On Intergovernmental Communication: A Tale of Two Decentralization Reforms[*]

Shiyu Bo,[†] Liuchun Deng,[‡] Yufeng Sun, [§] Boqun Wang[¶]

October 27, 2018

Abstract: Motivated by the contrasting experience following the two waves of decentralization in China, we develop a formal 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, the political incentive to follow the policy prescriptions from the central. We show that the loyalty concern impacts the economic outcome of decentralization by distorting inter-governmental transmission of information as well as final policy-making. A strict adherence to the central could render decentralization welfare-reducing, causing low output and high volatility. We offer several theoretical extensions to highlight the underlying mechanisms and demonstrate the robustness of our results.

Keywords: Decentralization, authoritarian regime, output, volatility, communication

JEL Classification Numbers: H10, H70, O50, P20.

[*]We are grateful to Ying Chen for helpful conversations. We thank Pravin Krishna, Yu-Hsiang Lei, Jiwei Qian, Zhiyong Yao, Sezer Yasar, Miaojie Yu, our discussant, Vanesa Pesqué-Cela, and participants at Georgetown Center for Economic Research 2017 Conference, Econometric Society Asian Meeting, Econometric Society China Meeting, Henan University Symposium on Institutional and Development Economics, CES North America Conference, International Workshop on Economic Analysis of Institutions at Xiamen University, International Conference on Game Theory at Stony Brook University, and SCERI-NSD Conference for helpful comments. This paper subsumes an earlier version entitled "A Tale of Two Decentralizations: Volatility and Economic Regimes". All errors are ours.

[†]Institute for Economic and Social Research, Jinan University.

[‡]Halle Institute for Economic Research (IWH) and Friedrich Schiller University Jena.

[§]Corresponding author. School of Public Economics and Administration, Shanghai University of Finance and Economics, Shanghai, 200433, China, **Email**: sun.yufeng@mail.shufe.edu.cn.

[¶]School of Banking & Finance, University of International Business and Economics; China Financial Policy Research Center, Renmin University of China.

1 Introduction

Local information has long been argued as a key reason for decentralization (Hayek, 1945).[1] It is empirically identified as a driving force of decentralization in the reform era of China, a period in which the country undergoes very large-scale decentralization of decision making (Huang et al., 2017). Coupled with local competition of economic performance (Li and Zhou, 2005), the information advantage of the local governments is one of the main explanations of why China's reform in 1978 was so successful. In sharp contrast with the huge success of the 1978 reform is China's first major decentralization reform in the late 1950s. This wave of decentralization, which seen by outsider observers appears to share many ingredients with the 1978 reform, produced disastrous outcomes including the great famine, claiming millions of lives (Wu and Reynolds, 1988).

Why these two decentralization reforms led to drastically different results? More specifically, why local information seems to be of little use, if not misused, in the 1950s? The existing work suggests that career incentives of local officials played a crucial role. Since political loyalty paid off, officials became blind followers rather critics of the wishful thinking at the very top (Kung and Chen, 2011; Li and Yang, 2005), and therefore, the benefits of local information cannot be 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. More fundamentally, it does not clarify the nature of how career incentives, loyalty concern in particular,[2] distorts acquisition and use of local information in an authoritarian government.[3] This paper attempts to fill this void.

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 possibly more broadly, the mixed outcomes of decentralization in authoritarian countries during 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 the decision making of the local governments. In the extreme case like China's 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 perhaps being less appreciated is the second channel: Loyalty concern alters endogenous allocation of efforts between information acquisition and transmission. The information advantage of being local, as forcefully argued by Hayek (1945), could completely be squandered when the local bureaucrats are strongly motivated to decipher policy documents from the central rather than acquire useful information about the economy.[4] The success of China's 1978 reform can then be attributed to the bundling of promotion with economic performance. The incentive of the local bureaucrats to signal loyalty by

[1]Decentralization, being political or economic, has become a catchword in the discussion about structural reforms in the developing world. As one of most prominent features in their policy reform packages, many emerging market countries, transition economies formerly or continuously under the authoritarian regime in particular, have decentralized, to various degrees, their economic decision-making (Wetzel, 2001; Gadenne and Singhal, 2014).

[2]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 be viewed important in promotion, but due to fiscal decentralization, ideological shifts, and various economic motivations, its relative weight became smaller.

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

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

sacrificing the economy through either channel has been substantially dampened.

To be sure, we are not the first to juxtapose these two reforms in China and examine the nexus between political career incentive and decentralization. Based on similar motivation, Che et al. (2017) construct an overlapping generation model to characterizes the two-way relationship between decentralization and career concern. Abstracting from the dynamic aspects and taking loyalty concern as a parameter, our model opens up the black box of information transmission between the local and central governments, one of the fundamental mechanisms we believe, which governs the functioning of decentralization in a non-democracy.

In our model, a central government and a local government engage in policymaking subject to uncertainty. Decentralization shifts the decision power of policy-making from the central government to the local who 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. It first decides how to allocate its resources between directly acquiring information from the economy and indirectly seeking policy advice from the central government.[5] Given the information it has obtained, the local government then makes the policy decision. We demonstrate that the economic outcome of decentralization depends crucially on the degree of the loyalty concern. 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.

To check the robustness of our model prediction, we provide several 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. This means the distortion on the margin of information processing itself is not sufficient to generate differential outcomes of decentralization. Thus, the inter-governmental communication channel is essential to generating 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 own 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. Our second and third extensions further weaken the assumptions in the baseline model, but our main results still hold.

Our work joins a long-standing debate over centralization versus decentralization.[6] In a seminal work, Tiebout (1956) first pointed 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. The trade-off between conflicts of interests under centralization and externality problems under decentralization is further formalized in a political-economic framework (Besley and Coate, 2003). 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

[5]Formally, we model the decision problem 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.

[6]For a recent review, see Bardhan (2016).

paper contributes to this literature by offering a new perspective of information transmission with a particular focus on the authoritarian regime. 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). 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. Our framework could serve as a building block for a fully-fledged information-based theory of decentralization.

The empirical work on the consequences of decentralization predominantly focuses on the level terms with relatively limited causal evidence (Mookherjee, 2015). It is until very recently that a burgeoning literature starts to document the relationship between decentralization and volatility (Akai et al., 2009; Wang and Yang, 2016; Cheng et al., 2018), which is partly due to the lack of theoretical underpinnings.[7] One advantage of our information-based framework is that the economic performance can be both measured by the output level and volatility. It has the potential to understand the consequences of decentralization beyond the first moment.[8]

Our model adds to the literature on information transmission in hierarchical organizations. According to the "yes-man" theory of Prendergast (1993), an incentive contract could endogenously give rise to inefficient conformity of subordinates to the leaders. We take the distorted incentive of the subordinates as our model primitives and examine its implications beyond the scope of profit-maximizing firms. Also close in spirit to our work is Aghion and Tirole (1997) which distinguishes two types of authorities in an organization: formal authority with the decision power and informal authority with the implementation power. Our model shares several similar ingredients with their principal-agent framework but with a main departure. Instead of having a first-order difference in preferred outcomes, the conflict of interests in our model lies in the loyalty concern of the local bureaucrats as opposed to a benevolent central government. Hence, one major risk of decentralization is its potential of resulting in excessive efforts made to inter-governmental communication.

The rest of this paper is organized as follows. Section 2 discusses the motivating evidence from China to set up an institutional background of our theoretical framework. It also briefly covers the cross country experience of decentralization. Section 3 describes the baseline model. Section 4 presents the main results of the baseline model. Section 5 discusses theoretical extensions. Section 6 offers concluding remarks.

2 The Motivating Evidence

In this section, we present a case study of China to motivate our theoretical framework.

Ever since its establishment in 1949, People's Republic of China 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

[7] 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.

[8] In this sense, our work also contributes to a large literature that investigates the relationship between volatility and development (Ramey and Ramey, 1995; Koren and Tenreyro, 2007, 2013).

a socialist economy. Very soon the Communist Party leaders realized the issue of over-concentration of decision power in this new central planning regime.[9] 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 of 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, ed, 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 at the local level. Local bureaucrats were rewarded for following instructions from the central government (Kung and Chen, 2011).

Following an extended period of political and economic turmoil,[10] 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 performance in promotion criterion, which stands in sharp contrast with the earlier reforms which stigmatized the single-minded pursuit of economic goals (Li and Zhou, 2005).

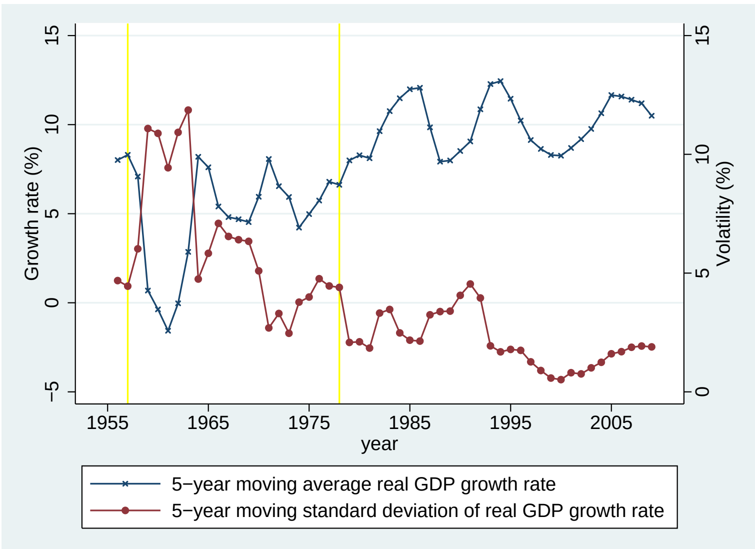


Figure 1: China's Economic Growth and Volatility

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.

[9]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)

[10]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).

Figure 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. Somewhat being paid less attention is the output volatility, but the same, contrasting dynamics followed two waves of decentralization:[11] volatility skyrocketed in the late 1950s and it steadily went down after 1978. Figure 2 plots the evolution of GDP growth rate and output volatility using the data from Liaoning, Henan, and Guangdong, three provinces from northern, central, and southern China. We find similar trends across these provinces, suggesting that the pattern of volatility at the national level can be at least partly attributed to within-province changes over time.

Why do 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. But this answer still does not clarify the nature of the differences, which, we believe, lies on the incentive of local bureaucrats. The great failure of the first wave of decentralization and several followup attempts in the 1960s and 1970s can be viewed as a manifestation of the distorted incentive at the bottom through the inter-governmental interaction.

Before turning to our theoretical specification, it is worth emphasizing that even though we motivate our discussion with the evidence from China, reforms featuring decentralization of decision making are widely observed across countries under the authoritarian regime since late 1980s. Among Asian countries, Viet Nam launched its large-scale reform ("Doi Moi") in 1986 which shares 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).[12] 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.[13] The experience is more mixed if we take into account the decentralization reforms implemented in most of the Eastern European economies. The pre- and post-reform comparisons cannot be made for the Post-Soviet states because of the missing data prior to the collapse of Soviet Union. For Albania, Bulgaria, Albania, and Poland which we have data, the impact of decentralization in 1990s is somewhat unclear.

[11]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 tease out conventional economic factors that determine volatility such as financial development, openness, inventory management, and monetary policy.

[12]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.

[13]We plot how output growth and volatility evolves around the reform period for these four Asian countries in the appendix. See Figures 8 and 9.

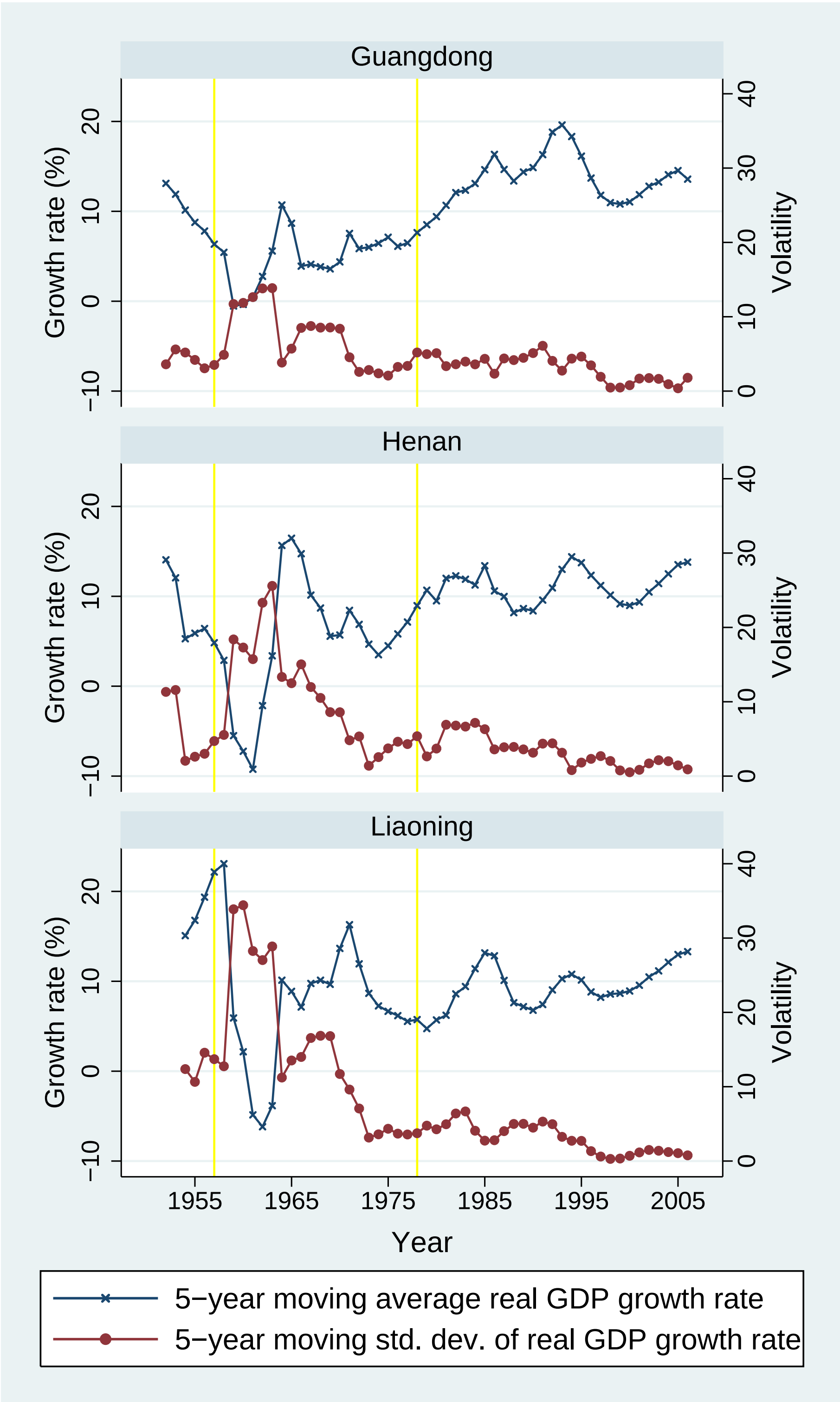


Figure 2: Economic Growth and Volatility at the Provincial Level

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.

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.[14] 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. Figure 3 illustrates 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 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:

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

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. The central government receives a signal s'_ℓ with $s'_\ell = s_\ell + \epsilon$, where ϵ is an exogenous communication friction with $\epsilon \sim$

[14] In the baseline setting, the communication friction is endogenously determined. We will present a version of the model with exogenously given communication friction in the discussion section, highlighting under what condition the endogenous communication channel is essential to our main results.

[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.

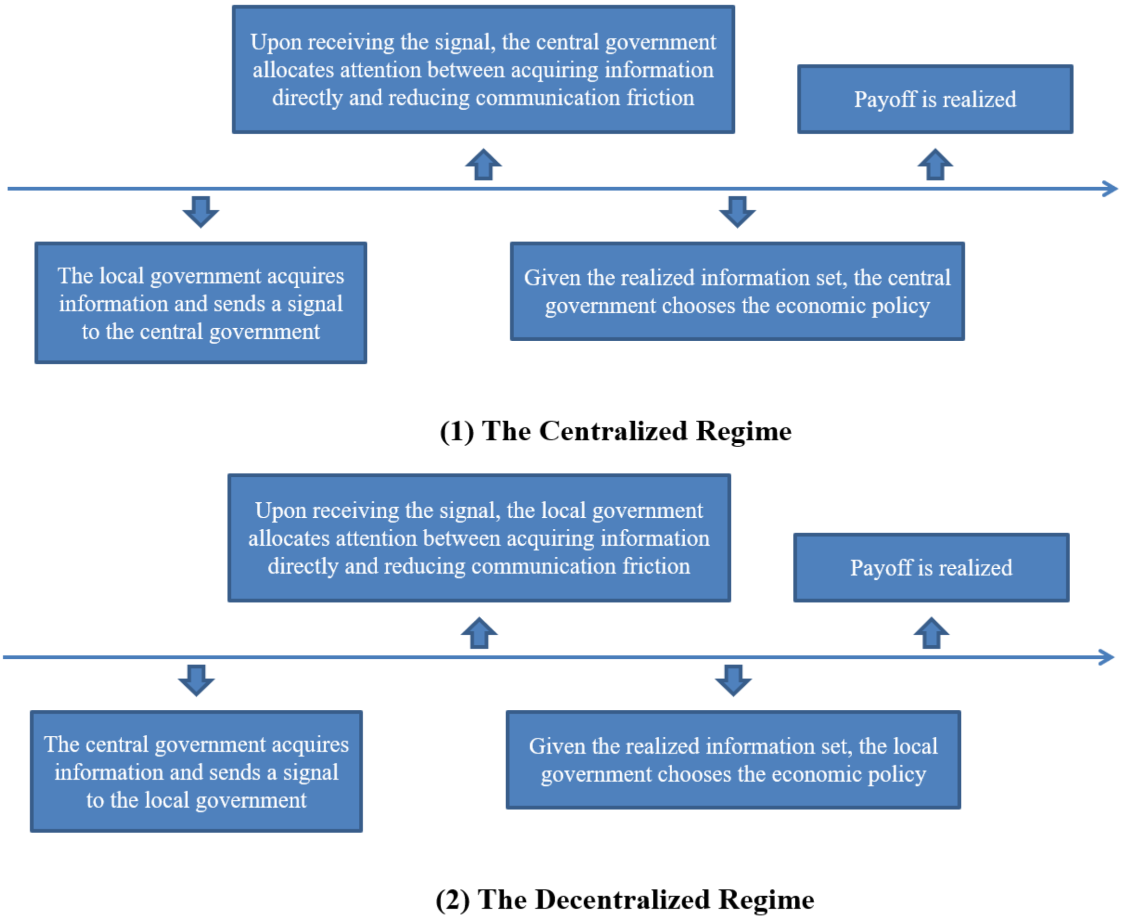


Figure 3: Timeline: The Centralized and Decentralized Regimes

$\mathcal{N}(0, \sigma_\epsilon^2)$. Upon receiving the signal, the central government can acquire additional information from the local government s''_ℓ with $s''_\ell = s_\ell + \epsilon_c$ and $\epsilon_c \sim \mathcal{N}(0, \sigma^2_{\epsilon c})$, where $\sigma^2_{\epsilon c}$ will be endogenously determined as an outcome of the trade-off between two information channels which will be formally specified later. Based on its private information θ_c and two signals received,[16] the central government makes the policy choice a_c.

Similarly, under the decentralized regime, the central government sends a signal s_c. The local government receives a signal s'_c with $s'_c = s_c + \epsilon$. The local government also decides how much resource to be spent on the second signal $s''_c = s_c + \epsilon_\ell$ with $\epsilon_\ell \sim \mathcal{N}(0, \sigma^2_{\epsilon\ell})$. Given the resulting information set, the local government then picks its preferred policy a_ℓ based on θ_ℓ, s'_c, and s''_c.

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,[17] not to mention

[16] The specification with two signals facilitates the introduction of entropy reduction constraint in what follows. One can think of the first signal as the signal receiver's prior of the signal sent by the other government. Alternatively, we can introduce only one inter-governmental signal with its variance bounded above by σ_ϵ^2, but the decision problem for the government is effectively the same, that is, to decide the preciseness of inter-governmental communication.

[17] 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

the complication of coupling policy prescription with political propaganda.[18] 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 the discussion section, we will allow the signal sender to strategically introduce noise into the inter-governmental communication. As will be explained later, it turns out in our setting, the signal sender always has incentive to truthfully reveal its information.[19]

We assume all the white noises z_c, z_ℓ, ϵ, ϵ_ℓ, and ϵ_c are independent.

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.[20]

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 dimensionality 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

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

and disseminating official documents, explaining why even specific wording or quotation could have deep reflections.

[18]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, significantly 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.

[19]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.

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

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 $H(X) - H(X|Y) < \kappa$. We now specialize this constraint to our specific setting.

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

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

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

$$H(\theta,\epsilon|s'_\ell) - H(\theta,\epsilon|\theta_c, s'_\ell, s''_\ell) \leq \kappa_c,$$

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 $(\theta_c, s'_\ell, s''_\ell)$.[21] The following lemma provides a closed form solution to this information flow constraint.[22]

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

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

However complicated Equation 2 appears, the choice variables, σ_c^2 and $\sigma_{\epsilon c}^2$ for the central government are multiplicatively separable in the information flow constraint, which is very important for a sharp characterization of the attention allocation problem. Moreover, we have $K_\ell \geq \left(1/\sigma_\epsilon^2 + 1/\sigma_\ell^2\right)\left(1/\sigma^2 + 1/\sigma_\ell^2\right)$ with the equality if and only if $\kappa_c = 0$. That said, when the central government has zero information capacity, then it is impossible to acquire any information $(\sigma_c^2 = \sigma_{\epsilon c}^2 = \infty)$. It should be noted that σ_ℓ^2, the 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 if it would not cause any confusion.

Symmetrically, under the decentralized regime, the information flow constraint for the signal

[21] 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.

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

sender, the central government, is given by

$$H(\theta) - H(\theta|\theta_c) \leq \kappa_c \quad \Leftrightarrow \quad \frac{1}{2}\log_2\left(\frac{\mathrm{Var}(\theta)}{\mathrm{Var}(\theta|\theta_c)}\right) = \frac{1}{2}\log_2\left(\frac{\sigma^2+\sigma_c^2}{\sigma_c^2}\right) \leq \kappa_c. \tag{3}$$

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

$$H(\theta, \epsilon|s_c') - H(\theta, \epsilon|\theta_\ell, s_c', s_c'') \leq \kappa_\ell.$$

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

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

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.[23]

We now impose the key assumption of 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 seems to be a reasonable assumption in our setting for two main reasons. First, since there is only one local government in our model, θ should be interpreted as the state of the *local* economy, and according to Hayek (1945), the local government tends to have the intrinsic advantage of obtaining local information. In the Chinese context, 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 in 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

[23] 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 $\sigma_{\epsilon c}^2 = \infty$) than the decentralized regime, while the local government can get a more precise signal θ_ℓ under the decentralized regime (setting $\sigma_{\epsilon\ell}^2 = \infty$) 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.

$E(Y)$ and its variance $Var(Y)$.

3.5 The Decision Problem for Each Government

The second layer of the government's decision problem is to pick the desired economic policy. 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

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

subject to Constraint 2.

The local government, who sends the signal to the central government under this regime, cares about both the economic output and whether its policy suggestion is actually implemented by the government. More precisely, its decision problem is given by

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

subject to Constraint 1, where $0 \leq \gamma \leq 1$. The first term in the payoff function is the utility that the local government directly derives from the economic output. The second term captures the fact that the local government also cares about how closely the central government follows its policy suggestion. This is a reduced-form way to incorporate additional promotion incentive beyond the merit-based rules.[24] Therefore, the local government faces a trade-off between economic welfare and career concern. The parameter γ measures the relative importance of career concern.[25] 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.

Under the decentralized regime, the central government, who now becomes the sender of the signal, has the same payoff function but with different choice variables. Its decision problem is given by

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

subject to Constraint 3. In words, the central government chooses the signal and its precision in order to induce the local government to maximize the expected economic output.

The local government now has a two-layer decision problem, which is given by

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

subject to Constraint 4. Despite having a similar form, the second term of the payoff function

[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.

[25] For simplicity, we introduce only one parameter to capture the degree of political career concern relative to economic motives. Even though we stick to the interpretation of γ as loyalty concern, our comparative static results can also be interpreted as a change of economic motives.

entails a different interpretation. We assume that the local government has the incentive to stick to the policy prescription of the central, signaling its political loyalty. This additional career concern seems to be a common characteristics in many of the authoritarian regimes,[26] and it turns out to be a key driving force of our main results.

The equilibrium of this model under each regime is characterized by solving the constrained optimization problem of the signal receiver and sender sequentially. We now turn to the main results of the baseline model.[27]

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 fully 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 and Bayesian update with normal distributions, the optimal policy of the final policy maker is always a linear combination of the signals it receives.

Lemma 3. Under the centralized regime, the optimal policy of the central government is given by

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

Since the central government is assumed to be benevolent, Lemma 3 says it always targets its policy to the expected state of the economy conditional on all the information it gathers. We can write the distance between the policy and the true state of the economy as

$$E(a_c - \theta)^2 = Var(\theta|\theta_c, s'_\ell, s''_\ell) = \left(\frac{1}{\sigma^2} + \frac{1}{\sigma_c^2} + \frac{1}{\sigma_\ell^2 + (1/\sigma_\epsilon^2 + 1/\sigma_{\epsilon c}^2)^{-1}}\right)^{-1}. \tag{5}$$

Lemma 4. Under the decentralized regime, the optimal policy of the local government is given by

$$a_\ell = (1-\gamma)E(\theta|\theta_\ell, s'_c, s''_c) + \gamma E(\theta_c|\theta_\ell, s'_c, s''_c) = k_1\theta_\ell + k_2 s'_c + k_3 s''_c,$$

[26] 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.

[27] The formal definition of the equilibrium can be found in the appendix.

with $s'_c = \theta_c + \epsilon$, $s''_c = \theta_c + \epsilon_\ell$, and

$$
\begin{aligned}
k_1 &\equiv \frac{\frac{1-\gamma}{\sigma_\ell^2}}{\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_c^2+(1/\sigma_\epsilon^2+1/\sigma_{\epsilon\ell}^2)^{-1}}} + \frac{\frac{\gamma}{\sigma_c^2+\sigma_\ell^2+\sigma_c^2\sigma_\ell^2/\sigma^2}}{\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_{\epsilon\ell}^2}+\frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}}, \\
k_2 &\equiv \frac{\frac{1-\gamma}{\sigma_c^2+(1/\sigma_\epsilon^2+1/\sigma_{\epsilon\ell}^2)^{-1}}\frac{1/\sigma_\epsilon^2}{1/\sigma_\epsilon^2+1/\sigma_{\epsilon\ell}^2}}{\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_c^2+(1/\sigma_\epsilon^2+1/\sigma_{\epsilon\ell}^2)^{-1}}} + \frac{\frac{\gamma}{\sigma_\epsilon^2}}{\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_{\epsilon\ell}^2}+\frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}}, \\
k_3 &\equiv \frac{\frac{1-\gamma}{\sigma_c^2+(1/\sigma_\epsilon^2+1/\sigma_{\epsilon\ell}^2)^{-1}}\frac{1/\sigma_{\epsilon\ell}^2}{1/\sigma_\epsilon^2+1/\sigma_{\epsilon\ell}^2}}{\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_c^2+(1/\sigma_\epsilon^2+1/\sigma_{\epsilon\ell}^2)^{-1}}} + \frac{\frac{\gamma}{\sigma_{\epsilon\ell}^2}}{\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_{\epsilon\ell}^2}+\frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}}.
\end{aligned}
$$

The local government's policy choice is a weighted average of the benevolent policy and the local government's conditional expectation of the signal from the central. The loyalty concern of the local government distorts the economy through the policy-making margin. In the absence of loyalty concern ($\gamma = 0$), the distance between the policy and the true state of the economy is similar to what we have obtained under the centralized regime:

$$
E(a_\ell - \theta)^2\Big|_{\gamma=0} = Var(\theta|\theta_\ell, s'_c, s''_c) = \left(\frac{1}{\sigma^2} + \frac{1}{\sigma_\ell^2} + \frac{1}{\sigma_c^2 + (1/\sigma_\epsilon^2 + 1/\sigma_{\epsilon\ell}^2)^{-1}}\right)^{-1} \tag{6}
$$

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 same form of harmonic mean:

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

Since the optimal action is always a linear combination of the signal obtained by the signal receiver, we obtain a tight relationship between the expected output and output volatility.

Lemma 5. Let $a_i = m_{i1}\theta_\ell + m_{i2}\theta_c + m_{i3}\epsilon + m_{i4}\epsilon_c + m_{i5}\epsilon_\ell$ $(i = c, \ell)$. The expected output is given by

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

Moreover, the output volatility strictly decreases with the expected output, which is given by

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

The lemma provides the closed-form solution of the expected output level and its variance. More importantly, it 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, $E(a_i - \theta)^2$, given its simpler expression.

4.2 The Equilibrium under the Centralized Regime

According to Lemma 3 and Equation 5, the resource allocation problem for the central government can be simplified to

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

subject to Constraint 2. By inspection, it is observed that 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 2. Since $K_c > 1/\sigma_\ell^4$, the maximum is attained when σ_c^2 attains its minimum under the constraint, or equivalently, $\sigma_{\epsilon c}^2=\infty$. More intuitively, the best possible signal the central government would get if it spent all its attention budget on the direct information acquisition channel is θ_ℓ with variance of σ_ℓ^2 while directly inquiring information from the economy has a first-order impact on the economic output. Therefore, we obtain the following characterization of the central government's strategy under the centralized regime.

Lemma 6. Under the centralized regime, for any given σ_ℓ^2, the central government completely devotes itself to the direct information acquisition with $\sigma_{\epsilon c}^2=\infty$.

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 7. Under the centralized regime, for any γ, the local government always spends all of its attention resource on information acquisition (Constraint 1 is binding), which leads to

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

The intuition behind this sharp characterization is twofold. 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 its signal. 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.

Collecting the results from Lemmas 3, 6, and 7, we obtain the following equilibrium characterization for 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 $\sigma_\ell^2=\sigma^2/(2^{2\kappa_\ell}-1)$, $\sigma_{\epsilon c}^2=\infty$, and

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

4.3 The Equilibrium under the Decentralized Regime

We start with the two extreme cases of the resource allocation problem for the local government: (i) $\gamma = 0$; (ii) $\gamma = 1$.

4.3.1 No Loyalty Concern ($\gamma = 0$)

In the absence of loyalty concern ($\gamma = 0$), we know from Equation 6,

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

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

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

subject to Constraint 4. This is symmetric to the central government's decision problem under the centralized regime. For the same intuition, we obtain a simple characterization of the local government's strategy under the decentralized regime with no loyalty concern.

Lemma 8. 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 $\sigma_{\epsilon\ell}^2 = \infty$.

4.3.2 Pure Loyalty Concern ($\gamma = 1$)

If the local government is purely loyalty driven ($\gamma = 1$), we know from Equation 7 that

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

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

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

subject to Constraint 4. Similarly, by inspection, the constraint must be binding. Given the binding constraint, we can further simplify the constrained optimization problem to

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

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

[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 less efficient.

Lemma 9. Under the decentralized regime with $\gamma = 1$, for any given σ_c^2, the local government completely devotes to the inter-governmental communication with $\sigma_\ell^2 = \infty$.

According to Lemmas 8 and 9, the two extreme cases of γ lead to two corner solutions for the local government's strategy, which demonstrates that 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.

We now turn to the case with a general $\gamma \in [0, 1]$ under the decentralized regime.

4.3.3 The General Case: $\gamma \in [0, 1]$

For the general case, given Lemma 4, we solve the first layer of the decision problem for the local government with respect to σ_ℓ^2 and $\sigma_{\epsilon\ell}^2$. We first show 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. The following lemma states our finding formally.

Lemma 10. 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 ($\sigma_{\epsilon\ell}^2 = \infty$). If $\gamma \ge \bar{\gamma}$, 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 $\sigma_{\epsilon\ell}^2 < \infty$) with

$$\frac{1}{\sigma_\ell^2} + \frac{1}{\sigma^2} + \frac{1}{\sigma_c^2} = \frac{K_\ell^{1/2}(1-\gamma)}{\gamma}, \tag{8}$$

which implies that $\partial\sigma_\ell^2/\partial\gamma > 0$. Moreover, $\underline{\gamma}$ and $\bar{\gamma}$ are the unique roots to the following two equations, respectively.

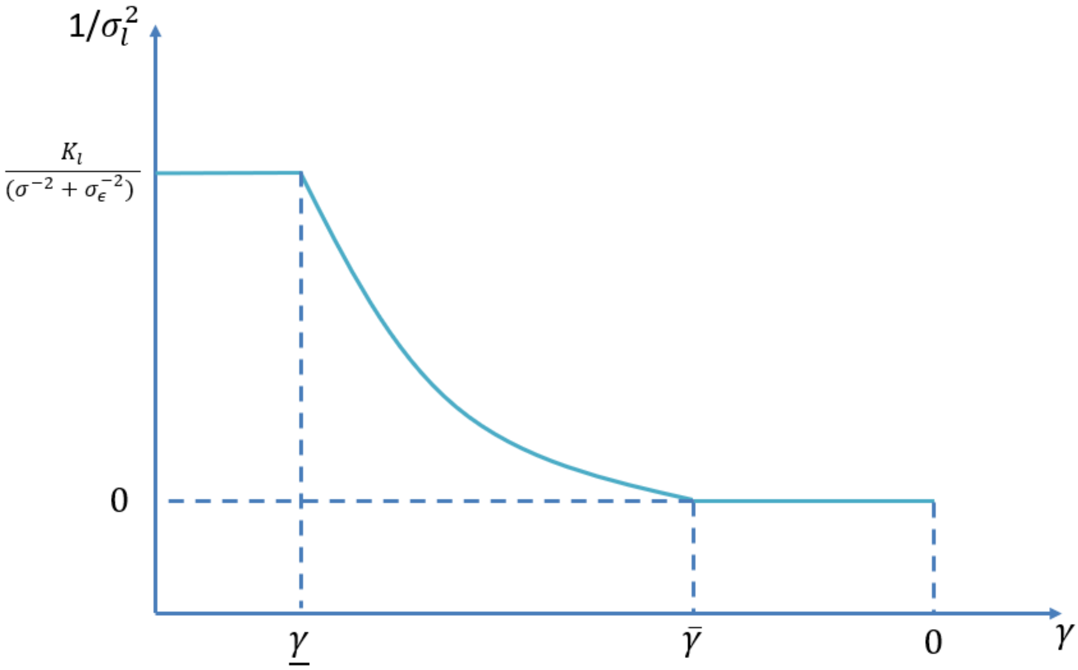
$$\begin{aligned} \underline{\gamma}^2 - (1-\underline{\gamma})^2 K_\ell^{-1}(1/\sigma_\epsilon^2 + 1/\sigma_c^2)^2 &= 0 \\ \bar{\gamma}^2 - (1-\bar{\gamma})^2 K_\ell (1/\sigma^2 + 1/\sigma_c^2)^{-2} &= 0 \end{aligned}$$

The optimal resource allocation of the local government generalizes our observations under the two extreme cases. The local government focuses exclusively on direct information acquisition provided that the economic motive is sufficiently strong ($\gamma \le \underline{\gamma}$), while it focuses exclusively on intergovernmental communication if its loyalty concern is sufficiently strong ($\gamma \ge \bar{\gamma}$). If its loyalty concern is in the intermediate range, then the attention resource will be allocated to both dimensions with the effort on intergovernmental communication strictly increasing with the intensity of the loyalty concern.

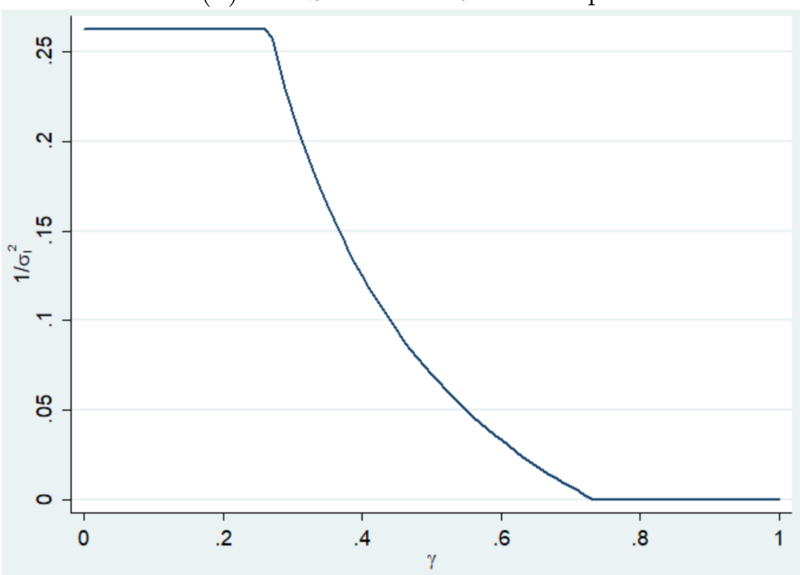
We illustrate the relationship between σ_ℓ^2 $(1/\sigma_\ell^2))$ with γ in Figure 4. Clearly seen in the figure, an immediate implication of Lemma 10 is that σ_ℓ^2 only changes with γ at the middle range while when γ is sufficiently small or large, additional increase of decrease of γ does not have further bearing on information acquisition. It is in the middle range of γ that both margins on which the loyalty concerns distorts the economy are active.

We turn to the strategy of the central government under the decentralized regime. The actual proof is tedious because the strategy of the local government is not differentiable with respect to γ and the two cutoffs $\underline{\gamma}$ and $\bar{\gamma}$ are functions of σ_c^2, but the basic idea is very simple. A better

(a) The Analytical Form



(b) The Simulated Relationship



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

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

signal from the central helps the local target the true state of the economy, but it is 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. However, the first channel dominates, so the central government always strives for a better signal.

Lemma 11. Under the decentralized regime, for any γ, the central government spends all of its attention resource on information acquisition (Constraint 3 is binding), which leads to

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

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

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

In words, the stronger is the loyalty concern, the further away is the economic policy from the true state of the 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. 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 become noisier. To predict the true state of the economy, it is better to have one high quality signal rather 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 crucial for this result.[30]

Lemma 12. $E(a_c - \theta)^2 > E(a_\ell - \theta)^2\Big|_{\gamma=0}$.

On the other hand, decentralization with $\gamma = 1$ always worsens economic performance. This result is more straightforward as decentralization in this case leads to strictly less informative signal for the decision maker and introduces additional distortion in setting the policy.

Lemma 13. $E(a_c - \theta)^2 < E(a_\ell - \theta)^2\Big|_{\gamma=1}$.

Following Proposition 2 and Lemmas 5, 12, and 13, we now obtain the main result of the paper.

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

[29] See Propositions 6 and 7 in the appendix.

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

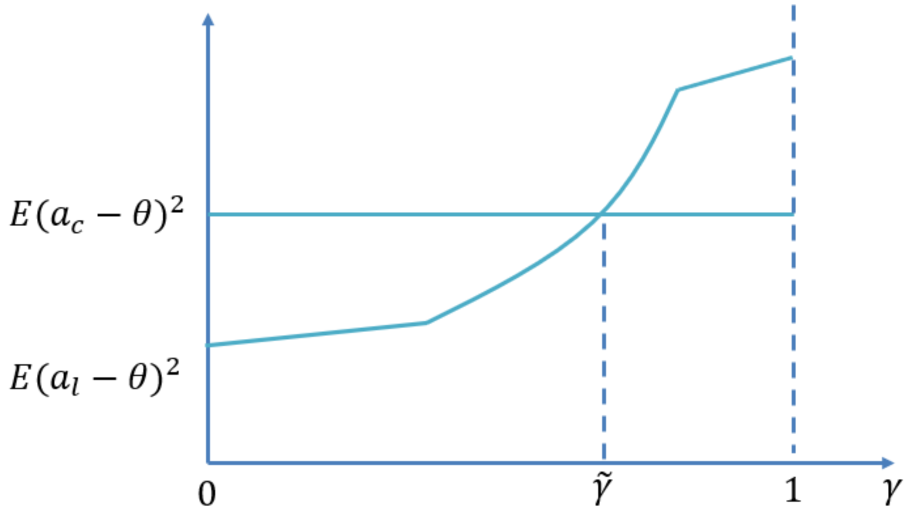


Figure 5: The Comparison between Two Regimes

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.

Figure 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.

Moreover, we have the following simple corollary.

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 13, the devotion of local bureaucrats to understanding and deciphering the policy message from the top guarantees the failure of decentralization.

5 Extension and Discussion

In this section, we present a few extensions of the model and discuss how the model is related to the two waves of decentralization in China.

5.1 Exogenous Communication Frictions

The first extension shuts down the endogenous communication channel. It concerns whether the model is able to deliver our main result when loyalty concern only operates through the final decision-

[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 $\tilde{\gamma} \approx 0.26$.

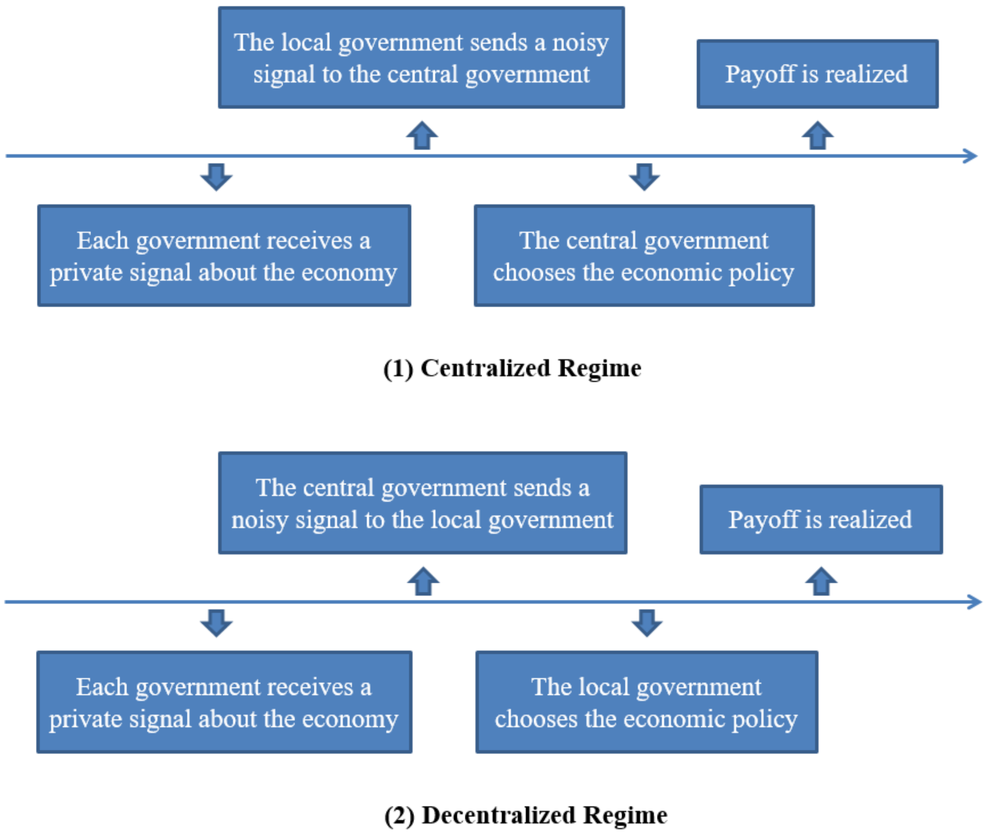


Figure 6: Timeline without Rational Inattention

making margin and under what condition, the endogenous communication channel is essential to generating differential outcomes of decentralization.

The basic setup is the same. 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. Figure 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

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

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.[32] According to Lemma 5, to analyze economic volatility and output, we just need to focus on $E(a_c - \theta)^2$ and $E(a_\ell - \theta)^2$.[33]

Lemma 14. In the absence of the loyalty concern ($\gamma = 0$), decentralization improves the economic performance.

Following the same intuition as in the benchmark model, the assumptions $\sigma_\ell^2 < \sigma_c^2$ 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 15. 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 own, more precise signal and simply follow the policy prescription from the central government, thus leading to 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 13 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 20 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 always worsens 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.*

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

[32] See Lemma 18 in the appendix.

[33] For the explicit expressions of $E(a_c - \theta)^2$ and $E(a_\ell - \theta)^2$ under $\gamma = 0$ and $\gamma = 1$, see Lemma 20 in the appendix.

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.

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 $s_c = \theta_\ell$ 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 $\sigma_{\delta c}^2$) without incurring any cost.

In an equilibrium, the signal receiver correctly expects the variance of the white noise added by the signal sender and acts accordingly. Therefore, 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.

Proposition 4. In the case of strategic communication, we have $\sigma_{\delta\ell}^2 = \sigma_{\delta c}^2 = 0$.

The above result follows immediately from Lemmas 7 and 11. In the benchmark model, we have shown that the information flow constraint is always binding for the signal sender. Therefore, even allowing the signal sender to strategically obscure the signal, it will not have the incentive to do so.

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 4 with

$$\left(\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_{\epsilon\ell}^2}\right)\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{\lambda}{\sigma_\ell^2}\right) \leq \frac{2^{2\kappa}(\sigma^2+\sigma_c^2+\sigma_\epsilon^2)}{\sigma^2\sigma_c^2\sigma_\epsilon^2}+\frac{1}{\sigma_c^4} \equiv K_\ell(\sigma_c^2), \tag{9}$$

and replace Constraint 1 with

$$\frac{1}{\sigma^2}+\frac{\lambda}{\sigma_\ell^2} \leq \frac{2^{2\kappa}}{\sigma^2}, \tag{10}$$

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.[34]

We proceed to characterize the equilibrium under this alternative setting.

First, notice that the optimal policy does not depend on λ, so Lemmas 3 and 4 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 10), 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

[34] 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, s_c', and s_c'' (or θ_c, s_ℓ', and s_ℓ''.

impose the following regularity condition on λ throughout this subsection.

$$\lambda(1/\lambda-1)^2<\frac{2^{4\kappa}\left(\frac{1}{2^{2\kappa}-1}+\frac{2^{2\kappa}\sigma^2}{(2^{2\kappa}-1)^2\sigma_\epsilon^2}\right)^2}{2^{2\kappa}\left(\frac{1}{2^{2\kappa}-1}+\frac{2^{2\kappa}\sigma^2}{(2^{2\kappa}-1)^2\sigma_\epsilon^2}\right)+1}. \tag{11}$$

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 16. If $\lambda \geq 1/2$, then the regularity condition 11 holds for any κ, σ^2, and σ_ϵ^2.

The following lemma is the counterpart of Lemma 10. It can be seen that the the characterization of the optimal policy for a general γ under the decentralized regime is qualitatively unchanged in this alternative setting.

Lemma 17. 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 ($\sigma_{\epsilon\ell}^2=\infty$). 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 $\sigma_{\epsilon\ell}^2<\infty$) 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.

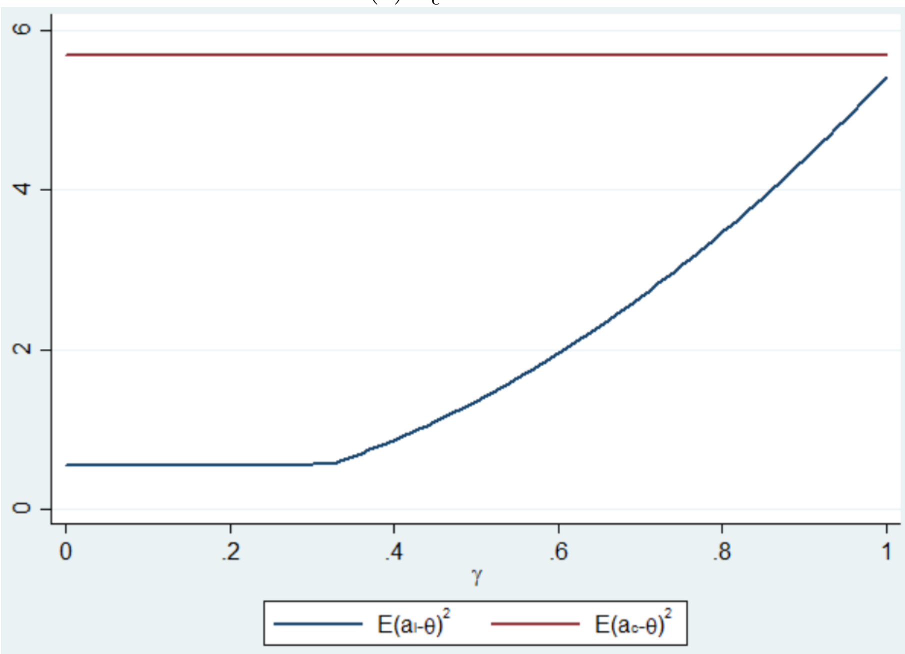
Albeit not being formally established, our simulation results suggest that the comparison between two regimes seems very similar to what has been illustrated by Figure 5: economic volatility increases monotonically with γ and there are two kinks on the curve of $E(a_\ell-\theta)^2$, representing the 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 11 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 Figure 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 setting with exogenous communication friction.

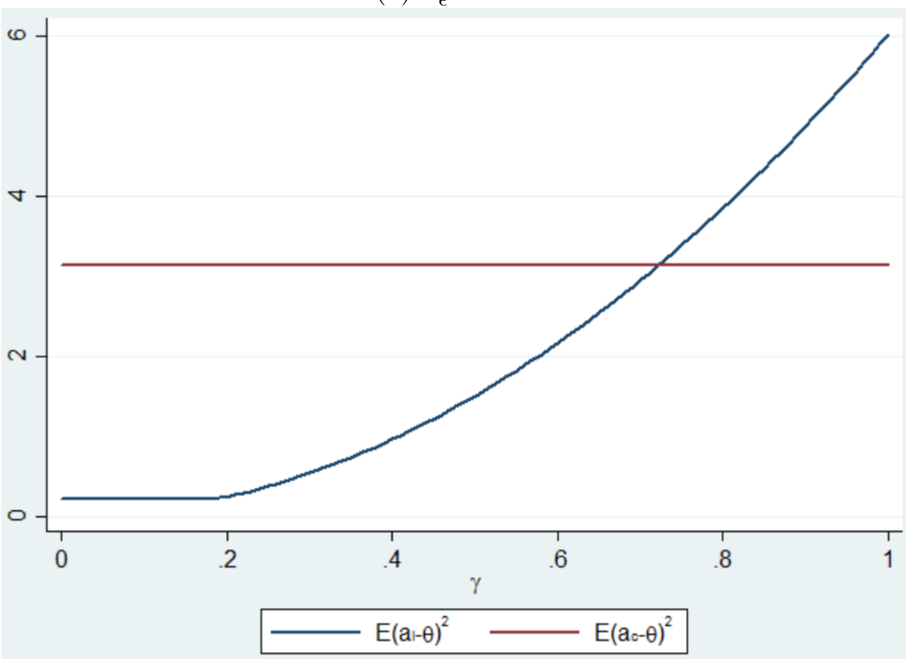
5.4 Discussion

In light of our theory, the difference in γ could be one explanation of the contrasting experience following the two decentralization reforms in China. In the 1950s, pursuit of economic welfare, even in the interest of the public, could 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

(1) $\sigma_\epsilon^2 = 1000$



(2) $\sigma_\epsilon^2 = 100$



(Note: $\sigma^2 = 100; \kappa = 2.$)

Figure 7: Comparison between Two Regimes When Condition 11 Is Violated

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. Inevitably, the local bureaucrats cannot be fully freed from the policy suggestions from the central in an authoritarian regime, evidence has convincingly showed 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, 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 everyone want to be more radical than the other, thus pushing γ towards its upper bound. In the post-reform era, since economic development became priority one, 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.

6 Concluding Remarks

To understand the contrasting dynamics following the two waves of decentralization in China, we propose a model of inter-governmental information transmission. The model demonstrates that the impact of decentralization on economic performance hinges on the degree of loyalty concern of the local government. Decentralization could be welfare-reducing, leading to lower output and higher volatility, if the local government is a loyal follower of the central government. Even though our story is mainly motivated by the experience from China, we believe it could also shed light on decentralization experience in other authoritarian regime and more broadly, large organizations with strict hierarchy.

In our model, the central government is a benevolent government when it comes to the economic policy, while the local government attempts to signal the political loyalty through its economic policy. The underlying presumption is that the central government could separate its political consideration from economic policy-making. A natural question is what if the central government's preference is not separable, that is, it also values local government's loyalty in economic policy-making. Our results would then depend on the comparison two γs, one for the local and the other for the central. More fundamentally, one might ask why the central government would choose to decentralize despite its detrimental effects to the economy and what the deep and non-economic roots of decentralization are in an authoritarian regime. We think these are the fruitful avenues for future research.

A Proofs

A.1 Proof of Lemma 1

Proof. Since $H(\theta,\epsilon|s'_\ell) = H(\theta,\epsilon,s'_\ell) - H(s'_\ell)$ and $H(\theta,\epsilon|\theta_c,s'_\ell,s''_\ell) = H(\theta,\epsilon,\theta_c,s'_\ell,s''_\ell) - H(\theta_c,s'_\ell,s''_\ell)$, we have[35]

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

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

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

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

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

Plugging the three determinants into the information flow constraint, we obtain

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

Simplifying the expression above, we obtain the desired conclusion. ☐

A.2 Proof of Lemma 2

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

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

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

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

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

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

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

Plugging the three determinants into the information flow constraint, we obtain

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

Simplifying the expression above, 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.

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 sextuplet $(\sigma_\ell^2, \sigma_c^2, \sigma_{\epsilon c}^2; \theta_c, s'_\ell, s''_\ell) \in \mathbb{R}_+^3 \times \mathbb{R}^3$,

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

for any $\sigma_\ell^2 \in \mathbb{R}_+$,

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

subject to Constraint 2; and

$$\sigma_\ell^{*2} \in \arg\max_{\sigma_\ell^2} \quad \Big\{(1-\gamma)(Y^* - E(a_c^*(\sigma_\ell^2, \sigma_c^{*2}(\sigma_\ell^2), {\sigma_{\epsilon c}^*}^2(\sigma_\ell^2); \theta_c, s'_\ell, s''_\ell) - \theta)^2)$$
$$-\gamma E(\theta_\ell - a_c^*(\sigma_\ell^2, \sigma_c^{*2}(\sigma_\ell^2), {\sigma_{\epsilon c}^*}^2(\sigma_\ell^2); \theta_c, s'_\ell, s''_\ell))^2\Big\}$$

subject to Constraint 1.

The equilibrium under the decentralized regime is defined as a quadruplet $(\sigma_c^{*2}, \sigma_\ell^{*2}(\cdot), {\sigma_{\ell c}^*}^2(\cdot), a_\ell^*(\cdot))$ such that for any sextuplet $(\sigma_c^2, \sigma_\ell^2, \sigma_{\epsilon\ell}^2; \theta_\ell, s'_c, s''_c) \in \mathbb{R}_+^3 \times \mathbb{R}^3$,

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

for any $\sigma_c^2 \in \mathbb{R}_+$,

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

subject to Constraint 4; and

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

subject to Constraint 3.

In both regimes, we require the belief updating follows the Bayes' rule.[36]

A.4 Proof of Lemma 3

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

$$a_c = E(\theta|\theta_c, s_\ell', s_\ell'').$$

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

$$\begin{aligned} f(\theta, \theta_c, \theta_\ell, s_\ell', s_\ell'') &= f(\theta) f(\theta_c, \theta_\ell, s_\ell', s_\ell''|\theta) = f(\theta) f(\theta_c|\theta) f(\theta_\ell, s_\ell', s_\ell''|\theta) \\ &= f(\theta) f(\theta_c|\theta) f(\theta_\ell|\theta) f(s_\ell', s_\ell''|\theta_\ell) = f(\theta) f(\theta_c|\theta) f(\theta_\ell|\theta) f(s_\ell'|\theta_\ell) f(s_\ell''|\theta_\ell), \end{aligned}$$

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[37]

$$\begin{aligned} f(\theta, \theta_c, \theta_\ell, s_\ell', s_\ell'') &\sim \exp\left\{-\frac{1}{2}\left[\frac{\theta^2}{\sigma^2} + \frac{(\theta_c - \theta)^2}{\sigma_c^2} + \frac{(\theta_\ell - \theta)^2}{\sigma_\ell^2} + \frac{(s_\ell' - \theta_\ell)^2}{\sigma_\epsilon^2} + \frac{(s_\ell'' - \theta_\ell)^2}{\sigma_{\epsilon c}^2}\right]\right\} \\ &= \exp\left\{-\frac{1}{2}\left[\left(\frac{1}{\sigma_\ell^2} + \frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma_{\epsilon c}^2}\right)\theta_\ell^2 - 2\left(\frac{\theta}{\sigma_\ell^2} + \frac{s_\ell'}{\sigma_\epsilon^2} + \frac{s_\ell''}{\sigma_{\epsilon c}^2}\right)\theta_\ell \right.\right. \\ &\quad \left.\left. + \frac{\theta^2}{\sigma^2} + \frac{(\theta_c - \theta)^2}{\sigma_c^2} + \frac{\theta^2}{\sigma_\ell^2} + \frac{s_\ell'^2}{\sigma_\epsilon^2} + \frac{s_\ell''^2}{\sigma_{\epsilon c}^2}\right]\right\} \\ &= \exp\left\{-\frac{1}{2}\left[\left(\frac{1}{\sigma_\ell^2} + \frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma_{\epsilon c}^2}\right)\left(\theta_\ell - \frac{\frac{\theta}{\sigma_\ell^2} + \frac{s_\ell'}{\sigma_\epsilon^2} + \frac{s_\ell''}{\sigma_{\epsilon c}^2}}{\frac{1}{\sigma_\ell^2} + \frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma_{\epsilon c}^2}}\right)^2 \right.\right. \end{aligned}$$

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

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

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

$$+\frac{\theta^2}{\sigma^2}+\frac{(\theta_c-\theta)^2}{\sigma_c^2}+\frac{\theta^2}{\sigma_\ell^2}+\frac{{s'_\ell}^2}{\sigma_\epsilon^2}+\frac{{s''_\ell}^2}{\sigma_{\epsilon c}^2}-\frac{\left(\frac{\theta}{\sigma_\ell^2}+\frac{s'_\ell}{\sigma_\epsilon^2}+\frac{s''_\ell}{\sigma_{\epsilon c}^2}\right)^2}{\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_{\epsilon c}^2}}\Bigg]\Bigg\}$$

Integrating out θ_ℓ, we obtain

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

This leads to

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

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

A.5 Proof of Lemma 4

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

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

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

$$\begin{aligned}f(\theta,\theta_c,\theta_\ell,s'_c,s''_c) &= f(\theta)f(\theta_c|\theta)f(\theta_\ell|\theta)f(s'_c|\theta_c)f(s''_c|\theta_c)\\
&\sim \exp\left\{-\frac{1}{2}\left[\frac{\theta^2}{\sigma^2}+\frac{(\theta_c-\theta)^2}{\sigma_c^2}+\frac{(\theta_\ell-\theta)^2}{\sigma_\ell^2}+\frac{(s'_c-\theta_c)^2}{\sigma_\epsilon^2}+\frac{(s''_c-\theta_c)^2}{\sigma_{\epsilon\ell}^2}\right]\right\}\\
&= \exp\left\{-\frac{1}{2}\left[\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2}\right)\theta^2-2\left(\frac{\theta_c}{\sigma_c^2}+\frac{\theta_\ell}{\sigma_\ell^2}\right)\theta+\frac{\theta_c^2}{\sigma_c^2}+\frac{\theta_\ell^2}{\sigma_\ell^2}+\frac{(s'_c-\theta_c)^2}{\sigma_\epsilon^2}+\frac{(s''_c-\theta_c)^2}{\sigma_{\epsilon\ell}^2}\right]\right\}\\
&= \exp\left\{-\frac{1}{2}\left[\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2}\right)\left(\theta-\frac{\frac{\theta_c}{\sigma_c^2}+\frac{\theta_\ell}{\sigma_\ell^2}}{\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2}}\right)^2\right.\right.\\
&\quad\left.\left.+\frac{\theta_c^2}{\sigma_c^2}+\frac{\theta_\ell^2}{\sigma_\ell^2}+\frac{(s'_c-\theta_c)^2}{\sigma_\epsilon^2}+\frac{(s''_c-\theta_c)^2}{\sigma_{\epsilon\ell}^2}-\frac{\left(\frac{\theta_c}{\sigma_c^2}+\frac{\theta_\ell}{\sigma_\ell^2}\right)^2}{\frac{1}{\sigma^2}+\frac{1}{\sigma_c^2}+\frac{1}{\sigma_\ell^2}}\right]\right\}.\end{aligned}$$

Integrating out θ, we obtain

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

This leads to

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

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

A.6 Proof of Lemma 5

Proof. Writing $(a_i - \theta)$ 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.7 Proof of Lemma 7

Proof. According to Lemma 6, under the centralized regime, $\sigma_{\epsilon c}^2 = \infty$, 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

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

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 1. 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 $\sigma_{\epsilon c}^2=\infty$, Constraint 2 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 arises 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 nature of our information flow constraint.

Using the binding constraint, then the objective function can be rewritten as

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

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

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, $\sigma^2/(2^{2\kappa_\ell}-1)$. □

A.8 Proof of Lemma 10

Proof. According to Lemma 4, we can rewrite the constrained optimization problem as

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

with $a_\ell=(1-\gamma)E(\theta|\theta_\ell,s_c',s_c'')+\gamma E(\theta_c|\theta_\ell,s_c',s_c'')=k_1\theta_\ell+k_2s_c'+k_3s_c''$, subject to Constraint 4. Given the expression of a_ℓ, we have

$$\begin{aligned}
F(\sigma_\ell^2,\sigma_{\epsilon\ell}^2) &= E\left\{(1-\gamma)E[((1-\gamma)(E(\theta|\theta_\ell,s_c',s_c'')-\theta)+\gamma(E(\theta_c|\theta_\ell,s_c',s_c'')-\theta))^2|\theta_\ell,s_c',s_c'']\right.\\
&\quad\left.+\gamma E[((1-\gamma)(E(\theta|\theta_\ell,s_c',s_c'')-\theta_c)+\gamma(E(\theta_c|\theta_\ell,s_c',s_c'')-\theta_c))^2|\theta_\ell,s_c',s_c'']\right\}\\
&= E\left\{(1-\gamma)[(1-\gamma)^2Var(\theta|\theta_\ell,s_c',s_c'')+2\gamma(1-\gamma)Var(\theta|\theta_\ell,s_c',s_c'')\right.\\
&\quad+\gamma^2E((E(\theta_c|\theta_\ell,s_c',s_c'')-\theta_c+\theta_c-\theta)^2|\theta_\ell,s_c',s_c'')]\\
&\quad+\gamma[\gamma^2Var(\theta_c|\theta_\ell,s_c',s_c'')+2\gamma(1-\gamma)Var(\theta_c|\theta_\ell,s_c',s_c'')\\
&\quad\left.+(1-\gamma)^2E((E(\theta|\theta_\ell,s_c',s_c'')-\theta+\theta-\theta_c)^2|\theta_\ell,s_c',s_c'')]\right\}\\
&= (1-\gamma)^2(1+\gamma)Var(\theta|\theta_\ell,s_c',s_c'')+\gamma^2(2-\gamma)Var(\theta_c|\theta_\ell,s_c',s_c'')\\
&\quad+(1-\gamma)\gamma^2[Var(\theta_c|\theta_\ell,s_c',s_c'')+2E((E(\theta_c|\theta_\ell,s_c',s_c'')-\theta_c)(\theta_c-\theta))+\sigma_c^2]\\
&\quad+(1-\gamma)^2\gamma[Var(\theta|\theta_\ell,s_c',s_c'')+2E((E(\theta|\theta_\ell,s_c',s_c'')-\theta)(\theta-\theta_c))+\sigma_c^2]\\
&= (1-\gamma)^2(1+2\gamma)Var(\theta|\theta_\ell,s_c,s_c'')+\gamma^2(3-2\gamma)Var(\theta_c|\theta_\ell,s_c',s_c'')+(1-\gamma)\gamma\sigma_c^2 \qquad (12)\\
&\quad-2\sigma_c^2\left\{(1-\gamma)\gamma^2\frac{\frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}}{\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_{\epsilon\ell}^2}+\frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}}+(1-\gamma)^2\gamma\frac{\frac{1}{\sigma_c^2+(1/\sigma_\epsilon^2+1/\sigma_{\epsilon\ell}^2)^{-1}}}{\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_c^2+(1/\sigma_\epsilon^2+1/\sigma_{\epsilon\ell}^2)^{-1}}}\right\}
\end{aligned}$$

To see that Constraint 4 has to be binding, we rewrite the objective function as

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

which strictly increases with σ_ℓ^2 or $\sigma_{\epsilon\ell}^2$.

Since Constraint 4 is binding, from Equation 12, the objective function can be further simplified to be

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

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

It is easy to see that when $\gamma = 0$, $x^* = K_\ell/(1/\sigma_\epsilon^2 + 1/\sigma_c^2)$ with $\sigma_{\epsilon\ell}^2 = \infty$, echoing Lemma 8, and when $\gamma = 1$, $x^* = 1/\sigma^2 + 1/\sigma_c^2$ with $\sigma_\ell^2 = \infty$, echoing Lemma 9. Moreover, we have

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

Since $\frac{dG}{d\gamma}(\gamma; x) = 2\gamma + 2(1-\gamma)K_\ell/x^2 > 0$ for any $\gamma \in [0,1]$ and it is easy to see that $G(0;x) < 0$ 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 $x = K_\ell/(1/\sigma_\epsilon^2 + 1/\sigma_c^2)$ 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

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

which implies $\underline{\gamma} < \bar{\gamma}$.

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

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

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

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

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

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

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

$$\begin{aligned} H'(1/\sigma^2 + 1/\sigma_c^2) &= G(\gamma; 1/\sigma^2 + 1/\sigma_c^2) < G(\bar{\gamma}; 1/\sigma^2 + 1/\sigma_c^2) = 0 \\ H'(K_\ell/(1/\sigma^2 + 1/\sigma_c^2)) &= G(\gamma; K_\ell/(1/\sigma^2 + 1/\sigma_c^2)) > G(\underline{\gamma}; K_\ell/(1/\sigma^2 + 1/\sigma_c^2)) = 0. \end{aligned}$$

Further, $H''(x) > 0$ for any $x \in [1/\sigma^2 + 1/\sigma_c^2, K_\ell/(1/\sigma_\epsilon^2 + 1/\sigma_c^2)]$. Therefore, there exists a unique $x^* \in (1/\sigma^2 + 1/\sigma_c^2, K_\ell/(1/\sigma_\epsilon^2 + 1/\sigma_c^2))$ with $\sigma_\ell^2 < \infty$ and $\sigma_{\epsilon\ell}^2 < \infty$: 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. $\square$

A.9 Proof of Lemma 11

Proof. Following a similar derivation of Equation 12 as in the proof of Lemma 10, we have

$$
\begin{aligned}
E(a_\ell-\theta)^2 &= E\left\{E[((1-\gamma)(E(\theta|\theta_\ell,s'_c,s''_c)-\theta)+\gamma(E(\theta_c|\theta_\ell,s'_c,s''_c)-\theta))^2|\theta_\ell,s'_c,s''_c]\right\}\\
&= (1-\gamma)^2Var(\theta|\theta_\ell,s'_c,s''_c)+2\gamma(1-\gamma)Var(\theta|\theta_\ell,s'_c,s''_c)\\
&\quad +\gamma^2E\{E[(E(\theta_c|\theta_\ell,s'_c,s''_c)-\theta_c+\theta_c-\theta)^2|\theta_\ell,s'_c,s''_c]\}\\
&= (1-\gamma^2)Var(\theta|\theta_\ell,s'_c,s''_c)+\gamma^2Var(\theta_c|\theta_\ell,s'_c,s''_c)+\gamma^2\sigma_c^2-2\gamma^2\sigma_c^2\frac{\frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}}{\frac{1}{\sigma_\epsilon^2}+\frac{1}{\sigma_{\epsilon\ell}^2}+\frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}}\\
&= \frac{1}{K_\ell-1/\sigma_c^4}\left[\frac{(1-\gamma^2)K_\ell}{\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_c^2}}+\gamma^2\left(\frac{1}{\sigma_c^2}-\frac{1}{\sigma^2}-\frac{1}{\sigma_\ell^2}\right)\right]+\gamma^2\sigma_c^2\\
&= \frac{1}{K_\ell-1/\sigma_c^4}\left[\frac{(1-\gamma^2)K_\ell}{\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}+\frac{1}{\sigma_c^2}}+K_\ell\gamma^2\sigma_c^2-\gamma^2\left(\frac{1}{\sigma^2}+\frac{1}{\sigma_\ell^2}\right)\right]. \qquad (13)
\end{aligned}
$$

where the second to last equality follows from the fact that Constraint 4 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 3, where it should be emphasized that both σ_ℓ^2 and K_ℓ are functions of σ_c^2.

According to Lemma 10, $\underline{\gamma}$ 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 the given γ belongs to.

We arbitrarily pick a σ_c^2 subject to Constraint 3 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 10, we have $\sigma_{\epsilon\ell}^2=\infty$, which implies that Constraint 4 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 the definition of K_ℓ ($K_\ell = 2^{2\kappa_\ell}(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)/(\sigma_c^2\sigma^2\sigma_\epsilon^2)+1/\sigma_c^4$). 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) &= \left\{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]\right\} \\
&\quad\cdot(\sigma_c^2+\sigma^2+\sigma_\epsilon^2)(\sigma_c^2+\sigma_\epsilon^2) \\
&\quad-\left\{[(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)\right\}(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) \\
&\quad+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) \\
&\quad-[(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 \\
&\quad-\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) \\
&\quad+\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) \\
&\quad-\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 \\
&\quad+\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) \\
&\quad-\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 \\
&\quad+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 \\
&\quad+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 10, 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

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

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 10, $\sigma_\ell^2 = \infty$. 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\}\\
&= \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\}
\end{aligned}
$$

$$= \frac{2^{2\kappa_\ell}}{\sigma_c^4}\frac{(\sigma_c^2+\sigma^2)^2}{\sigma_c^4\sigma^4\sigma_\epsilon^4} + \frac{\gamma^2 2^{2\kappa_\ell}}{\sigma_c^4\sigma^4}\frac{\sigma_c^2}{\sigma_\epsilon^2\sigma^2} > 0$$

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 3 must be binding and $\sigma_c^2 = \sigma^2/(2^{2\kappa_c} - 1)$. We have obtained the desired conclusion. ☐

A.10 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 $\sigma_c^2 = \sigma^2/(2^{2\kappa_c} - 1)$, $\sigma_{\epsilon\ell}^2 = \infty$, 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 8 that $\sigma_{\epsilon\ell}^2 = \infty$ and from Lemma 11 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 4. We know the expression of $E(a_\ell - \theta)^2\Big|_{\gamma=0}$ from Equation 6. ☐

A.11 Equilibrium Characterization under the Decentralized Regime with $\gamma = 1$

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

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

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

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

Proof. It directly follows from Lemma 9 that $\sigma_\ell^2 = \infty$ and from Lemma 11 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 4. We can derive the expression of $E(a_\ell-\theta)^2\Big|_{\gamma=0}$ by invoking the formula in Lemma 5 and specializing it with the expression of a_ℓ for $\gamma=1$ in Lemma 4:[38]

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

□

A.12 Proof of Proposition 2

Proof. In the proof of Lemma 11, we have obtained Equation 13:

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

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

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

where the last equality follows again from the fact that Constraint 4 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 10 that σ_ℓ^2 weakly increases with γ, we conclude that $E(a_\ell-\theta)^2$ strictly increases with γ. □

A.13 Proof of Lemma 12

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 $\kappa_\ell>\kappa_c$. □

[38] The expression can alternatively be obtained from Equation 13.

A.14 Proof of Lemma 13

Proof. According to Proposition 7, we know

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

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

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

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

A.15 Proof of Corollary 1

Proof. Using Equation 13 in the proof of Lemma 11, we obtain

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

where $\sigma_c=\sigma^2/(2^{2\kappa_c}-1)$ (Lemma 11), the second inequality follows from the fact that $\sigma_\ell^2=\infty$ when $\gamma=\bar\gamma$ (Lemma 10), 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.16 Optimal Policy in a Non-strategic Environment

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

$$a_c = E(\theta|\theta_c, s_\ell') = \frac{\frac{\theta_c}{\sigma_c^2} + \frac{s_\ell'}{(\sigma_\ell^2+\sigma_\epsilon^2)}}{\frac{1}{(\sigma_\ell^2+\sigma_\epsilon^2)} + \frac{1}{\sigma_c^2} + \frac{1}{\sigma^2}}, \tag{14}$$

with $s_\ell' = \theta_\ell + \epsilon$.

Under the decentralized regime, the optimal policy for the local government is given by

$$a_\ell = (1-\gamma)E(\theta|\theta_\ell, s_c') + \gamma E(\theta_c|\theta_\ell, s_c') = k_1'\theta_\ell + k_2's_c', \tag{15}$$

with $s_c' = \theta_c + \epsilon$ and

$$\begin{aligned} k_1' &\equiv \frac{\frac{1-\gamma}{\sigma_\ell^2}}{\frac{1}{\sigma_\ell^2} + \frac{1}{\sigma_c^2+\sigma_\epsilon^2} + \frac{1}{\sigma^2}} + \frac{\frac{\gamma}{\sigma_\ell^2+\sigma_c^2+\sigma_c^2\sigma_\ell^2/\sigma^2}}{\frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}} \\ k_2' &\equiv \frac{\frac{1-\gamma}{\sigma_c^2+\sigma_\epsilon^2}}{\frac{1}{\sigma_\ell^2} + \frac{1}{\sigma_c^2+\sigma_\epsilon^2} + \frac{1}{\sigma^2}} + \frac{\frac{\gamma}{\sigma_\epsilon^2}}{\frac{1}{\sigma_\epsilon^2} + \frac{1}{\sigma_c^2+(1/\sigma^2+1/\sigma_\ell^2)^{-1}}}. \end{aligned}$$

Proof. The proof directly follows from Lemmas 3 and 4 by letting $\sigma_{\epsilon\ell}^2 = \sigma_{\epsilon c}^2 = \infty$.

Alternatively, to derive the expressions of $E(\theta|\theta_c, s_\ell')$ and $E(\theta|\theta_\ell, s_c')$, we can invoke the following Bayesian updating rule with a normal prior.

Lemma 19. Let $\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$ and $x_i|\mu \sim \mathcal{N}(\mu, \sigma_i^2)$ with $i = 1, 2, ..., n.$. Conditional on μ, x_1, x_2,..., and x_n are independent. If μ_0, σ_0^2, and σ_i^2 are known, then[39]

$$\mu|x_1, x_2, ..., x_n \sim \mathcal{N}\left(\frac{\sum_{i=1}^n(x_i/\sigma_i^2)+\mu_0/\sigma_0^2}{\sum_{i=1}^n(1/\sigma_i^2)+1/\sigma_0^2}, \left(\sum_{i=1}^n \frac{1}{\sigma_i^2} + \frac{1}{\sigma_0^2}\right)^{-1}\right).$$

[39]To see this result, writing the probability density function explicitly, we have

$$f(x_1, x_2, ..., x_n, \mu) = f(x_1, x_2, ..., x_n|\mu)f(\mu) \propto \exp\left\{-\sum_{i=1}^n (x_i-\mu)^2/2\sigma_i^2 - (\mu-\mu_0)^2/2\sigma_0^2\right\}.$$

Given $f(\mu|x_1, x_2, ..., x_n) = f(x_1, x_2, ..., x_n, \mu)/\int f(x_1, x_2, ..., x_n, \mu)d\mu$, ignoring the normalizing constant, then

$$f(\mu|x_1, x_2, ..., x_n) \propto f(x_1, x_2, ..., x_n, \mu) \propto \exp\left\{-\frac{\left(\mu - \frac{\sum_{i=1}^n(x_i/\sigma_i^2)+\mu_0/\sigma_0^2}{\sum_{i=1}^n(1/\sigma_i^2)+1/\sigma_0^2}\right)^2}{2\left(\sum_{i=1}^n(1/\sigma_i^2)+1/\sigma_0^2\right)^{-1}}\right\},$$

which implies the posterior distribution in the lemma. Moreover, if $\sigma_i^2 = \sigma^2$ for $i = 1, 2, ..., n$, then we obtain the familiar posterior distribution, $\mu|x_1, x_2, ..., x_n \sim \mathcal{N}(\frac{n\bar{x}/\sigma^2+\mu_0/\sigma_0^2}{n/\sigma^2+1/\sigma_0^2}, (\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2})^{-1})$. with $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$.

Since $\theta \sim \mathcal{N}(0, \sigma^2)$, $\theta_c|\theta \sim \mathcal{N}(\theta, \sigma_c^2)$, $s'_\ell|\theta \sim \mathcal{N}(\theta, \sigma_\ell^2 + \sigma_\epsilon^2)$, and θ_c and s'_ℓ are conditionally independent, applying Lemma 19, we obtain

$$\theta|\theta_c, s'_\ell \sim \mathcal{N}\left(\frac{\theta_c/\sigma_c^2 + s'_\ell/(\sigma_\ell^2+\sigma_\epsilon^2)}{1/\sigma_c^2 + 1/(\sigma_\ell^2+\sigma_\epsilon^2) + 1/\sigma^2}, \left(\frac{1}{\sigma_c^2} + \frac{1}{\sigma_\ell^2+\sigma_\epsilon^2} + \frac{1}{\sigma^2}\right)^{-1}\right).$$

Similarly, given that $\theta_\ell|\theta \sim \mathcal{N}(\theta, \sigma_\ell^2)$, $s'_c|\theta \sim \mathcal{N}(\theta, \sigma_c^2 + \sigma_\epsilon^2)$, and θ_ℓ and s'_c are conditionally independent, we have

$$\theta|\theta_\ell, s'_c \sim \mathcal{N}\left(\frac{\theta_\ell/\sigma_\ell^2 + s'_c/(\sigma_c^2+\sigma_\epsilon^2)}{1/\sigma_\ell^2 + 1/(\sigma_c^2+\sigma_\epsilon^2) + 1/\sigma^2}, \left(\frac{1}{\sigma_\ell^2} + \frac{1}{\sigma_c^2+\sigma_\epsilon^2} + \frac{1}{\sigma^2}\right)^{-1}\right).$$

$\square$

A.17 $E(a_c - \theta)^2$, $E(a_\ell - \theta)^2\Big|_{\gamma=0}$, and $E(a_\ell - \theta)^2\Big|_{\gamma=1}$ in a Non-strategic Environment

Lemma 20. 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,

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

Proof. The expressions are obtained from taking the expressions of a_c and a_ℓ from Lemma 18 and applying the results in Lemma 5. Under the decentralized regime, when $\gamma = 1$, we have

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

$\square$

A.18 Proof of Lemma 14

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 20. $\square$

A.19 Proof of Lemma 15

Proof. According to Lemma 20, we have

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

with the numerator given by

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

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 $\sigma_\epsilon^2<\bar{\sigma}_\epsilon^2$. $\square$

A.20 Proof of Proposition 3

Proof. According to Lemma 20, 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. $\square$

A.21 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

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

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.22 Proof of Lemma 16

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 11 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 11 to hold, it suffices to have

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

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

A.23 Proof of Lemma 17

Proof. Following the proof of Lemma 10, 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 9, with $F(\sigma_\ell^2, \sigma_{\epsilon\ell}^2)$ given by

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

We know Constraint 9 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 \\
&\quad - \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} \\
&\quad - \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)} \\
&\quad +(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)} \\
&\quad - \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 \\
&\quad + \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^*.

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

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

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

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

which implies that x^* must be unique.

Following the proof of Lemma 10, we define

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

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

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

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 11 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

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

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 $\lim_{\gamma\to 0} L(\gamma) \to \infty$. 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 10. When $\gamma \in (\underline{\gamma}', \bar{\gamma}')$, we have $H'(x^*) = 0$, or equivalently

$$\begin{aligned} \frac{\gamma^2\sigma_c^4}{K_\ell\sigma_c^4 - \lambda} &= \frac{B^2\frac{\gamma^2 K_\ell\sigma_c^8\lambda}{K_\ell\sigma_c^4 - \lambda} + 2(1-\gamma)\gamma K_\ell\sigma_c^6 B + (1-\gamma)^2 K_\ell\sigma_c^4\left(\frac{K_\ell\sigma_c^4}{\lambda} - 1\right)}{\left[(K_\ell\sigma_c^4/\lambda - 1)x^* - K_\ell\sigma_c^4 B\right]^2} \\ \left[(K_\ell\sigma_c^4 - \lambda)x^* - \lambda K_\ell\sigma_c^4 B\right]^2 &= \lambda\left(B^2 K_\ell\lambda^2\sigma_c^4 + \frac{2(1-\gamma)}{\gamma}BK_\ell\lambda\sigma_c^2(K_\ell\sigma_c^4 - \lambda) + \left(\frac{1-\gamma}{\gamma}\right)^2 K_\ell(K_\ell\sigma_c^4 - \lambda)^2\right) \\ \left[(K_\ell\sigma_c^4 - \lambda)x^* - \lambda K_\ell\sigma_c^4 B\right]^2 &= \lambda K_\ell\left(B\lambda\sigma_c^2 + \left(\frac{1-\gamma}{\gamma}\right)(K_\ell\sigma_c^4 - \lambda)\right)^2 \\ (K_\ell\sigma_c^4 - \lambda)x^* - \lambda K_\ell\sigma_c^4 B &= \lambda^{1/2}K_\ell^{1/2}\left(B\lambda\sigma_c^2 + \left(\frac{1-\gamma}{\gamma}\right)(K_\ell\sigma_c^4 - \lambda)\right) \\ x^* &= \frac{1-\gamma}{\gamma}(\lambda K_\ell)^{1/2} + \frac{\lambda BK_\ell^{1/2}\sigma_c^2(K_\ell^{1/2}\sigma_c^2 + \lambda^{1/2})}{K_\ell\sigma_c^4 - \lambda} \\ x^* &= \frac{1-\gamma}{\gamma}(\lambda K_\ell)^{1/2} + \frac{\lambda BK_\ell^{1/2}\sigma_c^2}{K_\ell^{1/2}\sigma_c^2 - \lambda^{1/2}}, \end{aligned} \tag{16}$$

where the last equation nests Equation 8 as a special case. □

A.24 Proof of Theorem 3

We first prove a weaker counterpart of Proposition 2.

Lemma 21. Under the decentralized regime, $E(a_\ell - \theta)^2$ strictly increases with σ_c^2 for σ_c^2 and γ such that $\gamma < \underline{\gamma}'$ or $\gamma > \bar{\gamma}'$.

Proof. Following the proof of Lemma 11, we first obtain

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

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

where the second to last equality follows from the fact that Constraint 9 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 3, 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 10, $\sigma_\ell^2=\infty$. 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 10, we have $\sigma_{\epsilon\ell}^2=\infty$, which implies that Constraint 4 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

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

with

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

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)} \\
&\quad -\frac{\{\sigma_\epsilon^2[(1+\lambda)\sigma^2+(1-\lambda)\sigma_c^2]\left(\sigma_c^2+\sigma_\epsilon^2\right)-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)} \\
&\quad -\frac{\sigma_\epsilon^2\{[(1+\lambda)\sigma^2+(1-\lambda)\sigma_c^2]\left(\sigma_c^2+\sigma_\epsilon^2\right)-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} \\
&= \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

$$
F'(\sigma_c^2)=(1-\gamma^2)F_1'(\sigma_c^2)+\gamma^2F_2'(\sigma_c^2)>0.
$$

$\square$

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

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

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

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 13, we can show that $E(a_\ell - \theta)^2\Big|_{\gamma=1} > \min\{\sigma^2, \sigma_c^2\} > E(a_c - \theta)^2$. $\square$

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

B Additional Figures

References

Aghion, Philippe and Jean Tirole, "Formal and real authority in organizations," *Journal of Political Economy*, 1997, *105* (1), 1–29.

Akai, Nobuo, Masayo Hosoi, and Yukihiro Nishimura, "Fiscal decentralization and economic volatility: evidence from state-level cross-section data of the USA," *Japanese Economic Review*, 2009, *60* (2), 223–235.

Bardhan, Pranab, "Decentralization of governance and development," *Journal of Economic Perspectives*, 2002, *16* (4), 185–205.

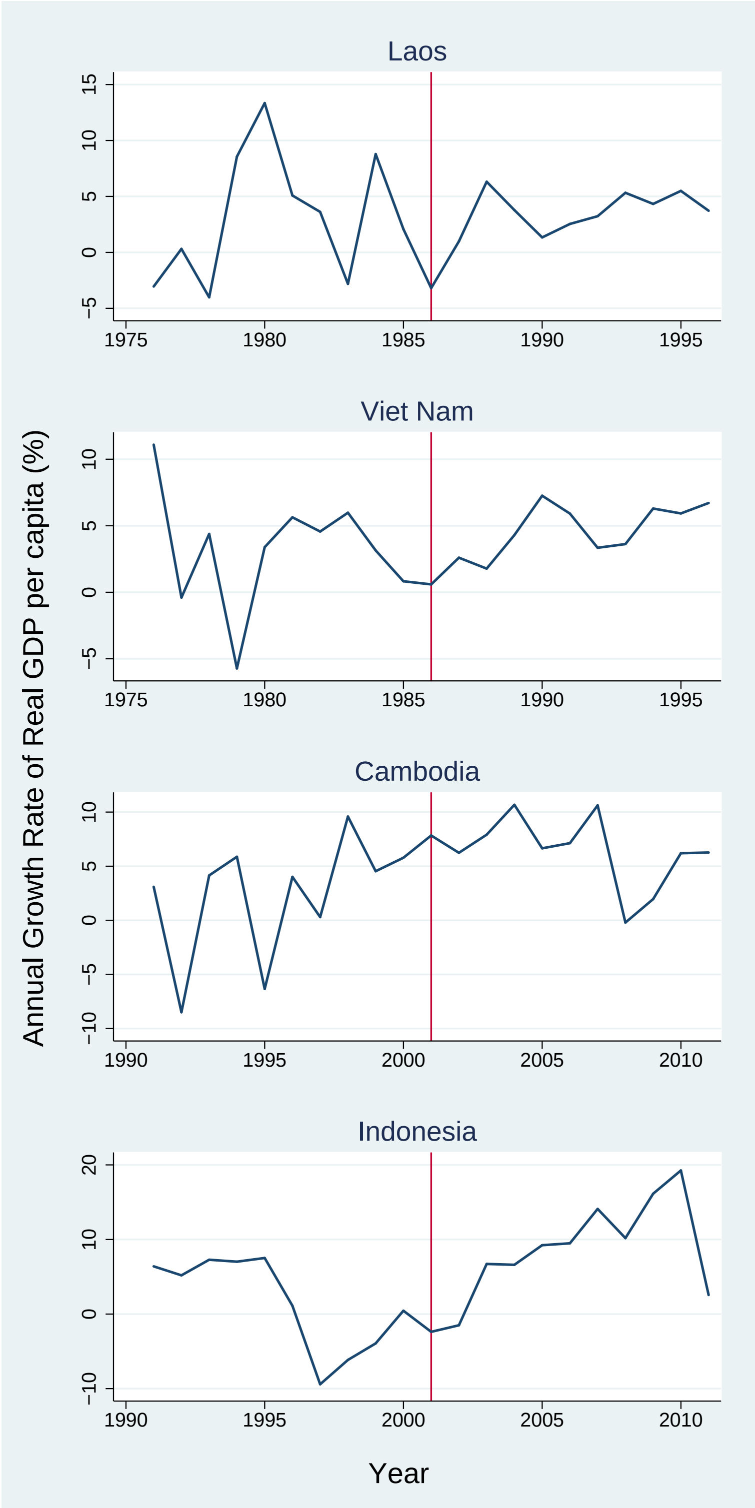


Figure 8: Decentralization and Economic Growth

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.

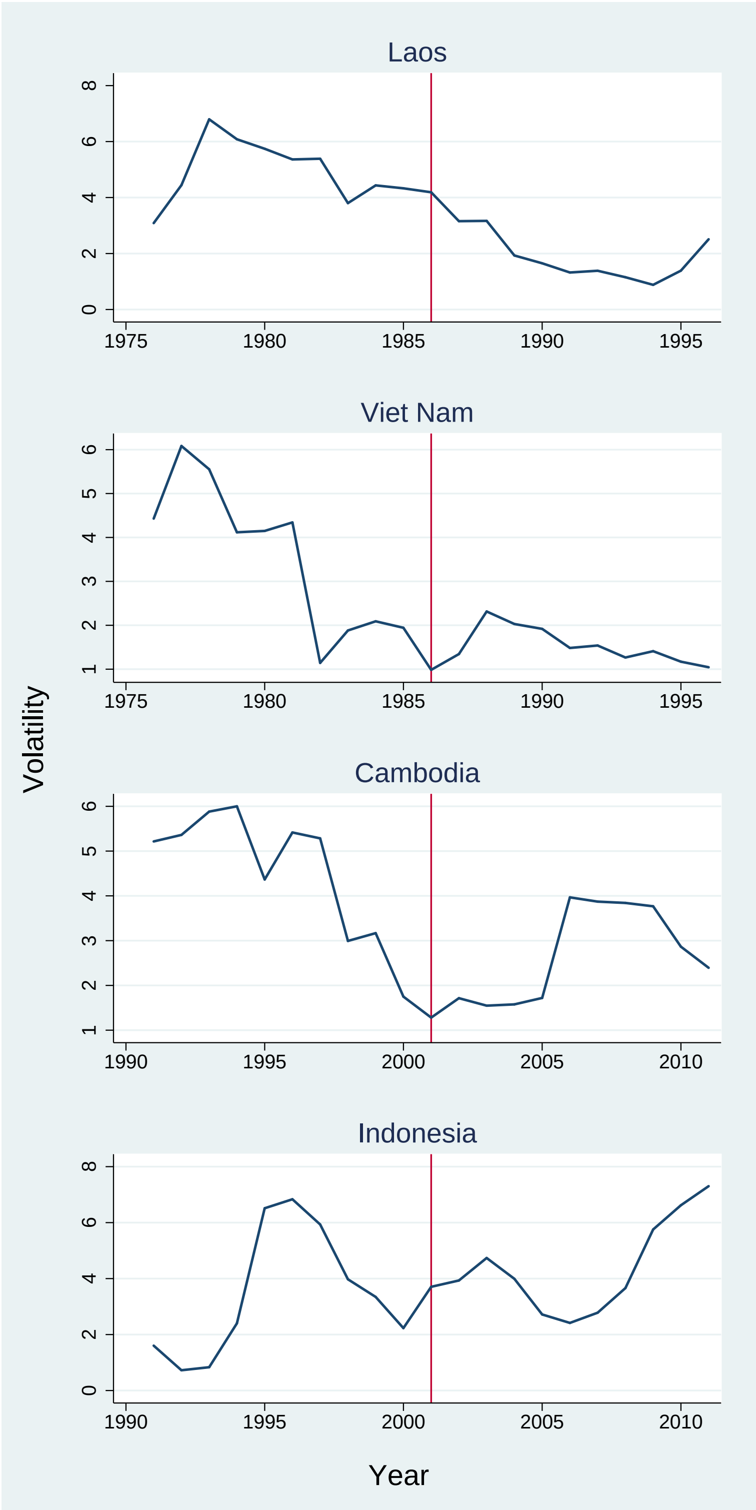


Figure 9: Decentralization and Output Volatility

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

_ , “State and development: The need for a reappraisal of the current literature,” *Journal of Economic Literature*, 2016, *54* (3), 862–92.

Besley, Timothy and Stephen Coate, “Centralized versus decentralized provision of local public goods: a political economy approach,” *Journal of Public Economics*, 2003, *87* (12), 2611–2637.

Board, Oliver J, Andreas Blume, and Kohei Kawamura, “Noisy talk,” *Theoretical Economics*, 2007, *2* (4), 395–440.

Bolton, Patrick, Markus K Brunnermeier, and Laura Veldkamp, “Leadership, coordination, and corporate culture,” *Review of Economic Studies*, 2012, *80* (2), 512–537.

Bowie, Robert R, *Communist China 1955-1959: Policy Documents with Analysis*, Vol. 1, Harvard University Press, 1962.

Cai, Hongbin and Daniel Treisman, “State corroding federalism,” *Journal of Public Economics*, 2004, *88* (3-4), 819–843.

Che, Jiahua, Kim-Sau Chung, and Yang K Lu, “Decentralization and political career concerns,” *Journal of Public Economics*, 2017, *145*, 201–210.

Cheng, Chen, Christopher Li, and Weifeng Zhong, “Governmental system and economic volatility under democracy,” *Working Paper*, 2018.

Cover, Thomas M and Joy A Thomas, *Elements of Information Theory*, John Wiley & Sons, 2012.

Dessein, Wouter, Andrea Galeotti, and Tano Santos, “Rational inattention and organizational focus,” *American Economic Review*, 2016, *106* (6), 1522–36.

Enikolopov, Ruben and Ekaterina Zhuravskaya, “Decentralization and political institutions,” *Journal of Public Economics*, 2007, *91* (11-12), 2261–2290.

Fan, Ziying, Wei Xiong, and Li-An Zhou, “Information distortion in hierarchical organization: a study of China’s great famine,” *Working Paper*, 2016.

Feenstra, Robert C, Robert Inklaar, and Marcel P Timmer, “The next generation of the Penn World Table,” *American Economic Review*, 2015, *105* (10), 3150–82.

Gadenne, Lucie and Monica Singhal, “Decentralization in developing economies,” *Annual Review of Economics*, 2014, *6* (1), 581–604.

Hale-Dorrell, Aaron Todd, “Khrushchev’s Corn Crusade: The Industrial Ideal and Agricultural Practice in the Era of Post-Stalin Reform, 1953-1964.” PhD dissertation, University of North Carolina at Chapel Hill 2014.

Hayek, Friedrich August, “The use of knowledge in society,” *American Economic Review*, 1945, *35* (4), 519–530.

Huang, Zhangkai, Lixing Li, Guangrong Ma, and Lixin Colin Xu, "Hayek, local information, and commanding heights: Decentralizing state-owned enterprises in china," *American Economic Review*, 2017, *107* (8), 2455–78.

Jin, Hehui, Yingyi Qian, and Barry R Weingast, "Regional decentralization and fiscal incentives: Federalism, Chinese style," *Journal of Public Economics*, 2005, *89* (9-10), 1719–1742.

John, Ronald Bruce St, "End of the beginning: Economic reform in Cambodia, Laos, and Vietnam," *Contemporary Southeast Asia*, 1997, *19* (2), 172.

Keen, Michael and Maurice Marchand, "Fiscal competition and the pattern of public spending," *Journal of Public Economics*, 1997, *66* (1), 33–53.

Koren, Miklós and Silvana Tenreyro, "Volatility and development," *The Quarterly Journal of Economics*, 2007, *122* (1), 243–287.

_ and _ , "Technological diversification," *American Economic Review*, 2013, *103* (1), 378–414.

Kung, James Kai-Sing and Shuo Chen, "The tragedy of the nomenklatura: Career incentives and political radicalism during China's Great Leap famine," *American Political Science Review*, 2011, *105* (1), 27–45.

Li, Hongbin and Li-An Zhou, "Political turnover and economic performance: the incentive role of personnel control in China," *Journal of Public Economics*, 2005, *89* (9), 1743–1762.

Li, Wei and Dennis Tao Yang, "The great leap forward: Anatomy of a central planning disaster," *Journal of Political Economy*, 2005, *113* (4), 840–877.

Lin, Justin Yifu, "Collectivization and China's agricultural crisis in 1959-1961," *Journal of Political Economy*, 1990, *98* (6), 1228–1252.

_ , Ran Tao, and Mingxing Liu, "Decentralization and local governance in Chinas economic transition," *Decentralization and Local Governance in Developing Countries: A Comparative Perspective*, 2006, pp. 305–327.

Mackowiak, Bartosz and Mirko Wiederholt, "Optimal sticky prices under rational inattention," *American Economic Review*, 2009, *99* (3), 769–803.

Meisner, Mitch, "Dazhai: the mass line in practice," *Modern China*, 1978, *4* (1), 27–62.

Mookherjee, Dilip, "Political decentralization," *Annual Review of Economics*, 2015, *7* (1), 231–249.

Oates, Wallace E, *Fiscal Federalism*, Edward Elgar Publishing, 1972.

Prendergast, Canice, "A theory of "yes men"," *The American Economic Review*, 1993, pp. 757–770.

Qian, Yingyi and Gerard Roland, "Federalism and the soft budget constraint," *American Economic Review*, 1998, pp. 1143–1162.

Ramey, Garey and Valerie Ramey, "Cross-Country Evidence on the Link between Volatility and Growth," *American Economic Review*, 1995, *85* (5), 1138–51.

Seabright, Paul, "Accountability and decentralisation in government: An incomplete contracts model," *European Economic Review*, 1996, *40* (1), 61–89.

Sims, Christopher A, "Implications of rational inattention," *Journal of Monetary Economics*, 2003, *50* (3), 665–690.

Stuart-Fox, Martin, "Politics and Reform in the Lao People's Democratic Republic," *Working Paper*, 2005.

Tiebout, Charles M, "A pure theory of local expenditures," *Journal of Political Economy*, 1956, *64* (5), 416–424.

ud-din Kassum, Jemal, Andrew Steer, Homi Kharas, and Bert Hofman, "Decentralizing Indonesia," *The World Bank Report*, 2003.

Un, Kheang and Judy Ledgerwood, "Cambodia in 2002: Decentralization and its effects on party politics," *Asian Survey*, 2003, *43* (1), 113–119.

Wang, Boqun and Tao Yang, "Volatility and economic systems: evidence from a large transitional economy," *Working Paper*, 2016.

Weitzman, Martin L, "The new Soviet incentive model," *The Bell Journal of Economics*, 1976, pp. 251–257.

Wetzel, Deborah, "Decentralization in the transition economies: Challenges and the road ahead," 2001.

World Bank, "Viet Nam Transition to the Market: An Economic Report," *The World Bank Official Report*, 1993.

_ , "Cambodia: Description of Legislation/Policy," *The World Bank Country Summary Brief*, 2015.

Wu, Guoguang, "Documentary Politics: Hypotheses, Process, and Case Studies," in "Decision-Making in Deng's China, Perspectives from Insiders," M.E. Sharpe, 1995.

Wu, Jinglian and Bruce L Reynolds, "Choosing a strategy for China's economic reform," *The American Economic Review*, 1988, *78* (2), 461–466.

Xu, Chenggang, "The fundamental institutions of China's reforms and development," *Journal of Economic Literature*, 2011, *49* (4), 1076–1151.

Zhou, Taihe, ed., *Dangdai Zhongguo de jing ji tizhi gaige*, China Social Science Press, 1984.

Zhuravskaya, Ekaterina V, "Incentives to provide local public goods: fiscal federalism, Russian style," *Journal of Public Economics*, 2000, *76* (3), 337–368.