$. The curve is much steeper in the middle range as both margins of loyalty-driven distortion are effectively at work.

Corollary 1. $\bar{\gamma} > \tilde{\gamma}$.

The corollary suggests that for decentralization to be welfare-improving, we should expect local bureaucrats to at least spend some effort on direct information acquisition.[31] In light of the intuition behind Lemma 7, the devotion of local bureaucrats to understanding and deciphering the policy message from the top guarantees the failure of decentralization.

4.5. Discussion

4.5.1. Revisiting China's reform history

In light of our theory, the difference in γ could help explain the contrasting experience following the two decentralization reforms in China and possibly the heterogeneous outcomes across regions. In the 1950s, the pursuit of economic welfare, even in the interest of the public, can bear great political risks. The inherent unpredictability of the policy choices at the very top, as well as the strong tendency to politicalize the (economic) policy mistakes, left the lower-level bureaucrats with little incentive to deviate from the policy prescriptions from the central. This means the loyalty concern, γ, could be very close to one in the 50s, leading to the great failure of the first decentralization. In contrast, the political environment during the 1978 reform was fundamentally changed. Due to ideological shifts, the central government put great emphasis on economic development, which engaged all levels of local governments in the tournament of GDP growth. Although the local bureaucrats cannot be fully freed from the policy suggestions from the central in an authoritarian regime, anecdotal and empirical evidence points to the fact that γ substantially decreased during the 1978 reform period, contributing to the success of the second great decentralization in China.

By having only one local government and focusing exclusively on the inter-governmental communication, our model abstracts from the competition between different local governments. In fact, introducing inter-regional competition could strengthen our argument. In the 1950s, knowing that political loyalty would pay off, competition among the lower level bureaucrats made them behave more radically, thus pushing γ towards its upper bound. In the post-reform era, since economic development became the new priority, signaling political loyalty by sacrificing the local economy could backfire. Inter-regional competition of economic performance acted as a disciplinary device in the authoritarian regime to put some downward pressure on γ.

It should be noted that variation in loyalty concern is not the only source of heterogeneity that could rationalize differential outcomes of decentralization, especially when it comes to the regional differences in output drop during the Great Leap Forward. An alternative explanation is that different localities may start with different levels of capabilities to implement economic policy during a decentralization reform (Besley and Persson, 2014). This state-capacity channel, which has been empirically validated by Lu et al. (2020), and our communication channel complement each other. If we interpret the local government's information capacity κ_ℓ as a proxy for state capacity, then the state capacity story says a higher κ_ℓ would improve decentralization outcomes, whereas our analysis enriches this argument by imposing a qualification: decentralization is only welfare improving when the loyalty concern is not too strong; otherwise building the local government's information capacity would be in vain.

[31] On the other hand, there is no clear relationship between $\tilde{\gamma}$ and $\underline{\gamma}$. Consider two numerical example. First, we let $\sigma^2 = \sigma_\epsilon^2 = 100$, $\kappa_\ell = 2\kappa_c = 2$. We find that $\underline{\gamma} \approx 0.27$ and $\tilde{\gamma} \approx 0.47$. Then, we reduce σ_ϵ^2 to be 10, which leads to $\underline{\gamma} \approx 0.33$ and $\tilde{\gamma} \approx 0.26$.

4.5.2. Reinterpreting the model in the context of democracies

Throughout our modeling exercise, we focus exclusively on a hierarchical authoritarian regime, but our formulation and results can also be reinterpreted in the context of democracies, thus shedding light on the role of voters' information-processing capacity in representative democracies.[32] To reinterpret the model, we relabel the central government as voters and the local government as the elected politicians. The centralized regime in our model can be viewed as classical Athenian democracies, in which voters directly deliberate and decide on policy-making while the politicians' role is to help voters make informed decisions. The decentralized regime can be viewed as representative democracies, in which it is the elected politicians who enact legislation. Our Theorem 1 suggests that, representative democracies tend to produce better outcomes than classical democracies, which is consistent with the fact that modern democracies are predominantly representative democracies; however, perhaps more importantly, in the presence of communication frictions and information processing constraints, when voters possess less precise information, larger inefficiencies may arise from representative democracies (for example, in the form of populist parties) when the politicians have very strong incentives to pander to voters. Policy-making mostly driven by voter opinion erodes the informational advantage of the representative democracies.

Closely related to this reinterpretation is the recent work by Matějka and Tabellini (2020). In their model, they consider an electoral competition between two opportunistic politicians for rationally inattentive voters. Due to the interplay between endogenous information acquisition and opportunistic policy making, greater availability of information may have negative welfare consequences. Albeit in different settings, both models point to the critical importance of information processing constraints in understanding the distortions in political processes. More broadly, the reinterpretation connects our model to a larger literature examining electoral accountability through the lens of contractual theory. As reviewed by Ashworth (2012), a key insight from this literature is that politicians' incentive to "impress the voters" may be conflicted with "the normative imperative to advance the voter's interests". There is a direct parallelism between this tension in the models of political agency and the loyalty concern in our formulation.

5. Extensions

To check the robustness of our model prediction, we consider three extensions. The first extension shuts down the channel of endogenous information transmission between governments. In this non-strategic environment, if the exogenous communication friction is large enough, then decentralization always benefits the economy. Thus, the inter-governmental communication channel is essential to generate differential outcomes of decentralization. Moreover, in this simplified setting, a seemingly paradoxical relationship arises: the economic performance improves with the noisiness of the inter-governmental communication friction, as higher exogenous communication friction incentivizes the local government to focus on its self-obtained and more precise information in the policy marking. This finding echoes the earlier result due to Board et al. (2007) in a cheap-talk environment. The second extension introduces strategic communication and the third extension relaxes the assumption of information advantage of the local. We discuss how the main insights of this paper adapt to those extensions.

5.1. Exogenous communication frictions

The first extension shuts down the endogenous communication channel. It concerns whether the model will deliver our main results when loyalty concern only operates through the final decision-making margin and, under what condition, the endogenous communication channel is essential to generating differential outcomes of decentralization.

The basic setup is very similar. The main departure is that σ_ℓ^2 and σ_c^2 are exogenously given. In particular, we assume that each government receives a private signal about θ from the nature. The private signal of the central government θ_c follows $\mathcal{N}(\theta, \sigma_c^2)$, while the private signal of the local government θ_ℓ follows $\sim \mathcal{N}(\theta, \sigma_\ell^2)$ with σ_ℓ^2 and σ_c^2 exogenously given. We assume $\sigma_c^2 > \sigma_\ell^2$, that is, the local government receives a more precise signal than the central government. The quality of the signal (σ_ℓ^2 or σ_c^2) is the same under both regimes.

Under the centralized regime, the local government sends its signal $s_\ell = \theta_\ell$ to the central. The central government receives a signal s_ℓ' with $s_\ell' = s_\ell + \epsilon$. Upon receiving the signal, the central government makes the policy choice a_c based on the private information θ_c and the signal received s_ℓ'. Similarly, under the decentralized regime, the central government sends a signal $s_c = \theta_c$. The local government receives a signal s_c' with $s_c' = s_c + \epsilon$. The local government then picks its preferred policy a_ℓ based on θ_ℓ and s_c'. We assume $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ with σ_ϵ^2 being exogenous. The model otherwise follows the baseline setting. Fig. 6 illustrates the timeline of the model without rational inattention.

Abstracting from the strategic behavior in this simplified framework, the decision problem of the signal receiver is only to determine the desired economic policy.

Under the centralized regime, the decision problem of the central government is given by

$$\max_{a_c} E(Y|\theta_c, s_\ell') = Y^* - E[(a_c - \theta)^2|\theta_c, s_\ell'],$$

[32] We gratefully acknowledge one of our referees for his/her suggestion and advice to write this subsection. It is to bring out an insightful reinterpretation and an intriguing connection elucidated in the referee report between our model and the issues arising from informational asymmetry in a democratic environment.

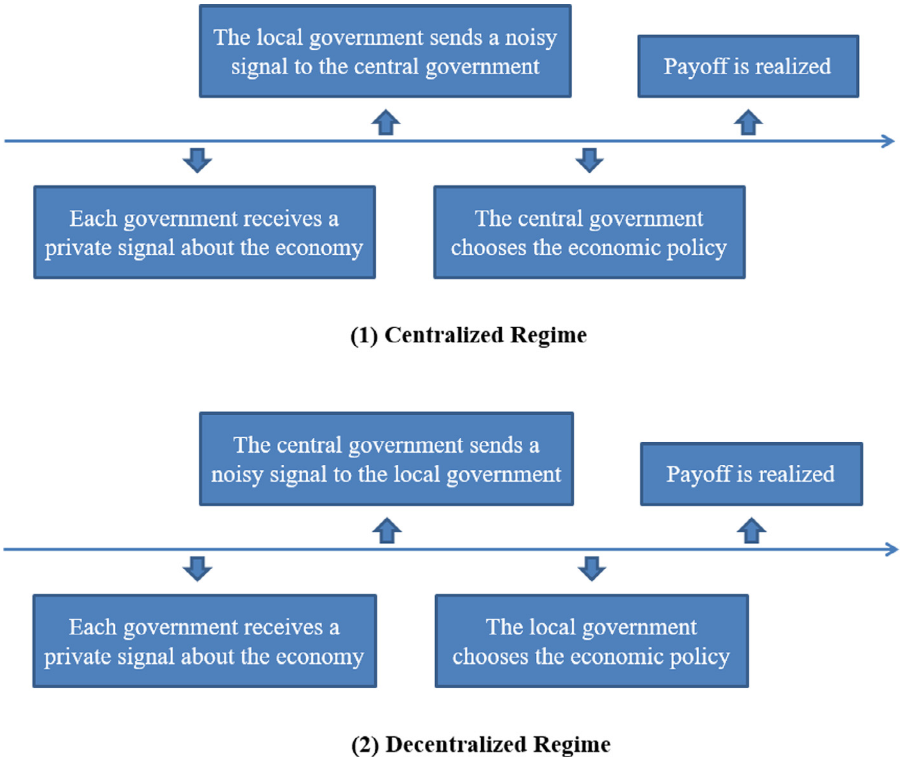


Fig. 6. Timeline without Rational Inattention.

Under the decentralized regime, the decision problem of the local government is given by

$$\max_{a_\ell} (1-\gamma)\left(Y^* - E[(a_\ell-\theta)^2|\theta_\ell, s_c']\right) - \gamma E[(a_\ell - s_c)^2|\theta_\ell, s_c'].$$

Solving for the optimal policy, we again find that the government policy is always a linear combination of two private signals, θ_c and θ_ℓ, and the communication friction ε.[33] According to Lemma 3, to analyze economic output and volatility, we only need to focus on $E(a_c-\theta)^2$ and $E(a_\ell-\theta)^2$.[34]

Lemma 8. *In the absence of the loyalty concern ($\gamma = 0$), decentralization improves the economic performance.*

Following the same intuition as in the benchmark model, the assumption $\sigma_\ell^2 < \sigma_c^2$ and the presence of communication friction play a crucial role in the above result. However, the outcome of decentralization under $\gamma = 1$ is now less clear.

Lemma 9. *Under the decentralized regime with $\gamma = 1$, there exists a unique $\bar{\sigma}_\epsilon^2 > 0$ such that $E(a_\ell-\theta)^2 = E(a_c-\theta)^2$ for $\sigma_\epsilon^2 = \bar{\sigma}_\epsilon^2$ and $E(a_\ell-\theta)^2 > E(a_c-\theta)^2$ if and only if $\sigma_\epsilon^2 < \bar{\sigma}_\epsilon^2$.*

To understand this result, consider two thought experiments. If the communication friction vanishes, this non-strategic environment coincides with the benchmark model and therefore, our earlier results carry over. The local government ignores its self-obtained and more precise signal and simply follows the policy prescription from the central government, thus leading to a worse economic outcome. To the other extreme, if the communication is prohibitively noisy ($\sigma_\epsilon^2 \to \infty$), the local government cannot rely on the signal it receives from the central government to predict s_c. Instead, it has to rely on its own signal, which is correlated with the original signal sent by the central, s_c, via the mutual component θ. In this case, decentralization could be welfare improving even in the presence of pure loyalty concern.

This result stands in sharp contrast with Lemma 7 in which decentralization under $\gamma = 1$ always leads to worse economic performance. The difference precisely stems from the missing endogenous communication margin.

It can be easily seen from Lemma 19 that the expected output strictly decreases with σ_ϵ^2 under the centralized regime and the decentralized regime with $\gamma = 0$. When governments attempt to maximize the expected output, additional communication friction worsens the economic outcome. However, the intuition gets reversed when we turn to a purely loyalty-driven local government. In fact, we can prove the following seemingly paradoxical result: *higher communication friction could be welfare improving under decentralization.*

[33] See Lemma 18 in the appendix.

[34] For the explicit expressions of $E(a_c-\theta)^2$ and $E(a_\ell-\theta)^2$ under $\gamma = 0$ and $\gamma = 1$, see Lemma 19 in the appendix.

Proposition 3. *Under the decentralized regime with* $\gamma = 1$, $\partial E(a_\ell - \theta)^2/\partial(\sigma_\epsilon^2) < 0$.

This result underscores the insights behind the two thought experiments conducted above. Higher communication friction, on the one hand, makes information transmission more difficult, but on the other hand, makes the local government effectively more independent from the central in policy-making. The second channel dominates when $\gamma = 1$.

We now prove a counterpart of Proposition 2 in this alternative setup. In the absence of strategic considerations, the proof turns out to be much simpler.

Corollary 2. *Under the decentralized regime in a non-strategic environment,* $E(a_\ell - \theta)^2$ *strictly increases with* γ.

Now we are ready to provide the main theorem in this non-strategic environment, a complete characterization of the relative economic performance under two regimes.

Theorem 2. *In a non-strategic environment, there exists a unique* $\bar{\sigma}_\epsilon^2 > 0$ *such that*

1. *if* $\sigma_\epsilon^2 > \bar{\sigma}_\epsilon^2$, *decentralization always improves economic performance*
2. *if* $\sigma_\epsilon^2 < \bar{\sigma}_\epsilon^2$, *there exists a unique* $\hat{\gamma} \in (0, 1)$ *such that (i) if* $\gamma < \hat{\gamma}$, *decentralization improves economic performance; (ii) if* $\gamma > \hat{\gamma}$, *decentralization worsens economic performance.*

The differential outcomes of decentralization emerge only when the exogenous communication friction is sufficiently small. When inter-governmental communication is sufficiently noisy, this non-strategic environment predicts that decentralization is always welfare improving, even for $\gamma = 1$. This is because distortion in the information acquisition margin is muted and thus loyalty concern only impacts the policy making margin. When there is substantial exogenous communication friction, distortion in the policy making margin is not large enough to alter the sign of the outcome under decentralization.

5.2. Strategic communication

In the baseline setting, we assume that the signal sender always reveals its information truthfully: $s_\ell = \theta_\ell$ under the centralized regime and $s_c = \theta_c$ under the decentralized regime. We now relax this assumption by allowing the signal sender to introduce additional noise to its signal. In particular, under the decentralized regime, we assume the local government sends a signal of the form

$$s_\ell = \theta_\ell + \delta_\ell$$

with $\delta_\ell \sim \mathcal{N}(0, \sigma_{\delta\ell}^2)$. Under the decentralized regime, the central government sends a signal of the form

$$s_c = \theta_c + \delta_c$$

with $\delta_c \sim \mathcal{N}(0, \sigma_{\delta c}^2)$. We assume that the signal sender can choose any $\sigma_{\delta\ell}^2$ (or $\sigma_{\delta c}^2$) costlessly.

The local government's decision problem can be modeled in two different ways.[35] First, if the local government attempts to target its policy to the central government's original signal θ_c, then the objective function is given by

$$\max_{\sigma_\ell^2, \sigma_{\epsilon\ell}^2} E\left\{ \max_{a_\ell} (1-\gamma)\left(Y^* - E[(a_\ell - \theta)^2|\theta_\ell, \mathbf{s}_c']\right) - \gamma E[(a_\ell - \theta_c)^2|\theta_\ell, \mathbf{s}_c'] \right\}.$$

In this case, the local government's information acquisition decision hinges on $\sigma_{\delta c}^2$, which is summarized by the following proposition.

Proposition 4. *Under the decentralized regime with* $\gamma = 1$, *there exist two cutoffs for* $\sigma_{\delta c}^2$: $\bar{\sigma}_{\delta c}^2$ *and* $\underline{\sigma}_{\delta c}^2$. *If* $\sigma_{\delta c}^2 \geq \bar{\sigma}_{\delta c}^2$, *then* $\sigma_{\epsilon\ell}^2 = \sigma_\epsilon^2$; *If* $\sigma_{\delta c}^2 \leq \underline{\sigma}_{\delta c}^2$, $\sigma_\ell^2 = \infty$; *if* $\sigma_{\delta c}^2 \in (\underline{\sigma}_{\delta c}^2, \bar{\sigma}_{\delta c}^2)$, *then* σ_ℓ^2 *strictly decreases with* $\sigma_{\delta c}^2$.

The proposition suggests a (weakly) monotonically decreasing relationship between σ_ℓ^2 and $\sigma_{\delta c}^2$. For a large $\sigma_{\delta c}^2$, the return of acquiring information about the central government's private signal θ_c through communication is lower and thus the local government has stronger incentives for direct information acquisition. As a result, strategic communication gives rise to a new type of equilibrium: knowing that the local government may divert their attention resources to recover the central government's private signal, the central government may intentionally send an extremely noisy signal to deter any inter-governmental communication effort. This equilibrium resembles the babbling equilibrium in the cheap-talk literature (Crawford and Sobel, 1982). It should nevertheless be noted that, unlike the cheap-talk setting in which there is a much richer message space, the communication instrument in our model is a single variable capturing the precision of the signal. Despite this simplification, strategic communication still leads to multiple equilibria because if the local believes the central to send a non-informative signal ($\sigma_{\delta c}^2 = \infty$), a babbling-type equilibrium would always emerge.[36] In a babbling-type equilibrium, the local government decides on economic policy entirely based on its own signal, as opposed to the centralized regime, the central government makes its policy based on its own (less precise) signal and the local government's

[35] We are indebted to one of the referees for the rewriting of this subsection. The second case is incorporated in light of his/her insightful comment.

[36] Due to the complication arising from strategic communication in a rational inattention framework, we are unable to provide a sharp equilibrium characterization as in the benchmark setting.

signal-cum-communication-friction. Our main results will continue to hold, that is, decentralization being welfare-reducing for $\gamma = 1$, provided that either the difference in information capacity between the two governments (κ_ℓ and κ_c) or the inter-governmental communication friction σ_ϵ^2 is not too large.

Second, if the local government attempts to signal its loyalty by orienting its policy target towards the signal s_c sent by the central government rather than the central government's original signal θ_c, then the objective function is given by

$$\max_{\sigma_\ell^2, \sigma_{\epsilon\ell}^2} E\left\{\max_{a_\ell} (1-\gamma)\left(Y^* - E[(a_\ell - \theta)^2|\theta_\ell, \mathbf{s}_c']\right) - \gamma E[(a_\ell - s_c)^2|\theta_\ell, \mathbf{s}_c']\right\}.$$

In this case, the decision problem of the signal receiver is the same as that in the baseline setting by simply replacing σ_ℓ^2 with $\sigma_\ell^2 + \sigma_{\delta\ell}^2$ under the centralized regime and replacing σ_c^2 with $\sigma_c^2 + \sigma_{\delta c}^2$ under the decentralized regime. Immediately following from Lemmas 5 and 15, we obtain $\sigma_{\delta\ell}^2 = \sigma_{\delta c}^2 = 0$. In the benchmark model, we have shown that the information flow constraint is always binding for the signal sender. Therefore, even if the signal sender is allowed to strategically obscure the signal, it will not have the incentive to do so. However, if we further allow the central government to conceal its signal, the central government will have to compare sending a noiseless signal ($\sigma_{\delta c}^2 = 0$) with not sending the signal at all. In this case, a babbling-type equilibrium may again emerge and the equilibrium discussion will follow the first case then.

5.3. Comparative advantage of information acquisition

In our baseline setting, we assume that the local government enjoys absolute advantage in information acquisition over the central government ($\kappa_\ell > \kappa_c$). This assumption may appear too strong under certain economic situations. We now assume, instead, the local government only enjoys comparative advantage in the direct acquisition of economic information. In particular, we assume that $\kappa_\ell = \kappa_c = \kappa$, and we replace Constraint 5 with

$$\left(\frac{1}{\sigma_c^2} + \frac{1}{\sigma_{\epsilon\ell}^2}\right)\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2} + \frac{\lambda}{\sigma_\ell^2}\right) \leq \frac{2^{2\kappa}(\sigma^2 + \sigma_c^2 + \sigma_\epsilon^2)}{\sigma^2\sigma_c^2\sigma_\epsilon^2} + \frac{1}{\sigma_c^4} \equiv K_\ell(\sigma_c^2), \tag{10}$$

and replace Constraint 2 with

$$\frac{1}{\sigma^2} + \frac{\lambda}{\sigma_\ell^2} \leq \frac{2^{2\kappa}}{\sigma^2}, \tag{11}$$

where $0 < \lambda < 1$. Under this assumption, the lowest attainable $\sigma_{\epsilon\ell}^2$ under the decentralized regime is the same as the lowest attainable $\sigma_{\epsilon c}^2$ under the centralized regime, while the local government has information advantage over the central government if both governments devote their attention to direct information acquisition.[37]

First, notice that the optimal policy does not depend on λ, so Lemmas 12 and 13 concerning the optimal policy for the decision maker under two regimes carry over. Under the centralized regime, the decision problem of the central government is unchanged. The only departure from the baseline setting is the introduction of the scaling parameter λ into the constraint of the signal sender (Constraint 11), so the following result immediately follows from Proposition 1.

Proposition 5. *Under the centralized regime, we have*

$$E(a_c - \theta)^2 = \left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2} + \frac{1}{\sigma_\ell^2 + \sigma_\epsilon^2}\right)^{-1}$$

with $\sigma_\ell^2 = \lambda\sigma^2/(2^{2\kappa} - 1)$ *and*

$$\sigma_c^2 = \left[K_c\left(\frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma_\ell^2}\right)^{-1} - \frac{1}{\sigma^2} - \frac{1}{\sigma_\ell^2}\right]^{-1} = \frac{\sigma^2(\sigma_\epsilon^2 + \sigma_\ell^2)}{(2^{2\kappa} - 1)(\sigma^2 + \sigma_\epsilon^2 + \sigma_\ell^2)}.$$

Under the decentralized regime, since the local government, which is the signal receiver, enjoys information comparative advantage, the trade-off of the resource allocation for the local government hinges on λ. If λ is sufficiently close to zero, the local government always devotes itself to direct information acquisition. To rule out this less interesting case, unless explicitly stated, we always impose the following regularity condition on λ throughout this subsection.

$$\lambda(1/\lambda - 1)^2 < \frac{2^{4\kappa}\left(\frac{1}{2^{2\kappa}-1} + \frac{2^{2\kappa}\sigma^2}{(2^{2\kappa}-1)^2\sigma_\epsilon^2}\right)^2}{2^{2\kappa}\left(\frac{1}{2^{2\kappa}-1} + \frac{2^{2\kappa}\sigma^2}{(2^{2\kappa}-1)^2\sigma_\epsilon^2}\right) + 1}. \tag{12}$$

[37] We introduce λ in a rather reduced-form way, which transparently conveys intuition, but the drawback is that we deviate from the standard entropy reduction framework. Strictly speaking, what we deal with is no longer an entropy. In fact, it is difficult to introduce the notion of comparative advantage in information acquisition into this framework without deviating from the formal definition of entropy, because as we point out earlier, θ and ϵ are not independent conditional on θ_ℓ and $\mathbf{s}_c'$ (or θ_c and $\mathbf{s}_\ell'$).

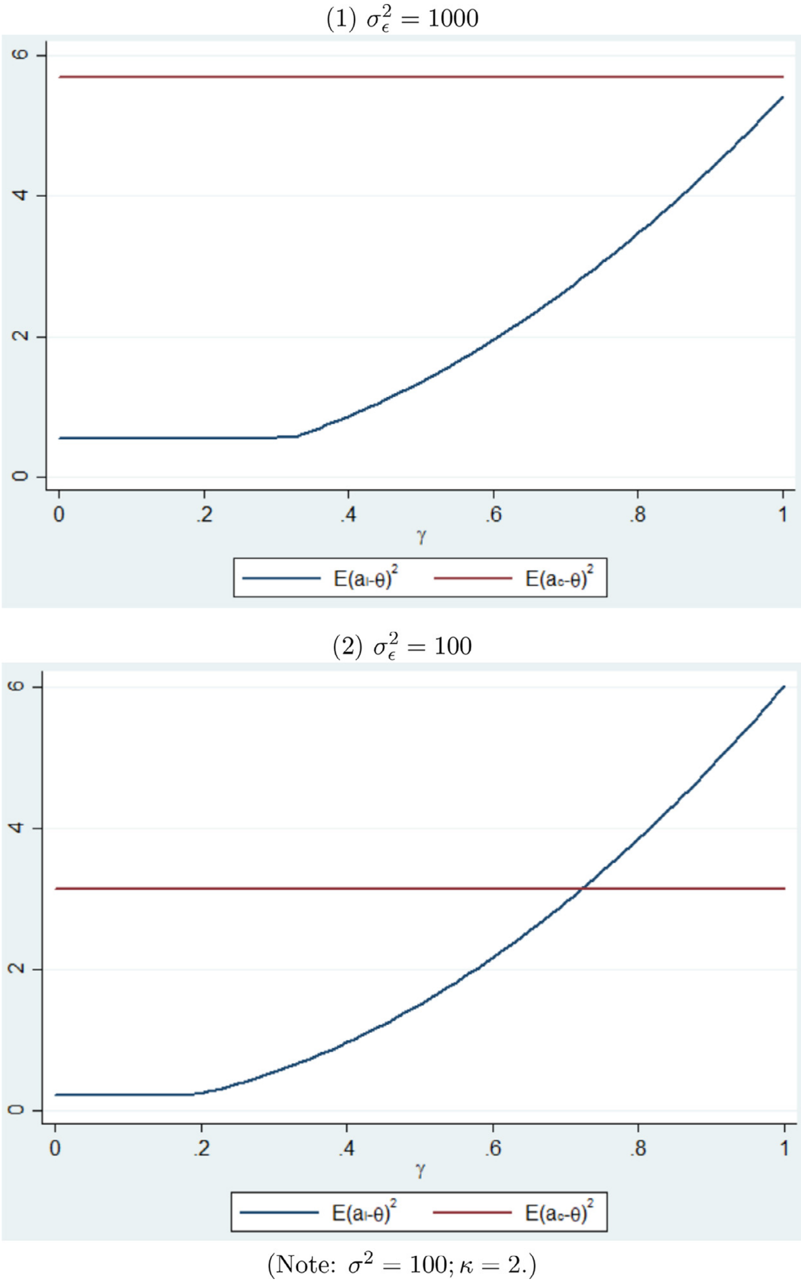


Fig. 7. Comparison between Two Regimes When Condition 12 Is Violated.

Since λ has to be in (0,1), this condition is equivalent to imposing a lower bound on λ. The condition may appear complicated, so we provide the following technical result that sharpens the lower bound for λ.

Lemma 10. *If $\lambda \geq 1/2$, then the regularity condition 12 holds for any κ, σ^2, and σ_ϵ^2.*

The following lemma is the counterpart of Lemma 4. It can be seen that the characterization of the optimal policy for a general γ under the decentralized regime is qualitatively unchanged in this alternative setting.

Lemma 11. *Under the decentralized regime, there exist two cutoffs $\underline{\gamma}'$ and $\bar{\gamma}'$ such that $0 < \underline{\gamma}' < \bar{\gamma}' < 1$. If $\gamma \leq \underline{\gamma}'$, the local government specializes in direct information acquisition ($\sigma_{\epsilon\ell}^2 = \sigma_\epsilon^2$). If $\gamma \geq \bar{\gamma}'$, the local government specializes in intergovernmental communication ($\sigma_\ell^2 = \infty$). If $\gamma \in (\underline{\gamma}', \bar{\gamma}')$, the local government allocates its budget to both activities ($\sigma_\ell^2 < \infty$ and $\sigma_{\epsilon\ell}^2 < \sigma_\epsilon^2$) with $\partial\sigma_\ell^2/\partial\gamma > 0$.*

However, due to the complication for γ in the middle range, it is very challenging to establish the counterpart of Theorem 1 in this setting. Instead, we provide a slightly weaker result which nevertheless captures the main insight.

Theorem 3. *If γ is sufficiently close to one, decentralization worsens economic performance; if γ is sufficiently close to zero, decentralization improves economic performance.*

Albeit not being formally established, our simulation results suggest that the comparison between two regimes seems very similar to what has been illustrated by Fig. 5: economic volatility increases monotonically with γ and there are two kinks on the curve of $E(a_\ell - \theta)^2$, representing the structural changes of the local government's optimal strategy when $\gamma = \underline{\gamma}'$ and $\gamma = \bar{\gamma}'$.

To close this subsection, we consider the case the local government has very strong comparative advantage such that Condition 12 does not hold. In this case, $\bar{\gamma}'$ disappears. Since there is no equilibrium characterization for the middle range of γ, we perform a battery of numerical experiments. As shown in Fig. 7, other things equal, when σ_ϵ^2 is relatively large, then decentralization always leads to improvement of economic performance; when σ_ϵ^2 is relatively small, then the economic outcome of decentralization hinges on γ. The numerical results are qualitatively similar to what we have shown for the setting with exogenous communication friction.

6. Conclusion

In this paper, we propose a model of inter-governmental information transmission to understand the political economy of decentralization. The model demonstrates that the impact of decentralization on economic performance hinges on the degree of loyalty concern of the local government. Decentralization leads to higher volatility and lower output, if the local government is a loyal follower of the central government. A closer examination of the underlying mechanism reveals that loyalty concern impacts the economic outcome of decentralization through two margins: policy making and information acquisition. The latter turns out to be crucial for welfare-reducing decentralization to emerge from our theoretical construction. The main result sheds light on the sub-context of our model, the contrasting dynamics following the two waves of decentralization in China. It highlights the complex interplay between loyalty concern, inter-governmental communication, and friction therein. Therefore, for policy makers, evaluation of the past decentralization reforms requires a better understanding of political distortion deeply embedded in the system.

Our work can be extended in several directions. Empirically, it would be interesting to directly test the model prediction by applying text analysis of official government documents. Does a loyalty-driven local government tend to spend more time on inter-governmental communication? Is there a general trend of how lower level governments study and interpret directives from their superordinates from the Maoist era to the reform era? Theoretically, the model can be extended to accommodate multiple local governments. Once competition across localities is put into play, under what condition will information distortion through inter-governmental communication be further amplified? More importantly, if we allow the local governments to be heterogeneous in their information capacity, could we further incorporate the insights from the state capacity literature into our model? Moreover, given its generality, our analytical framework can be reframed to help understand decentralization experience in other authoritarian countries and more broadly, large organizations with internal hierarchies. We leave all these avenues for future research.

Declaration of Competing Interest

For the manuscript: Intergovernmental Communication under Decentralization.

The author Shiyu BO declares that this research is supported by China National Natural Science Foundation No. 71803064, where he acts as the PI; and Guangzhou Philosophy and Social Science Planning Grant No. 2019GZYB24, where he acts as the PI. There is no conflict of interests.

The author Liuchun Deng declares that he has no relevant or material financial interests that relate to the research described in this paper.

The author Yufeng SUN declares that he has no relevant or material financial interests that relate to the research described in this paper.

The author Boqun Wang declares that he has no relevant or material financial interests that relate to the research described in this paper.

Appendix A. Proofs and Some Unreported Results

A.1. Proof of lemma 1

Proof. Since $H(\theta, \epsilon | s'_{\ell 0}) = H(\theta, \epsilon, s'_{\ell 0}) - H(s'_{\ell 0})$ and $H(\theta, \epsilon | \theta_c, \mathbf{s}'_\ell) = H(\theta, \epsilon, \theta_c, \mathbf{s}'_\ell) - H(\theta_c, \mathbf{s}'_\ell)$, we have[38]

[38] Alternatively, for two multivariate normal distributions X and Y, we have $|\Sigma_{X|Y}||\Sigma_Y| = |\Sigma_X|$.

$$\begin{aligned} H(\theta,\epsilon|s'_{\ell 0}) - H(\theta,\epsilon|\theta_c,\mathbf{s}'_\ell) &= H(\theta,\epsilon,s'_{\ell 0}) + H(\theta_c,\mathbf{s}'_\ell) - H(s'_{\ell 0}) - H(\theta,\epsilon,\theta_c,\mathbf{s}'_\ell) \\ &= \frac{1}{2}\Big(\log_2|\Sigma_{\theta,\epsilon,s'_{\ell 0}}| + \log_2|\Sigma_{\theta_c,\mathbf{s}'_\ell}| - \log_2|\Sigma_{s'_{\ell 0}}| - \log_2|\Sigma_{\theta,\epsilon,\theta_c,\mathbf{s}'_\ell}|\Big) \le \kappa_c. \end{aligned}$$

Under the assumption of the model, we have $|\Sigma_{s'_{\ell 0}}| = \sigma^2 + \sigma_\ell^2 + \sigma_\epsilon^2$ and

$$|\Sigma_{\theta,\epsilon,s'_{\ell 0}}| = \begin{vmatrix} \sigma^2 & 0 & \sigma^2 \\ 0 & \sigma_\epsilon^2 & \sigma_\epsilon^2 \\ \sigma^2 & \sigma_\epsilon^2 & \sigma^2+\sigma_\ell^2+\sigma_\epsilon^2 \end{vmatrix} = \sigma^2\sigma_\epsilon^2\sigma_\ell^2,$$

$$|\Sigma_{\theta_c,\mathbf{s}'_\ell}| = \begin{vmatrix} \sigma^2+\sigma_c^2 & \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2+\sigma_\ell^2+\sigma_\epsilon^2 & \sigma^2+\sigma_\ell^2 \\ \sigma^2 & \sigma^2+\sigma_\ell^2 & \sigma^2+\sigma_\ell^2+\tilde\sigma_{\epsilon c}^2 \end{vmatrix} = (\sigma^2+\sigma_c^2)(\sigma_\ell^2\tilde\sigma_{\epsilon c}^2 + \sigma_\epsilon^2\tilde\sigma_{\epsilon c}^2 + \sigma_\epsilon^2\sigma_\ell^2) + \sigma_c^2\sigma^2(\tilde\sigma_{\epsilon c}^2 + \sigma_\epsilon^2),$$

$$|\Sigma_{\theta,\epsilon,\theta_c,\mathbf{s}'_\ell}| = \begin{vmatrix} \sigma^2 & 0 & \sigma^2 & \sigma^2 & \sigma^2 \\ 0 & \sigma_\epsilon^2 & 0 & \sigma_\epsilon^2 & 0 \\ \sigma^2 & 0 & \sigma^2+\sigma_c^2 & \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma_\epsilon^2 & \sigma^2 & \sigma^2+\sigma_\ell^2+\sigma_\epsilon^2 & \sigma^2+\sigma_\ell^2 \\ \sigma^2 & 0 & \sigma^2 & \sigma^2+\sigma_\ell^2 & \sigma^2+\sigma_\ell^2+\tilde\sigma_{\epsilon c}^2 \end{vmatrix} = \sigma^2\sigma_\epsilon^2\sigma_c^2\sigma_\ell^2\tilde\sigma_{\epsilon c}^2.$$

Plugging the terms into the information flow constraint, we obtain

$$\left(\frac{1}{\sigma_c^2\sigma_\epsilon^2} + \frac{1}{\sigma^2\sigma_\epsilon^2} + \frac{1}{\sigma_c^2\sigma_\ell^2} + \frac{1}{\sigma^2\sigma_\ell^2} + \frac{1}{\sigma_c^2\tilde\sigma_{\epsilon c}^2} + \frac{1}{\sigma^2\tilde\sigma_{\epsilon c}^2} + \frac{1}{\sigma_\epsilon^2\sigma_\ell^2} + \frac{1}{\sigma_\ell^2\tilde\sigma_{\epsilon c}^2}\right) \le 2^{2\kappa_c}\frac{\sigma^2+\sigma_\ell^2+\sigma_\epsilon^2}{\sigma^2\sigma_\epsilon^2\sigma_\ell^2}.$$

Using $\sigma_{\epsilon c}^2 = (\sigma_\epsilon^{-2} + \tilde\sigma_{\epsilon c}^{-2})^{-1}$ and simplifying the equation, we obtain the desired conclusion. □

A.2. Proof of lemma 2

Proof. Since $H(\theta,\epsilon|s'_{c0}) = H(\theta,\epsilon,s'_{c0}) - H(s'_{c0})$ and $H(\theta,\epsilon|\theta_\ell,\mathbf{s}'_c) = H(\theta,\epsilon,\theta_\ell,\mathbf{s}'_c) - H(\theta_\ell,\mathbf{s}'_c)$,

$$\begin{aligned} H(\theta,\epsilon|s'_{c0}) - H(\theta,\epsilon|\theta_\ell,\mathbf{s}'_c) &= H(\theta,\epsilon,s'_{c0}) + H(\theta_\ell,\mathbf{s}'_c) - H(s'_{c0}) - H(\theta,\epsilon,\theta_\ell,\mathbf{s}'_c) \\ &= \frac{1}{2}\Big(\log_2|\Sigma_{\theta,\epsilon,s'_{c0}}| + \log_2|\Sigma_{\theta_\ell,\mathbf{s}'_c}| - \log_2|\Sigma_{s'_{c0}}| - \log_2|\Sigma_{\theta,\epsilon,\theta_\ell,\mathbf{s}'_c}|\Big) \le \kappa_\ell. \end{aligned}$$

Under the assumption of the model, we have $|\Sigma_{s'_{c0}}| = \sigma^2 + \sigma_c^2 + \sigma_\epsilon^2$ and

$$|\Sigma_{\theta,\epsilon,s'_{c0}}| = \begin{vmatrix} \sigma^2 & 0 & \sigma^2 \\ 0 & \sigma_\epsilon^2 & \sigma_\epsilon^2 \\ \sigma^2 & \sigma_\epsilon^2 & \sigma^2+\sigma_c^2+\sigma_\epsilon^2 \end{vmatrix} = \sigma^2\sigma_\epsilon^2\sigma_c^2,$$

$$|\Sigma_{\theta_\ell,\mathbf{s}'_c}| = \begin{vmatrix} \sigma^2+\sigma_\ell^2 & \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2+\sigma_c^2+\sigma_\epsilon^2 & \sigma^2+\sigma_c^2 \\ \sigma^2 & \sigma^2+\sigma_c^2 & \sigma^2+\sigma_c^2+\tilde\sigma_{\epsilon\ell}^2 \end{vmatrix} = (\sigma^2+\sigma_\ell^2)(\sigma_c^2\tilde\sigma_{\epsilon\ell}^2 + \sigma_\epsilon^2\tilde\sigma_{\epsilon\ell}^2 + \sigma_\epsilon^2\sigma_c^2) + \sigma_\ell^2\sigma^2(\tilde\sigma_{\epsilon\ell}^2 + \sigma_\epsilon^2),$$

$$|\Sigma_{\theta,\epsilon,\theta_\ell,\mathbf{s}'_c}| = \begin{vmatrix} \sigma^2 & 0 & \sigma^2 & \sigma^2 & \sigma^2 \\ 0 & \sigma_\epsilon^2 & 0 & \sigma_\epsilon^2 & 0 \\ \sigma^2 & 0 & \sigma^2+\sigma_\ell^2 & \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma_\epsilon^2 & \sigma^2 & \sigma^2+\sigma_c^2+\sigma_\epsilon^2 & \sigma^2+\sigma_c^2 \\ \sigma^2 & 0 & \sigma^2 & \sigma^2+\sigma_c^2 & \sigma^2+\sigma_c^2+\tilde\sigma_{\epsilon\ell}^2 \end{vmatrix} = \sigma^2\sigma_\epsilon^2\sigma_c^2\sigma_\ell^2\tilde\sigma_{\epsilon\ell}^2.$$

Plugging the terms into the information flow constraint, we obtain

$$\left(\frac{1}{\sigma_\epsilon^2} + \frac{1}{\tilde\sigma_{\epsilon\ell}^2} + \frac{1}{\sigma_c^2}\right)\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_\ell^2}\right) + \frac{1}{\sigma_c^2}\left(\frac{1}{\tilde\sigma_{\epsilon\ell}^2} + \frac{1}{\sigma_\epsilon^2}\right) + \frac{1}{\sigma_c^4} \le 2^{2\kappa_\ell}\frac{\sigma^2+\sigma_c^2+\sigma_\epsilon^2}{\sigma^2\sigma_\epsilon^2\sigma_c^2} + \frac{1}{\sigma_c^4}.$$

Using $\sigma_{\epsilon\ell}^2 = (\sigma_\epsilon^{-2} + \tilde\sigma_{\epsilon\ell}^{-2})^{-1}$ and simplifying the equation, we obtain the desired conclusion. □

A.3. Equilibrium definition

We now formally define the (perfect Bayesian Nash) equilibrium of the sequential game under each regime. In both regimes, we require the belief updating follows the Bayes' rule.[39] The equilibrium under the centralized regime is defined as a quadruplet $(\sigma_\ell^{*2}, \sigma_c^{*2}(\cdot), \sigma_{\epsilon c}^{*2}(\cdot), a_c^*(\cdot))$ such that for any quintuplet $(\sigma_\ell^2, \sigma_c^2, \sigma_{\epsilon c}^2; \theta_c, \mathbf{s}'_\ell)$,

$$a_c^*(\sigma_\ell^2, \sigma_c^2, \sigma_{\epsilon c}^2; \theta_c, \mathbf{s}'_\ell) \in \arg\max_{a_c} E(Y^* - (a_c - \theta)^2 | \theta_c, \mathbf{s}'_\ell);$$

For any $\sigma_\ell^2 > 0$,

$$(\sigma_c^{*2}(\sigma_\ell^2), \sigma_{\epsilon c}^{*2}(\sigma_\ell^2)) \in \arg\max_{\sigma_c^2, \sigma_{\epsilon c}^2} E(Y^* - (a_c^*(\sigma_\ell^2, \sigma_c^2, \sigma_{\epsilon c}^2; \theta_c, \mathbf{s}'_\ell) - \theta)^2)$$

subject to Constraint 3 and $\sigma_{\epsilon c}{}^2 \le \sigma_\epsilon^2$;

$$\begin{aligned}\sigma_\ell^{*2} \in \arg\max_{\sigma_\ell^2} \ & \left\{(1-\gamma)(Y^* - E(a_c^*(\sigma_\ell^2, \sigma_c^{*2}(\sigma_\ell^2), \sigma_{\epsilon c}^{*2}(\sigma_\ell^2); \theta_c, \mathbf{s}'_\ell) - \theta)^2)\right. \\ & \left. -\gamma E(\theta_\ell - a_c^*(\sigma_\ell^2, \sigma_c^{*2}(\sigma_\ell^2), \sigma_{\epsilon c}^{*2}(\sigma_\ell^2); \theta_c, \mathbf{s}'_\ell))^2\right\}\end{aligned}$$

subject to Constraint 2.

The equilibrium under the decentralized regime is defined as a quadruplet $(\sigma_c^{*2}, \sigma_\ell^{*2}(\cdot), \sigma_{\epsilon\ell}^{*2}(\cdot), a_\ell^*(\cdot))$ such that for any quintuplet $(\sigma_c^2, \sigma_\ell^2, \sigma_{\epsilon\ell}^2; \theta_\ell, \mathbf{s}'_c)$,

$$a_\ell^*(\sigma_c^2, \sigma_\ell^2, \sigma_{\epsilon\ell}^2; \theta_\ell, \mathbf{s}'_c) \in \arg\max_{a_\ell} (1-\gamma) E(Y^* - (a_\ell - \theta)^2 | \theta_\ell, \mathbf{s}'_c) - \gamma E[(a_\ell - \theta_c)^2 | \theta_\ell, \mathbf{s}'_c];$$

For any $\sigma_c^2 > 0$,

$$\begin{aligned}(\sigma_\ell^{*2}(\sigma_c^2), \sigma_{\epsilon\ell}^{*2}(\sigma_c^2)) \in \arg\max_{\sigma_\ell^2, \sigma_{\epsilon\ell}^2} \ & \left\{(1-\gamma) E(Y^* - (a_\ell^*(\sigma_c^2, \sigma_\ell^2, \sigma_{\epsilon\ell}^2; \theta_\ell, \mathbf{s}'_c) - \theta)^2)\right. \\ & \left. -\gamma E(a_\ell^*(\sigma_c^2, \sigma_\ell^2, \sigma_{\epsilon\ell}^2; \theta_\ell, \mathbf{s}'_c) - \theta_c)^2\right\}\end{aligned}$$

subject to Constraint 5 and $\sigma_{\epsilon\ell}{}^2 \le \sigma_\epsilon^2$;

$$\sigma_c^{*2} \in \arg\max_{\sigma_c^2} (Y^* - E(a_\ell^*(\sigma_c^2, \sigma_\ell^{*2}(\sigma_c^2), \sigma_{\epsilon\ell}^{*2}(\sigma_c^2); \theta_\ell, \mathbf{s}'_c) - \theta)^2)$$

subject to Constraint 4.

A.4. The optimal policy under the two regimes

Lemma 12. *Under the centralized regime, the optimal policy of the central government is given by*

$$a_c = E(\theta | \theta_c, \mathbf{s}'_\ell) = \frac{\frac{\theta_c}{\sigma_c^2} + \frac{s'_\ell}{\sigma_\ell^2 + \sigma_{\epsilon c}^2}}{\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2} + \frac{1}{\sigma_\ell^2 + \sigma_{\epsilon c}^2}}.$$

Proof. Under the centralized regime, given the quadratic form of the objective function, the optimal policy for the central government is given by

$$a_c = E(\theta | \theta_c, \mathbf{s}'_\ell).$$

To find the expression of the conditional expectation above, we first notice the probability density function of the joint distribution of $(\theta, \theta_c, \theta_\ell, \mathbf{s}'_\ell)$ can be written as

$$\begin{aligned} f(\theta, \theta_c, \theta_\ell, \mathbf{s}'_\ell) &= f(\theta) f(\theta_c, \theta_\ell, \mathbf{s}'_\ell | \theta) = f(\theta) f(\theta_c | \theta) f(\theta_\ell, \mathbf{s}'_\ell | \theta) \\ &= f(\theta) f(\theta_c | \theta) f(\theta_\ell | \theta) f(\mathbf{s}'_\ell | \theta_\ell) = f(\theta) f(\theta_c | \theta) f(\theta_\ell | \theta) f(s'_{\ell 0} | \theta_\ell) f(s'_{\ell 1} | \theta_\ell), \end{aligned}$$

[39] For simplicity, we omit the prior of θ whenever we state the information set.

where the second to last equation stems from the fact that conditional on θ_ℓ, ϵ and ϵ_c are independent of θ. More explicitly, we have[40]

$$f(\theta,\theta_c,\theta_\ell,\mathbf{s}'_\ell) \sim \exp\left\{-\frac{1}{2}\left[\frac{\theta^2}{\sigma^2}+\frac{(\theta_c-\theta)^2}{\sigma_c^2}+\frac{(\theta_\ell-\theta)^2}{\sigma_\ell^2}+\frac{(s'_{\ell 0}-\theta_\ell)^2}{\sigma_\epsilon^2}+\frac{(s'_{\ell 1}-\theta_\ell)^2}{\tilde{\sigma}_{\epsilon c}^2}\right]\right\}$$

$$= \exp\left\{-\frac{1}{2}\left[\left(\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_\epsilon^2}+\frac{1}{\tilde{\sigma}_{\epsilon c}^2}\right)\theta_\ell^2 - 2\left(\frac{\theta}{\sigma_\ell^2}+\frac{s'_{\ell 0}}{\sigma_\epsilon^2}+\frac{s'_{\ell 1}}{\tilde{\sigma}_{\epsilon c}^2}\right)\theta_\ell + \frac{\theta^2}{\sigma^2}+\frac{(\theta_c-\theta)^2}{\sigma_c^2}+\frac{\theta^2}{\sigma_\ell^2}+\frac{{s'_{\ell 0}}^2}{\sigma_\epsilon^2}+\frac{{s'_{\ell 1}}^2}{\tilde{\sigma}_{\epsilon c}^2}\right]\right\}$$

$$= \exp\left\{-\frac{1}{2}\left[\left(\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_\epsilon^2}+\frac{1}{\tilde{\sigma}_{\epsilon c}^2}\right)\left(\theta_\ell - \frac{\frac{\theta}{\sigma_\ell^2}+\frac{s'_{\ell 0}}{\sigma_\epsilon^2}+\frac{s'_{\ell 1}}{\tilde{\sigma}_{\epsilon c}^2}}{\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_\epsilon^2}+\frac{1}{\tilde{\sigma}_{\epsilon c}^2}}\right)^2 + \frac{\theta^2}{\sigma^2}+\frac{(\theta_c-\theta)^2}{\sigma_c^2}+\frac{\theta^2}{\sigma_\ell^2}+\frac{{s'_{\ell 0}}^2}{\sigma_\epsilon^2}+\frac{{s'_{\ell 1}}^2}{\tilde{\sigma}_{\epsilon c}^2} - \frac{\left(\frac{\theta}{\sigma_\ell^2}+\frac{s'_{\ell 0}}{\sigma_\epsilon^2}+\frac{s'_{\ell 1}}{\tilde{\sigma}_{\epsilon c}^2}\right)^2}{\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_\epsilon^2}+\frac{1}{\tilde{\sigma}_{\epsilon c}^2}}\right]\right\}$$

Integrating out θ_ℓ and using $s'_\ell = \frac{s'_{\ell 0}/\sigma_\epsilon^2 + s'_{\ell 1}/\tilde{\sigma}_{\epsilon c}^2}{1/\sigma_\epsilon^2+1/\tilde{\sigma}_{\epsilon c}^2}$ and $\sigma_{\epsilon c}^2 = (\sigma_\epsilon^{-2}+\tilde{\sigma}_{\epsilon c}^{-2})^{-1}$, we obtain

$$f(\theta,\theta_c,\mathbf{s}'_\ell) \sim \exp\left\{-\frac{1}{2}\left[\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2+\sigma_{\epsilon c}^2}\right)\theta^2 - 2\left(\frac{\theta_c}{\sigma_c^2}+\frac{s'_\ell}{\sigma_\ell^2+\sigma_{\epsilon c}^2}\right)\theta\right]\right\}.$$

This leads to

$$f(\theta|\theta_c,\mathbf{s}'_\ell) \sim f(\theta,\theta_c,\mathbf{s}'_\ell) \sim \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2+\sigma_{\epsilon c}^2}\right)\left(\theta - \frac{\frac{\theta_c}{\sigma_c^2}+\frac{s'_\ell}{\sigma_\ell^2+\sigma_{\epsilon c}^2}}{\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2+\sigma_{\epsilon c}^2}}\right)^2\right\},$$

which yields the closed-form solution to $E(\theta|\theta_c,\mathbf{s}'_\ell)$. □

Lemma 13. *Under the decentralized regime, the optimal policy of the local government is given by*

$$a_\ell = (1-\gamma)E(\theta|\theta_\ell,\mathbf{s}'_c) + \gamma E(\theta_c|\theta_\ell,\mathbf{s}'_c) = (1-\gamma)\frac{\frac{\theta_\ell}{\sigma_\ell^2}+\frac{s'_c}{\sigma_c^2+\sigma_{\epsilon\ell}^2}}{\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_c^2+\sigma_{\epsilon\ell}^2}} + \gamma\frac{\frac{\theta_\ell}{\sigma_c^2+\sigma_\ell^2+\sigma_c^2\sigma_\ell^2/\sigma^2}+\frac{s'_c}{\sigma_{\epsilon\ell}^2}}{\frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}+\frac{1}{\sigma_{\epsilon\ell}^2}}.$$

Proof. Under the decentralized regime, given the quadratic form of the objective function, the optimal policy for the local government is given by

$$a_\ell = (1-\gamma)E(\theta|\theta_\ell,\mathbf{s}'_c) + \gamma E(\theta_c|\theta_\ell,\mathbf{s}'_c).$$

$E(\theta|\theta_\ell,\mathbf{s}'_c)$ can be obtained similarly as $E(\theta|\theta_c,\mathbf{s}'_\ell)$. For $E(\theta_c|\theta_\ell,\mathbf{s}'_c)$, we have

$$f(\theta,\theta_c,\theta_\ell,\mathbf{s}'_c) = f(\theta)f(\theta_c|\theta)f(\theta_\ell|\theta)f(s'_{c0}|\theta_c)f(s'_{c1}|\theta_c)$$
$$\sim \exp\left\{-\frac{1}{2}\left[\frac{\theta^2}{\sigma^2}+\frac{(\theta_c-\theta)^2}{\sigma_c^2}+\frac{(\theta_\ell-\theta)^2}{\sigma_\ell^2}+\frac{(s'_{c0}-\theta_c)^2}{\sigma_\epsilon^2}+\frac{(s'_{c1}-\theta_c)^2}{\tilde{\sigma}_{\epsilon\ell}^2}\right]\right\}$$

[40] Alternatively, we know $f(\theta,\theta_c,\theta_\ell,\mathbf{s}'_\ell) \sim \exp\left(-\frac{1}{2}(\theta,\theta_c,\theta_\ell,\mathbf{s}'_\ell)\Sigma^{-1}_{\theta,\theta_c,\theta_\ell,\mathbf{s}'_\ell}(\theta,\theta_c,\theta_\ell,\mathbf{s}'_\ell)^T\right)$ with

$$\Sigma^{-1}_{\theta,\theta_c,\theta_\ell,\mathbf{s}'_\ell} = \begin{pmatrix} \sigma^2 & \sigma^2 & \sigma^2 & \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2+\sigma_c^2 & \sigma^2 & \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 & \sigma^2+\sigma_\ell^2 & \sigma^2+\sigma_\ell^2 & \sigma^2+\sigma_\ell^2 \\ \sigma^2 & \sigma^2 & \sigma^2+\sigma_\ell^2 & \sigma^2+\sigma_\ell^2+\sigma_\epsilon^2 & \sigma^2+\sigma_\ell^2 \\ \sigma^2 & \sigma^2 & \sigma^2+\sigma_\ell^2 & \sigma^2+\sigma_\ell^2 & \sigma^2+\sigma_\ell^2+\tilde{\sigma}_{\epsilon c}^2 \end{pmatrix}^{-1}$$
$$= \begin{pmatrix} \frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2} & -\frac{1}{\sigma_c^2} & -\frac{1}{\sigma_\ell^2} & 0 & 0 \\ -\frac{1}{\sigma_c^2} & \frac{1}{\sigma_c^2} & 0 & 0 & 0 \\ -\frac{1}{\sigma_\ell^2} & 0 & \frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_\epsilon^2}+\frac{1}{\tilde{\sigma}_{\epsilon c}^2} & -\frac{1}{\sigma_\epsilon^2} & -\frac{1}{\tilde{\sigma}_{\epsilon c}^2} \\ 0 & 0 & -\frac{1}{\sigma_\epsilon^2} & \frac{1}{\sigma_\epsilon^2} & 0 \\ 0 & 0 & -\frac{1}{\tilde{\sigma}_{\epsilon c}^2} & 0 & \frac{1}{\tilde{\sigma}_{\epsilon c}^2} \end{pmatrix}.$$

$$= \exp\left\{-\frac{1}{2}\left[\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2}\right)\theta^2 - 2\left(\frac{\theta_c}{\sigma_c^2}+\frac{\theta_\ell}{\sigma_\ell^2}\right)\theta + \frac{\theta_c^2}{\sigma_c^2}+\frac{\theta_\ell^2}{\sigma_\ell^2}+\frac{(s'_{c0}-\theta_c)^2}{\sigma_\epsilon^2}+\frac{(s'_{c1}-\theta_c)^2}{\tilde{\sigma}_{\epsilon\ell}^2}\right]\right\}$$

$$= \exp\left\{-\frac{1}{2}\left[\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2}\right)\left(\theta - \frac{\frac{\theta_c}{\sigma_c^2}+\frac{\theta_\ell}{\sigma_\ell^2}}{\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2}}\right)^2\right.\right.$$

$$\left.\left. + \frac{\theta_c^2}{\sigma_c^2}+\frac{\theta_\ell^2}{\sigma_\ell^2}+\frac{(s'_{c0}-\theta_c)^2}{\sigma_\epsilon^2}+\frac{(s'_{c1}-\theta_c)^2}{\tilde{\sigma}_{\epsilon\ell}^2} - \frac{\left(\frac{\theta_c}{\sigma_c^2}+\frac{\theta_\ell}{\sigma_\ell^2}\right)^2}{\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2}}\right]\right\}.$$

Integrating out θ, we obtain

$$f(\theta_c,\theta_\ell,\mathbf{s}'_c) \sim \exp\left\{-\frac{1}{2}\left[\frac{\theta_c^2}{\sigma_c^2}+\frac{\theta_\ell^2}{\sigma_\ell^2}+\frac{(s'_{c0}-\theta_c)^2}{\sigma_\epsilon^2}+\frac{(s'_{c1}-\theta_c)^2}{\tilde{\sigma}_{\epsilon\ell}^2} - \frac{(\theta_c/\sigma_c^2+\theta_\ell/\sigma_\ell^2)^2}{1/\sigma^2+1/\sigma_c^2+1/\sigma_\ell^2}\right]\right\}.$$

Using $s'_c = \frac{s'_{c0}/\sigma_\epsilon^2 + s'_{c1}/\tilde{\sigma}_{\epsilon\ell}^2}{1/\sigma_\epsilon^2+1/\tilde{\sigma}_{\epsilon\ell}^2}$ and $\sigma_{\epsilon\ell}^2 = (\sigma_\epsilon^{-2}+\tilde{\sigma}_{\epsilon\ell}^{-2})^{-1}$, we obtain

$$f(\theta_c|\theta_\ell,\mathbf{s}'_c) \sim f(\theta_c,\theta_\ell,\mathbf{s}'_c) \sim \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma_{\epsilon\ell}^2}+\frac{1}{\sigma_c^2+\frac{1}{1/\sigma^2+1/\sigma_\ell^2}}\right)\left(\theta_c - \frac{\frac{s'_c}{\sigma_{\epsilon\ell}^2}+\frac{\theta_\ell}{\sigma_c^2+\sigma_\ell^2+\sigma_c^2\sigma_\ell^2/\sigma^2}}{\frac{1}{\sigma_{\epsilon\ell}^2}+\frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}}\right)^2\right\},$$

which implies the closed form solution to $E(\theta_c|\theta_\ell,\mathbf{s}'_c)$. Plugging the expressions of $E(\theta|\theta_\ell,\mathbf{s}'_c)$ and $E(\theta_c|\theta_\ell,\mathbf{s}'_c)$ into the equation for a_ℓ, we obtain the desired conclusion. □

A.5. Proof of lemma 3

Proof. Since a_i is a linear combination of all the signals θ_ℓ, θ_c, s'_ℓ, and s'_c, we can write $(a_i-\theta)$ as a linear combination of six independent normal random variables with mean zero,

$$a_i - \theta = m_{i1}(\theta_\ell-\theta) + m_{i2}(\theta_c-\theta) + m_{i3}\epsilon + m_{i4}\epsilon_c + m_{i5}\epsilon_\ell + (m_{i1}+m_{i2}-1)\theta.$$

Then according to the central moments of a normal distribution, we have

$$E(a_i-\theta)^2 = m_{i1}^2\sigma_\ell^2 + m_{i2}^2\sigma_c^2 + m_{i3}^2\sigma_\epsilon^2 + m_{i4}^2\sigma_{\epsilon c}^2 + m_{i5}^2\sigma_{\epsilon\ell}^2 + (1-m_{i1}-m_{i2})^2\sigma^2,$$

$$E(a_i-\theta)^4 = 3\left(m_{i1}^2\sigma_\ell^2 + m_{i2}^2\sigma_c^2 + m_{i3}^2\sigma_\epsilon^2 + m_{i4}^2\sigma_{\epsilon c}^2 + m_{i5}^2\sigma_{\epsilon\ell}^2 + (1-m_{i1}-m_{i2})^2\sigma^2\right)^2 = 3\left(E(a_i-\theta)^2\right)^2.$$

Therefore, by definition,

$$E(Y) \equiv Y^* - E(a_i-\theta)^2 = Y^* - \left(m_{i1}^2\sigma_\ell^2 + m_{i2}^2\sigma_c^2 + m_{i3}^2\sigma_\epsilon^2 + m_{i4}^2\sigma_{\epsilon c}^2 + m_{i5}^2\sigma_{\epsilon\ell}^2 + (1-m_{i1}-m_{i2})^2\sigma^2\right),$$

$$Var(Y) = E(a_i-\theta)^4 - \left(E(a_i-\theta)^2\right)^2 = 2\left(E(a_i-\theta)^2\right)^2 = 2(Y^*-E(Y))^2.$$

□

A.6. The equilibrium under the centralized regime

According to Lemma 12 and Eq. 6, the resource allocation problem for the central government can be simplified to

$$\max_{\sigma_c^2,\sigma_{\epsilon c}^2} \frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2+\sigma_{\epsilon c}^2},$$

subject to Constraint 3 and $\sigma_{\epsilon c}^2 \le \sigma_\epsilon^2$. By inspection, the constraint has to be binding. Given the binding constraint, we can further simplify the constrained optimization problem to

$$\max_{\sigma_c^2,\sigma_{\epsilon c}^2} \left(1-\frac{1}{\sigma_\ell^4 K_c}\right)\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_c^2}\right)$$

subject to Constraint 3. Since $K_c > 1/\sigma_\ell^4$, the maximum is attained when σ_c^2 attains its minimum under the constraint, or equivalently, $\sigma_{\epsilon c}^2 = \sigma_\epsilon^2$. Therefore, we obtain the following characterization of the central government's strategy under the centralized regime.

Lemma 14. *Under the centralized regime, for any given σ_ℓ^2, the central government completely devotes itself to the direct information acquisition with $\sigma_{\epsilon c}^2 = \sigma_\epsilon^2$.*

Our next lemma suggests that the career concern of the local bureaucrats does not distort the economic outcome under the centralized regime, as long as the central government is benevolent.

Lemma 15. *Under the centralized regime, for any γ, the local government always spends all of its attention resource on information acquisition (Constraint 2 is binding), which leads to*

$$\sigma_\ell^2 = \sigma^2/(2^{2\kappa_\ell} - 1).$$

Proof. According to Lemma 14, under the centralized regime, $\sigma_{\epsilon c}^2 = \sigma_\epsilon^2$, so we can simplify a_c as

$$a_c = \frac{\frac{\theta_c}{\sigma_c^2} + \frac{s_\ell'}{\sigma_\ell^2+\sigma_\epsilon^2}}{\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2} + \frac{1}{\sigma_\ell^2+\sigma_\epsilon^2}},$$

Using the backward induction, the local government solves its decision problem

$$\max_{\sigma_\ell^2} (1-\gamma)\big(Y^* - E(a_c-\theta)^2\big) - \gamma E(\theta_\ell - a_c)^2,$$

or equivalently,

$$\min_{\sigma_\ell^2} (1-\gamma)E(a_c-\theta)^2 + \gamma E(\theta_\ell - a_c)^2 \equiv F(\sigma_\ell^2),$$

subject to Constraint 2. Plugging in the expression of a_c, we can write the objective function explicitly as

$$\begin{aligned} F(\sigma_\ell^2) &= (1-\gamma)E(a_c-\theta)^2 + \gamma E(\theta_\ell - a_c)^2 \\ &= E(a_c-\theta)^2 + 2\gamma E(\theta_\ell-\theta)(\theta-a_c) + \gamma E(\theta_\ell-\theta)^2 \\ &= \frac{1}{\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2+\sigma_\epsilon^2}} - \frac{\frac{2\gamma\sigma_\ell^2}{\sigma_\ell^2+\sigma_\epsilon^2}}{\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2+\sigma_\epsilon^2}} + \gamma\sigma_\ell^2 \end{aligned}$$

Since we know $\sigma_{\epsilon c}^2 = \sigma_\epsilon^2$, Constraint 3 gives us

$$\left(\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_\ell^2}\right)\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_c^2}\right) = K_c(\sigma_\ell^2) = 2^{2\kappa_c}\frac{\sigma^2+\sigma_\ell^2+\sigma_\epsilon^2}{\sigma^2\sigma_\ell^2\sigma_\epsilon^2} + \frac{1}{\sigma_\ell^4}.$$

The difficulty of this optimization problem can be seen from the equation above. Despite the fact that the central government always devotes to direct information acquisition, the resulting σ_c^2 is still a function of σ_ℓ^2 due to the information flow constraint. Using the binding constraint, then the objective function can be rewritten as

$$\begin{aligned} F(\sigma_\ell^2) &= \frac{1}{\frac{K_c}{1/\sigma_\epsilon^2+1/\sigma_\ell^2} - \frac{1}{\sigma_\ell^2} + \frac{1}{\sigma_\ell^2+\sigma_\epsilon^2}} - \frac{\frac{2\gamma\sigma_\ell^2}{\sigma_\ell^2+\sigma_\epsilon^2}}{\frac{K_c}{1/\sigma_\epsilon^2+1/\sigma_\ell^2} - \frac{1}{\sigma_\ell^2} + \frac{1}{\sigma_\ell^2+\sigma_\epsilon^2}} + \gamma\sigma_\ell^2 \\ &= \frac{(1-2\gamma)\sigma_\ell^4 + \sigma_\epsilon^2\sigma_\ell^2}{K_c\sigma_\epsilon^2\sigma_\ell^4 - \sigma_\epsilon^2} + \gamma\sigma_\ell^2 = \frac{(1-2\gamma)\sigma_\ell^4 + \sigma_\epsilon^2\sigma_\ell^2}{2^{2\kappa_c}\sigma_\epsilon^2\big((1/\sigma^2+1/\sigma_\epsilon^2)\sigma_\ell^2 + \sigma_\ell^4/(\sigma^2\sigma_\epsilon^2)\big)} + \gamma\sigma_\ell^2. \end{aligned}$$

$$\begin{aligned} F'(\sigma_\ell^2) &= \frac{[(2-4\gamma)\sigma_\ell^2+\sigma_\epsilon^2]\left[\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\epsilon^2}\right)\sigma_\ell^2 + \frac{\sigma_\ell^4}{\sigma^2\sigma_\epsilon^2}\right] - [(1-2\gamma)\sigma_\ell^4+\sigma_\epsilon^2\sigma_\ell^2]\left[\frac{1}{\sigma^2}+\frac{1}{\sigma_\epsilon^2}+\frac{2\sigma_\ell^2}{\sigma^2\sigma_\epsilon^2}\right]}{2^{2\kappa_c}\sigma_\epsilon^2\left[\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\epsilon^2}\right)\sigma_\ell^2+\frac{\sigma_\ell^4}{\sigma^2\sigma_\epsilon^2}\right]^2} + \gamma \\ &= \frac{\left[\frac{1-2\gamma}{\sigma_\epsilon^2} - \frac{2\gamma}{\sigma^2}\right]\sigma_\ell^4}{2^{2\kappa_c}\sigma_\epsilon^2\left[\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\epsilon^2}\right)\sigma_\ell^2+\frac{\sigma_\ell^4}{\sigma^2\sigma_\epsilon^2}\right]^2} + \gamma \\ &\geq \frac{2^{2\kappa_c}\gamma\sigma_\epsilon^2\left[2\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\epsilon^2}\right)\frac{\sigma_\ell^6}{\sigma^2\sigma_\epsilon^2}+\frac{\sigma_\ell^8}{\sigma^4\sigma_\epsilon^4}\right] + \left[\frac{1}{\sigma_\epsilon^2}+\gamma\sigma_\epsilon^2\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\epsilon^2}\right)^2 - \frac{2\gamma}{\sigma_\epsilon^2} - \frac{2\gamma}{\sigma^2}\right]\sigma_\ell^4}{2^{2\kappa_c}\sigma_\epsilon^2\left[\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\epsilon^2}\right)\sigma_\ell^2+\frac{\sigma_\ell^4}{\sigma^2\sigma_\epsilon^2}\right]^2} \\ &= \frac{2^{2\kappa_c}\gamma\sigma_\epsilon^2\left[2\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\epsilon^2}\right)\frac{\sigma_\ell^6}{\sigma^2\sigma_\epsilon^2}+\frac{\sigma_\ell^8}{\sigma^4\sigma_\epsilon^4}\right] + \left[\frac{1-\gamma}{\sigma_\epsilon^2}+\frac{\gamma\sigma_\epsilon^2}{\sigma^4}\right]\sigma_\ell^4}{2^{2\kappa_c}\sigma_\epsilon^2\left[\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\epsilon^2}\right)\sigma_\ell^2+\frac{\sigma_\ell^4}{\sigma^2\sigma_\epsilon^2}\right]^2} > 0 \end{aligned}$$

where the first inequality follows from $\kappa_c > 0$ and the last inequality follows from $\gamma \in [0, 1]$. Therefore, the objective function is minimized if σ_ℓ^2 attains its minimum, $\sigma^2/(2^{2\kappa_\ell} - 1)$. □

Collecting the results from Lemmas 12, 14, and 15, we have the equilibrium characterization for the centralized regime.

A.7. Proof of lemma 4

To gain some intuition, we first show the optimal resource allocation for the local government when $\gamma = 0$ or $\gamma = 1$ under the decentralized regime.

In the absence of loyalty concern ($\gamma = 0$), we know from Eq. 7,

$$E(a_\ell - \theta)^2 = \left(\frac{1}{\sigma^2} + \frac{1}{\sigma_\ell^2} + \frac{1}{\sigma_c^2 + \sigma_{\epsilon\ell}^2}\right)^{-1} = Var(\theta|\theta_\ell, \mathbf{s}_c').$$

The resource allocation problem of the local government can now be written as

$$\max_{\sigma_\ell^2, \sigma_{\epsilon\ell}^2} \frac{1}{\sigma^2} + \frac{1}{\sigma_\ell^2} + \frac{1}{\sigma_c^2 + \sigma_{\epsilon\ell}^2}$$

subject to Constraint 5 and $\sigma_{\epsilon\ell}^2 \le \sigma_\epsilon^2$. This is symmetric to the central government's decision problem under the centralized regime. Following the same intuition, we obtain a simple characterization of the local government's strategy under the decentralized regime with no loyalty concern.

Lemma 16. *Under the decentralized regime with $\gamma = 0$, for any given σ_c^2, the local government completely devotes itself to the direct information acquisition with $\sigma_{\epsilon\ell}^2 = \sigma_\epsilon^2$.*

If the local government is purely loyalty driven ($\gamma = 1$), we know from Eq. 8 that

$$E(a_\ell - \theta_c)^2 = \left(\frac{1}{\sigma_{\epsilon\ell}^2} + \frac{1}{\sigma_c^2 + (1/\sigma^2 + 1/\sigma_\ell^2)^{-1}}\right)^{-1} = Var(\theta_c|\theta_\ell, \mathbf{s}_c').$$

The resource allocation of the local government can now be written as

$$\max_{\sigma_\ell^2, \sigma_{\epsilon\ell}^2} \frac{1}{\sigma_{\epsilon\ell}^2} + \frac{1}{\sigma_c^2 + (1/\sigma^2 + 1/\sigma_\ell^2)^{-1}}$$

subject to Constraint 5 and $\sigma_{\epsilon\ell}^2 \le \sigma_\epsilon^2$. Similarly, by inspection, the constraint must be binding. Given the binding constraint, we can further simplify the constrained optimization problem to

$$\max_{\sigma_\ell^2, \sigma_{\epsilon\ell}^2} \left(1 - \frac{1}{\sigma_c^4 K_\ell}\right)\left(\frac{1}{\sigma_c^2} + \frac{1}{\sigma_{\epsilon\ell}^2}\right)$$

subject to Constraint 5 and $\sigma_{\epsilon\ell}^2 \le \sigma_\epsilon^2$. Since $K_\ell > 1/\sigma_c^4$, the maximum is attained when $\sigma_{\epsilon\ell}^2$ attains its minimum under the constraint, or equivalently, $\sigma_\ell^2 = \infty$.

Lemma 17. *Under the decentralized regime with $\gamma = 1$, for any given σ_c^2, the local government completely devotes to the inter-governmental communication with $\sigma_\ell^2 = \infty$.*

We now turn to the proof of Lemma 4 that nests the two lemmas above as special cases.

Proof. According to Lemma 13, we can rewrite the constrained optimization problem as

$$\min_{\sigma_\ell^2, \sigma_{\epsilon\ell}^2} E\left\{(1-\gamma)E[(a_\ell - \theta)^2|\theta_\ell, \mathbf{s}_c'] + \gamma E[(a_\ell - \theta_c)^2|\theta_\ell, \mathbf{s}_c']\right\} \equiv F(\sigma_\ell^2, \sigma_{\epsilon\ell}^2),$$

with $a_\ell = (1-\gamma)E(\theta|\theta_\ell, \mathbf{s}_c') + \gamma E(\theta_c|\theta_\ell, \mathbf{s}_c')$, subject to Constraint 5 and $\sigma_{\epsilon\ell}^2 \le \sigma_\epsilon^2$. Given the expression of a_ℓ, we have

$$\begin{aligned}
F(\sigma_\ell^2, \sigma_{\epsilon\ell}^2) &= E\Big\{(1-\gamma)E[((1-\gamma)(E(\theta|\theta_\ell, \mathbf{s}_c') - \theta) + \gamma(E(\theta_c|\theta_\ell, \mathbf{s}_c') - \theta))^2|\theta_\ell, \mathbf{s}_c'] \\
&\quad + \gamma E[((1-\gamma)(E(\theta|\theta_\ell, \mathbf{s}_c') - \theta_c) + \gamma(E(\theta_c|\theta_\ell, \mathbf{s}_c') - \theta_c))^2|\theta_\ell, \mathbf{s}_c']\Big\} \\
&= E\Big\{(1-\gamma)[(1-\gamma)^2 Var(\theta|\theta_\ell, \mathbf{s}_c') + 2\gamma(1-\gamma)Var(\theta|\theta_\ell, \mathbf{s}_c') \\
&\quad + \gamma^2 E((E(\theta_c|\theta_\ell, \mathbf{s}_c') - \theta_c + \theta_c - \theta)^2|\theta_\ell, \mathbf{s}_c')] \\
&\quad + \gamma[\gamma^2 Var(\theta_c|\theta_\ell, \mathbf{s}_c') + 2\gamma(1-\gamma)Var(\theta_c|\theta_\ell, \mathbf{s}_c') \\
&\quad + (1-\gamma)^2 E((E(\theta|\theta_\ell, \mathbf{s}_c') - \theta + \theta - \theta_c)^2|\theta_\ell, \mathbf{s}_c')]\Big\} \\
&= (1-\gamma)^2(1+\gamma)Var(\theta|\theta_\ell, \mathbf{s}_c') + \gamma^2(2-\gamma)Var(\theta_c|\theta_\ell, \mathbf{s}_c') \\
&\quad + (1-\gamma)\gamma^2[Var(\theta_c|\theta_\ell, \mathbf{s}_c') + 2E((E(\theta_c|\theta_\ell, \mathbf{s}_c') - \theta_c)(\theta_c - \theta)) + \sigma_c^2] \\
&\quad + (1-\gamma)^2\gamma[Var(\theta|\theta_\ell, \mathbf{s}_c') + 2E((E(\theta|\theta_\ell, \mathbf{s}_c') - \theta)(\theta - \theta_c)) + \sigma_c^2] \\
&= (1-\gamma)^2(1+2\gamma)Var(\theta|\theta_\ell, \mathbf{s}_c') + \gamma^2(3-2\gamma)Var(\theta_c|\theta_\ell, \mathbf{s}_c') + (1-\gamma)\gamma\sigma_c^2 \\
&\quad - 2\sigma_c^2\left\{(1-\gamma)\gamma^2 \frac{\frac{1}{\sigma_c^2 + (1/\sigma^2 + 1/\sigma_\ell^2)^{-1}}}{\frac{1}{\sigma_{\epsilon\ell}^2} + \frac{1}{\sigma_c^2 + (1/\sigma^2 + 1/\sigma_\ell^2)^{-1}}} + (1-\gamma)^2\gamma \frac{\frac{1}{\sigma_c^2 + \sigma_{\epsilon\ell}^2}}{\frac{1}{\sigma^2} + \frac{1}{\sigma_\ell^2} + \frac{1}{\sigma_c^2 + \sigma_{\epsilon\ell}^2}}\right\}
\end{aligned} \tag{13}$$

To see that Constraint 5 has to be binding, we further write the objective function as

$$F(\sigma_\ell^2,\sigma_{\epsilon\ell}^2) = \frac{(1-\gamma)^2}{\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_c^2+\sigma_{\epsilon\ell}^2}} + \frac{\gamma^2}{\frac{1}{\sigma_{\epsilon\ell}^2}+\frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}} + (1-\gamma)\gamma\sigma_c^2$$
$$+ \frac{\frac{2\gamma(1-\gamma)^2\sigma_{\epsilon\ell}^2}{\sigma_c^2+\sigma_{\epsilon\ell}^2}}{\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_c^2+\sigma_{\epsilon\ell}^2}} + \frac{\frac{2(1-\gamma)\gamma^2(1/\sigma^2+1/\sigma_\ell^2)^{-1}}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}}{\frac{1}{\sigma_{\epsilon\ell}^2}+\frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}},$$

which strictly increases with σ_ℓ^2 or $\sigma_{\epsilon\ell}^2$. Since Constraint 5 is binding, from Eq. 13, the objective function can be further simplified

$$F(\sigma_\ell^2,\sigma_{\epsilon\ell}^2) = \frac{1}{K_\ell-1/\sigma_c^4}\left[(1-\gamma)^2(1+2\gamma)\left(\frac{1}{\sigma_{\epsilon\ell}^2}+\frac{1}{\sigma_c^2}\right)+\gamma^2(3-2\gamma)\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_c^2}\right)\right]$$
$$+(1-\gamma)\gamma\sigma_c^2 - 2(1-\gamma)\gamma^2\frac{1/\sigma^2+1/\sigma_\ell^2}{K_\ell-1/\sigma_c^4} - 2(1-\gamma)^2\gamma\frac{1/\sigma_{\epsilon\ell}^2}{K_\ell-1/\sigma_c^4}$$
$$= \frac{1}{K_\ell-1/\sigma_c^4}\left[(1-\gamma)^2K_\ell\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_c^2}\right)^{-1}+\gamma^2\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_c^2}\right)\right]$$
$$+(1-\gamma)\gamma\sigma_c^2 - \frac{2(1-\gamma)\gamma}{\sigma_c^2(K_\ell-1/\sigma_c^4)}.$$

Since $K_\ell > 1/\sigma_c^4$ and K_ℓ is constant with respect to σ_ℓ^2 and $\sigma_{\epsilon\ell}^2$, we can simplify the optimization problem as

$$\min_x \gamma^2 x + (1-\gamma)^2 K_\ell/x \equiv H(x)$$

with $x \equiv \frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_c^2}$ taking value from $[\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}, \frac{K_\ell}{1/\sigma_\epsilon^2+1/\sigma_c^2}]$. Let the minimizer be x^*. It is easy to see that when $\gamma=0$, $x^* = \frac{K_\ell}{1/\sigma_\epsilon^2+1/\sigma_c^2}$ with $\sigma_{\epsilon\ell}^2=\sigma_\epsilon^2$, as in Lemma 16, and when $\gamma=1$, $x^*=\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}$ with $\sigma_\ell^2=\infty$, as in Lemma 17. Moreover, we have

$$H'(x) = \gamma^2-(1-\gamma)^2K_\ell/x^2 \equiv G(\gamma;x).$$

Since $\frac{dG}{d\gamma}(\gamma;x) = 2\gamma+2(1-\gamma)K_\ell/x^2 > 0$ for any $\gamma\in[0,1]$ and $G(0;x)<0$ and $G(1;x)>1$, there exists a unique $\gamma\in(0,1)$ such that $G(\gamma;x)=0$ for any given $x>0$. Define $\underline{\gamma}$ such that $G(\underline{\gamma};x)=0$ for $x=\frac{K_\ell}{1/\sigma_\epsilon^2+1/\sigma_c^2}$ and $\bar{\gamma}$ such that $G(\bar{\gamma};x)=0$ for $x=\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}$. Since $\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2} < \frac{K_\ell}{1/\sigma_\epsilon^2+1/\sigma_c^2}$, by construction, we have

$$\left(\frac{\bar{\gamma}}{1-\bar{\gamma}}\right)^2 = K_\ell\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}\right)^{-2} > K_\ell\left(\frac{K_\ell}{1/\sigma_\epsilon^2+1/\sigma_c^2}\right)^{-2} = \left(\frac{\underline{\gamma}}{1-\underline{\gamma}}\right)^2,$$

which implies $\underline{\gamma}<\bar{\gamma}$. Since $\frac{dG}{d\gamma}(\gamma;x)>0$, for any $\gamma\le\underline{\gamma}$ and $x<\frac{K_\ell}{1/\sigma_\epsilon^2+1/\sigma_c^2}$, we must have

$$H'(x) = G(\gamma;x) \le G(\underline{\gamma};x) = \underline{\gamma}^2-(1-\underline{\gamma})^2K_\ell/x^2 < \underline{\gamma}^2-(1-\underline{\gamma})^2\frac{K_\ell}{(K_\ell/(1/\sigma_\epsilon^2+1/\sigma_c^2))^2} = 0.$$

Therefore, for $\gamma\le\underline{\gamma}$, $x^*=K_\ell/(1/\sigma_\epsilon^2+1/\sigma_c^2)$ with $\sigma_{\epsilon\ell}^2=\sigma_\epsilon^2$: The local government specializes in direct information acquisition provided that γ is sufficiently small.

Similarly, for any $\gamma\ge\bar{\gamma}$ and $x>\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}$,

$$H'(x) = G(\gamma;x) \ge G(\bar{\gamma};x) = \bar{\gamma}^2-(1-\bar{\gamma})^2K_\ell/x^2 > \bar{\gamma}^2-(1-\bar{\gamma})^2\frac{K_\ell}{(1/\sigma^2+1/\sigma_c^2)^2} = 0.$$

Therefore, for $\gamma\ge\bar{\gamma}$, $x^*=\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}$ with $\sigma_\ell^2=\infty$: The local government devotes all its attention budget to intergovernmental communication provided that γ is sufficiently large.

For $\gamma\in(\underline{\gamma},\bar{\gamma})$, we have

$$H'\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}\right) = G\left(\gamma;\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}\right) < G\left(\bar{\gamma};\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}\right) = 0$$
$$H'\left(\frac{K_\ell}{1/\sigma^2+1/\sigma_c^2}\right) = G\left(\gamma;\frac{K_\ell}{1/\sigma^2+1/\sigma_c^2}\right) > G\left(\underline{\gamma};\frac{K_\ell}{1/\sigma^2+1/\sigma_c^2}\right) = 0.$$

Further, $H''(x)>0$ for any $x\in[\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}, \frac{K_\ell}{1/\sigma_\epsilon^2+1/\sigma_c^2}]$. Therefore, there exists a unique $x^*\in(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}, \frac{K_\ell}{1/\sigma_\epsilon^2+1/\sigma_c^2})$ with $\sigma_\ell^2<\infty$ and $\sigma_{\epsilon\ell}^2<\sigma_\epsilon^2$: When γ is in the intermediate range, the government allocates its attention budget to both dimensions.

Furthermore, if $\gamma \in (\underline{\gamma}, \bar{\gamma})$, we have

$$H'(x^*) = 0 \Leftrightarrow x^* = K_\ell^{1/2}(1-\gamma)/\gamma.$$

Clearly, x^* strictly decreases with γ for $\gamma \in (\underline{\gamma}, \bar{\gamma})$, or equivalently, σ_ℓ^2 strictly increases with γ. Thus, we have obtained the desired conclusion. □

A.8. Proof of lemma 5

Proof. Following a similar derivation of Eq. 13 as in the proof of Lemma 4, we have

$$\begin{aligned} E(a_\ell-\theta)^2 &= E\Big\{E[((1-\gamma)(E(\theta|\theta_\ell,\mathbf{s}_c')-\theta)+\gamma(E(\theta_c|\theta_\ell,\mathbf{s}_c')-\theta))^2|\theta_\ell,\mathbf{s}_c']\Big\} \\ &= (1-\gamma)^2Var(\theta|\theta_\ell,\mathbf{s}_c')+2\gamma(1-\gamma)Var(\theta|\theta_\ell,\mathbf{s}_c') \\ &\quad +\gamma^2E\{E[(E(\theta_c|\theta_\ell,\mathbf{s}_c')-\theta_c+\theta_c-\theta)^2|\theta_\ell,\mathbf{s}_c']\} \\ &= (1-\gamma^2)Var(\theta|\theta_\ell,\mathbf{s}_c')+\gamma^2Var(\theta_c|\theta_\ell,\mathbf{s}_c')+\gamma^2\sigma_c^2-2\gamma^2\sigma_c^2\frac{\frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}}{\frac{1}{\sigma_{\epsilon\ell}^2}+\frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}} \\ &= \frac{1}{K_\ell-1/\sigma_c^4}\left[\frac{(1-\gamma^2)K_\ell}{\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_c^2}}+\gamma^2\left(\frac{1}{\sigma_c^2}-\frac{1}{\sigma^2}-\frac{1}{\sigma_\ell^2}\right)\right]+\gamma^2\sigma_c^2 \\ &= \frac{1}{K_\ell-1/\sigma_c^4}\left[\frac{(1-\gamma^2)K_\ell}{\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_c^2}}+K_\ell\gamma^2\sigma_c^2-\gamma^2\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}\right)\right]. \end{aligned} \tag{14}$$

where the second to last equality follows from the fact that Constraint 5 is binding. Then the optimization problem of the central government can be rewritten as

$$\min_{\sigma_c^2} E(a_\ell-\theta)^2 = \frac{1}{K_\ell-1/\sigma_c^4}\left[\frac{(1-\gamma^2)K_\ell}{\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_c^2}}+K_\ell\gamma^2\sigma_c^2-\gamma^2\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}\right)\right] \equiv F(\sigma_c^2)$$

subject to Constraint 4, where it should be emphasized that both σ_ℓ^2 and K_ℓ are functions of σ_c^2.

According to Lemma 4, $\underline{\gamma}$ and $\bar{\gamma}$ are, by construction, continuous functions of σ_c^2. For any given σ_c^2, we can divide the [0,1] interval for γ into three regions: $[0,\underline{\gamma})$, $(\underline{\gamma},\bar{\gamma})$, and $(\bar{\gamma},1]$. Since $\underline{\gamma}$ and $\bar{\gamma}$ are continuous in σ_c^2, for a given γ that is in any of three regions, a small change of σ_c^2 will not change the region that a given γ belongs to.

We arbitrarily pick a σ_c^2 subject to Constraint 4 and consider four possible cases: (1) $\gamma < \underline{\gamma}$; (2) $\underline{\gamma} < \gamma < \bar{\gamma}$; (3) $\gamma > \bar{\gamma}$; (4) $\gamma = \bar{\gamma}$ or $\gamma = \underline{\gamma}$.

Case (1): $\gamma < \underline{\gamma}$.

According to Lemma 4, we have $\sigma_{\epsilon\ell}^2 = \sigma_\epsilon^2$, which implies that Constraint 5 can be rewritten as

$$\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_c^2} = K_\ell\left(\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_c^2}\right)^{-1}.$$

The objective function of the central government can then be written as

$$\begin{aligned} F(\sigma_c^2) &= \frac{1}{K_\ell-1/\sigma_c^4}\left[(1-\gamma^2)\left(\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_c^2}\right)+K_\ell\gamma^2\sigma_c^2-\gamma^2\left(K_\ell\left(\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_c^2}\right)^{-1}-\frac{1}{\sigma_c^2}\right)\right] \\ &= \frac{1}{K_\ell-1/\sigma_c^4}\left[(1-\gamma^2)\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_c^2}+\frac{\frac{\gamma^2K_\ell\sigma_c^2}{\sigma_\epsilon^2}}{\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_c^2}}\right] \\ &= \frac{(1-\gamma^2)\sigma^2\sigma_c^2+\sigma^2\sigma_\epsilon^2+\frac{\gamma^2(2^{2\kappa_\ell}(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)\sigma_c^4+\sigma^2\sigma_\epsilon^2\sigma_c^2)}{\sigma_c^2+\sigma_\epsilon^2}}{2^{2\kappa_\ell}(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)} \\ &= \frac{[(1-\gamma^2)\sigma^2\sigma_c^2+\sigma^2\sigma_\epsilon^2](\sigma_c^2+\sigma_\epsilon^2)+\gamma^2(2^{2\kappa_\ell}(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)\sigma_c^4+\sigma^2\sigma_\epsilon^2\sigma_c^2)}{2^{2\kappa_\ell}(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)(\sigma_c^2+\sigma_\epsilon^2)} \end{aligned}$$

where the second to last inequality follows from $K_\ell = 2^{2\kappa_\ell}(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)/(\sigma_c^2\sigma^2\sigma_\epsilon^2)+1/\sigma_c^4$. Since $\gamma < \underline{\gamma}$ continues to hold for a small change of σ_c^2, the objective function is differentiable and its first derivative is given by

$$F'(\sigma_c^2) = \frac{G_1(\sigma_c^2)}{2^{2\kappa_\ell}(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)^2(\sigma_c^2+\sigma_\epsilon^2)^2}$$

with the numerator $G_1(\sigma_c^2)$ given by

$$\begin{aligned}
G_1(\sigma_c^2) =\ & \Big\{2(1-\gamma^2)\sigma^2\sigma_c^2 + \sigma^2\sigma_\epsilon^2 + (1-\gamma^2)\sigma^2\sigma_\epsilon^2 + \gamma^2[2^{2\kappa_\ell}(3\sigma_c^4 + 2(\sigma^2+\sigma_\epsilon^2)\sigma_c^2) + \sigma^2\sigma_\epsilon^2]\Big\} \\
& \cdot(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)(\sigma_c^2+\sigma_\epsilon^2) \\
& - \Big\{[(1-\gamma^2)\sigma^2\sigma_c^2 + \sigma^2\sigma_\epsilon^2](\sigma_c^2+\sigma_\epsilon^2) + \gamma^2(2^{2\kappa_\ell}(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)\sigma_c^4 + \sigma^2\sigma_\epsilon^2\sigma_c^2)\Big\}(2\sigma_c^2+\sigma^2+2\sigma_\epsilon^2) \\
=\ & 2(1-\gamma^2+\gamma^2 2^{2\kappa_\ell})\sigma^2\sigma_c^2(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)(\sigma_c^2+\sigma_\epsilon^2) + 2\sigma^2\sigma_\epsilon^2(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)(\sigma_c^2+\sigma_\epsilon^2) \\
& + 3\gamma^2 2^{2\kappa_\ell}\sigma_c^4(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)(\sigma_c^2+\sigma_\epsilon^2) + 2\gamma^2 2^{2\kappa_\ell}\sigma_\epsilon^2\sigma_c^2(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)(\sigma_c^2+\sigma_\epsilon^2) \\
& - [(1-\gamma^2)\sigma^2\sigma_c^2 + \sigma^2\sigma_\epsilon^2](\sigma_c^2+\sigma^2+\sigma_\epsilon^2)(\sigma_c^2+\sigma_\epsilon^2) - [(1-\gamma^2)\sigma^2\sigma_c^2 + \sigma^2\sigma_\epsilon^2](\sigma_c^2+\sigma_\epsilon^2)^2 \\
& - \gamma^2 2^{2\kappa_\ell}\sigma_c^4(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)(2\sigma_c^2+\sigma^2+2\sigma_\epsilon^2) - \gamma^2\sigma^2\sigma_\epsilon^2\sigma_c^2(2\sigma_c^2+\sigma^2+2\sigma_\epsilon^2) \\
=\ & (2\gamma^2 2^{2\kappa_\ell}+1-\gamma^2)\sigma^2\sigma_c^2(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)(\sigma_c^2+\sigma_\epsilon^2) + \sigma^2\sigma_\epsilon^2(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)(\sigma_c^2+\sigma_\epsilon^2) \\
& + \gamma^2 2^{2\kappa_\ell}\sigma_c^4(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)(\sigma_c^2-\sigma^2+\sigma_\epsilon^2) + 2\gamma^2 2^{2\kappa_\ell}\sigma_\epsilon^2\sigma_c^2(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)(\sigma_c^2+\sigma_\epsilon^2) \\
& - \sigma^2[(1-\gamma^2)\sigma_c^2+\sigma_\epsilon^2](\sigma_c^2+\sigma_\epsilon^2)^2 - \gamma^2\sigma_\epsilon^2\sigma^2\sigma_c^2(2\sigma_c^2+\sigma^2+2\sigma_\epsilon^2) \\
=\ & 2\gamma^2 2^{2\kappa_\ell}\sigma^2\sigma_c^2(\sigma_c^2+\sigma_\epsilon^2)^2 + (2\gamma^2 2^{2\kappa_\ell}+1-\gamma^2)\sigma^4\sigma_c^2(\sigma_c^2+\sigma_\epsilon^2) + (1-\gamma^2)\sigma^4\sigma_\epsilon^2\sigma_c^2 + \sigma^4\sigma_\epsilon^4 \\
& + \gamma^2 2^{2\kappa_\ell}\sigma_c^4(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)(\sigma_c^2+\sigma_\epsilon^2) + 2\gamma^2 2^{2\kappa_\ell}\sigma_\epsilon^2\sigma_c^2(\sigma_c^2+\sigma_\epsilon^2)^2 + 2\gamma^2(2^{2\kappa_\ell}-1)\sigma_\epsilon^2\sigma^2\sigma_c^2(\sigma_c^2+\sigma_\epsilon^2) \\
& - \gamma^2 2^{2\kappa_\ell}\sigma_c^4\sigma^2(\sigma_c^2+\sigma^2+\sigma_\epsilon^2) \\
>\ & 2\gamma^2 2^{2\kappa_\ell}\sigma^2\sigma_c^4(\sigma_c^2+\sigma_\epsilon^2) + (2\gamma^2 2^{2\kappa_\ell}+1-\gamma^2)\sigma^4\sigma_c^4 + (1-\gamma^2)\sigma^4\sigma_\epsilon^2\sigma_c^2 + \sigma^4\sigma_\epsilon^4 \\
& + 2\gamma^2 2^{2\kappa_\ell}\sigma_\epsilon^2\sigma_c^2(\sigma_c^2+\sigma_\epsilon^2)^2 + 2\gamma^2(2^{2\kappa_\ell}-1)\sigma_\epsilon^2\sigma^2\sigma_c^2(\sigma_c^2+\sigma_\epsilon^2) - \gamma^2 2^{2\kappa_\ell}\sigma_c^4\sigma^2(\sigma_c^2+\sigma^2+\sigma_\epsilon^2) \\
=\ & \gamma^2 2^{2\kappa_\ell}\sigma^2\sigma_c^4(\sigma_c^2+\sigma_\epsilon^2) + (\gamma^2 2^{2\kappa_\ell}+1-\gamma^2)\sigma^4\sigma_c^4 + (1-\gamma^2)\sigma^4\sigma_\epsilon^2\sigma_c^2 + \sigma^4\sigma_\epsilon^4 \\
& + 2\gamma^2 2^{2\kappa_\ell}\sigma_\epsilon^2\sigma_c^2(\sigma_c^2+\sigma_\epsilon^2)^2 + 2\gamma^2(2^{2\kappa_\ell}-1)\sigma_\epsilon^2\sigma^2\sigma_c^2(\sigma_c^2+\sigma_\epsilon^2) > 0,
\end{aligned}$$

where the last inequality follows from $\kappa_\ell > 0$ and $\gamma \in [0,1]$. Since $G_1(\sigma_c^2) > 0$, $F'(\sigma_c^2) > 0$.

Case (2): $\underline{\gamma} < \gamma < \bar{\gamma}$.

According to Lemma 4, we obtain the first order condition for the local government,

$$\frac{1}{\sigma^2} + \frac{1}{\sigma_\ell^2} + \frac{1}{\sigma_c^2} = \frac{1-\gamma}{\gamma}K_\ell^{1/2}.$$

Plugging in the expression of $1/\sigma_\ell^2$, the objective function F can be simplified as

$$\begin{aligned}
F(\sigma_c^2) &= \frac{\gamma(1+\gamma)K_\ell^{1/2} + K_\ell\gamma^2\sigma_c^2 - \gamma(1-\gamma)K_\ell^{1/2} + \gamma^2/\sigma_c^2}{K_\ell - 1/\sigma_c^4} \\
&= \gamma^2\frac{\sigma_c^2K_\ell + 2K_\ell^{1/2} + 1/\sigma_c^2}{K_\ell - 1/\sigma_c^4} = \gamma^2\left(\sigma_c^2 + \frac{2K_\ell^{1/2} + 2/\sigma_c^2}{K_\ell - 1/\sigma_c^4}\right) \\
&= \gamma^2\left(\sigma_c^2 + \frac{2}{K_\ell^{1/2} - 1/\sigma_c^2}\right)
\end{aligned}$$

Again, since γ is still in the middle range for a small change of σ_c^2, the objective function is differentiable and its first derivative is given by

$$\begin{aligned}
F'(\sigma_c^2) &= \gamma^2\left(1 - \frac{K_\ell^{-1/2}dK_\ell/d(\sigma_c^2) + 2/\sigma_c^4}{(K_\ell^{1/2} - 1/\sigma_c^2)^2}\right) \\
&= \gamma^2\left(\frac{K_\ell - 2K_\ell^{1/2}/\sigma_c^2 - K_\ell^{-1/2}dK_\ell/d(\sigma_c^2) - 1/\sigma_c^4}{(K_\ell^{1/2} - 1/\sigma_c^2)^2}\right) \\
&= \gamma^2\left(\frac{2^{2\kappa_\ell}\frac{\sigma^2+\sigma_c^2+\sigma_\epsilon^2}{\sigma^2\sigma_c^2\sigma_\epsilon^2} - \frac{2K_\ell^{1/2}}{\sigma_c^2} + \frac{2^{2\kappa_\ell}(\sigma^2+\sigma_\epsilon^2)\sigma_c^2+2\sigma^2\sigma_\epsilon^2}{K_\ell^{1/2}\sigma_c^6\sigma^2\sigma_\epsilon^2}}{(K_\ell^{1/2} - 1/\sigma_c^2)^2}\right) \\
&= \frac{\gamma^2}{K_\ell^{1/2}(K_\ell^{1/2} - 1/\sigma_c^2)^2}\left(\frac{K_\ell^{1/2}2^{2\kappa_\ell}(\sigma^2+\sigma_c^2+\sigma_\epsilon^2) - 2\sigma^2\sigma_\epsilon^2K_\ell + 2^{2\kappa_\ell}(\sigma^2+\sigma_\epsilon^2)/\sigma_c^2 + 2\sigma^2\sigma_\epsilon^2/\sigma_c^4}{\sigma^2\sigma_c^2\sigma_\epsilon^2}\right) \\
&= \frac{2^{2\kappa_\ell}\gamma^2}{K_\ell^{1/2}(K_\ell^{1/2} - 1/\sigma_c^2)^2}\left(\frac{K_\ell^{1/2}(\sigma^2+\sigma_c^2+\sigma_\epsilon^2) - (\sigma^2/\sigma_c^2 + \sigma_\epsilon^2/\sigma_c^2 + 2)}{\sigma^2\sigma_c^2\sigma_\epsilon^2}\right) \\
&> \frac{2^{2\kappa_\ell}\gamma^2}{K_\ell^{1/2}(K_\ell^{1/2} - 1/\sigma_c^2)^2}\left(\frac{(1/\sigma_\epsilon^2 + 1/\sigma_c^2)^{1/2}(1/\sigma^2 + 1/\sigma_c^2)^{1/2}(\sigma^2+\sigma_c^2+\sigma_\epsilon^2) - (\sigma^2/\sigma_c^2 + \sigma_\epsilon^2/\sigma_c^2 + 2)}{\sigma^2\sigma_c^2\sigma_\epsilon^2}\right),
\end{aligned}$$

where the last inequality follows from $\kappa_\ell > 0$. There are two possibilities. If $1/\sigma_\epsilon^2 \geq 1/\sigma^2$, then

$$F'(\sigma_c^2) \geq \frac{2^{2\kappa_\ell}\gamma^2}{K_\ell^{1/2}(K_\ell^{1/2} - 1/\sigma_c^2)^2}\left(\frac{(1/\sigma^2 + 1/\sigma_c^2)(\sigma^2 + \sigma_c^2 + \sigma_\epsilon^2) - (\sigma^2/\sigma_c^2 + \sigma_\epsilon^2/\sigma_c^2 + 2)}{\sigma^2\sigma_c^2\sigma_\epsilon^2}\right) > 0.$$

If $1/\sigma_\epsilon^2 < 1/\sigma^2$, then

$$F'(\sigma_c^2) > \frac{2^{2\kappa_\ell}\gamma^2}{K_\ell^{1/2}(K_\ell^{1/2} - 1/\sigma_c^2)^2}\left(\frac{(1/\sigma_\epsilon^2 + 1/\sigma_c^2)(\sigma^2 + \sigma_c^2 + \sigma_\epsilon^2) - (\sigma^2/\sigma_c^2 + \sigma_\epsilon^2/\sigma_c^2 + 2)}{\sigma^2\sigma_c^2\sigma_\epsilon^2}\right) > 0.$$

Hence, we must have $F'(\sigma_c^2) > 0$.

Case (3): $\gamma > \bar{\gamma}$.

In this case, according to Lemma 4, $\sigma_\ell^2 = \infty$. Then we have

$$\begin{aligned}
F(\sigma_c^2) &= \frac{1}{K_\ell - 1/\sigma_c^4}\left[\frac{(1-\gamma^2)K_\ell}{\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2}} + K_\ell\gamma^2\sigma_c^2 - \frac{\gamma^2}{\sigma^2}\right] \\
&= \frac{K_\ell + K_\ell\gamma^2\sigma_c^2/\sigma^2 - \gamma^2(1/\sigma^2 + 1/\sigma_c^2)/\sigma^2}{(K_\ell - 1/\sigma_c^4)(1/\sigma^2 + 1/\sigma_c^2)}
\end{aligned}$$

Again, since $\gamma > \bar{\gamma}$ continues to hold for a small change of σ_c^2, the objective function is differentiable and its first derivative is given by

$$F'(\sigma_c^2) = \frac{G_2(\sigma_c^2)}{\left(K_\ell - \frac{1}{\sigma_c^4}\right)^2\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2}\right)^2}$$

with the numerator $G_2(\sigma_c^2)$ given by

$$\begin{aligned}
G_2(\sigma_c^2) &= \left[\frac{dK_\ell}{d(\sigma_c^2)}\left(1 + \frac{\gamma^2\sigma_c^2}{\sigma^2}\right) + \frac{K_\ell\gamma^2}{\sigma^2} + \frac{\gamma^2}{\sigma_c^4\sigma^2}\right]\left(K_\ell - \frac{1}{\sigma_c^4}\right)\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2}\right) \\
&\quad - \left[K_\ell\left(1 + \frac{\gamma^2\sigma_c^2}{\sigma^2}\right) - \frac{\gamma^2}{\sigma^2}\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2}\right)\right]\left[\frac{dK_\ell}{d(\sigma_c^2)}\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2}\right) + \frac{2}{\sigma_c^6\sigma^2} - \frac{K_\ell}{\sigma_c^4} + \frac{3}{\sigma_c^8}\right] \\
&= \left(\gamma^2\left(\frac{1}{\sigma^4} + \frac{2}{\sigma^2\sigma_c^2}\right) + \frac{1}{\sigma_c^4}\right)K_\ell^2 + \left[\frac{\gamma^2}{\sigma_c^4\sigma^2}\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2}\right) - \frac{\gamma^2}{\sigma^2\sigma_c^4}\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2}\right)\right. \\
&\quad \left. - \left(\frac{2}{\sigma_c^6\sigma^2} + \frac{3}{\sigma_c^8}\right)\left(1 + \frac{\gamma^2\sigma_c^2}{\sigma^2}\right) - \frac{\gamma^2}{\sigma_c^4\sigma^2}\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2}\right)\right]K_\ell + \left[-\frac{1}{\sigma_c^4}\left(1 + \frac{\gamma^2\sigma_c^2}{\sigma^2}\right)\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2}\right)\right. \\
&\quad \left. + \frac{\gamma^2}{\sigma^2}\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2}\right)^2\right]\frac{dK_\ell}{d(\sigma_c^2)} - \frac{\gamma^2}{\sigma_c^8\sigma^2}\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2}\right) + \frac{\gamma^2}{\sigma^2}\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2}\right)\left(\frac{2}{\sigma_c^6\sigma^2} + \frac{3}{\sigma_c^8}\right) \\
&= \frac{K_\ell^2}{\sigma_c^4} - \left(\frac{2}{\sigma_c^6\sigma^2} + \frac{3}{\sigma_c^8}\right)K_\ell - \frac{1}{\sigma_c^4}\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2}\right)\frac{dK_\ell}{d(\sigma_c^2)} + \gamma^2\left\{\left(\frac{1}{\sigma^4} + \frac{2}{\sigma^2\sigma_c^2}\right)K_\ell^2 - \left(\frac{3}{\sigma_c^4\sigma^4} + \frac{4}{\sigma_c^6\sigma^2}\right)K_\ell\right. \\
&\quad \left. + \frac{1}{\sigma^4}\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2}\right)\frac{dK_\ell}{d(\sigma_c^2)} + \left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2}\right)\left(\frac{2}{\sigma_c^6\sigma^4} + \frac{2}{\sigma_c^8\sigma^2}\right)\right\} \\
&= \frac{1}{\sigma_c^4}\left\{\left[\frac{2^{2\kappa_\ell}}{\sigma_c^2\sigma_\epsilon^2\sigma^2}(\sigma_c^2 + \sigma^2 + \sigma_\epsilon^2) - \frac{2}{\sigma_c^2}\left(\frac{1}{\sigma_c^2} + \frac{1}{\sigma^2}\right)\right]\left[\frac{2^{2\kappa_\ell}}{\sigma_c^2\sigma_\epsilon^2\sigma^2}(\sigma_c^2 + \sigma^2 + \sigma_\epsilon^2) + \frac{1}{\sigma_c^4}\right]\right. \\
&\quad \left. + \left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2}\right)\left[\frac{2^{2\kappa_\ell}}{\sigma_c^4}\left(\frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma^2}\right) + \frac{2}{\sigma_c^6}\right]\right\} \\
&\quad + \gamma^2\left\{\left(\frac{2^{2\kappa_\ell}}{\sigma_c^2\sigma_\epsilon^2\sigma^2}(\sigma_c^2 + \sigma^2 + \sigma_\epsilon^2) + \frac{1}{\sigma_c^4}\right)\left[\left(\frac{1}{\sigma^4} + \frac{2}{\sigma^2\sigma_c^2}\right)\frac{2^{2\kappa_\ell}(\sigma_c^2 + \sigma^2 + \sigma_\epsilon^2)}{\sigma_c^2\sigma_\epsilon^2\sigma^2} - \frac{2}{\sigma_c^4\sigma^4} - \frac{2}{\sigma_c^6\sigma^2}\right]\right. \\
&\quad \left. - \frac{1}{\sigma^4}\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2}\right)\left(\frac{2^{2\kappa_\ell}}{\sigma_c^4}\left(\frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma^2}\right) + \frac{2}{\sigma_c^6}\right) + \left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2}\right)\left(\frac{2}{\sigma_c^6\sigma^4} + \frac{2}{\sigma_c^8\sigma^2}\right)\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sigma_c^4}\left\{\frac{2^{4\kappa_\ell}}{\sigma_c^4\sigma_\epsilon^4\sigma^4}(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)^2 - \frac{1}{\sigma_c^2}\left(\frac{1}{\sigma_c^2}+\frac{2}{\sigma^2}\right)\cdot\frac{2^{2\kappa_\ell}}{\sigma_c^2\sigma_\epsilon^2\sigma^2}(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)\right.\\
&\quad\left. + \frac{2^{2\kappa_\ell}}{\sigma_c^4}\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}\right)\left(\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma^2}\right)\right\}\\
&\quad + \gamma^2\left\{\left(\frac{2^{2\kappa_\ell}}{\sigma_c^2\sigma_\epsilon^2\sigma^2}(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)+\frac{1}{\sigma_c^4}\right)\left(\frac{1}{\sigma^4}+\frac{2}{\sigma^2\sigma_c^2}\right)\frac{2^{2\kappa_\ell}(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)}{\sigma_c^2\sigma_\epsilon^2\sigma^2}\right.\\
&\quad\left. - \frac{2^{2\kappa_\ell}}{\sigma_c^4\sigma^4}\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma^2}\right)\left(\frac{3}{\sigma_\epsilon^2}+\frac{2}{\sigma_c^2}+\frac{1}{\sigma^2}+\frac{2\sigma^2}{\sigma_c^2\sigma_\epsilon^2}\right)\right\}\\
&= \frac{2^{2\kappa_\ell}}{\sigma_c^4}\left\{\frac{2^{2\kappa_\ell}(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)^2}{\sigma_c^4\sigma_\epsilon^4\sigma^4} - \frac{2\sigma_c^2+2\sigma^2+\sigma_\epsilon^2}{\sigma_c^4\sigma^4\sigma_\epsilon^2}\right\}\\
&\quad + \gamma^2 2^{2\kappa_\ell}\left\{\left(\frac{2^{2\kappa_\ell}}{\sigma_c^2\sigma_\epsilon^2\sigma^2}(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)+\frac{1}{\sigma_c^4}\right)\left(\frac{1}{\sigma^4}+\frac{2}{\sigma^2\sigma_c^2}\right)\frac{\sigma_c^2+\sigma^2+\sigma_\epsilon^2}{\sigma_c^2\sigma_\epsilon^2\sigma^2}\right.\\
&\quad\left. - \frac{1}{\sigma_c^4\sigma^4}\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma^2}\right)\left(\frac{3}{\sigma_\epsilon^2}+\frac{2}{\sigma_c^2}+\frac{1}{\sigma^2}+\frac{2\sigma^2}{\sigma_c^2\sigma_\epsilon^2}\right)\right\}\\
&> \frac{2^{2\kappa_\ell}}{\sigma_c^4}\left\{\frac{(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)^2}{\sigma_c^4\sigma_\epsilon^4\sigma^4} - \frac{2\sigma_c^2+2\sigma^2+\sigma_\epsilon^2}{\sigma_c^4\sigma^4\sigma_\epsilon^2}\right\} + \gamma^2 2^{2\kappa_\ell}\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma^2}\right)\\
&\quad\cdot\left\{\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\epsilon^2}\right)\left(\frac{1}{\sigma^4}+\frac{2}{\sigma^2\sigma_c^2}\right)\frac{\sigma_c^2+\sigma^2+\sigma_\epsilon^2}{\sigma_c^2\sigma_\epsilon^2\sigma^2} - \frac{1}{\sigma_c^4\sigma^4}\left(\frac{3}{\sigma_\epsilon^2}+\frac{2}{\sigma_c^2}+\frac{1}{\sigma^2}+\frac{2\sigma^2}{\sigma_c^2\sigma_\epsilon^2}\right)\right\}\\
&> \frac{2^{2\kappa_\ell}}{\sigma_c^4}\left\{\frac{(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)^2}{\sigma_c^4\sigma_\epsilon^4\sigma^4} - \frac{2\sigma_c^2+2\sigma^2+\sigma_\epsilon^2}{\sigma_c^4\sigma^4\sigma_\epsilon^2}\right\} + \gamma^2 2^{2\kappa_\ell}\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma^2}\right)\\
&\quad\cdot\left\{\frac{1}{\sigma_c^2}\left(\frac{1}{\sigma^4}+\frac{2}{\sigma^2\sigma_c^2}\right)\frac{\sigma_c^2+\sigma^2+\sigma_\epsilon^2}{\sigma_c^2\sigma_\epsilon^2\sigma^2} - \frac{1}{\sigma_c^4\sigma^4}\left(\frac{3}{\sigma_\epsilon^2}+\frac{2}{\sigma_c^2}+\frac{1}{\sigma^2}+\frac{2\sigma^2}{\sigma_c^2\sigma_\epsilon^2}\right)\right\}\\
&= \frac{2^{2\kappa_\ell}}{\sigma_c^4}\frac{(\sigma_c^2+\sigma^2)^2}{\sigma_c^4\sigma^4\sigma_\epsilon^4} + \frac{\gamma^2 2^{2\kappa_\ell}}{\sigma_c^4\sigma^4}\frac{\sigma_c^2}{\sigma_\epsilon^2\sigma^2} > 0
\end{aligned}
$$

where the first inequality follows from $\kappa_\ell > 0$. Therefore, we have $F'(\sigma_c^2) > 0$.

Case (4): $\gamma = \underline{\gamma}$ or $\gamma = \bar{\gamma}$.

Suppose $\gamma = \underline{\gamma}$. A small change of σ_c^2 will make $\gamma < \underline{\gamma}$ or $\gamma \in (\underline{\gamma}, \bar{\gamma})$. Since whether γ ends up in Case (1) or (2) depends on the direction of the change of σ_c^2, the left or right derivatives of F at σ_c^2 may not be equal to each other. If an infinitesimal negative change of σ_c^2 leads to Case (1), we know that $F'_-(\sigma_c^2)$ is equal to $F'(\sigma_c^2)$ for Case (1) and therefore $F'_-(\sigma_c^2) > 0$. If an infinitesimal negative of σ_c^2 leads to Case (2), we know that $F'_-(\sigma_c^2)$ is equal to $F'(\sigma_c^2)$ for Case (2) and therefore $F'_-(\sigma_c^2) > 0$. The same argument applies to $F'_+(\sigma_c^2)$ and we have $F'_+(\sigma_c^2) > 0$. Similarly, we can also show that $F'_-(\sigma_c^2) > 0$ and $F'_+(\sigma_c^2) > 0$ for $\gamma = \bar{\gamma}$.

In sum, for an arbitrarily picked σ_c^2, we have shown that the objective function is strictly increasing for any $\gamma \in [0, 1]$. Then the optimal strategy for the central government is to minimize σ_c^2. Therefore, Constraint 4 must be binding and $\sigma_c^2 = \sigma^2/(2^{2\kappa_c} - 1)$. We have obtained the desired conclusion. □

A.9. Equilibrium characterization under the decentralized regime with $\gamma = 0$

Proposition 6. *Under the decentralized regime with $\gamma = 0$, we have*

$$
E(a_\ell - \theta)^2\Big|_{\gamma=0} = \left(\frac{1}{\sigma^2} + \frac{1}{\sigma_\ell^2} + \frac{1}{\sigma_c^2 + \sigma_\epsilon^2}\right)^{-1}
$$

with $\sigma_c^2 = \sigma^2/(2^{2\kappa_c} - 1)$, $\sigma_{\epsilon\ell}^2 = \sigma_\epsilon^2$, and

$$
\sigma_\ell^2 = \left[K_\ell\left(\frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma_c^2}\right)^{-1} - \frac{1}{\sigma^2} - \frac{1}{\sigma_c^2}\right]^{-1} = \frac{\sigma^2(\sigma_\epsilon^2 + \sigma_c^2)}{(2^{2\kappa_\ell} - 1)(\sigma^2 + \sigma_\epsilon^2 + \sigma_c^2)}.
$$

Proof. It directly follows from Lemma 16 that $\sigma_{\epsilon\ell}^2 = \sigma_\epsilon^2$ and from Lemma 5 that $\sigma_c^2 = \sigma^2/(2^{2\kappa_c} - 1)$. The expression of σ_ℓ^2 can then be derived from the binding Constraint 5. The expression of $E(a_\ell - \theta)^2\Big|_{\gamma=0}$ from Eq. 7. □

A.10. Equilibrium characterization under the decentralized regime with $\gamma = 1$

Proposition 7. *Under the decentralized regime with $\gamma = 1$, we have*

$$E(a_\ell - \theta)^2\Big|_{\gamma=1} = \frac{\sigma_c^2/\sigma_{\epsilon\ell}^4 + 1/\sigma_{\epsilon\ell}^2 + \sigma^2/(\sigma_c^2 + \sigma^2)^2}{\left[1/\sigma_{\epsilon\ell}^2 + 1/(\sigma_c^2 + \sigma^2)\right]^2}$$

with $\sigma_c^2 = \sigma^2/(2^{2\kappa_c} - 1)$, $\sigma_\ell^2 = \infty$, and

$$\sigma_{\epsilon\ell}^2 = \left[K_\ell\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2}\right)^{-1} - \frac{1}{\sigma_c^2}\right]^{-1} = \frac{\sigma_\epsilon^2(\sigma^2 + \sigma_c^2)}{2^{2\kappa_\ell}(\sigma^2 + \sigma_c^2) + (2^{2\kappa_\ell} - 1)\sigma_\epsilon^2}.$$

Proof. It directly follows from Lemma 17 that $\sigma_\ell^2 = \infty$ and from Lemma 5 that $\sigma_c^2 = \sigma^2/(2^{2\kappa_c} - 1)$. The expression of $\sigma_{\epsilon\ell}^2$ can then be derived from the binding Constraint 5. We can derive the expression of $E(a_\ell - \theta)^2\Big|_{\gamma=0}$ by invoking the formula in Lemma 3 and specializing it with the expression of a_ℓ for $\gamma = 1$ in Lemma 13 (or alternatively from Eq. 14):

$$E(a_\ell - \theta)^2\Big|_{\gamma=1} = \frac{\sigma_c^2/\sigma_{\epsilon\ell}^4 + 1/\sigma_{\epsilon\ell}^2 + \sigma^2/(\sigma_c^2 + \sigma^2)^2}{\left[1/\sigma_{\epsilon\ell}^2 + 1/(\sigma_c^2 + \sigma^2)\right]^2} = \frac{\sigma_c^2/\sigma_{\epsilon\ell}^2 + \sigma^2/(\sigma_c^2 + \sigma^2)}{1/\sigma_{\epsilon\ell}^2 + 1/(\sigma_c^2 + \sigma^2)}.$$

□

A.11. Proof of proposition 2

Proof. In the proof of Lemma 5, we have obtained Eq. 14:

$$E(a_\ell - \theta)^2 = \frac{1}{K_\ell - 1/\sigma_c^4}\left[\frac{(1-\gamma^2)K_\ell}{\frac{1}{\sigma^2} + \frac{1}{\sigma_\ell^2} + \frac{1}{\sigma_c^2}} + K_\ell\gamma^2\sigma_c^2 - \gamma^2\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_\ell^2}\right)\right] \equiv F(\sigma_\ell^2(\gamma), \gamma),$$

with $\sigma_c^2 = \sigma^2/(2^{2\kappa_c} - 1)$ being constant with respect to γ.

It is easy to see that $\partial F/\partial(\sigma_\ell^2) > 0$, and

$$\begin{aligned}\frac{\partial F}{\partial \gamma} &= \frac{2\gamma}{K_\ell - 1/\sigma_\ell^4}\left(-\frac{K_\ell}{\frac{1}{\sigma^2} + \frac{1}{\sigma_\ell^2} + \frac{1}{\sigma_c^2}} + K_\ell\sigma_c^2 - \frac{1}{\sigma^2} - \frac{1}{\sigma_\ell^2}\right) \\ &= \frac{2\gamma}{K_\ell - 1/\sigma_\ell^4}\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_\ell^2}\right)\left(\frac{K_\ell\sigma_c^2}{\frac{1}{\sigma^2} + \frac{1}{\sigma_\ell^2} + \frac{1}{\sigma_c^2}} - 1\right) = \frac{2\gamma\sigma_c^2}{K_\ell - 1/\sigma_\ell^4}\left(\frac{1}{\sigma^2} + \frac{1}{\sigma_\ell^2}\right)\cdot\frac{1}{\sigma_{\epsilon\ell}^2}\end{aligned}$$

where the last equality follows again from the fact that Constraint 5 is binding. Then $\partial F/\partial\gamma \geq 0$ with the equality if and only if $\gamma = 0$. Since σ_ℓ^2 is a function of γ and we know from Lemma 4 that σ_ℓ^2 weakly increases with γ, we conclude that $E(a_\ell - \theta)^2$ strictly increases with γ. □

A.12. Proof of lemma 6

Proof. According to Propositions 1 and 6, $E(a_c - \theta)^2 > E(a_\ell - \theta)^2\Big|_{\gamma=0}$ if and only if

$$\frac{(2^{2\kappa_\ell} - 1)(\sigma^2 + \sigma_\epsilon^2 + \sigma_c^2)}{\sigma^2(\sigma_\epsilon^2 + \sigma_c^2)} + \frac{1}{\sigma_c^2 + \sigma_\epsilon^2} > \frac{(2^{2\kappa_c} - 1)(\sigma^2 + \sigma_\epsilon^2 + \sigma_\ell^2)}{\sigma^2(\sigma_\epsilon^2 + \sigma_\ell^2)} + \frac{1}{\sigma_\ell^2 + \sigma_\epsilon^2},$$

where $\sigma_c^2 = \sigma^2/(2^{2\kappa_c} - 1)$ and $\sigma_\ell^2 = \sigma^2/(2^{2\kappa_\ell} - 1)$. Simplifying the expression above, we obtain

$$\frac{\sigma^2 + \sigma_\epsilon^2 + \sigma_c^2}{\sigma_\ell^2(\sigma_\epsilon^2 + \sigma_c^2)} + \frac{1}{\sigma_c^2 + \sigma_\epsilon^2} > \frac{\sigma^2 + \sigma_\epsilon^2 + \sigma_\ell^2}{\sigma_c^2(\sigma_\epsilon^2 + \sigma_\ell^2)} + \frac{1}{\sigma_\ell^2 + \sigma_\epsilon^2}.$$

The inequality holds if and only if $\sigma_c^2 > \sigma_\ell^2$, which follows from $\kappa_\ell > \kappa_c$. □

A.13. Proof of lemma 7

Proof. According to Proposition 7, we know

$$E(a_\ell-\theta)^2\Big|_{\gamma=1} = \frac{\sigma_c^2/\sigma_{\epsilon\ell}^4+1/\sigma_{\epsilon\ell}^2+\sigma^2/(\sigma_c^2+\sigma^2)^2}{\left[1/\sigma_{\epsilon\ell}^2+1/(\sigma_c^2+\sigma^2)\right]^2}$$
$$= \frac{\sigma_c^2/\sigma_{\epsilon\ell}^2+\sigma^2/(\sigma_c^2+\sigma^2)}{1/\sigma_{\epsilon\ell}^2+1/(\sigma_c^2+\sigma^2)} = \sigma_c^2+\frac{\sigma^2-\sigma_c^2}{\sigma_c^2+\sigma^2}\left(\frac{1}{\sigma_{\epsilon\ell}^2}+\frac{1}{\sigma_c^2+\sigma^2}\right)^{-1}.$$

If $\sigma_c^2\le\sigma^2$, we have $E(a_\ell-\theta)^2\Big|_{\gamma=1}\ge\sigma_c^2$. If $\sigma_c^2>\sigma^2$, then $E(a_\ell-\theta)^2\Big|_{\gamma=1}$ strictly decreases with $\sigma_{\epsilon\ell}^2$. We know $\lim_{\sigma_{\epsilon\ell}^2\to\infty}E(a_\ell-\theta)^2\Big|_{\gamma=1}=\sigma^2$, so when $\sigma_c^2>\sigma^2$, $E(a_\ell-\theta)^2\Big|_{\gamma=1}>\sigma^2$. Therefore, we must have $E(a_\ell-\theta)^2\Big|_{\gamma=1}\ge\min\{\sigma^2,\sigma_c^2\}$ where $\sigma_c^2=\sigma^2/(2^{2\kappa_c}-1)$.

According to Proposition 1, with $\sigma_\ell^2=\sigma^2/(2^{2\kappa_\ell}-1)$, we have

$$E(a_c-\theta)^2=\left(\frac{1}{\sigma^2}+\frac{\sigma^2+\sigma_\epsilon^2+\sigma_\ell^2}{\sigma_c^2(\sigma_\epsilon^2+\sigma_\ell^2)}+\frac{1}{\sigma_\ell^2+\sigma_\epsilon^2}\right)^{-1}=\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\sigma^2+\sigma_c^2}{\sigma_c^2(\sigma_\epsilon^2+\sigma_\ell^2)}\right)^{-1},$$

where $\sigma_c^2=\sigma^2/(2^{2\kappa_c}-1)$. It is easy to see that $E(a_c-\theta)^2<1/\sigma^2$ and $E(a_c-\theta)^2<1/\sigma_c^2$. Therefore, $E(a_c-\theta)^2<\min\{\sigma^2,\sigma_c^2\}\le E(a_\ell-\theta)^2\Big|_{\gamma=1}$. We have obtained the desired conclusion. □

A.14. Proof of corollary 1

Proof. Using Eq. 14 in the proof of Lemma 5, we obtain

$$E(a_\ell-\theta)^2\Big|_{\gamma=\bar\gamma} = \frac{1}{K_\ell-1/\sigma_c^4}\left[\frac{(1-\bar\gamma^2)K_\ell}{\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_c^2}}+K_\ell\bar\gamma^2\sigma_c^2-\bar\gamma^2\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}\right)\right]$$
$$= \frac{1}{K_\ell-1/\sigma_c^4}\left[\frac{(1-\bar\gamma^2)K_\ell}{\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}}+K_\ell\bar\gamma^2\sigma_c^2-\bar\gamma^2\left(\frac{1}{\sigma^2}\right)\right]$$
$$= \frac{1}{K_\ell-1/\sigma_c^4}\,\frac{K_\ell+\frac{\bar\gamma^2\sigma_c^2}{\sigma^2}\left(K_\ell-\frac{1}{\sigma_c^2}\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}\right)\right)}{\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}}$$
$$> \frac{K_\ell}{K_\ell-1/\sigma_c^4}\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}\right)^{-1}>\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}\right)^{-1}$$

where $\sigma_c^2=\sigma^2/(2^{2\kappa_c}-1)$ (Lemma 5), the second inequality follows from the fact that $\sigma_\ell^2=\infty$ when $\gamma=\bar\gamma$ (Lemma 4), and the first inequality follows from the definition of K_ℓ. From Proposition 1, we know

$$E(a_c-\theta)^2=\left(\frac{1}{\sigma^2}+\frac{(2^{2\kappa_c}-1)(\sigma^2+\sigma_\epsilon^2+\sigma_\ell^2)}{\sigma^2(\sigma_\epsilon^2+\sigma_\ell^2)}+\frac{1}{\sigma_\ell^2+\sigma_\epsilon^2}\right)^{-1}$$

with $\sigma_\ell^2=\sigma^2/(2^{2\kappa_\ell}-1)$. Hence, we have

$$E(a_c-\theta)^2=\left(\frac{1}{\sigma^2}+\frac{\sigma^2+\sigma_\epsilon^2+\sigma_\ell^2}{\sigma_c^2(\sigma_\epsilon^2+\sigma_\ell^2)}+\frac{1}{\sigma_\ell^2+\sigma_\epsilon^2}\right)^{-1}<\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}\right)^{-1}<E(a_\ell-\theta)^2\Big|_{\gamma=\bar\gamma}.$$

By definition, $E(a_c-\theta)^2=E(a_\ell-\theta)^2\Big|_{\gamma=\tilde\gamma}$. Then $\bar\gamma>\tilde\gamma$ directly follows from Proposition 2. □

A.15. Optimal policy in a non-strategic environment

Lemma 18. *Under the centralized regime, the optimal policy for the central government is given by*

$$a_c=E(\theta|\theta_c,s'_\ell)=\frac{\frac{\theta_c}{\sigma_c^2}+\frac{s'_\ell}{(\sigma_\ell^2+\sigma_\epsilon^2)}}{\frac{1}{(\sigma_\ell^2+\sigma_\epsilon^2)}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma^2}}, \tag{15}$$

with $s'_\ell = \theta_\ell + \epsilon$. Under the decentralized regime, the optimal policy for the local government is given by

$$a_\ell = (1-\gamma)E(\theta|\theta_\ell, s'_c) + \gamma E(\theta_c|\theta_\ell, s'_c) = k'_1\theta_\ell + k'_2 s'_c, \tag{16}$$

with $s'_c = \theta_c + \epsilon$ and

$$k'_1 \equiv \frac{\frac{1-\gamma}{\sigma_\ell^2}}{\frac{1}{\sigma_\ell^2} + \frac{1}{\sigma_c^2+\sigma_\epsilon^2} + \frac{1}{\sigma^2}} + \frac{\frac{\gamma}{\sigma_\ell^2+\sigma_c^2+\sigma_c^2\sigma_\ell^2/\sigma^2}}{\frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}}$$

$$k'_2 \equiv \frac{\frac{1-\gamma}{\sigma_c^2+\sigma_\epsilon^2}}{\frac{1}{\sigma_\ell^2} + \frac{1}{\sigma_c^2+\sigma_\epsilon^2} + \frac{1}{\sigma^2}} + \frac{\frac{\gamma}{\sigma_\epsilon^2}}{\frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}}.$$

Proof. The proof directly follows from Lemmas 12 and 13 by letting $\sigma_{\epsilon\ell}^2 = \sigma_{\epsilon c}^2 = \sigma_\epsilon^2$. □

A.16. $E(a_c-\theta)^2$, $E(a_\ell-\theta)^2\Big|_{\gamma=0}$, And $E(a_\ell-\theta)^2\Big|_{\gamma=1}$ in a non-strategic environment

Lemma 19. *Under the centralized regime,*

$$E(a_c-\theta)^2 = Var(\theta|\theta_c, s'_\ell) = \left(\frac{1}{\sigma_c^2} + \frac{1}{\sigma_\ell^2+\sigma_\epsilon^2} + \frac{1}{\sigma^2}\right)^{-1}.$$

Under the decentralized regime,

$$E(a_\ell-\theta)^2\Big|_{\gamma=0} = Var(\theta|\theta_\ell, s'_c) = \left(\frac{1}{\sigma_\ell^2} + \frac{1}{\sigma_c^2+\sigma_\epsilon^2} + \frac{1}{\sigma^2}\right)^{-1}$$

$$E(a_\ell-\theta)^2\Big|_{\gamma=1} = \frac{\frac{\sigma_c^2}{\sigma_\epsilon^2} + \frac{(1/\sigma^2+1/\sigma_\ell^2)^{-1}}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}}{\frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}}$$

Proof. The expressions are obtained from taking the expressions of a_c and a_ℓ from Lemma 18 and applying the results in Lemma 3. Under the decentralized regime, when $\gamma = 1$, we have

$$E(a_\ell-\theta)^2\Big|_{\gamma=1} = \frac{\frac{\sigma_\ell^2}{(\sigma_c^2+\sigma_\ell^2+\sigma_c^2\sigma_\ell^2/\sigma^2)^2} + \frac{\sigma_c^2+\sigma_\epsilon^2}{\sigma_\epsilon^4} + \frac{\sigma_\ell^4/\sigma^2}{(\sigma_c^2+\sigma_\ell^2+\sigma_c^2\sigma_\ell^2/\sigma^2)^2}}{\left(\frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}\right)^2}$$

$$= \frac{\frac{\sigma_c^2}{\sigma_\epsilon^4} + \frac{1}{\sigma_\epsilon^2} + \frac{(1/\sigma^2+1/\sigma_\ell)^{-1}}{\left(\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}\right)^2}}{\left(\frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}\right)^2} = \frac{\frac{\sigma_c^2}{\sigma_\epsilon^2} + \frac{(1/\sigma^2+1/\sigma_\ell^2)^{-1}}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}}{\frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell)^{-1}}}.$$

□

A.17. Proof of lemma 8

Proof. Since we assume $\sigma_\epsilon^2 > 0$ and $\sigma_c^2 > \sigma_\ell^2$, we have

$$\frac{1}{\sigma_c^2\sigma_\ell^2} > \frac{1}{(\sigma_c^2+\sigma_\epsilon^2)(\sigma_\ell^2+\sigma_\epsilon^2)} \Leftrightarrow \frac{1}{\sigma_\ell^2} - \frac{1}{\sigma_c^2} = \frac{\sigma_c^2-\sigma_\ell^2}{\sigma_c^2\sigma_\ell^2} > \frac{\sigma_c^2-\sigma_\ell^2}{(\sigma_c^2+\sigma_\epsilon^2)(\sigma_\ell^2+\sigma_\epsilon^2)} = \frac{1}{\sigma_\ell^2+\sigma_\epsilon^2} - \frac{1}{\sigma_c^2+\sigma_\epsilon^2}.$$

Therefore, we have

$$\left(\frac{1}{\sigma_c^2} + \frac{1}{\sigma_\ell^2+\sigma_\epsilon^2} + \frac{1}{\sigma^2}\right)^{-1} > \left(\frac{1}{\sigma_\ell^2} + \frac{1}{\sigma_c^2+\sigma_\epsilon^2} + \frac{1}{\sigma^2}\right)^{-1}.$$

The desired conclusion directly follows from Lemma 19. □

A.18. Proof of lemma 9

Proof. According to Lemma 19, we have

$$E(a_\ell-\theta)^2\Big|_{\gamma=1}-E(a_c-\theta)^2=\frac{\frac{\sigma_c^2}{\sigma_\epsilon^2}+\frac{(1/\sigma^2+1/\sigma_\ell^2)^{-1}}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}}{\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}}-\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2+\sigma_\epsilon^2}+\frac{1}{\sigma^2}\right)^{-1}$$
$$=\frac{F(\sigma_\epsilon^2)}{\left(\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}\right)\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2+\sigma_\epsilon^2}+\frac{1}{\sigma^2}\right)},$$

with the numerator given by

$$F(\sigma_\epsilon^2)=\left(\frac{\sigma_c^2}{\sigma_\epsilon^2}+\frac{(1/\sigma^2+1/\sigma_\ell^2)^{-1}}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}\right)\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2+\sigma_\epsilon^2}+\frac{1}{\sigma^2}\right)-\left(\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}\right)$$
$$=\frac{\sigma_c^2}{\sigma_\epsilon^2}\left(\frac{1}{\sigma_\ell^2+\sigma_\epsilon^2}+\frac{1}{\sigma^2}\right)+\frac{1/\sigma_c^2-1/\sigma_\ell^2+1/(\sigma_\ell^2+\sigma_\epsilon^2)}{\sigma_c^2(1/\sigma^2+1/\sigma_\ell^2)+1}.$$

We have $F'(\sigma_\epsilon^2)<0$, $\lim_{\sigma_\epsilon^2\to 0}F(\sigma_\epsilon^2)>0$ and

$$\lim_{\sigma_\epsilon^2\to\infty}F(\sigma_\epsilon^2)=\frac{1/\sigma_c^2-1/\sigma_\ell^2}{\sigma_c^2(1/\sigma^2+1/\sigma_\ell^2)+1}<0,$$

where the inequality follows from $\sigma_c^2>\sigma_\ell^2$. Applying the intermediate value theorem, there must exist a unique $\bar{\sigma}_\epsilon^2>0$ such that $F(\bar{\sigma}_\epsilon^2)=0$, or equivalently, $E(a_\ell-\theta)^2\Big|_{\gamma=1}=E(a_c-\theta)^2$. Given the monotonicity of F, we know that $F(\sigma_\epsilon^2)>0$, or equivalently, $E(a_\ell-\theta)^2\Big|_{\gamma=1}>E(a_c-\theta)^2$ if and only if $\sigma_\epsilon^2<\bar{\sigma}_\epsilon^2$. □

A.19. Proof of proposition 3

Proof. According to Lemma 19, we have

$$E(a_\ell-\theta)^2\Big|_{\gamma=1}=\frac{\frac{\sigma_c^2}{\sigma_\epsilon^2}+\frac{(1/\sigma^2+1/\sigma_\ell)^{-1}}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell)^{-1}}}{\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell)^{-1}}}=\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}\right)^{-1}+\frac{\sigma_c^2-(1/\sigma^2+1/\sigma_\ell^2)^{-1}}{1+\frac{\sigma_\epsilon^2}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}}.$$

Since $\sigma_c^2>\sigma_\ell^2$, $\sigma_c^2>(1/\sigma^2+1/\sigma_\ell^2)^{-1}$, which suggests that $E(a_\ell-\theta)\Big|_{\gamma=1}$ strictly decreases with σ_ϵ^2. □

A.20. Proof of corollary 2

Proof. According to Lemma 18, $a_\ell=(1-\gamma)E(\theta|\theta_\ell,s_c')+\gamma E(\theta_c|\theta_\ell,s_c')$. Then we have

$$\begin{aligned}E(a_\ell-\theta)^2&=E\big([E(\theta|\theta_\ell,s_c')-\theta]+\gamma[E(\theta_c|\theta_\ell,s_c')-E(\theta|\theta_\ell,s_c')]\big)^2\\&=E[E(\theta|\theta_\ell,s_c')-\theta]^2+\gamma^2E[E(\theta_c|\theta_\ell,s_c')-E(\theta|\theta_\ell,s_c')]^2\\&\quad+2\gamma E\big([E(\theta|\theta_\ell,s_c')-\theta]\cdot[E(\theta_c|\theta_\ell,s_c')-E(\theta|\theta_\ell,s_c')]\big)\\&=Var(\theta|\theta_\ell,s_c')+\gamma^2E((\theta_c-\theta)|\theta_\ell,s_c')^2\end{aligned}$$

Since $E((\theta_c-\theta)|\theta_\ell,s_c')^2>0$, we must have $\partial E(a_\ell-\theta)^2/\partial(\gamma^2)>0$. □

A.21. Proof of proposition 4

Proof. In the first case of strategic communication, the information constraint for the local government, Constraint 5 should be replaced with

$$\left(\frac{1}{\sigma_c^2+\sigma_{\delta c}^2}+\frac{1}{\sigma_{\epsilon\ell}^2}\right)\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2+\sigma_{\delta c}^2}+\frac{1}{\sigma_\ell^2}\right)\le\frac{2^{2\kappa_\ell}(\sigma^2+\sigma_c^2+\sigma_{\delta c}^2+\sigma_\epsilon^2)}{\sigma^2(\sigma_c^2+\sigma_{\delta c}^2)\sigma_\epsilon^2}+\frac{1}{(\sigma_c^2+\sigma_{\delta c}^2)^2}\equiv\mathcal{K}_\ell.$$

Explicitly expressing the objective function with $\gamma = 1$, we obtain the following optimization problem for the local government:

$$\min_{\sigma_\ell^2,\sigma_{\epsilon\ell}^2}\left(\frac{1}{\sigma_{\epsilon\ell}^2+\sigma_{\delta c}^2}+\frac{1}{\sigma_c^2+(\sigma^{-2}+\sigma_\ell^{-2})^{-1}}\right)^{-1}$$

subject to the above constraint and $\sigma_{\epsilon\ell}^2 \le \sigma_\epsilon^2$. For $\sigma_{\delta c}^2 = 0$, this decision problem is reduced to the one in the baseline setting. Using the constraint above, letting $x \equiv \sigma^{-2}+\sigma_\ell^{-2}$, we rewrite the constrained optimization problem as

$$\max_x F(x) \equiv \left[\mathcal{K}_\ell(\sigma_c^2+\sigma_{\delta c}^2)-\frac{1}{\sigma_c^2+\sigma_{\delta c}^2}\right]\cdot\frac{x+\frac{1}{\sigma_c^2+\sigma_{\delta c}^2}}{\left[\mathcal{K}_\ell\sigma_{\delta c}^2+\frac{\sigma_c^2}{\sigma_c^2+\sigma_{\delta c}^2}x+\frac{\sigma_c^2}{(\sigma_c^2+\sigma_{\delta c}^2)^2}\right](\sigma_c^2x+1)}$$

subject to $x \ge \sigma^{-2}$ and $\sigma_{\epsilon\ell}^2 \le \sigma_\epsilon^2$. Let $y \equiv x+\frac{1}{\sigma_c^2+\sigma_{\delta c}^2}$. We further have

$$G(y) \equiv \frac{\mathcal{K}_\ell(\sigma_c^2+\sigma_{\delta c}^2)-\frac{1}{\sigma_c^2+\sigma_{\delta c}^2}}{F(y)} = \frac{\sigma_c^4}{\sigma_c^2+\sigma_{\delta c}^2}y+\mathcal{K}_\ell\frac{\sigma_{\delta c}^4}{\sigma_c^2+\sigma_{\delta c}^2}y^{-1}+\frac{\sigma_c^2\sigma_{\delta c}^2}{(\sigma_c^2+\sigma_{\delta c}^2)^2}+\mathcal{K}_\ell\sigma_{\delta c}^2\sigma_c^2.$$

$G(\cdot)$ is strictly decreasing for $y \le \sqrt{\mathcal{K}_\ell\sigma_{\delta c}^4/\sigma_c^4}$ and is strictly increasing for $y > \sqrt{\mathcal{K}_\ell\sigma_{\delta c}^4/\sigma_c^4}$. Or equivalently, $F(\cdot)$ is strictly increasing for $x \le \sqrt{\mathcal{K}_\ell\sigma_{\delta c}^4/\sigma_c^4}-\frac{1}{\sigma_c^2+\sigma_{\delta c}^2}$ and is strictly decreasing for $x > \sqrt{\mathcal{K}_\ell\sigma_{\delta c}^4/\sigma_c^4}-\frac{1}{\sigma_c^2+\sigma_{\delta c}^2}$. The turning point $t(\sigma_{\delta c}^2) \equiv \sqrt{\mathcal{K}_\ell\sigma_{\delta c}^4/\sigma_c^4}-\frac{1}{\sigma_c^2+\sigma_{\delta c}^2}$ strictly increases with $\sigma_{\delta c}^2$. There exists a unique $\underline{\sigma}_{\delta c}^2$ such that $t(\underline{\sigma}_{\delta c}^2) = \sigma^{-2}$ and for $\sigma_{\delta c}^2 \le \underline{\sigma}_{\delta c}^2$, $F(x)$ attains its maximum at $x = \sigma^{-2}$ with $\sigma_\ell^2 = \infty$. There exists a unique $\bar{\sigma}_{\delta c}^2 > \underline{\sigma}_{\delta c}^2$ such that $t(\bar{\sigma}_{\delta c}^2) = \frac{\mathcal{K}_\ell}{\frac{1}{\sigma_c^2+\sigma_{\delta c}^2}+\frac{1}{\sigma_\epsilon^2}}-\frac{1}{\sigma_c^2+\sigma_{\delta c}^2}$ and for $\sigma_{\delta c}^2 \ge \bar{\sigma}_{\delta c}^2$, $F(x)$ attains its maximum when σ_ℓ^2 attains its minimum with $\sigma_{\epsilon\ell}^2 = \sigma_\epsilon^2$. Moreover, for $\sigma_{\delta c}^2 \in (\underline{\sigma}_{\delta c}^2, \bar{\sigma}_{\delta c}^2)$, then $F(x)$ is maximized for $x = t(\sigma_{\delta c}^2)$ which implies σ_ℓ^2 strictly decreases with $\sigma_{\delta c}^2$. Thus, we have obtained the desired conclusion. □

A.22. Proof of lemma 10

Proof. Let $x \equiv \frac{\sigma^2}{(2^{2\kappa}-1)\sigma_\epsilon^2} > 0$ and $k \equiv 2^{2\kappa} > 1$, we can write the right hand side of Condition 12 as

$$F(k,x) \equiv \frac{k^2\left(\frac{1}{k-1}+\frac{kx}{k-1}\right)^2}{k\left(\frac{1}{k-1}+\frac{kx}{k-1}\right)+1} = \frac{k^2(kx+1)^2}{(k-1)(k(kx+1)+k-1)}.$$

It is easy to see that $F(k,x)$ strictly increases with x for $k > 1$. So we must have

$$F(k,x) > F(k,0) = \frac{k^2}{(k-1)(2k-1)}.$$

It is to show that $F(k,0)$ attains its minimum on $(1,\infty)$ when $k \to \infty$, which implies $F(k,0) > 1/2$. Therefore, for the regularity condition 12 to hold, it suffices to have

$$\lambda(1/\lambda-1)^2 \le 1/2.$$

Solving the inequality with the constraint that $\lambda \in (0,1)$, we then obtain $\lambda \ge 1/2$. □

A.23. Proof of lemma 11

Proof. Following the proof of Lemma 4, we can write the decision problem of the local government as

$$\min_{\sigma_\ell^2,\sigma_{\epsilon\ell}^2} F(\sigma_\ell^2,\sigma_{\epsilon\ell}^2),$$

subject to Constraint 10, with $F(\sigma_\ell^2,\sigma_{\epsilon\ell}^2)$ given by

$$\begin{aligned}F(\sigma_\ell^2,\sigma_{\epsilon\ell}^2) = {} & \frac{(1-\gamma)^2(1+2\gamma)}{\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_c^2+\sigma_{\epsilon\ell}^2}}+\frac{\gamma^2(3-2\gamma)}{\frac{1}{\sigma_{\epsilon\ell}^2}+\frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}}+(1-\gamma)\gamma\sigma_c^2\\ & -\frac{2(1-\gamma)\gamma^2\sigma_c^2}{\frac{1}{\sigma_{\epsilon\ell}^2}\cdot\left(\sigma_c^2+\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}\right)^{-1}\right)+1}-\frac{2(1-\gamma)^2\gamma\sigma_c^2}{\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}\right)\left(\sigma_c^2+\sigma_{\epsilon\ell}^2\right)+1}\end{aligned}$$

We know Constraint 10 must be binding, so we can rewrite the objective function as

$$
\begin{aligned}
F(\sigma_\ell^2,\sigma_{\epsilon\ell}^2) = {} & \frac{(1-\gamma)^2(1+2\gamma)}{\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}+\frac{\frac{K_\ell}{\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}}-\frac{1}{\sigma_c^2}}{\sigma_c^2\left(\frac{K_\ell}{\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}}-\frac{1}{\sigma_c^2}\right)+1}} + \frac{\gamma^2(3-2\gamma)}{\frac{K_\ell}{\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}}-\frac{1}{\sigma_c^2}+\frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}} + (1-\gamma)\gamma\sigma_c^2 \\
& - \frac{2(1-\gamma)\gamma^2\sigma_c^2}{\left(\frac{K_\ell}{\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}}-\frac{1}{\sigma_c^2}\right)\left(\sigma_c^2+\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}\right)^{-1}\right)+1} \\
& - \frac{2(1-\gamma)^2\gamma\sigma_c^2\left(\frac{K_\ell}{\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}}-\frac{1}{\sigma_c^2}\right)}{\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}\right)\left(\sigma_c^2\left(\frac{K_\ell}{\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}}-\frac{1}{\sigma_c^2}\right)+1\right)+\left(\frac{K_\ell}{\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}}-\frac{1}{\sigma_c^2}\right)} \\
= {} & \frac{(1-\gamma)^2(1+2\gamma)K_\ell\sigma_c^4}{K_\ell\sigma_c^4\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2}\right)-\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}\right)} + \frac{\gamma^2(3-2\gamma)\sigma_c^4\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}\right)\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2}\right)}{K_\ell\sigma_c^4\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2}\right)-\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}\right)} \\
& + (1-\gamma)\gamma\sigma_c^2 - \frac{2(1-\gamma)\gamma^2\sigma_c^4\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}\right)\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}\right)}{K_\ell\sigma_c^4\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2}\right)-\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}\right)} \\
& - \frac{2(1-\gamma)^2\gamma\left(\sigma_c^4K_\ell-\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}\right)\sigma_c^2\right)}{K_\ell\sigma_c^4\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2}\right)-\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}\right)} \\
= {} & \frac{(1-\gamma)^2\left(K_\ell\sigma_c^4+2\gamma\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}\right)\sigma_c^2\right)}{K_\ell\sigma_c^4\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2}\right)-\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}\right)} + (1-\gamma)\gamma\sigma_c^2 \\
& + \frac{\gamma^2\sigma_c^4\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}\right)\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2}+\frac{2(1-\gamma)}{\sigma_c^2}\right)}{K_\ell\sigma_c^4\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2}\right)-\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}\right)}.
\end{aligned}
$$

Dropping terms that are constant with respect to σ_ℓ^2 and $\sigma_{\epsilon\ell}^2$ and letting $x=\lambda/\sigma_\ell^2+1/\sigma^2+1/\sigma_c^2$, we can rewrite the decision problem as

$$
\min_x \frac{(1-\gamma)^2\left(K_\ell\sigma_c^4+2\gamma\sigma_c^2x\right)+\gamma^2\sigma_c^4x\left(\frac{1-1/\lambda}{\sigma^2}+\frac{1-1/\lambda}{\sigma_c^2}+x/\lambda+\frac{2(1-\gamma)}{\sigma_c^2}\right)}{K_\ell\sigma_c^4\left(\frac{1-1/\lambda}{\sigma^2}+\frac{1-1/\lambda}{\sigma_c^2}+x/\lambda\right)-x} \equiv H(x)
$$

with x taking value from $[1/\sigma^2+1/\sigma_c^2, K_\ell/(1/\sigma_\epsilon^2+1/\sigma_c^2)]$. Denote the minimizer be x^*.

$$
\begin{aligned}
H'(x) = {} & \frac{\left[2(1-\gamma)^2\gamma\sigma_c^2+\gamma^2\sigma_c^4\left(\frac{1-1/\lambda}{\sigma^2}+\frac{1-1/\lambda}{\sigma_c^2}+2x/\lambda+\frac{2(1-\gamma)}{\sigma_c^2}\right)\right]\left[K_\ell\sigma_c^4\left(\frac{1-1/\lambda}{\sigma^2}+\frac{1-1/\lambda}{\sigma_c^2}+x/\lambda\right)-x\right]}{\left[K_\ell\sigma_c^4\left(\frac{1-1/\lambda}{\sigma^2}+\frac{1-1/\lambda}{\sigma_c^2}+x/\lambda\right)-x\right]^2} \\
& - \frac{\left[(1-\gamma)^2\left(K_\ell\sigma_c^4+2\gamma\sigma_c^2x\right)+\gamma^2\sigma_c^4x\left(\frac{1-1/\lambda}{\sigma^2}+\frac{1-1/\lambda}{\sigma_c^2}+x/\lambda+\frac{2(1-\gamma)}{\sigma_c^2}\right)\right]\left(K_\ell\sigma_c^4/\lambda-1\right)}{\left[K_\ell\sigma_c^4\left(\frac{1-1/\lambda}{\sigma^2}+\frac{1-1/\lambda}{\sigma_c^2}+x/\lambda\right)-x\right]^2} \\
= {} & \frac{\frac{\gamma^2\sigma_c^4}{\lambda}\left(\frac{K_\ell\sigma_c^4}{\lambda}-1\right)x^2+\gamma^2K_\ell\sigma_c^8\left(\frac{1-1/\lambda}{\sigma^2}+\frac{1-1/\lambda}{\sigma_c^2}+2x/\lambda+\frac{2(1-\gamma)}{\sigma_c^2}\right)\left(\frac{1-1/\lambda}{\sigma^2}+\frac{1-1/\lambda}{\sigma_c^2}\right)}{\left[K_\ell\sigma_c^4\left(\frac{1-1/\lambda}{\sigma^2}+\frac{1-1/\lambda}{\sigma_c^2}+x/\lambda\right)-x\right]^2}
\end{aligned}
$$

$$+\frac{2(1-\gamma)^2\gamma K_\ell\sigma_c^6\left(\frac{1-1/\lambda}{\sigma^2}+\frac{1-1/\lambda}{\sigma_c^2}\right)-(1-\gamma)^2K_\ell\sigma_c^4(K_\ell\sigma_c^4/\lambda-1)}{\left[K_\ell\sigma_c^4\left(\frac{1-1/\lambda}{\sigma^2}+\frac{1-1/\lambda}{\sigma_c^2}+x/\lambda\right)-x\right]^2}$$
$$=\frac{\gamma^2\sigma_c^4}{K_\ell\sigma_c^4-\lambda}-\frac{A}{\left[K_\ell\sigma_c^4\left(\frac{1-1/\lambda}{\sigma^2}+\frac{1-1/\lambda}{\sigma_c^2}+x/\lambda\right)-x\right]^2},$$

with A, being constant with respect to x, is given by

$$A\equiv B^2\frac{\gamma^2K_\ell\sigma_c^8\lambda}{K_\ell\sigma_c^4-\lambda}+2(1-\gamma)\gamma K_\ell\sigma_c^6B+(1-\gamma)^2K_\ell\sigma_c^4\left(\frac{K_\ell\sigma_c^4}{\lambda}-1\right),$$

where $B\equiv(1/\lambda-1)(1/\sigma^2+1/\sigma_c^2)>0$. Since $B>0$ and $K_\ell\sigma_c^4>1>\lambda$, $A>0$. We then have

$$H''(x)=\frac{2A(K_\ell\sigma_c^4/\lambda-1)}{\left[(K_\ell\sigma_c^4/\lambda-1)x-K_\ell\sigma_c^4B\right]^3}>0,$$

which implies that x^* must be unique.

Following the proof of Lemma 4, we define

$$G(\gamma;x)\equiv\frac{\gamma^2\sigma_c^4}{K_\ell\sigma_c^4-\lambda}-\frac{B^2\frac{\gamma^2K_\ell\sigma_c^8\lambda}{K_\ell\sigma_c^4-\lambda}+2(1-\gamma)\gamma K_\ell\sigma_c^6B+(1-\gamma)^2K_\ell\sigma_c^4\left(\frac{K_\ell\sigma_c^4}{\lambda}-1\right)}{\left[(K_\ell\sigma_c^4/\lambda-1)x-K_\ell\sigma_c^4B\right]^2}=H'(x)$$

with x taking value from $[1/\sigma^2+1/\sigma_c^2, K_\ell/(1/\sigma_\epsilon^2+1/\sigma_c^2)]$.

Claim 1. *For any x in $[1/\sigma^2+1/\sigma_c^2, K_\ell/(1/\sigma_\epsilon^2+1/\sigma_c^2)]$, $G(1;x)>0$.*

Given the definition of G, it is equivalent to show

$$\frac{\sigma_c^4}{K_\ell\sigma_c^4-\lambda}>\frac{B^2\frac{K_\ell\sigma_c^8\lambda}{K_\ell\sigma_c^4-\lambda}}{\left[(K_\ell\sigma_c^4/\lambda-1)x-K_\ell\sigma_c^4B\right]^2}\Leftrightarrow\left[(K_\ell\sigma_c^4/\lambda-1)x-K_\ell\sigma_c^4B\right]^2>B^2K_\ell\sigma_c^4\lambda,$$

for any x in $[1/\sigma^2+1/\sigma_c^2, K_\ell/(1/\sigma_\epsilon^2+1/\sigma_c^2)]$. Since the left hand side increases with x, it is equivalent to have the inequality with x being replaced with $1/\sigma^2+1/\sigma_c^2$. Then we have

$$\frac{(K_\ell\sigma_c^4-1)^2}{K_\ell\sigma_c^4}>\lambda(1/\lambda-1)^2.$$

Since $K_\ell\sigma_c^4>1$, $(K_\ell\sigma_c^4-1)^2/(K_\ell\sigma_c^4)$ increases with $K_\ell\sigma_c^4$ which itself increases with σ_c^2. Then it suffices for the inequality above to hold if it holds for the smallest $\sigma_c^2=\sigma^2/(2^{2\kappa}-1)$, that is,

$$\lambda(1/\lambda-1)^2<\frac{2^{4\kappa}\left(\frac{1}{2^{2\kappa}-1}+\frac{2^{2\kappa}\sigma^2}{(2^{2\kappa}-1)^2\sigma_\epsilon^2}\right)^2}{2^{2\kappa}\left(\frac{1}{2^{2\kappa}-1}+\frac{2^{2\kappa}\sigma^2}{(2^{2\kappa}-1)^2\sigma_\epsilon^2}\right)+1},$$

which coincides with regularity condition 12 for λ.

We now have shown that $G(1;x)>0$. This implies that $x^*=1/\sigma^2+1/\sigma_c^2$ or equivalently $\sigma_\ell^2=\infty$ when $\gamma=1$, which echoes the result in the baseline setting.

Since $G(1;x)>0$, we claim that there must exist a unique $\gamma\in(0,1)$ such that $G(\gamma;x)=0$ for any x in $[1/\sigma^2+1/\sigma_c^2, K_\ell/(1/\sigma_\epsilon^2+1/\sigma_c^2)]$. To see this, $G(\gamma;x)=0$ is equivalent to

$$L(\gamma)\equiv2K_\ell\sigma_c^6B\left(\frac{1-\gamma}{\gamma}\right)+K_\ell\sigma_c^4\left(\frac{K_\ell\sigma_c^4}{\lambda}-1\right)\left(\frac{1-\gamma}{\gamma}\right)^2$$
$$=\frac{\sigma_c^4\left[(K_\ell\sigma_c^4/\lambda-1)x-K_\ell\sigma_c^4B\right]^2}{K_\ell\sigma_c^4-\lambda}-B^2\frac{K_\ell\sigma_c^8\lambda}{K_\ell\sigma_c^4-\lambda}>0,$$

where the inequality follows from $G(1;x)>0$. Given $\gamma\in[0,1]$, it is easy to see that $L(\gamma)$ strictly decreases with γ and $L(1)=0$ and $\lim_{\gamma\to0}L(\gamma)\to\infty$. Therefore, there must exist a unique $\gamma\in(0,1)$ such that $G(\gamma;x)=0$. We now define $\underline{\gamma}'$

such that $G(\underline{\gamma}';x)=0$ for $x=K_\ell/(1/\sigma_\epsilon^2+1/\sigma_c^2)$ and $\bar{\gamma}'$ such that $G(\bar{\gamma}';x)=0$ for $x=1/\sigma^2+1/\sigma_c^2$. Since $1/\sigma^2+1/\sigma_c^2 < K_\ell/(1/\sigma_\epsilon^2+1/\sigma_c^2)$, by construction, we have $L(\underline{\gamma}')>L(\bar{\gamma}')$. Further, we know L is strictly decreasing, so $\underline{\gamma}'<\bar{\gamma}'$.

The rest of the proof directly follows from the proof of Lemma 4. When $\gamma\in(\underline{\gamma}',\bar{\gamma}')$, we have $H'(x^*)=0$, or equivalently

$$
\begin{aligned}
\frac{\gamma^2\sigma_c^4}{K_\ell\sigma_c^4-\lambda} &= \frac{B^2\frac{\gamma^2K_\ell\sigma_c^8\lambda}{K_\ell\sigma_c^4-\lambda}+2(1-\gamma)\gamma K_\ell\sigma_c^6B+(1-\gamma)^2K_\ell\sigma_c^4\left(\frac{K_\ell\sigma_c^4}{\lambda}-1\right)}{\left[(K_\ell\sigma_c^4/\lambda-1)x^*-K_\ell\sigma_c^4B\right]^2}\\
\left[(K_\ell\sigma_c^4-\lambda)x^*-\lambda K_\ell\sigma_c^4B\right]^2 &= \lambda\left(B^2K_\ell\lambda^2\sigma_c^4+\frac{2(1-\gamma)}{\gamma}BK_\ell\lambda\sigma_c^2(K_\ell\sigma_c^4-\lambda)+\left(\frac{1-\gamma}{\gamma}\right)^2K_\ell(K_\ell\sigma_c^4-\lambda)^2\right)\\
\left[(K_\ell\sigma_c^4-\lambda)x^*-\lambda K_\ell\sigma_c^4B\right]^2 &= \lambda K_\ell\left(B\lambda\sigma_c^2+\left(\frac{1-\gamma}{\gamma}\right)(K_\ell\sigma_c^4-\lambda)\right)^2\\
(K_\ell\sigma_c^4-\lambda)x^*-\lambda K_\ell\sigma_c^4B &= \lambda^{1/2}K_\ell^{1/2}\left(B\lambda\sigma_c^2+\left(\frac{1-\gamma}{\gamma}\right)(K_\ell\sigma_c^4-\lambda)\right)\\
x^* &= \frac{1-\gamma}{\gamma}(\lambda K_\ell)^{1/2}+\frac{\lambda BK_\ell^{1/2}\sigma_c^2(K_\ell^{1/2}\sigma_c^2+\lambda^{1/2})}{K_\ell\sigma_c^4-\lambda}\\
x^* &= \frac{1-\gamma}{\gamma}(\lambda K_\ell)^{1/2}+\frac{\lambda BK_\ell^{1/2}\sigma_c^2}{K_\ell^{1/2}\sigma_c^2-\lambda^{1/2}},
\end{aligned}
\tag{17}
$$

where the last equation nests Eq. 9 as a special case. □

A.24. Proof of theorem 3

We first prove a weaker counterpart of Proposition 2.

Lemma 20. *Under the decentralized regime, $E(a_\ell-\theta)^2$ strictly increases with σ_c^2 for γ such that $\gamma<\underline{\gamma}'$ or $\gamma>\bar{\gamma}'$.*

Proof. Following the proof of Lemma 5, we first obtain

$$
\begin{aligned}
E(a_\ell-\theta)^2 &= (1-\gamma^2)Var(\theta|\theta_\ell,\mathbf{s}_c')+\gamma^2Var(\theta_c|\theta_\ell,\mathbf{s}_c')+\gamma^2\sigma_c^2-2\gamma^2\sigma_c^2\frac{\frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}}{\frac{1}{\sigma_{\epsilon\ell}^2}+\frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}}\\
&= \frac{(1-\gamma^2)K_\ell\sigma_c^4}{K_\ell\sigma_c^4\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2}\right)-\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}\right)}+\frac{\gamma^2\sigma_c^4\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}\right)\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2}\right)}{K_\ell\sigma_c^4\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2}\right)-\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}\right)}\\
&\quad+\gamma^2\sigma_c^2-\frac{2\gamma^2\sigma_c^4\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}\right)\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}\right)}{K_\ell\sigma_c^4\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2}\right)-\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}\right)}\\
&= \frac{(1-\gamma^2)K_\ell\sigma_c^4+\gamma^2\sigma_c^4\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}\right)\left(\frac{1}{\sigma_c^2}-\frac{1}{\sigma^2}-\frac{1}{\sigma_\ell^2}\right)}{K_\ell\sigma_c^4\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2}\right)-\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}\right)}+\gamma^2\sigma_c^2
\end{aligned}
$$

where the second to last equality follows from the fact that Constraint 10 is binding. Then the optimization problem of the central government can be rewritten as

$$
\min_{\sigma_c^2}E(a_\ell-\theta)^2=\frac{(1-\gamma^2)K_\ell\sigma_c^4+\gamma^2\sigma_c^4\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}\right)\left(\frac{1}{\sigma_c^2}-\frac{1}{\sigma^2}-\frac{1}{\sigma_\ell^2}\right)}{K_\ell\sigma_c^4\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2}\right)-\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}\right)}+\gamma^2\sigma_c^2\equiv F(\sigma_c^2)
$$

subject to Constraint 4, where it should be emphasized that both σ_ℓ^2 and K_ℓ are functions of σ_c^2.

Consider two cases: (1) $\gamma>\bar{\gamma}'$; (2) $\gamma<\underline{\gamma}'$.

Case (1): $\gamma > \bar{\gamma}'$. In this case, according to Lemma 4, $\sigma_\ell^2 = \infty$. Then we have

$$F(\sigma_c^2) = \frac{K_\ell + K_\ell \gamma^2 \sigma_c^2/\sigma^2 - \gamma^2(1/\sigma^2 + 1/\sigma_c^2)/\sigma^2}{(K_\ell - 1/\sigma_c^4)(1/\sigma^2 + 1/\sigma_c^2)},$$

which does not depend on λ and the objective function coincides with that in the baseline setting. Then we know $F'(\sigma_c^2) > 0$.

Case (2): $\gamma < \underline{\gamma}'$. According to Lemma 4, we have $\sigma_{\epsilon\ell}^2 = \sigma_\epsilon^2$, which implies that Constraint 5 can be rewritten as

$$\frac{1}{\sigma^2} + \frac{\lambda}{\sigma_\ell^2} + \frac{1}{\sigma_c^2} = K_\ell \left(\frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma_c^2} \right)^{-1}.$$

The objective function of the central government can then be written as

$$\begin{aligned}
F(\sigma_c^2) &= \frac{(1-\gamma^2)K_\ell\sigma_c^4 + \frac{\gamma^2\sigma_c^4 K_\ell}{\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_c^2}} \left(\frac{1}{\sigma_c^2} - \frac{1}{\sigma^2} - \frac{1}{\lambda}\left(\frac{K_\ell}{\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_c^2}} - \frac{1}{\sigma^2} - \frac{1}{\sigma_c^2} \right) \right)}{K_\ell\sigma_c^4 \left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2} + \frac{1}{\lambda}\left(\frac{K_\ell}{\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_c^2}} - \frac{1}{\sigma^2} - \frac{1}{\sigma_c^2} \right) \right) - \frac{K_\ell}{\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_c^2}}} + \gamma^2\sigma_c^2 \\
&= \frac{(1-\gamma^2)K_\ell\sigma_c^4\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\epsilon^2}\right)^2 + \gamma^2\sigma_c^4K_\ell\left(\left(\frac{1+\lambda}{\lambda\sigma_c^2}+\frac{1-\lambda}{\lambda\sigma^2}\right)\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\epsilon^2}\right) - \frac{K_\ell}{\lambda}\right)}{K_\ell\sigma_c^4\left(\frac{\lambda-1}{\lambda}\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}\right)\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\epsilon^2}\right)^2 + \frac{K_\ell}{\lambda}\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\epsilon^2}\right)\right) - K_\ell\left(\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_c^2}\right)} + \gamma^2\sigma_c^2 \\
&= \frac{(1-\gamma^2)\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\epsilon^2}\right)^2 + \gamma^2\left(\left(\frac{1+\lambda}{\lambda\sigma_c^2}+\frac{1-\lambda}{\lambda\sigma^2}\right)\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\epsilon^2}\right) - \frac{K_\ell}{\lambda}\right)}{\left(\frac{K_\ell}{\lambda}-\frac{1}{\sigma_c^4}\right)\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\epsilon^2}\right) - \frac{1-\lambda}{\lambda}\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}\right)\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\epsilon^2}\right)^2} + \gamma^2\sigma_c^2 \\
&\equiv (1-\gamma^2)F_1(\sigma_c^2) + \gamma^2 F_2(\sigma_c^2)
\end{aligned}$$

with

$$\begin{aligned}
F_1(\sigma_c^2) &\equiv \frac{\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\epsilon^2}\right)^2}{\left(\frac{K_\ell}{\lambda}-\frac{1}{\sigma_c^4}\right)\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\epsilon^2}\right) - \frac{1-\lambda}{\lambda}\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}\right)\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\epsilon^2}\right)^2} \\
&= \left(\frac{\frac{K_\ell}{\lambda}-\frac{1}{\sigma_c^4}}{\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\epsilon^2}} - \frac{1-\lambda}{\lambda}\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}\right) \right)^{-1} \\
&= \frac{\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\epsilon^2}}{\frac{2^{2\kappa}-1+\lambda}{\lambda}\left(\frac{\sigma_c^2+\sigma^2+\sigma_\epsilon^2}{\sigma_c^2\sigma^2\sigma_\epsilon^2}\right)} = \frac{\lambda\sigma^2(\sigma_c^2+\sigma_\epsilon^2)}{(2^{2\kappa}-1+\lambda)(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)}, \\
F_2(\sigma_c^2) &\equiv \frac{\left(\frac{1+\lambda}{\lambda\sigma_c^2}+\frac{1-\lambda}{\lambda\sigma^2}\right)\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\epsilon^2}\right) - \frac{K_\ell}{\lambda}}{\left(\frac{K_\ell}{\lambda}-\frac{1}{\sigma_c^4}\right)\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\epsilon^2}\right) - \frac{1-\lambda}{\lambda}\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}\right)\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\epsilon^2}\right)^2} + \sigma_c^2 \\
&= \frac{\sigma_\epsilon^2[(1+\lambda)\sigma^2 + (1-\lambda)\sigma_c^2]\big(\sigma_c^2+\sigma_\epsilon^2\big) - K_\ell\sigma^2\sigma_\epsilon^4\sigma_c^4}{(2^{2\kappa}-1+\lambda)(\sigma_c^2+\sigma_\epsilon^2+\sigma^2)(\sigma_c^2+\sigma_\epsilon^2)} + \sigma_c^2
\end{aligned}$$

It is easy to see that $F_1'(\sigma_c^2) > 0$ and

$$\begin{aligned}
F_2'(\sigma_c^2) = 1 &+ \frac{\sigma_\epsilon^2[(1+\lambda)\sigma^2 + (1-\lambda)\sigma_\epsilon^2 + 2(1-\lambda)\sigma_c^2] - 2^{2\kappa}\sigma_\epsilon^2(\sigma^2+2\sigma_c^2+\sigma_\epsilon^2)}{(2^{2\kappa}-1+\lambda)(\sigma_c^2+\sigma_\epsilon^2+\sigma^2)(\sigma_c^2+\sigma_\epsilon^2)} \\
&- \frac{\{\sigma_\epsilon^2[(1+\lambda)\sigma^2+(1-\lambda)\sigma_c^2]\big(\sigma_c^2+\sigma_\epsilon^2\big) - K_\ell\sigma^2\sigma_\epsilon^4\sigma_c^4\}(2\sigma_c^2+2\sigma_\epsilon^2+\sigma^2)}{(2^{2\kappa}-1+\lambda)(\sigma_c^2+\sigma_\epsilon^2+\sigma^2)^2(\sigma_c^2+\sigma_\epsilon^2)^2} \\
= &\frac{2\lambda\sigma^2\sigma_\epsilon^2 + (2^{2\kappa}-1+\lambda)\sigma_c^2(\sigma_c^2+\sigma^2)}{(2^{2\kappa}-1+\lambda)(\sigma_c^2+\sigma_\epsilon^2+\sigma^2)(\sigma_c^2+\sigma_\epsilon^2)} \\
&- \frac{\sigma_\epsilon^2\{[(1+\lambda)\sigma^2+(1-\lambda)\sigma_c^2]\big(\sigma_c^2+\sigma_\epsilon^2\big) - K_\ell\sigma^2\sigma_\epsilon^2\sigma_c^4\}(2\sigma_c^2+2\sigma_\epsilon^2+\sigma^2)}{(2^{2\kappa}-1+\lambda)(\sigma_c^2+\sigma_\epsilon^2+\sigma^2)^2(\sigma_c^2+\sigma_\epsilon^2)^2}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\sigma_c^2(\sigma_c^2+\sigma^2)}{(\sigma_c^2+\sigma_\epsilon^2+\sigma^2)(\sigma_c^2+\sigma_\epsilon^2)} + \frac{2\lambda\sigma^4\sigma_\epsilon^2}{(2^{2\kappa}-1+\lambda)(\sigma_c^2+\sigma_\epsilon^2+\sigma^2)^2(\sigma_c^2+\sigma_\epsilon^2)} \\
&\quad + \frac{\sigma_\epsilon^2\{K_\ell\sigma^2\sigma_\epsilon^2\sigma_c^4-[\sigma^2+(1-\lambda)\sigma_c^2]\left(\sigma_c^2+\sigma_\epsilon^2\right)\}(2\sigma_c^2+2\sigma_\epsilon^2+\sigma^2)}{(2^{2\kappa}-1+\lambda)(\sigma_c^2+\sigma_\epsilon^2+\sigma^2)^2(\sigma_c^2+\sigma_\epsilon^2)^2} \\
&> \frac{\sigma_c^2(\sigma_c^2+\sigma^2)}{(\sigma_c^2+\sigma_\epsilon^2+\sigma^2)(\sigma_c^2+\sigma_\epsilon^2)} + \frac{2\lambda\sigma^4\sigma_\epsilon^2}{(2^{2\kappa}-1+\lambda)(\sigma_c^2+\sigma_\epsilon^2+\sigma^2)^2(\sigma_c^2+\sigma_\epsilon^2)} \\
&\quad + \frac{\sigma_\epsilon^4\sigma^2\sigma_c^4\left\{K_\ell-\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}\right)\left(\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\epsilon^2}\right)\right\}(2\sigma_c^2+2\sigma_\epsilon^2+\sigma^2)}{(2^{2\kappa}-1+\lambda)(\sigma_c^2+\sigma_\epsilon^2+\sigma^2)^2(\sigma_c^2+\sigma_\epsilon^2)^2} > 0,
\end{aligned}
$$

where the first inequality follows from $\lambda > 0$ and the second inequality follows from $K_\ell > (1/\sigma^2 + 1/\sigma_c^2)(1/\sigma_c^2 + 1/\sigma_\epsilon^2)$. Since $\gamma < \underline{\gamma}$ continues to hold for a small change of σ_c^2, the objective function is differentiable and its first derivative is given by $F'(\sigma_c^2) = (1-\gamma^2)F_1'(\sigma_c^2) + \gamma^2 F_2'(\sigma_c^2) > 0$. We thus have obtained the desired conclusion. □

Notice that $\underline{\gamma}'$ and $\bar{\gamma}'$ are functions of σ_c^2. Although according to the simulation, $E(a_\ell-\theta)^2$ appears to strictly increase with σ_c^2 for $\gamma \in [\underline{\gamma}', \bar{\gamma}']$, we are not able to formally establish this result. Since we know $0 < \underline{\gamma}' < \bar{\gamma}' < 1$, we then have a partial characterization of the optimal strategy of the central government.

Corollary 3. *Let $0 < \lambda < 1$. Under the decentralized regime, the central government devotes itself to information acquisition with $\sigma_c^2 = \sigma^2/(2^{2\kappa}-1)$ if $\gamma = 0$ or $\gamma = 1$. Moreover, we have*

$$
\begin{aligned}
E(a_c-\theta)^2 &[> E(a_\ell-\theta)^2\Big|_{\gamma=0}, \\
E(a_c-\theta)^2 &< E(a_\ell-\theta)^2\Big|_{\gamma=1}.
\end{aligned}
$$

Proof. We first consider $\gamma = 0$. We have

$$
E(a_\ell-\theta)^2\Big|_{\gamma=0} = \left(\frac{1}{\sigma^2} + \frac{\lambda(2^{2\kappa}-1)(\sigma^2+\sigma_\epsilon^2+\sigma_c^2)}{\sigma^2(\sigma_\epsilon^2+\sigma_c^2)} + \frac{1}{\sigma_c^2+\sigma_\epsilon^2}\right)^{-1},
$$

with $\sigma_c^2 = \sigma^2/(2^{2\kappa}-1)$.

$$
E(a_c-\theta)^2 = \left(\frac{1}{\sigma^2} + \frac{(2^{2\kappa}-1)(\sigma^2+\sigma_\epsilon^2+\sigma_\ell^2)}{\sigma^2(\sigma_\epsilon^2+\sigma_\ell^2)} + \frac{1}{\sigma_\ell^2+\sigma_\epsilon^2}\right)^{-1},
$$

with $\sigma_\ell^2 = \lambda\sigma^2/(2^{2\kappa}-1)$. Then to show $E(a_c-\theta)^2 > E(a_\ell-\theta)^2\Big|_{\gamma=0}$, it is equivalent to show

$$
\begin{aligned}
&\frac{1}{\sigma^2} + \frac{(2^{2\kappa}-1)(\sigma^2+\sigma_\epsilon^2+\sigma_\ell^2)}{\sigma^2(\sigma_\epsilon^2+\sigma_\ell^2)} + \frac{1}{\sigma_\ell^2+\sigma_\epsilon^2} < \frac{1}{\sigma^2} + \frac{\lambda(2^{2\kappa}-1)(\sigma^2+\sigma_\epsilon^2+\sigma_c^2)}{\sigma^2(\sigma_\epsilon^2+\sigma_c^2)} + \frac{1}{\sigma_c^2+\sigma_\epsilon^2} \\
\Leftrightarrow\ &\frac{\sigma^2+\sigma_\epsilon^2+\sigma_\ell^2+\sigma_c^2}{\sigma_c^2(\sigma_\epsilon^2+\sigma_\ell^2)} < \frac{\sigma^2+\sigma_\epsilon^2+\sigma_\ell^2+\sigma_c^2}{\sigma_\ell^2(\sigma_\epsilon^2+\sigma_c^2)},
\end{aligned}
$$

where the second inequality directly follows from $\lambda < 1$ (i.e., $\sigma_\ell^2 < \sigma_c^2$).

Now consider $\gamma = 1$. The expression of $E(a_\ell-\theta)^2\Big|_{\gamma=1}$ coincides with that in the baseline setting. Following the proof of Lemma 7, we can show that $E(a_\ell-\theta)^2\Big|_{\gamma=1} > \min\{\sigma^2, \sigma_c^2\} > E(a_c-\theta)^2$. □

Theorem 3 directly follows from the above result and continuity of $E(a_\ell-\theta)^2$ with respect to γ.

Appendix B. Additional Figures

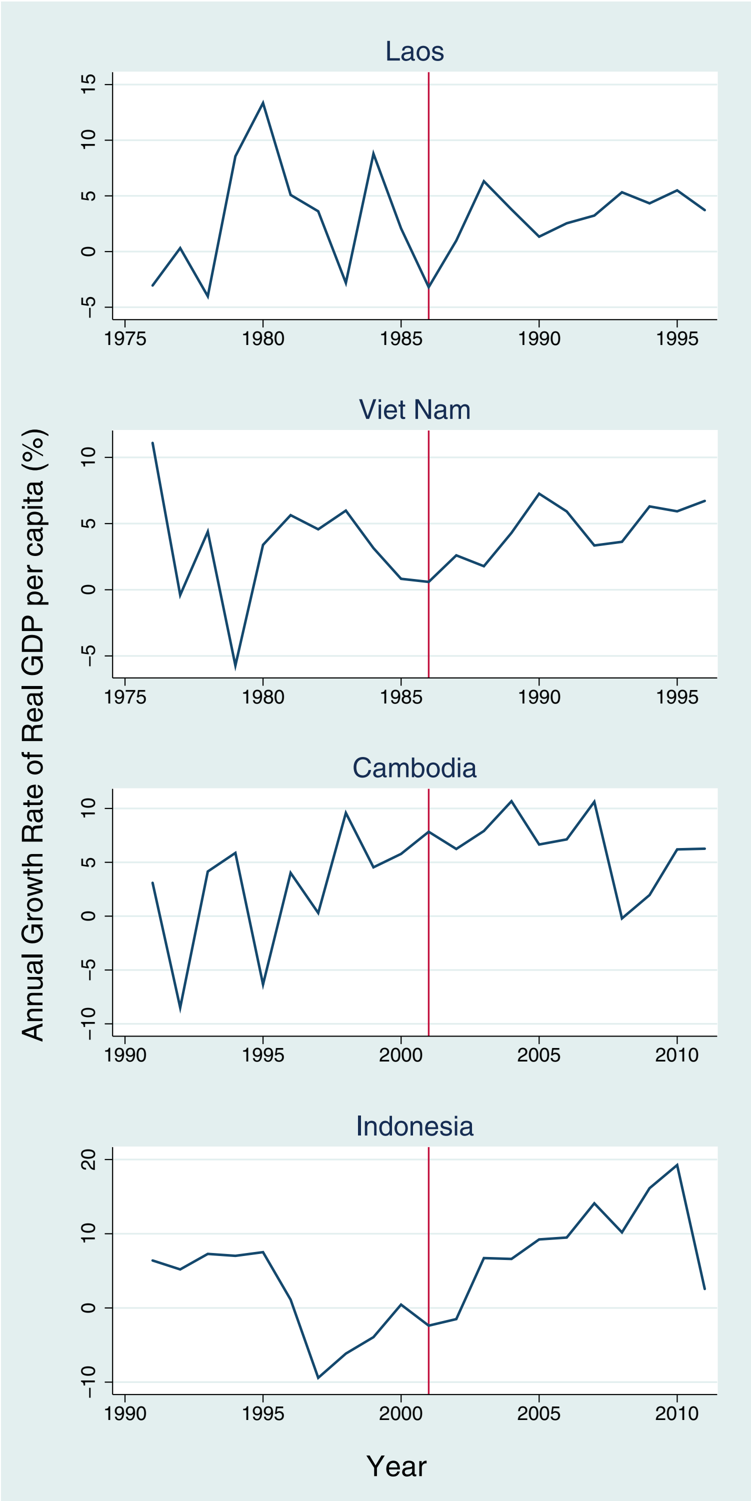


Fig. 8. Decentralization and Economic Growth.
Notes: (1) Data Source: Penn World Table 9.0 (Feenstra et al., 2015); (2) Our calculation is based on the expenditure-side real GDP at chained PPPs in 2011 US dollars; (3) The red line indicates the start year of the decentralization reform. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

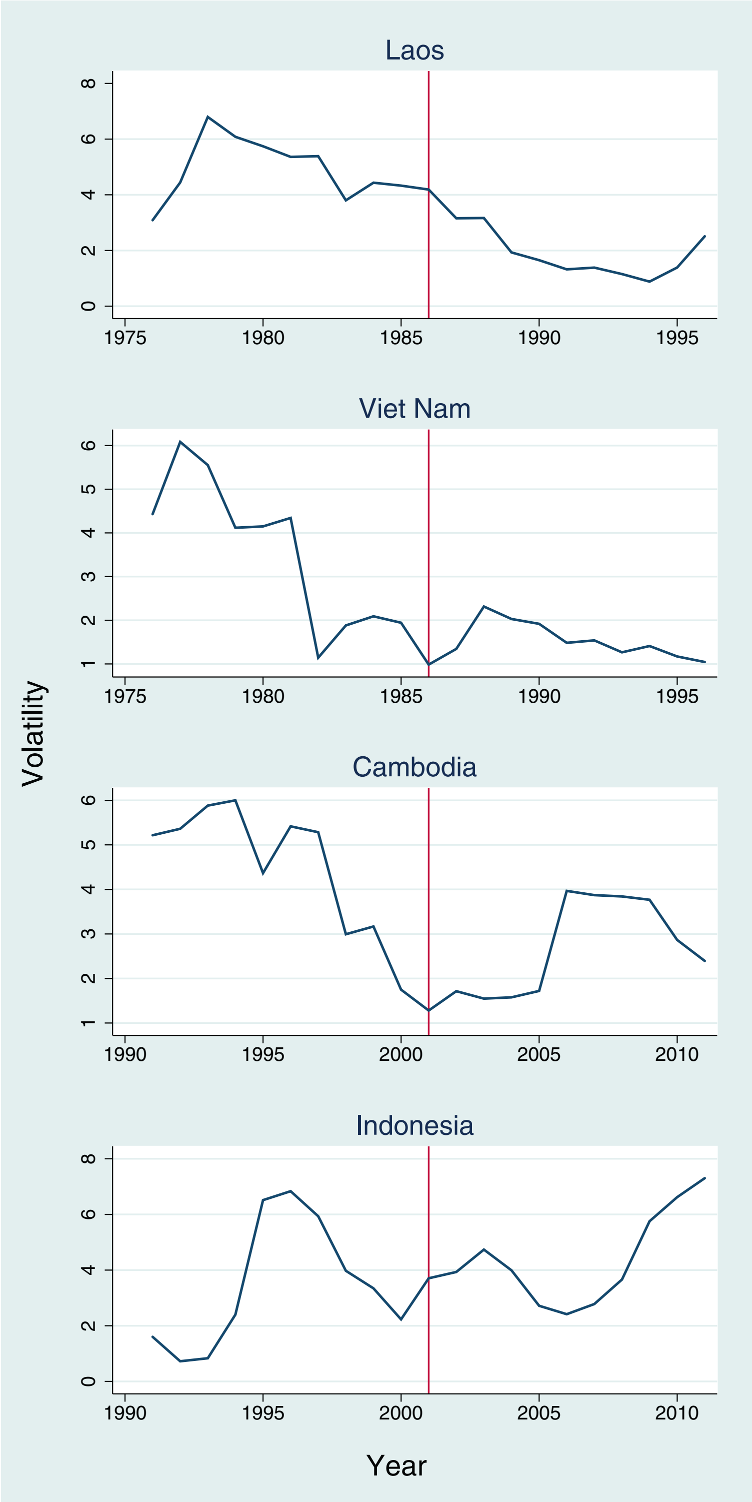


Fig. 9. Decentralization and Output Volatility *Notes:* Volatility is measured as the standard deviation of the growth rate of real GDP per capita for a five-year moving window.

References

Acemoglu, D., Johnson, S., Robinson, J.A., 2005. Institutions as a fundamental cause of long-run growth, 1. Elsevier, pp. 385–472.
Akai, N., Hosoi, M., Nishimura, Y., 2009. Fiscal decentralization and economic volatility: evidence from state-level cross-section data of the usa. Jpn. Econ. Rev. 60 (2), 223–235.
Ashworth, S., 2012. Electoral accountability: recent theoretical and empirical work. Annu. Rev. Polit. Sci. 15, 183–201.
Bardhan, P., 2002. Decentralization of governance and development. J. Econ. Perspect. 16 (4), 185–205.
Bardhan, P., 2016. State and development: the need for a reappraisal of the current literature. J. Econ. Lit. 54 (3), 862–892.
Bellofatto, A.A., Besfamille, M., 2018. Regional state capacity and the optimal degree of fiscal decentralization. J. Public Econ. 159, 225–243.
Besley, T., Coate, S., 2003. Centralized versus decentralized provision of local public goods: a political economy approach. J. Public Econ. 87 (12), 2611–2637.
Besley, T., Persson, T., 2014. The causes and consequences of development clusters: state capacity, peace, and income. Annu. Rev. Econ. 6 (1), 927–949.
Bo, S., 2020. Centralization and regional development: evidence from a political hierarchy reform to create cities in china. J. Urban Econ. 115, 103182.
Board, O.J., Blume, A., Kawamura, K., 2007. Noisy talk. Theoret. Econ. 2 (4), 395–440.
Bolton, P., Brunnermeier, M.K., Veldkamp, L., 2012. Leadership, coordination, and corporate culture. Rev. Econ. Stud. 80 (2), 512–537.
Bowie, R.R., 1962. Communist China 1955–1959: Policy Documents with Analysis, 1. Harvard University Press.
Cai, H., Treisman, D., 2004. State corroding federalism. J. Public Econ. 88 (3–4), 819–843.
Che, J., Chung, K.-S., Lu, Y.K., 2017. Decentralization and political career concerns. J. Public Econ. 145, 201–210.
Chen, T., Kung, J.K.-S., Ma, C., 2020. Long live keju! the persistent effects of China's civil examination system. Econ. J. 130 (161), 2030–2064.
Cheng, C., Li, C., 2019. Laboratories of democracy: policy experimentation under decentralization. Am. Econ. J.: Microecon. 11 (3), 125–154.
Cheng, C., Li, C., Zhong, W., 2018. Governmental system and economic volatility under democracy. Working Paper.
Cover, T.M., Thomas, J.A., 2012. Elements of Information Theory. John Wiley & Sons.
Crawford, V.P., Sobel, J., 1982. Strategic information transmission. Econometrica 50 (6), 1431–1451.
Dessein, W., Galeotti, A., Santos, T., 2016. Rational inattention and organizational focus. Am. Econ. Rev. 106 (6), 1522–1536.
Dincecco, M., 2011. Political Transformations and Public Finances: Europe, 1650–1913. Cambridge University Press.
Egorov, G., Sonin, K., 2011. Dictators and their viziers: endogenizing the loyalty–competence trade-off. J. Eur. Econ. Assoc. 9 (5), 903–930.
Enikolopov, R., Zhuravskaya, E., 2007. Decentralization and political institutions. J. Public Econ. 91 (11–12), 2261–2290.
Fan, Z., Xiong, W., Zhou, L.-A., 2016. Information distortion in hierarchical organization: a study of China's great famine. Working Paper.
Feenstra, R.C., Inklaar, R., Timmer, M.P., 2015. The next generation of the Penn World Table. Am. Econ. Rev. 105 (10), 3150–3182.
Fisman, R., Shi, J., Wang, Y., Wu, W., 2020. Social ties and the selection of China's political elite. Am. Econ. Rev. 110 (6), 1752–1781.
Hale-Dorrell, A.T., 2014. Khrushchev's Corn Crusade: The Industrial Ideal and Agricultural Practice in the Era of Post-Stalin Reform, 1953–1964. University of North Carolina at Chapel Hill.
Hayek, F.A., 1945. The use of knowledge in society. Am. Econ. Rev. 35 (4), 519–530.
Hoffman, P.T., 2015. Why did Europe Conquer the World?. Princeton University Press.
Huang, Z., Li, L., Ma, G., Xu, L.C., 2017. Hayek, local information, and commanding heights: decentralizing state-owned enterprises in China. Am. Econ. Rev. 107 (8), 2455–2478.
Jia, R., Kudamatsu, M., Seim, D., 2015. Political selection in China: the complementary roles of connections and performance. J. Eur. Econ. Assoc. 13 (4), 631–668.
Jin, H., Qian, Y., Weingast, B.R., 2005. Regional decentralization and fiscal incentives: federalism, Chinese style. J. Public Econ. 89 (9–10), 1719–1742.
Karaman, K.K., Pamuk, Ş., 2013. Different paths to the modern state in europe: the interaction between warfare, economic structure, and political regime. Am.Polit. Sci. Rev. 107 (3), 603–626.
Kassum, J.-u.-d., Steer, A., Kharas, H., Hofman, B., 2003. Decentralizing indonesia. The World Bank Report.
Keen, M., Marchand, M., 1997. Fiscal competition and the pattern of public spending. J. Public Econ. 66 (1), 33–53.
Ko, C.Y., Koyama, M., Sng, T.-H., 2018. Unified China and divided Europe. Int. Econ. Rev. (Philadelphia) 59 (1), 285–327.
Koyama, M., Moriguchi, C., Sng, T.-H., 2018. Geopolitics and Asia's little divergence: state building in China and Japan after 1850. J. Econ. Behav. Org. 155, 178–204.
Kung, J.K.-S., Chen, S., 2011. The tragedy of the nomenklatura: career incentives and political radicalism during China's great leap famine. Am. Polit. Sci. Rev. 105 (1), 27–45.
Li, H., Zhou, L.-A., 2005. Political turnover and economic performance: the incentive role of personnel control in China. J. Public Econ. 89 (9), 1743–1762.
Li, W., Yang, D.T., 2005. The great leap forward: anatomy of a central planning disaster. J. Polit. Econ. 113 (4), 840–877.
Lin, J.Y., 1990. Collectivization and china's agricultural crisis in 1959–1961. J. Polit. Econ. 98 (6), 1228–1252.
Lin, J.Y., Tao, R., Liu, M., 2006. Decentralization and Local Governance in China's Economic Transition. In: Bardhan, P., Mookherjee, D. (Eds.), Decentralization and Local Governance in China's Economic Transition. MIT Press. Cambridge, MA, pp. 305–327.
Lu, Y., Luan, M., Sng, T.-H., 2020. Did the communists contribute to china's rural growth? Explor. Econ. Hist. 75, 101315.
Ma, D., Rubin, J., 2019. The paradox of power: principal-agent problems and administrative capacity in imperial china (and other absolutist regimes). J. Comp. Econ. 47 (2), 277–294.
Mackowiak, B., Wiederholt, M., 2009. Optimal sticky prices under rational inattention. Am. Econ. Rev. 99 (3), 769–803.
Mann, M., 1984. The autonomous power of the state: its origins, mechanisms and results. Eur. J. Sociol./Arch. Européennes de Sociol./Europäisches Arch. Soziol. 25 (2), 185–213.
Matějka, F., Tabellini, G., 2020. Electoral competition with rationally inattentive voters. J. Eur. Econ. Assoc. forthcoming.
Meisner, M., 1978. Dazhai: the mass line in practice. Mod. China 4 (1), 27–62.
Michalopoulos, S., Papaioannou, E., 2013. Pre-colonial ethnic institutions and contemporary african development. Econometrica 81 (1), 113–152.
Mookherjee, D., 2015. Political decentralization. Annu. Rev. Econom. 7 (1), 231–249.
Nishimura, Y., 2006. Human fallibility, complementarity, and fiscal decentralization. J. Public Econ. Theory 8 (3), 487–501.
North, D.C., 1990. Institutions, Institutional Change and Economic Performance. Cambridge University Press.
Oates, W.E., 1972. Fiscal Federalism. Edward Elgar Publishing.
Pomeranz, K., 2000. The Great Divergence: China, Europe, and the Making of the Modern World Economy. Princeton University Press.
Prendergast, C., 1993. A theory of "yes men". Am. Econ. Rev. 757–770.
Qian, Y., Roland, G., 1998. Federalism and the soft budget constraint. Am. Econ. Rev. 88 (5), 1143–1162.
Rosenthal, J.-L., Wong, R.B., 2011. Before and Beyond Divergence. Harvard University Press.
Seabright, P., 1996. Accountability and decentralisation in government: an incomplete contracts model. Eur. Econ. Rev. 40 (1), 61–89.
Sims, C.A., 2003. Implications of rational inattention. J. Monet. Econ. 50 (3), 665–690.
Sng, T.-H., 2014. Size and dynastic decline: the principal-agent problem in late imperial china, 1700–1850. Explor. Econ. Hist. 54, 107–127.
St John, R.B., 1997. End of the beginning: economic reform in cambodia, laos, and vietnam. Contemp. Southeast Asia 19 (2), 172.
Stuart-Fox, M., 2005. Politics and reform in the lao people's democratic republic. Working Paper.
Suzuki, Y., 2019. A contract theory analysis to fiscal relations between the central and local governments in china. Econ. Polit. Stud. 7 (3), 281–313.
Svolik, M.W., 2012. The Politics of Authoritarian Rule. Cambridge University Press.
Tiebout, C.M., 1956. A pure theory of local expenditures. J. Polit. Econ. 64 (5), 416–424.
Tilly, C., 1990. Coercion, Capital, and European States, AD 990–1992. Blackwell Oxford.
Un, K., Ledgerwood, J., 2003. Cambodia in 2002: decentralization and its effects on party politics. Asian Surv. 43 (1), 113–119.

Wang, B., Yang, T., 2016. Volatility and Economic Systems: Evidence from a Large Ttransitional Economy. University of Virginia Working Paper.
Weingast, B.R., 1995. The economic role of political institutions: market-preserving federalism and economic development. J.f Law, Econ. Org. 11 (1), 1–31.
Weitzman, M.L., 1976. The new soviet incentive model. Bell J. Econ. 7 (1), 251–257.
Wong, R.B., 1997. China Transformed: Historical Change and the Limits of European Experience. Cornell University Press.
World Bank, 1993. Viet Nam Transition to the Market: An Economic Report. The World Bank Official Report.
World Bank, 2015. Cambodia: description of legislation/policy. The World Bank Country Summary Brief.
Wu, G., 1995. Documentary politics: hypotheses, process, and case studies. Decision-Making in Deng's China, Perspectives from Insiders. M.E. Sharpe.
Wu, J., Reynolds, B.L., 1988. Choosing a strategy for china's economic reform. Am. Econ. Rev. 78 (2), 461–466.
Wu, L., 2019. Partially directed search in the labor market. Working Paper.
Xu, C., 2011. The fundamental institutions of china's reforms and development. J. Econ. Lit. 49 (4), 1076–1151.
Xue, M. M., 2021. Autocratic rule and social capital: evidence from imperial china. Working Paper.
Yang, J., 2009. Mu bei (Tombstone: The Great Famine). Tian Di Tu Shu.
Yap, O.F., 2018. Politics or the economy? weak economic performance and political support in east and southeast asia. Econ. Polit. Stud. 6 (1), 11–29.
Zhang, Q., Liu, M., 2016. Quan li jie gou, zheng zhi ji li, he jing ji zeng zhang (power structure, political incentives, and economic growth). Ge Zhi.
Zhang, Q., Liu, M., Shih, V., 2013. Guerrilla capitalism: revolutionary legacy, political cleavage, and the preservation of the private economy in zhejiang. J. East Asian Stud. 13 (3), 379–407.
Zhou, T. (Ed.), 1984. Dangdai Zhongguo de Jing ji Tizhi Gaige. China Social Science Press.
Zhuravskaya, E.V., 2000. Incentives to provide local public goods: fiscal federalism, russian style. J. Public Econ. 76 (3), 337–368.