Governmental Structure: Vertical or Horizontal?

Yunyun Lv†
Department of Economics,
University of Pittsburgh,
Pittsburgh PA

October 7, 2022

Abstract

This paper explores the interaction between the hierarchical structures of governments and career concerns of local officials in China using an infinite-horizon principal-agent model. Local officials either keep working to be promoted or shirk to collect benefits and get fired in the next period. The central government infers from the output of officials whether incumbents work or shirk and promotes those who work. The question is: what structure generates the highest total output? This paper shows that there exists a vertical structure that dominates all horizontal structures. Vertical structures generate less uncertainty in the promotion process and hence a clearer career path than horizontal structures. Therefore, vertical structures are more efficient in incentivizing local officials to work.

1 Introduction

To facilitate the economic reform in the late 1970s, subnational government structure in China has been modified through a series of reforms, including the nationwide creation of prefecture-level cities with subordinate counties since the

† ⟨yunyun.lv@pitt.edu⟩ The author is grateful to Kim-Sau Chung, Iza Ding, Luca Rigotti, Richard van Weelden, Yang Xie, and Charles Zheng for useful discussions.

1980s and the "province-managing-county" (PMC) reform since 2003. Local governance in China is currently an amalgam of the "city-managing-county" (CMC) and PMC systems, which are a deeper hierarchy with more layers and a flatter hierarchy with fewer layers, respectively. The hierarchical structure in China comprises five layers: the central government, provinces, prefectures, counties, and townships (Ma, 2005; Zhou, 2007; Li et al., 2016). However, the reforms caused much controversy among experts. The optimal government structure remains unsettled. Concerns are raised, such as the loss of control at the bottom of the hierarchy, the over-concentration of resources and authorities at the prefecture level, and the inadequate number of provincial units at the top.

I compare the total output generated in equilibrium across different structures and find that the vertical structure is optimal. In this model a Bureaucrat's ultimate goal is to climb up the ladder and become the Leader of the local government. Bureaucrats compete with each other by showing to the central government that they are not corrupted, that they are converting local public resources into output instead of stealing it. The central government needs to find the right amount of resources allocated to each Bureaucrat to maximize the total output without luring Bureaucrats into corruption by giving them too many resources. Vertical structures are better because the career path is clearer: There is no competition within the same layer, hence promotion happens more rapidly. Bureaucrats then have stronger incentives to work and therefore can be allocated with more resources than in flatter structures.

The following example illustrates this intuition. Consider two structures, a horizontal one and a vertical one depicted as follows:

Figure 1a: Horizontal Structure

Figure 1b: Vertical Structure

Figure 1: Comparison of Horizontal and Vertical Structure

In the horizontal structure, Bureaucrat 1 and 2 each has half the chance of being promoted every time a Leader is fired. The office of whoever is promoted will be filled with a random new entrant. In expectation a Bureaucrat has to wait for two Leaders to be fired before becoming the Leader. In the vertical structure, if the Leader is fired, Bureaucrat 1 becomes the new Leader immediately and Bureaucrat 2 is promoted to position 1. Notice that Bureaucrat 2 at the bottom only has to wait for two Leaders (the current Leader and Bureaucrat 1 in the future) to be fired before he becomes the Leader. Hence in the vertical structure the average time to become the Leader is 1.5 periods, shorter than 2 period in the horizontal structure. If the value of being the Leader in two structures are the same, then Bureaucrats in the vertical structure have stronger incentives to work on average, and hence the central government can allocate more resources to all Bureaucrats without inducing corruption.

The main results shed light on three important questions involved in the policy debates: (i) Should more prefectures/counties be arranged? This paper shows that arranging more positions in the government structure generates a higher total output. Therefore, setting up more prefectures/counties is beneficial as long as setting up an additional office does not cost much. (ii) Are symmetric structures better than asymmetric structures? Symmetric horizontal structures are optimal. If officials with different amount of authorities are indeed treated as the same when they compete for promotions, designing a complicated asymmetric structure is unnecessary. The simple and easy-to-execute symmetric structure is sufficient to stimulate good performance. (iii) Is the CMC system, the more vertical system that features more layers, better than the PMC system, the more horizontal system that features fewer layers? Yes, the CMC system stimulates better performance of local officials. A more vertical structure implies less uncertainty in the promotion process and is hence more efficient in providing incentives for Bureaucrats to work.

This study relates to the rich literature on optimal organization hierarchies in economics, which indicates a major trade-off among different structures. A hierarchy with more layers is subject to loss of control because every additional layer leads to extra agency costs (Qian, 1994; Rajan and Zingales, 2001; Besley and Ghatak, 2005; Mookherjee et al., 2013), whereas one with fewer layers is subject to loss of productivity because of excessive span of control at every layer (Williamson, 1975; Patacconi, 2009). Empirical findings show that the overall effect of flattening the government hierarchy in China (the PMC reform) is negative on economic performance (Li et al., 2016). The policy debates in China add new elements to the literature, and the most important among which is the interaction between the political career concerns of local officials and the hierarchical structures of governments.

Another related body of literature is that of federalism, which studies how resources and authorities should be allocated among different governments (Myerson, 2006). The traditional federalism literature adopts strong assumptions, such as benevolent governments, and identifies some technical reasons for the optimal degree of decentralization, including externalities among subnational governments. The more recent federalism literature relaxes or challenges some of these assumptions and studies how centralization/decentralization affects politicians' incentives (Coate and Knight, 2007; Boffa et al., 2016). This study benefits substantially from the most related paper within this literature, Che et al. (2017), which explores the interaction of political career concerns and decentralization. Che et al. (2017) use an overlapping principal-agent model and a two-layer hierarchy with one leader and two bureaucrats. They focus on how different allocations of authorities between the national and sub-national governments affect a bureaucrat's incentives. This model, however, focuses on the trade-offs in choosing among different hierarchical structures. In particular, the allocation of authorities/resources is part of the design of a hierarchical structure.

A key feature of local governance in China is the important role that the career concerns of politicians play in motivating them to perform. Local officials are not elected by the public but appointed by the central government. Accountability is top-down, as opposed to bottom-up. They are not subject to re-election pressure or checks and balances from local citizens. This paper shows that there is another trade-off among different structures because they create different career paths for local officials.

2 Model

Time is discrete with infinite horizon, denoted by $t = 0, 1, 2, \ldots$. Future payoffs are discounted at a common rate $\delta \in (0, 1)$. An n-th structure consists of $n + 1$ positions of different ranks, indexed by $i = 0, 1, \ldots, n$. One unit of resources can be allocated among these positions: $\sum_{i=0}^{n} z_i \leq 1$ where $z_i > 0$ represents the amount of resources allocated to position i with $i = 0, 1, \ldots, n$. This constraint is referred to as the budget constraint hereafter. One official is assigned to each position. The official at position 0 is the Leader, and other local officials are called Bureaucrats. Without loss of generality, their ranks satisfy $Rank(0) > Rank(1) \geq \ldots \geq Rank(n)$, and $\{Rank(0), \ldots, Rank(n)\} = \{1, \ldots, k, k+1\}$ for some $k \leq n, k \in \mathbb{N}$. Notice that $(k + 1)$ is the total number of different ranks, or simply the number of layers in the structure. One office is arranged at each position to accommodate the officials. The central government spends $\alpha > 0$ to maintain each office in each period. An official either works or shirks, but can randomize. Official i chooses strategy $w_i \in [0, 1]$ at the beginning of each period, where w_i denotes his probability of working.

The official possesses the technology to transform resources into output. The central government is risk-neutral. Expected output level y is determined by the effort level of the official and amount of resources available:

$$y_{it} = z_i w_{it},$$

where subscripts i and t stand for position and time, respectively. Official i collects perk z_i only if he shirks. In other words, all the resources allocated to this region is transformed into output if the official works. While he puts all the resource into his own pocket if he shirks. The official's payoff is $z_i - y_{it}$.

At the end of each period, output levels at each position are revealed, and shirking Bureaucrats are fired. If the Leader shirks, he is fired with probability $P \in [0, 1]$. The Leader remains at the top position if he is not terminated. The promotion process starts from filling the empty position(s) of the highest rank. Each empty position is filled with the working Bureaucrat of the highest rank among all Bureaucrats below the position. If there are multiple working Bureaucrats of the same highest rank below an empty position, ties are resolved by fair lottery. New players are randomly assigned to vacant positions after the promotion of all working officials. If no higher position is open, working officials stay at their current positions.

In period 0, the central government designs the hierarchical structure by choosing $n \in \mathbb{N}$, $P \in [0, 1]$, $z_0, \ldots, z_n \in [0, 1]$, the rank of each position, and proposes a strategy profile $w = (w_0, w_1, \ldots, w_n)$ where w_i is the suggested working probability for official at position i in every period.[1] The decided government structure and the proposed strategy are fixed over time and common knowledge to all officials. The central government maximizes the expected total output minus the maintenance cost of offices in each period:

$$\Pi = Y - \alpha(n + 1),$$

where $Y = \sum_{i=0}^{n} z_i w_i$ is the expected total output in one period and $\alpha(n + 1)$ is the cost of maintaining the office for the Leader and n Bureaucrats. The timeline is shown in Figure 2.

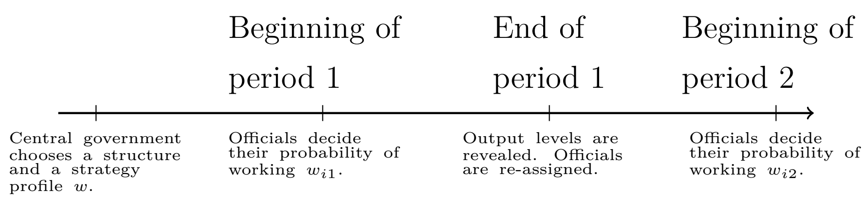


Figure 2: Timeline

[1]In other words, for the game of infinite horizon played by officials I only focus on stationary equilibria where officials assigned to the same position always use the same strategy in every period.

Tree diagrams are used to represent the hierarchical structures. The node at the top denotes the Leader who is connected to his direct subordinates. These positions are then connected with their direct subordinates. For example, a hierarchical structure of a provincial government consisting of four positions and three layers is represented by the following tree diagram:

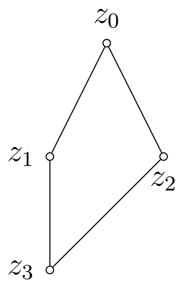


Figure 3: Example: Four-Official Three-Layer Structure

The top node represents a province, position 1 and 2 represent two prefectures in this province, and position 3 is a county. The lines depict possible career paths. When the official at position 1 or 2 is either fired or promoted, official 3 can be promoted to the prefecture level if he works.

The following sections start from analyzing two special categories of hierarchical structures: horizontal structures and vertical structures. An n-th horizontal structure h_n consists of only two ranks, whereas an n-th vertical structure v_n has only one position at each rank:

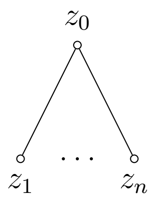


Figure 4a. horizontal structure h_n

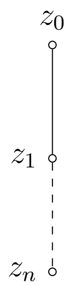


Figure 4b. vertical structure v_n

Figure 4: Graphic illustration

In the end I compare general structures with restricted attention to the case where the central government chooses a pure strategy profile. The solution concept is subgame perfect equilibrium.

3 Horizontal versus Vertical

First of all, there are two simple results:

- The Leader has no incentive to work because he cannot be further promoted, and because he cannot collect any benefit if he works. Therefore, $w_0 = 0$ in any equilibrium. The Leader's strategy is then omitted in the strategy profile in the following analysis.
- The incentives of officials are not influenced if all z_i's are multiplied by a positive constant simultaneously.

A direct application of the second result is that an equilibrium of a structure is also an equilibrium of another structure that is obtained by scaling up or down the original structure. Consequently, the budget constraint is always binding in optimal structures.

I first show that it cannot be optimal to have some Bureaucrats strictly prefer working or shirking in equilibrium.

Lemma 1. For any n and P, the structure that generates the highest total output must be one where all Bureaucrats are indifferent between working and shirking.

All omitted proofs are in the Appendix. The idea is simple: If a Bureaucrat strictly prefers to work (shirk), the central government can allocate more (fewer) resources to him without changing any official's incentives nor the equilibrium strategy and hence obtain a higher equilibrium output.

3.1 Horizontal Structures

This section shows that, given any n and P, a horizontal structure that generates the highest total output is symmetric and has a symmetric equilibrium in which all Bureaucrats work with probability 1 and are indifferent between working and shirking. Moreover, either $n = 0$ with any $P \in [0, 1]$ is optimal, or the optimal value of (n, P) is uniquely determined by α (cost of maintaining each office) and δ (discount rate).

Firstly, notice that it is not optimal to have any Bureaucrat shirking in equilibrium ($w_i = 0$ for some i). Dismissing all shirking Bureaucrats has no effect on total output or other Bureaucrats' incentives. For any equilibrium strategy profile $w \in (0,1]^n$, n, and P, the following algorithm pins down the unique horizontal structure, denoted by $h^*_{n,P}(w)$, that implements w.

- Let the amount of authority of the Leader be z_0. The continuation value of holding the Leader's position is then proportional to z_0.

- Find out for each Bureaucrat i his probability of being promoted using his opponents' strategy w_{-i}. This probability also depends on n and P.

- Let z_i equal the expected return of working, which is the probability of promotion times the value of being the Leader. Hence Bureaucrat i is indifferent between working and shirking. Moreover, $z_i > 0$ is proportional to z_0.

- Solve for z_0 such that the budget constraint binds.

Define the output generated by a structure s with equilibrium strategy profile w as $Y(s, w, n, P)$. The output generated by structure $h^*_{n,P}(w)$ in equilibrium can be rewritten as a function of w, denoted by $Y^*_{n,P}(w) := Y(h^*_{n,P}(w), w, n, P)$.

Next I show that it is optimal to make all Bureaucrats work with probability 1. This result is less straightforward because increasing the working probability of one Bureaucrat leads to a lower probability of promotion (and hence weaker incentives to work) for all other Bureaucrats. The key question is: What is the difference between $h^*_{n,P}(w)$ and $h^*_{n,P}(w')$, where $w' = ({w_i}', w_{-i})$ and ${w_i}' > w_i$? I modify $h^*_{n,P}(w)$ in two steps to obtain $h^*_{n,P}(w')$, and show that the total equilibrium output increases after each step: (i) Other Bureaucrats now face a more hard-working opponent and hence have lower chances of promotion if they work. Allocate less authorities to them such that they are still indifferent between working and shirking in equilibrium. (ii) Since the total amount of authorities is lower than before, the whole structure can be scaled up until the budget constraint is binding. Hence the total output is higher. To see why the first step also leads to a higher equilibrium total output, I compare the two effects of it:

- Gain in output of Bureaucrat i because he now works with a higher probability.
- Loss in output of other Bureaucrats because less authorities are assigned to them.

Intuitively, the gain is a direct effect of the increase in the working probability of Bureaucrat i, while the loss is triggered by other Bureaucrats' worries about their future payoffs because they face smaller chances of promotion. But this negative effect is diluted because (i) future values are discounted, (ii) promotion only happens when the Leader is fired, which happens with probability $P \leq 1$, and (iii) the competition between any two Bureaucrats exists only if they both work, which happens with probability less than 1. As a result the gain dominates the loss in output. It is then straightforward to conclude the following result:

Proposition 1. Given any n and P, the optimal horizontal structure $h^*_{n,P}$ implements an all-working strategy $w^1_n = (1, \dots, 1)$ with the following distribution of resources:

$$z_0 = \frac{(1-\delta(1-P))\left(1-\delta\left(1-\frac{P}{n}\right)\right)}{(1-\delta(1-P))\left(1-\delta\left(1-\frac{P}{n}\right)\right)+\delta P}, \tag{1}$$

and

$$z_1 = \ldots = z_n = \frac{\delta\frac{P}{n}}{(1-\delta(1-P))\left(1-\delta\left(1-\frac{P}{n}\right)\right)+\delta P}. \tag{2}$$

In particular, this is a symmetric structure. All Bureaucrats are assigned with the same amount of resources and are indifferent between working and shirking. Since z_0 is deceasing in n, the expected total output $Y^*_{n,P}(w^1_n) = \sum_{i=1} nz_i = 1 - z_0$ is increasing in n.

It is natural to ask what is the optimal horizontal structure if the central government is free to choose n, the number of local offices, and P, the Leader's probability of getting fired. Proposition 1 shows that given any P, a bigger n leads to a higher total output. But the increase in the total output brought by an additional position diminishes and goes to zero as n becomes bigger. This is because the net gain of adding another position to the structure decreases as other Bureaucrats work more. Therefore, the constant maintenance cost for each office ensures that the optimal number of offices for any given P is finite. Next I solve for the optimal value of n given P and optimal P given n, and then combine them to get the optimal (n, P). However I must ignore the restriction that n is an integer and find the solution, which is non-natural in general.

Proposition 2. For the optimal horizontal structure:

(i) The optimal value of n given P, denoted by $n^*(\alpha, \delta, P)$, is determined by three factors: α, the office cost, δ, the discount rate, and P. Moreover, $n^*(\alpha, \delta, P) \in (0, +\infty)$ if and only if $\alpha(1 - \delta(1 - P)) \leq 1$. Otherwise, $n^*(\alpha, \delta, P) = 0$.

(ii) The choice of P is irrelevant if $n = 0$. The optimal value of P given $n > 0$, denoted by $P^*(\delta, n)$, is determined by two factors: δ and n. $P^*(\delta, n)$ increases in n and decreases in δ. Moreover, $P^*(\delta, n) < 1$ if and only if $n < \tilde{n} = \left(\frac{\delta}{1-\delta}\right)^2$.

(iii) Given any α and δ, either $n^*(\alpha, \delta, \hat{P}) = 0$ for some $\hat{P} \in [0, 1]$, or there exists a unique pair $(\hat{n}, \hat{P})$ that maximizes the equilibrium total output, where $(\hat{n}, \hat{P})$ satisfies $\hat{n} = n^*(\alpha, \delta, \hat{P}) > 0$ and $\hat{P} = P^*(\delta, \hat{n}) \in (0, 1]$. In particular, $\hat{n}$ and $\hat{P}$ both decrease in α, $\hat{n}$ increases (decreases) in δ when δ is small (big), and $\hat{P}$ decreases in δ.

Given a fixed P, an increase in n affects total output through two channels: each Bureaucrat's amount of resources decreases as there are more Bureaucrats competing for promotion, and the number of working officials increases. The result is less interesting when α is big and/or P is big: If the output cannot cover the maintenance cost of one single office, it is optimal to not set up any local offices. I will focus on the case when α and P are small enough such that $n^*(\alpha, \delta, P) > 0$.

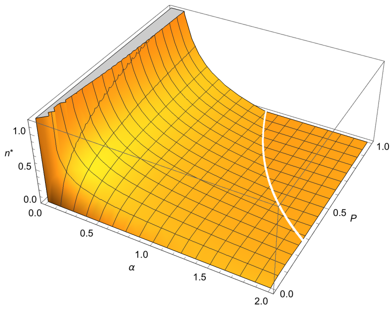


Figure 5a: $n^*(\alpha, 0.7, P)$

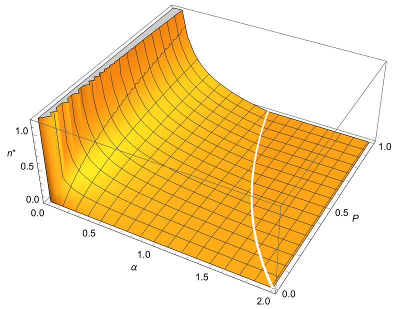


Figure 5b: $n^*(\alpha, 0.5, P)$

Figure 5: Example of $n^*(\alpha, \delta, P)$

Since P is small, promotion happens rarely. The decrease in a Bureaucrat's probability of promotion caused by adding another Bureaucrat into the competition is small. However, the additional output produced by the added Bureaucrat, which equals the amount of authority assigned to each Bureaucrat, is big when n is small, and is small when n is big. Therefore, the optimal n is positive and finite. Moreover, it is shown in the Appendix that central government's objective is concave in n. The optimal value is one of the two consecutive natural numbers that forms the interval of length 1 that contains the non-natural solution, $n^*(\alpha, \delta, P)$.

What is the optimal value of P when n is fixed? The office cost is fixed given that n is fixed. Thus, the optimal value of P is independent of α. An increase in P has two opposing effects on the optimal amount of authority allocated to each Bureaucrat: (i) It is less attractive to be the Leader because the Leader is more likely to be fired in each period. So fewer resources are allocated to Bureaucrats to keep them indifferent between working and shirking. (ii) Bureaucrats' chances of promotion increases. So more resources can be allocated to Bureaucrats. The second effect dominates when P is small, n is big, and δ is small. Because the value of being the Leader only matters when a promotion happens, which is less likely when P is smaller and/or n is bigger. Also future payoffs are less important when δ is smaller. So an intermediate P is optimal, unless either n is sufficiently big or δ is sufficiently small. The optimal P increases in n and decreases in δ.

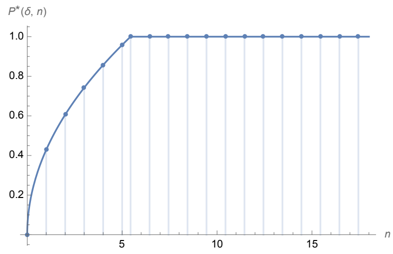


Figure 6a: $P^*(0.7, n)$

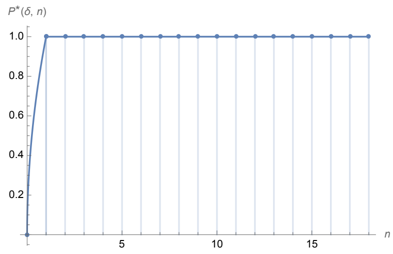


Figure 6b: $P^*(0.5, n)$

Figure 6: Example of $P^*(\delta, n)$

If n is sufficiently big, the Bureaucrat's chance of promotion is always sufficiently small and hence the positive effect always dominates. Therefore, $P^*(\delta, n) = 1$ when n is sufficiently big. If δ is small, people are impatient. It is better to make promotions happen sooner so as to incentivize Bureaucrats to work. Meanwhile, being able to hold the Leader's office for more periods is less attractive since people care less about future payoffs. The positive effect is stronger while the negative effect is weaker when δ is smaller. $P^*(\delta, n)$ decreases in δ.

Both $\hat{n}$ and $\hat{P}$ decrease in α: A higher maintenance cost leads to fewer offices at optimum, and hence also a lower $\hat{P}$ since $P^*(\delta, n)$ increases in n. To see why $\hat{n}$ increase in δ if and only if δ is small while $\hat{P}$ always decreases in δ: When δ is close to 0, people care little about future payoffs. It is optimal to always fire the Leader (picking $\hat{P} = 1$) so that promotion happens more frequently. At the same time $\hat{n}$ must be small so that the level of competition is not too high to discourage working and the office cost is not too high. As δ increases people become more patient, making it easier to incentivize the Bureaucrats to work. The government can decrease P while setting up more local offices (increasing n). However, as n grows, the marginal (positive) effect of n on the total output decreases and finally falls below the office cost α. Hence $\hat{n}$ decreases in δ when δ is big enough.

3.2 Vertical Structures

It is less obvious that the government wants to implement an all-working equilibrium in a vertical structure. Because allowing some Bureaucrats to shirk with a positive probability may also provide stronger incentives to work for Bureaucrats below. It may not be optimal to set only one promotion prize at the top. However, I find a subset of vertical structures that dominate the optimal horizontal structures. Let $v^*_{n,P}$ be the vertical structure such that

- all n Bureaucrats work with probability 1,
- all Bureaucrats are indifferent between working and shirking, and
- the budget constraint is binding.

The uniqueness of $v^*_{n,P}$ for any n and P is easy to verify. I provide the following algorithm to determine $v^*_{n,P}$.

- Suppose the amount of authority of the Leader is z_0.
- Assume all Bureaucrats work with probability 1 and believe their opponents also work with probability 1.
- For Bureaucrat 1, calculate the amount of authority at his position that makes him indifferent between working and shirking, z_1. z_1 is a function of e_{-1} and z_0. Moreover, z_1 is positive and proportional to z_0.
- Repeat the above step for Bureaucrat 2. Again, z_2 is positive and proportional to z_1, and hence also proportional to z_0.

...

- Repeat the above step for Bureaucrat n. z_n is also positive and proportional to z_0.
- Solve for z_0 such that the budget constraint is binding.

3.3 Main Result

Proposition 3. Given any $n > 0$ and $P \in [0, 1]$, the vertical structure $v^*_{n,P}$ generates a higher total output than the horizontal structure $h^*_{n,P}$.

In any structure, excessive resources make a Bureaucrat want to shirk, while inadequate resources harm output. In horizontal structures, it is optimal for the central government to make all Bureaucrats indifferent between working and shirking because this is exactly the edge case, where the total output is maximized subject to the constraint that Bureaucrats work. Notice that the uncertainty about whether the Leader will be fired exists in all structures. But there exists another source of uncertainty in the promotion process under horizontal structures: Every time a Leader is fired, a new Leader in the next period is drawn randomly from all working Bureaucrats. So all Bureaucrats have lower expected payoffs for working, and hence less resources in the optimal structure. However, in the all-working equilibrium of the vertical structure with the same parameters n and P, there is no uncertainty in the promotion process: Every Bureaucrat is promoted one position up when a Leader is fired. So the central government can allocate more authority to Bureaucrats in total in vertical structures than in horizontal structures.

The following thought experiment can help illustrate the intuition. Imagine that a Bureaucrat can choose which of the following two governments to work in. One government adopts the horizontal structure, $h^* = h^*_{n,P}$, while the other adopts the vertical structure, $v = v^*_{n,P} \times m$. The constant m ensures that $z_0^{h^*} = z_0^v$ so that the value of being the Leader in the two structures are the same. The Bureaucrat does not know which position he will be assigned to. He believes that he will be randomly drawn to one Bureaucrat's position in the government, regardless of his choice.

No matter what the government structure is, the Bureaucrat's only hope of benefit is to keep on working and wait to become the Leader so that he can keep collecting perk until he is terminated.

- In the horizontal structure, all Bureaucrats positions are the same. But it is unknown how many periods he has to wait to become the Leader. In expectation, he becomes the Leader after n Leaders are terminated.

- In the vertical structure, it matters where he starts. Once he knows the position he takes, he is aware of the number of Leaders that need to be terminated before he climbs up to the top position. The worst case is that he starts from the bottom position and must wait until n Leaders are fired. In expectation, he will become the Leader after $(n+1)/2$ Leaders are fired.

In expectation an average Bureaucrat becomes the leader sooner in the vertical government. In other words, the average value of working for a Bureaucrat is higher in the vertical structure. Given that all Bureaucrats are indifferent between working and shirking in both structures, the average expected value of shirking is also higher in the vertical structure, which implies that the average amount of resources of Bureaucrats is higher, which means the amount of total resources used in v is more than one. Now scale down v to $v^* = v^*_{n,P}$ such that the budget constraint binds. Then $z_0^{v^*} < z_0^v = z_0^{h^*}$ and hence the vertical structure $v^*_{n,P}$ generates a higher total output than the horizontal structure $h^*_{n,P}$.

Proposition 4. Focus on structures that have an equilibrium where all Bureaucrats work with probability one and are indifferent between working and shirking. For any n and P, the vertical structure $v^*_{n,P}$ generates the highest total output.

The intuition is the same: If there is more than one Bureaucrats of the same rank competing for promotion, put some of these positions into a new layer attached to the bottom. In the more vertical structure, there is less competition in each layer and working is more rewarding. Thus more resources can be allocated to Bureaucrats while all Bureaucrats keep working in equilibrium. Then scale down the whole structure to satisfy the budget constraint. The Leader takes fewer perks and hence the total output is higher.

4 Concluding Remarks

Although characterizing the optimal vertical structure remains an open question, it is shown that vertical structures dominate horizontal structures. Bureaucrats in the vertical structure $v^*_{n,P}$ wait in a queue to be promoted. The queue might be long, but each Bureaucrat is one step closer to the top every time a Leader is fired. The average waiting time is $\frac{n+1}{2}$. While Bureaucrats in the horizontal structure $h^*_{n,P}$ wait in a pool. Only one of them is randomly picked to be promoted every time a Leader is fired. Moreover, the size of the pool do not decrease as a new official joins the pool after each promotion. The average waiting time is n, which is the same as that of the Bureaucrat at the end of the queue in the vertical structure. So it is easier to incentivize working in the vertical structure than in the horizontal structure. As a result, more resources can be allocated to Bureaucrats and hence the total output generated in equilibrium is higher in the vertical structure.

Appendix A: Proof of Lemma 1

Proof. Consider a structure that implements an equilibrium strategy where at least one working Bureaucrat strictly prefers working to shirking. Modify the structure as follows:

1. Ignore the budget constraint. Increase the amount of authority for one bureaucrat who strictly prefers working to shirking by $\epsilon > 0$ with ϵ small enough such that working is still strictly better than shirking.

2. Scale down the structure by multiplying all z_i's by $\frac{1}{1+\epsilon}$ to make the budget constraint bind.

Notice that after the first step, the bureaucrat with more resources still work with probability one, and the value of holding this position is the same as under the old structure. Hence the incentives of all other bureaucrats remain unchanged. Since scaling does not change incentives either, the same strategy profile is still an equilibrium in the new structure. With a higher fraction of total resources distributed to working bureaucrats, the total output is therefore higher in the new structure.

The logic is the same if at least one shirking bureaucrat strictly prefers to shirk: Reduce the amount of resources of shirking bureaucrats by a small amount and then scale up the whole structure. More resources are then allocated to officials that work with a positive probability while the same strategy profile remains an equilibrium in the new structure. Therefore, a higher total output is achieved. $\square$

Appendix B: Proof of Proposition 1

A few lemmas are needed to prove Proposition 1:

Lemma 2. Any horizontal structure with a equilibrium strategy profile where some Bureaucrats work with zero probability in every period is not optimal.

Proof. Construct another structure by dismissing all Bureaucrats who work with zero probability and removing their offices. The incentives of Bureaucrats left are not affected because those who work with zero probability will be fired in next period for sure and never be part of the competition. Hence in the new structure the same equilibrium output level is achieved. Then scale up the new structure such that budget constraint is binding. The same equilibrium remains an equilibrium, and the equilibrium total output is strictly higher. □

Lemma 3. Given any $w = (w_1, \ldots, w_n) \in (0, 1]^n$, in the unique horizontal structure $h^*_{n,P}(w) = (z_0, z_1, \ldots, z_n)$, $\forall i \neq j$,

$$w_i > w_j \Leftrightarrow Q_i > Q_j \Leftrightarrow z_i > z_j,$$

where Q_i denotes Bureaucrat i's chance of promotion if his opponents use the strategy profile w_{-i}, conditional on that Leader is terminated and that Bureaucrat i works.

Proof. Understanding the equivalence among these three inequalities is easy.

- If $w_i > w_j$, then $Q_i(w_{-i}) > Q_j(w_{-j})$. Because, except for themselves, i and j have the same group of opponents, $\{1, \ldots, n\} \setminus \{i, j\}$; While i's another opponent, j, works with a probability lower than j's opponent, i. Thus, $Q_i(w_{-i}) > Q_j(w_{-j})$.

- If $Q_i(w_{-i}) > Q_j(w_{-j})$, then working brings a higher value to i than to j. Given that i and j are indifferent between working and shirking, the value of shirking for i, which equals the amount of perks he can collect, z_i, is also higher than that of j, which equals z_j.

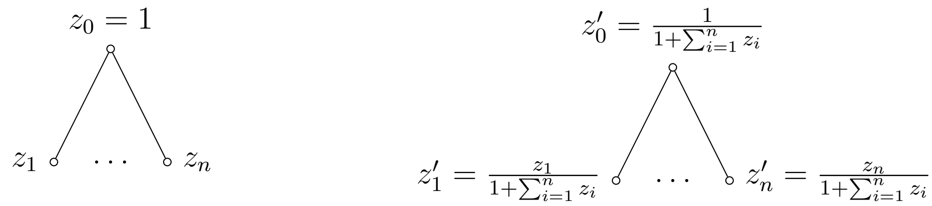
- If $z_i > z_j$, then working brings a higher value to i than j because they are indifferent between working and shirking. Since there is only one position that they can be promoted to, it implies that i has a higher probability of promotion if he works, compared with j. But they have the same group of opponents except for themselves $\{1, \ldots, n\} \setminus \{i, j\}$. Therefore, j works with lower or equal probability than i.

The three statements complete the circle. $\square$

Lemma 4. For any $w = (w_1, \ldots, w_n) \in (0, 1]^n$, let $m = |\{1 \le i \le n | w_i = w_n\}$ and denote by $\underline{w} = w_{n-m+1} = \ldots = w_n$. The equilibrium total output of the proper horizontal structure corresponding to w, $Y^*_{n,P}(w_1, \ldots, w_{n-m}, \underline{w}, \ldots, \underline{w})$ increases in $\underline{w}$.

Proof. Given the proposed strategy profile $w = (w_1, \ldots, w_n)$, I adopt the following algorithm, which is an alternative to the one proposed above, to determine the unique proper horizontal structure:

Step 1: calculate the ratios, Step 2: scale down the structure.



Assume without loss of generality that $w_1 \ge \ldots \ge w_n$, $Q_1 \ge \ldots \ge Q_n$, and $z_1 \ge \ldots \ge z_n$. Note that for each $i = 1, \ldots, n$, z_i represents the ratio of the amount of authority at position i to the Leader's authority. Since Bureaucrat i is indifferent between working and shirking, these two options brings the same value to him: $V_i = z_i + 0 = 0 + \delta(PQ_iV_L + (1 - PQ_i)V_i)$, where V_i and $V_L = \frac{1}{1-\delta(1-P)}$ are the expected value of holding position i and 0 (the Leader's position) in the structure in Step 1, respectively. Therefore,

$$z_i = \frac{\delta PQ_iV_L}{1 - \delta(1 - PQ_i)}. \tag{B.1}$$

Differentiating z_i with respect to Q_i twice gives

$$\frac{\partial z_i}{\partial Q_i} = \frac{(1 - \delta)\delta PV_L}{(1 - \delta(1 - PQ_i))^2} > 0 \tag{B.2}$$

and
$$\frac{\partial^2 z_i}{\partial Q_i^2} < 0. \tag{B.3}$$
As a function of w_{-i}, Q_i depends on the total number of working Bureaucrats. There are $(n+1)$ cases to be considered: all opponents work, all but one opponents work, ..., all but one opponents shirk, and all opponents shirk. For $i = 1, \ldots, n$,

$$\begin{aligned} Q_i = \frac{\prod_{j\neq i} w_j}{n} + &\frac{\sum_{k\neq i}(1-w_k)\prod_{j\neq i,k} w_j}{n-1} + \ldots \\ &+ \frac{\sum_{j\neq i} w_j \prod_{k\neq i,j}(1-w_k)}{2} + \prod_{k\neq i}(1-w_k) \end{aligned}$$

This generates some useful observations about derivatives:
$$\frac{\partial Q_i}{\partial w_j} < 0, \ \forall i \neq j \in \{1, \ldots, n\}. \tag{B.4}$$

Recall that m is defined as the number of Bureaucrats working with the lowest probability: $m = |\{1 \leq i \leq n | w_i = w_n\}|$. Consider the case $w_1 \geq \ldots \geq w_{n-1} > w_n$ (i.e. $m = 1$) first. Notice that
$$\begin{aligned} &\frac{\partial Q_n}{\partial w_{n-1}} = \frac{\partial Q_{n-1}}{\partial w_n} \\ &= -\frac{\prod_{j=1}^{n-2} w_j}{n(n-1)} - \frac{\sum_{k=1}^{n-2}(1-w_k)\prod_{j\neq n,n-1,k} w_j}{(n-1)(n-2)} - \ldots - \frac{\prod_{k=1}^{n-2}(1-w_k)}{2\times 1}, \end{aligned} \tag{B.5}$$

and for any $i \leq n-2$,
$$\begin{aligned} \frac{\partial^2 Q_i}{\partial w_n \partial w_{n-1}} =& \frac{2\prod_{j\neq n,n-1,i} w_j}{n(n-1)(n-2)} + \frac{2\sum_{k\neq n,n-1,i}(1-w_k)\prod_{j\neq n,n-1,i,k} w_j}{(n-1)(n-2)(n-3)} \\ &+ \ldots + \frac{2\prod_{k\neq n,n-1,i}(1-w_k)}{3\times 2\times 1} > 0. \end{aligned} \tag{B.6}$$

Equilibrium total output is $Y_n(w_1, \ldots, w_n) = \sum_{i=1}^n w_i z_i' = \frac{\sum_{i=1}^n w_i z_i}{1+\sum_{i=1}^n z_i}$. By (B.2) and (B.4), $\forall i \neq j \in \{1, \ldots, n\}$, the denominator decreases in w_n. It then suffices to show the numerator, $F_n(w_1, \ldots, w_n) := \sum_{i=1}^n w_i z_i$, increases in w_n. Notice that
$$\frac{\partial F_n}{\partial w_n} = z_n + \sum_{i=1}^{n-1} w_i \frac{\partial z_i}{\partial Q_i}\frac{\partial Q_i}{\partial w_n}$$

and

$$
\begin{aligned}
&\frac{\partial^2 F_n}{\partial w_n \partial w_{n-1}} \\
=&\frac{\partial z_n}{\partial Q_n}\frac{\partial Q_n}{\partial w_{n-1}} + \frac{\partial z_{n-1}}{\partial Q_{n-1}}\frac{\partial Q_{n-1}}{\partial w_n} + \sum_{i=1}^{n-2} w_i \frac{\partial z_i}{\partial Q_i}\frac{\partial^2 Q_i}{\partial w_n \partial w_{n-1}} + \sum_{i=1}^{n-2} w_i \frac{\partial^2 z_i}{\partial Q_i^2}\frac{\partial Q_i}{\partial w_n}\frac{\partial Q_i}{\partial w_{n-1}} \\
\leq&\frac{\partial z_n}{\partial Q_n}\frac{\partial Q_n}{\partial w_{n-1}} + \frac{\partial z_{n-1}}{\partial Q_{n-1}}\frac{\partial Q_{n-1}}{\partial w_n} + \sum_{i=1}^{n-2} w_i \frac{\partial z_i}{\partial Q_i}\frac{\partial^2 Q_i}{\partial w_n \partial w_{n-1}} \\
\leq&\left(\frac{\partial z_n}{\partial Q_n} + \frac{\partial z_{n-1}}{\partial Q_{n-1}}\right)\left(\frac{\partial Q_n}{\partial w_{n-1}} + \frac{1}{2}\sum_{i=1}^{n-2} w_i \frac{\partial^2 Q_i}{\partial w_n \partial w_{n-1}}\right)
\end{aligned}
$$

The first inequality is obtained from dropping the last term, which is negative according to equation (B.3) and (B.4). To see why the second inequality holds, notice that $\frac{\partial Q_n}{\partial w_{n-1}} = \frac{\partial Q_{n-1}}{\partial w_n}$, and that $\frac{1}{2}[\frac{\partial z_n}{\partial Q_n} + \frac{\partial z_{n-1}}{\partial Q_{n-1}}] \geq \frac{\partial z_i}{\partial Q_i}$ for all $i \leq n-2$ since $Q_1 \geq Q_2 \ldots \geq Q_n$ and $\frac{\partial z_i}{\partial Q_i}$ is decreasing in Q_i. The final step is to show $\frac{\partial Q_n}{\partial w_{n-1}} + \frac{1}{2}\sum_{i=1}^{n-2} w_i \frac{\partial^2 Q_i}{\partial w_n \partial w_{n-1}} \leq 0$. Plug in equation (B.5) and (B.6), it can be verified that the first $(n-2)$ terms on the right-hand side of equation (B.5) cancel out with each corresponding term of $\frac{1}{2}\sum_{i=1}^{n-2} w_i \frac{\partial^2 Q_i}{\partial w_n \partial w_{n-1}}$. For example, the second term of $\frac{1}{2}\sum_{i=1}^{n-2} w_i \frac{\partial^2 Q_i}{\partial w_n \partial w_{n-1}}$ is $\sum_{i=1}^{n-2} \frac{w_i \sum_{k\neq n,n-1,i}(1-w_k)\prod_{j\neq n,n-1,k,i} w_j}{(n-1)(n-2)(n-3)}$. The nuemrator is

$$
\begin{aligned}
&\sum_{i=1}^{n-2} w_i \sum_{k\neq n,n-1,i}(1-w_k)\prod_{j\neq n,n-1,k,i} w_j \\
=&\sum_{i\neq n,n-1}[\sum_{k\neq n,n-1,i}(1-w_k)\prod_{j\neq n,n-1,k} w_j] \\
=&\sum_{i\neq n,n-1}[\sum_{k\neq n,n-1}(1-w_k)\prod_{j\neq n,n-1,k} w_j - (1-w_i)\prod_{j\neq n,n-1,i} w_j] \\
=&[(n-2)\sum_{k\neq n,n-1}(1-w_k)\prod_{j\neq n,n-1,k} w_j - \sum_{i\neq n,n-1}(1-w_i)\prod_{j\neq n,n-1,i} w_j] \\
=&(n-3)\sum_{k\neq n,n-1}(1-w_k)\prod_{j\neq n,n-1,k} w_j.
\end{aligned}
$$

Therefore, the fraction can be simplified as $\frac{\sum_{k\neq n,n-1}(1-w_k)\prod_{j\neq n,n-1,k} w_j}{(n-1)(n-2)}$, which is exactly the additive inverse of the second term in $\frac{\partial Q_n}{\partial w_{n-1}}$. The logic is the same with the rest $(n-3)$ terms. So the sum reduces to only one term: the last term of $\frac{\partial Q_n}{\partial w_{n-1}}$, which is negative. In other words, in step 1, an increase in w_n leads to a smaller increase or a bigger decrease on total output F_n when w_{n-1} is higher. Therefore,

$$\begin{aligned}&\frac{\partial F_n(w_1,\ldots,w_n)}{\partial w_n}\\ \geq&\frac{\partial F_n(w_1,\ldots,w_{n-2},w_{n-2},w_n)}{\partial w_n}\\ =&z_n+2w_{n-2}\frac{\partial z_{n-2}}{\partial Q_{n-2}}\frac{\partial Q_{n-2}}{\partial w_n}+\sum_{i=1}^{n-3}w_i\frac{\partial z_i}{\partial Q_i}\frac{\partial Q_i}{\partial w_n}\end{aligned}$$

Repeat the above process to obtain similar results:

$$\frac{\partial}{\partial w_{n-2}}\frac{\partial F_n(w_1,\ldots,w_{n-2},w_{n-2},w_n)}{\partial w_n}\leq 0,$$

$$\vdots$$

and

$$\frac{\partial}{\partial w_1}\frac{\partial F_n(w_1,\ldots,w_1,w_n)}{\partial w_n}\leq 0.$$

Therefore,

$$\begin{aligned}\frac{\partial F_n}{\partial w_1}\geq&\frac{\partial F_n(w_1,\ldots,w_{n-2},w_{n-2},w_n)}{\partial w_n}\\ \geq\ldots\geq&\frac{\partial F_n(1,\ldots,1,w_n)}{\partial w_n}=(\frac{\delta P}{n(1-\delta(1-P/n))})^2V_L\geq 0\end{aligned}$$

For other $m<n$, such as $w_1\geq\ldots\geq w_{n-2}>w_{n-1}=w_n$ $(m=2)$, the logic is the same.

The last case to prove is when $m=n$, i.e. when $w_1=\ldots=w_n=w$, $z_1=\ldots=z_n=z$, and $Q_1=\ldots=Q_n=Q$. The only step is to show that $F_n=nwz$ is increasing in w. To see why this is true, observe that

$$Q=\sum_{i=0}^{n-1}\frac{1}{n-i}\binom{n-1}{i}(1-w)^iw^{n-1-i}, \tag{B.7}$$

and $\frac{\partial Q}{\partial w}<0$. Equation (B.1) and equation (B.2) together imply that

$$0<\frac{\partial z_i}{\partial Q_i}=\frac{(1-\delta)}{1-\delta(1-PQ_i)}\frac{z}{Q}<\frac{z}{Q}.$$

Therefore,

$$\frac{\partial nwz}{\partial w}=n(z+w\frac{\partial z}{\partial Q}\frac{\partial Q}{\partial w})>n\frac{z}{Q}\left(Q+w\frac{\partial Q}{\partial w}\right). \tag{B.8}$$

Hence it suffices to show that $Q + w\frac{\partial Q}{\partial w} > 0$. From equation (B.7) it follows that

$$\begin{aligned}\frac{\partial Q}{\partial w} =& \sum_{i=0}^{n-1} \frac{1}{n-i}\binom{n-1}{i}[-i(1-w)^{i-1}w^{n-1-i} + (n-1-i)(1-w)^i w^{n-2-i}] \\ =& \sum_{i=0}^{n-1} \frac{1}{n-i}\binom{n-1}{i}(1-w)^{i-1}w^{n-2-i}[-iw + (n-1-i)(1-w)].\end{aligned} \tag{B.9}$$

Recall the binomial formula, $(x+y)^{n-1} = \sum_{i=0}^{n-1}\binom{n-1}{i}x^i y^{n-1-i}$. Rearranging the terms of Q and $w\frac{\partial Q}{\partial w}$ based on their power ranks leads to two expressions that are (very close to) the binomial $((1-w)+w)^{n-1}$:

$$\begin{aligned}& Q + w\frac{\partial Q}{\partial w} \\ =& \sum_{i=0}^{n-1} \frac{1}{n-i}\binom{n-1}{i}(1-w)^i w^{n-1-i} \\ &+ \sum_{i=0}^{n-1} \frac{1}{n-i}\binom{n-1}{i}(1-w)^{i-1}w^{n-1-i}[-iw + (n-1-i)(1-w)] \\ =& \sum_{i=0}^{n-1} \frac{1}{n-i}\binom{n-1}{i}(1-w)^{i-1}w^{n-1-i}[(1-w) - iw + (n-1-i)(1-w)] \\ =& \sum_{i=0}^{n-1} \frac{1}{n-i}\binom{n-1}{i}(1-w)^{i-1}w^{n-1-i}[-iw + (n-i)(1-w)] \\ =& -\sum_{i=0}^{n-1} \frac{i}{(n-i)}\frac{(n-1)!}{i!(n-1-i)!}(1-w)^{i-1}w^{n-i} + \sum_{i=0}^{n-1}\binom{n-1}{i}(1-w)^i w^{n-1-i} \\ =& -\sum_{i=1}^{n-1} \frac{(n-1)!}{(i-1)!(n-i)!}(1-w)^{i-1}w^{n-i} + (1-w+w)^{n-1} \\ =& -\sum_{i=1}^{n-1}\binom{n-1}{i-1}(1-w)^{i-1}w^{n-i} + 1 \\ =& -\sum_{j=0}^{n-2}\binom{n-1}{j}(1-w)^j w^{n-1-j} + 1 \\ =& -\left[(1-w+w)^{n-1} - (1-w)^{n-1}\right] + 1 \\ =& (1-w)^{n-1} > 0.\end{aligned}$$

From (B.8) it then follows that $F_n = nwz$ is increasing in w, which finishes the proof. $\square$

Proof of Proposition 1. Given any equilibrium strategy profile $w = (w_1, \ldots, w_n) \in (0, 1]^n$ and its corresponding proper horizontal structure, assuming without loss of generality that $w_1 \geq w_2 \geq \ldots w_n$, go through the following process:

Step 1. Substitute w with $(w_1, \ldots, w_{n-1}, w_{n-1})$, the strategy profile obtained by increasing the smallest w_i to the second smallest one.

Step 2. Substitute the original structure with the proper horizontal structure corresponding to the new strategy profile.

Step 3. Stop the process if $w = (w_1, \ldots, w_1)$; Otherwise, repeat Step 1 and 2.

Step 4. Increase w_1 to 1.

The process stops within n rounds. Moreover, the total output increases after every round. Therefore, proposing the all-working strategy profile $w_n^1 := (1, \ldots, 1)$ is optimal for the central government. □

Appendix C: Proof of Proposition 2

Proof. I follow the notations used in proof of Lemma 4. Notice that if $n = 0$, $\Pi = -\alpha$. Now consider the case when $n > 0$. Since optimal horizontal structures are symmetric, for simplicity, let $z(n) = z_1 = \ldots = z_n = \frac{\delta P \frac{1}{n} V_L}{1-\delta(1-P\frac{1}{n})}$, where $V_L = \frac{1}{1-\delta(1-P)}$, be the ratio of amount of resources assigned to each Bureaucrat to the Leader's resources such that all Bureaucrats work with probability 1 at equilibrium. Denote by $\Pi(n) = \frac{nz(n)}{1+nz(n)} - \alpha(n+1)$ the government's payoff. Observe that $z(0) = V_L$ and $\Pi(0) = -\alpha$. So the expression $\Pi(n)$ also incorporates the case $n = 0$.

Given α, δ, P, the simplified central government's optimization problem is

$$\max_{n \geq 0} \Pi = \frac{nz(n)}{1 + nz(n)} - \alpha(n + 1). \tag{C.1}$$

First order condition:

$$\frac{\partial \Pi}{\partial n} = \frac{z(n) + nz'(n)}{(1 + nz(n))^2} - \alpha = 0. \tag{C.2}$$

Solving (C.2) leads to

$$n^*(\alpha, \delta, P) = \begin{cases} \frac{\delta P \sqrt{1-\delta(1-P)}\left(\frac{1}{\sqrt{\alpha}} - \sqrt{1-\delta(1-P)}\right)}{\delta P + (1-\delta)(1-\delta(1-P))}, & \alpha \leq \frac{1}{1-\delta(1-P)} \\ 0, & \alpha > \frac{1}{1-\delta(1-P)} \end{cases}.$$

It is easy to verify that the second order condition, $\frac{\partial^2 \Pi}{\partial n^2} < 0$, is satisfied.

Similarly, let $z(P) = z_1 = \ldots = z_n = \frac{\delta P \frac{1}{n} V_L}{1-\delta(1-P\frac{1}{n})}$ be the ratio of amount of authority assigned to each Bureaucrat to the Leader's authority such that all Bureaucrats work with probability 1 at equilibrium. Given α, δ, n, the central government's optimization problem is $\max_P \frac{nz(P)}{1+nz(P)} - \alpha(n+1)$. First order condition implies

$$\frac{\partial}{\partial P}\left(\frac{nz(P)}{1+nz(P)} - \alpha(n+1\right) = \frac{nz'(P)}{(1+nz(P))^2} = 0. \tag{C.3}$$

Solving (C.3) gives

$$P^*(\delta, n) = \begin{cases} \frac{1-\delta}{\delta}\sqrt{n}, & 0 < n \leq \tilde{n} \\ 1, & n > \tilde{n} \end{cases},$$

where $\tilde{n} = \left(\frac{\delta}{1-\delta}\right)^2$. Again, it is easy to check that the second order condition, $\frac{\partial^2 \Pi}{\partial P^2} < 0$ is satisfied.

If $n = 0$ then choice of P is irrelevant, and total output is zero. I will focus on the more interesting case where $n > 0$. Solve for $(\hat{n}, \hat{P})$ such that

$$\begin{cases} \hat{n} = n^*(\alpha, \delta, \hat{P}) > 0 \\ \hat{P} = P^*(\delta, \hat{n}) \end{cases}. \tag{C.4}$$

It is easy to verify that if and only if $\alpha \leq 1$ and $\alpha(1-\delta+\delta^2)^2 \leq (1-\delta)^4$, there exists a unique solution ($\hat{n} = \delta(\frac{1}{\sqrt{\alpha}} - 1), \hat{P} = 1$). When $\alpha(1-\delta+\delta^2)^2 > (1-\delta)^4$, it must be that $P^*(\delta, n) < 1$. Hence (C.4) implies

$$\alpha[(1-\delta)(\sqrt{\hat{n}}+1)^2 + n]^2 = (1-\delta)(1+\sqrt{\hat{n}}),$$

and

$$\alpha(1-\delta)(\sqrt{\hat{n}}+1) \leq 1.$$

Substitute $\sqrt{\hat{n}}$ with x. Then x satisfies

$$\alpha[(1-\delta)(x+1)^2+x^2]^2=(1-\delta)(1+x),$$

which is a quartic equation. Define

$$g(x):=\alpha[(1-\delta)(x+1)^2+x^2]^2-(1-\delta)(1+x).$$

Notice that $g(0)=\alpha(1-\delta)^2-(1-\delta)<0$ and $g(\frac{\delta}{1-\delta})=\frac{\alpha(1-\delta+\delta^2)^2}{(1-\delta)^4}-1>0$ together guarantee the existence of a solution to $g(x)=0$ in the interval $(0,\frac{\delta}{1-\delta})$. For uniqueness, observe that $g''(x)=4\alpha[(1-\delta)(3x+2)^2+3x^2]>0$. The sign of $g'(0)=4\alpha(1-\delta)^2-(1-\delta)$ is undetermined, but $g'(\frac{\delta}{1-\delta})>0$. Then in either case (either $g'(0)\geq 0$ or $g'(0)<0$) the solution to $g(x)=0$ in the interval $(0,\frac{\delta}{1-\delta})$ is unique. Denote this unique positive solution by $\hat{x}$. Then $\hat{n}=\hat{x}^2$ and $\hat{P}=\frac{1-\delta}{\delta}\hat{x}$. Also notice that $g'(\hat{x})\in(0,\infty)$ in both cases. Then by Envelop Theorem,

$$\frac{d\hat{x}}{d\alpha}=-\frac{[(1-\delta)(\hat{x}+1)^2+\hat{x}^2]^2}{g'(\hat{x})}<0, \tag{C.5}$$

and

$$\frac{d\hat{x}}{d\delta}=\frac{2\alpha[(1-\delta)(\hat{x}+1)^2+\hat{x}^2](\hat{x}+1)^2-(\hat{x}+1)}{g'(\hat{x})}.$$

By (C.5), $\frac{d\hat{n}}{d\alpha},\frac{d\hat{P}}{d\alpha}<0$. To check the sign of $\frac{d\hat{x}}{d\delta}$, it suffices to check the sign of the numerator, $2\alpha[(1-\delta)(\hat{x}+1)^2+\hat{x}^2](\hat{x}+1)^2-(\hat{x}+1)$. Notice that $g(\hat{x})=0$ implies $\alpha[(1-\delta)(\hat{x}+1)^2+\hat{x}^2]^2=(1-\delta)(1+\hat{x})$. Therefore, the numerator can be rewritten as

$$\begin{aligned}
&2\alpha[(1-\delta)(\hat{x}+1)^2+\hat{x}^2](\hat{x}+1)^2-(\hat{x}+1)\\
&=2\alpha[(1-\delta)(\hat{x}+1)^4+\hat{x}^2(\hat{x}+1)^2]-\frac{\alpha[(1-\delta)(\hat{x}+1)^2+\hat{x}^2]^2}{1-\delta}\\
&=2\alpha[(1-\delta)(\hat{x}+1)^4+\hat{x}^2(\hat{x}+1)^2]-\alpha[(1-\delta)(\hat{x}+1)^4+2(\hat{x}+1)^2\hat{x}^2+\frac{\hat{x}^4}{1-\delta}]\\
&=\frac{\alpha}{1-\delta}[(1-\delta)^2(\hat{x}+1)^2-\hat{x}^4]\\
&=\frac{\alpha}{1-\delta}[(1-\delta)(\hat{x}+1)^2+\hat{x}^2][\sqrt{1-\delta}(\hat{x}+1)+\hat{x}][\sqrt{1-\delta}(\hat{x}+1)-\hat{x}].
\end{aligned}$$

So it suffices to check the sign of $(\sqrt{1-\delta}(\hat{x}+1)-\hat{x})$. Check two special cases first: At $\delta=0$, the sign is clearly positive. So $\frac{d\hat{x}}{d\delta}>0$ when $\delta=0$. Similarly, $\frac{d\hat{x}}{d\delta}<0$

when $\delta = 1$. Moreover, the total number of points on the interval $(0, 1)$ where $\frac{d\hat{x}}{d\delta}$, a continuous function of δ, goes cross the horizontal axis must be odd. I prove that this number must be one by contradiction. Suppose, to the contrary, that there are more than one critical points where $\frac{d\hat{x}}{d\delta}$ intersects the horizontal axis. Then there exists at least one point, denoted by δ_0, where $\frac{d\hat{x}}{d\delta}|_{\delta_0} = 0$, and there exists an $\epsilon > 0$ such that $\frac{d\hat{x}}{d\delta} > 0$ for any $\delta \in (\delta_0, \delta_1)$, where $\delta_1 := \delta_0 + \epsilon$. Denote the solution to $g(x) = 0$ corresponding to δ_0 and δ_1 by $\hat{x}_0$ and $\hat{x}_1$, respectively. Notice that $\sqrt{1-\delta}(\hat{x}+1) - \hat{x} > 0$ is equivalent to $x < \frac{\sqrt{1-\delta}}{1-\sqrt{1-\delta}}$. The right-hand side is a decreasing function of δ. Thus $\hat{x}_0 = \frac{\sqrt{1-\delta_0}}{1-\sqrt{1-\delta_0}} > \frac{\sqrt{1-\delta_1}}{1-\sqrt{1-\delta_1}} > \hat{x}_1$. But $\hat{x}_0 < \hat{x}_1$ because $\frac{d\hat{x}}{d\delta} > 0$ for any $\delta \in (\delta_0, \delta_1)$. Contradiction! So there exists a unique $\bar{\delta} \in (0, 1)$ such that $\frac{d\hat{x}}{d\delta} > 0$ if and only if $\delta < \bar{\delta}$. It is then straightforward that $\frac{d\hat{n}}{d\delta} > 0$ if and only if $\delta < \bar{\delta}$.

Finally I show that $\frac{d\hat{P}}{d\delta} < 0$. Notice that $\hat{P} = \frac{1-\delta}{\delta}\hat{x}$. Rewrite $g(x)$ as a function of P and call this new function $\tilde{g}(P)$:

$$\tilde{g}(P) = \alpha[\frac{(1-\delta(1-P))^2}{1-\delta} + (\frac{\delta}{1-\delta})^2 P^2]^2 - (1-\delta(1-P)). \tag{C.6}$$

By exactly the same argument as above, there exists a unique solution to $\tilde{g}(P) = 0$ on interval $[0, 1]$, denoted as $\hat{P}$. Moreover, $\tilde{g}'(\hat{P}) > 0$. Then Envelop Theorem implies $\frac{d\hat{P}}{d\delta} = -\frac{\partial \tilde{g}}{\partial \delta}/\tilde{g}'(\hat{P})$. Therefore, $\frac{d\hat{P}}{d\delta} < 0$ is equivalent to $\frac{\partial \tilde{g}}{\partial \delta} > 0$. Take derivatives:

$$\begin{aligned}
\frac{\partial \tilde{g}}{\partial \delta} =& 2\alpha[\frac{(1-\delta(1-P))^2}{1-\delta} + (\frac{\delta}{1-\delta})^2 P^2] \times \\
& [\frac{(1-\delta(1-P))(1-(1-P)(2-\delta))}{(1-\delta)^2} + \frac{2\delta P^2}{(1-\delta)^3}] + (1-P) \\
\Rightarrow \frac{\partial^2 \tilde{g}}{\partial \delta^2} =& 2\alpha[\frac{(1-\delta(1-P))(1-(1-P)(2-\delta))}{(1-\delta)^2} + \frac{2\delta P^2}{(1-\delta)^3}]^2 + \\
& 4\alpha[\frac{(1-\delta(1-P))^2}{1-\delta} + (\frac{\delta}{1-\delta})^2 P^2] P^2 > 0
\end{aligned}$$

When $\delta \to 0^+$, $\hat{P} \to 1$, and hence $\lim_{\delta \to 0^+} \frac{\partial \tilde{g}}{\partial \delta}|_{\hat{P}} = 2\alpha(2\hat{P} - 1) + 1 - \hat{P} > 0$. So $\frac{\partial \tilde{g}}{\partial \delta} > 0$ for all $\delta \in (0, 1)$. $\square$

Appendix D: Proof of Proposition 3

Proof. The total output of $h^*_{n,P}$ with equilibrium strategy profile $w^1_n = (1,\dots,1) \in \mathbb{R}^n$ is $Y_{H,n} = Y(h^*_{n,P}, w^1_n, n, P) = 1 - z_{0,h}$, where

$$z_{0,h} = \frac{1}{1 + n \times \frac{1}{1-\delta(1-P)} \times \frac{\delta \frac{P}{n}}{1-\delta(1-\frac{P}{n})}} \tag{D.1}$$

is the amount of authority of the Leader in $h^*_{n,P}$.

The total output of $v^*_{n,P}$ with equilibrium strategy profile $w^1_n = (1,\dots,1) \in \mathbb{R}^n$ is $Y_{V,n} = Y(v^*_{n,P}, w^1_n, n, P) = 1 - z_{0,v}$, where $z_{0,v}$ is the amount of authority of the Leader in $v^*_{n,P}$. In the vertical structure, for each $i = 1,\dots,n$, the value of holding position $i+1$ is proportional to the value of holding position i: $V_{i+1} = \frac{\delta P}{1-\delta(1-P)} V_i$. And value of being a Leader is $V_0 = V_L = \frac{z_{0,v}}{1-\delta(1-P)}$. That budget constraint is binding implies

$$z_{0,v}\left[1+\left(\frac{\delta P}{1-\delta(1-P)}+\left(\frac{\delta P}{1-\delta(1-P)}\right)^2+\dots+\left(\frac{\delta P}{1-\delta(1-P)}\right)^n\right)\frac{1}{1-\delta(1-P)}\right] = 1. \tag{D.2}$$

To show $Y_{V,n} \geq Y_{H,n}$ is equivalent to show $\frac{z_{0,h}}{z_{0,v}} \geq 1$. Plug in (D.1) and (D.2), it is then equivalent to show

$$\begin{aligned}
&\frac{1}{1-\delta(1-P)}\left(\frac{\delta P}{1-\delta(1-P)} + \dots + \left(\frac{\delta P}{1-\delta(1-P)}\right)^n\right) \geq \frac{\delta P}{1-\delta(1-P)}\frac{1}{1-\delta(1-\frac{P}{n})} \\
\Leftrightarrow\ &\frac{1-\left(\frac{\delta P}{1-\delta(1-P)}\right)^n}{1-\delta} \geq \frac{1}{1-\delta(1-\frac{P}{n})} \\
\Leftrightarrow\ &((1-\delta)+\delta P)^n \geq (\delta P)^{n-1}(n(1-\delta)+\delta P) = (\delta P)^n + n(1-\delta)(\delta P)^{n-1}.
\end{aligned}$$

Notice that the two terms on the right-hand side of the last line are two of the $n+1$ terms of the binomial expansion of the left-hand side. □

Appendix E: Proof of Proposition 4

Proof. By the previous lemma, all Bureaucrats must be indifferent between working and shirking. Let $g^*_{(n_1,\dots,n_k),P}$ denote the structure that have an equilibrium where all bureaucrats work with probability one and are indifferent between working and shirking. There are k layers under the Leader in this structure: n_1 offices of rank k, n_2 of rank $k-1$, …, and n_k offices of rank 1, where $n_1 + \dots + n_k = n$

and $n_1, \ldots, n_k \geq 1$. Since all bureaucrats are indifferent between working and shirking, positions of the same rank must be allocated with the same amount of resource. Hence this structure must be symmetric. Ignore the budget constraint first and assume the amount of resources allocated to one office in each layer is $z_1, \ldots, z_k$ and that the Leader has one unit of resource. Then indifference indicates that for any $i = 1, \ldots, k$,

$$V_i = z_i + \delta \times 0 = 0 + \delta(P\frac{1}{n_i}V_{i-1} + (1 - P\frac{1}{n_i})V_i),$$

where V_i is the value of holding an office of rank i and $V_0 = V_L = \frac{1}{1-\delta(1-P)}$ is the value of holding the Leader's office. Therefore,

$$V_i = z_i = \frac{\delta P \frac{1}{n_i} V_{i-1}}{1 - \delta + \delta P \frac{1}{n_i}}.$$

The total amount of resources allocated to rank i offices is then

$$n_i z_i = \frac{\delta P V_{i-1}}{1 - \delta + \delta P \frac{1}{n_i}}.$$

So

$$n_1 z_1 = \frac{\delta P V_0}{1 - \delta + \delta P \frac{1}{n_1}},$$

$$n_2 z_2 = \frac{\delta P}{1 - \delta + \delta P \frac{1}{n_2}} \frac{\delta P \frac{1}{n_1} V_0}{1 - \delta + \delta P \frac{1}{n_1}},$$

$$\ldots,$$

and

$$n_k z_k = \frac{\delta P}{1 - \delta + \delta P \frac{1}{n_k}} \frac{\delta P \frac{1}{n_{k-1}}}{1 - \delta + \delta P \frac{1}{n_{k-1}}} \cdots \frac{\delta P \frac{1}{n_1} V_0}{1 - \delta + \delta P \frac{1}{n_1}}.$$

Notice that the budget constraint must bind in order to get the highest possible output. After scaling down the whole structure to satisfy the budget constraint the total output is $\frac{\sum_{i=1}^k n_i z_i}{1+\sum_{i=1}^k n_i z_i} = 1 - \frac{1}{1+\sum_{i=1}^k n_i z_i}$. So $\sum_{i=1}^k n_i z_i$ is bigger if and only if the total output is higher. For any given $g^*_{(n_1,\ldots,n_k),P}$,

- If $n_k \geq 2$, consider $g^*_{(n'_1,\ldots,n'_{k+1}),P}$ with $n'_1 = n_1, \ldots, n'_{k-1} = n_{k-1}, n'_k = 1, n'_{k+1} = n_k - 1$. This new structure is obtained by removing $(n_k - 1)$ offices from the bottom rank, create a new rank that is lower and put these offices in the new bottom rank. Since the two structures are the same in the top $k-1$ layers, the ratio of resources are also the same, as well as the values of holding an office in the first $(k-1)$ layers. Then

$$
\begin{aligned}
&\sum_{i=1}^{k+1} n_i' z_i' - \sum_{i=1}^{k} n_i z_i \\
=& n_k' z_k' + n_{k+1}' z_{k+1}' - n_k z_k \\
=& \frac{\delta P V_{k-1}}{1-\delta+\delta P} + \frac{\delta P V_k'}{1-\delta+\delta P \frac{1}{n_k-1}} - \frac{\delta P V_{k-1}}{1-\delta+\delta P \frac{1}{n_k}} \\
=& \frac{\delta P V_{k-1}}{1-\delta+\delta P} + \frac{\delta P}{1-\delta+\delta P \frac{1}{n_k-1}} \frac{\delta P V_{k-1}}{1-\delta+\delta P} - \frac{\delta P V_{k-1}}{1-\delta+\delta P \frac{1}{n_k}} \\
=& \delta P V_{k-1} [\frac{1}{1-\delta+\delta P} + \frac{\delta P}{1-\delta+\delta P \frac{1}{n_k-1}} \frac{1}{1-\delta+\delta P} - \frac{1}{1-\delta+\delta P \frac{1}{n_k}}] \\
=& \delta P V_{k-1} [\frac{(1-\delta)\delta P \frac{1}{n_k}}{(1-\delta+\delta P)(1-\delta+\delta P \frac{1}{n_k-1})(1-\delta+\delta P \frac{1}{n_k})}] \\
>& 0.
\end{aligned}
$$

So $g^*_{(n_1',\ldots,n_{k+1}'),P}$ is an improvement upon $g^*_{(n_1,\ldots,n_k),P}$.

- The logic is the same for any $n_i \geq 2$. Instead of moving one office down from the k-th layer we remove one from the layer closest to the bottom that has at least two offices and add another single-office layer at the bottom.

□

References

Besley, T., & Ghatak, M. (2005). Incentives, risk and accountability in organizations. Cambridge University Press.

Boffa, F., Piolatto, A., & Ponzetto, G. A. M. (2016). Political centralization and government accountability *. The Quarterly Journal of Economics, 131(1), 381.

Che, J., Chung, K.-S., & Lu, Y. K. (2017). Decentralization and political career concerns. Journal of Public Economics, 145, 201–210.

Coate, S., & Knight, B. (2007). Socially optimal districting: A theoretical and empirical exploration. The Quarterly Journal of Economics, 122(4), 1409–1471.

Li, P., Lu, Y., & Wang, J. (2016). Does flattening government improve economic performance? evidence from china. Journal of Development Economics, 123, 18–37.

Ma, L. J. (2005). Urban administrative restructuring, changing scale relations and local economic development in china. Political Geography, 24(4), 477–497.

Mookherjee, D., et al. (2013). Incentives in hierarchies. The Handbook of Organizational Economics, 764–98.

Myerson, R. B. (2006). Federalism and incentives for success of democracy. Quarterly Journal of Political Science, 1(1), 3–24.

Patacconi, A. (2009). Coordination and delay in hierarchies. The RAND Journal of Economics, 40(1), 190–208.

Qian, Y. (1994). Incentives and loss of control in an optimal hierarchy. The Review of Economic Studies, 61(3), 527–544.

Rajan, R. G., & Zingales, L. (2001). The firm as a dedicated hierarchy: A theory of the origins and growth of firms. The Quarterly Journal of Economics, 116(3), 805–851.

Williamson, O. E. (1975). Markets and hierarchies: Analysis and antitrust implications: A study in the economics of internal organization.

Zhou, L.-a. (2007). Governing china's local officials: An analysis of promotion tournament model [j]. Economic Research Journal, 7, 36–50.