Essays on Political Economy

by

Yunyun Lv

B.S.Sc. in Economics, The Chinese University of Hong Kong, 2015

M.Phil. in Economics, The Chinese University of Hong Kong, 2017

Submitted to the Graduate Faculty of

the Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Yunyun Lv

It was defended on

August 2nd, 2023

and approved by

Richard van Weelden, Professor of Economics

Luca Rigotti, Professor of Economics

Daniel Berkowitz, Professor of Economics

Arie Beresteanu, Associate Professor of Economics

Iza Ding, Assistant Professor of Political Science

Essays on Political Economy

Yunyun Lv, PhD

University of Pittsburgh,

This dissertation is a collection of essays focused on understanding: (i) the strategic interactions related to censorship; and (ii) the force behind changes across time in public policies, including emigration policies and governmental structures, in autocracies.

Essay 1 theoretically and empirically analyzes censorship on social media platforms in China. If the government sets a higher tax rate for the next tax year, it will benefit more from the profit made by the company. Hence the government has weaker incentives to punish the company in case of disobedience. Therefore, the company has weaker incentives to comply with the government's order to censor sensitive content.

The effort of censorship may be all in vain due to the paradox of censorship: The more the censor suppresses whatever the censor dislikes, the more attention that disliked subject receives. Essay 2 explains how the censor strategically times censorship when considering that it may backfire. In equilibrium, if the censor stops the discussion about a piece of news and the learning process sooner, the receiver believes he is more likely a bad type.

Essay 3 uses an infinite-horizon principal-agent model to explore the interaction between the hierarchical structures of governments and the career concerns of local officials in China. 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.

Essay 4 explains why autocrats and dictators draft the emigration policy so that emigration is sometimes easy for citizens, sometimes almost impossible. There are clear gains in opening the floodgates: If more citizens leave the country, fewer citizens will participate in revolutionary activities, making it more likely for the government to remain in power. Under certain conditions, this improvement in stability dominates the loss in GDP caused by emigration.

Table of Contents

List of Tables

List of Figures

Preface

First and foremost, I am extremely grateful to my supervisors, Prof. Van Weelden, Prof. Rigotti, Prof. Beresteanu, Prof. Ding, and Prof. Berkowitz, for their invaluable advice, continuous support, and patience during my Ph.D. study. Their immense knowledge and plentiful experience have encouraged me throughout my academic research and daily life. I also want to thank my fellow Ph.D. students and friends for their support as we struggle and survive together toward this milestone in our lives. I would like to thank all the members of the Department of Economics, especially Brian Deutsch. Their kind help and support have made my study and life at the University of Pittsburgh wonderful. Finally, I would like to express my gratitude to my parents and my cats. Without their tremendous understanding, encouragement, and emotional support over the past few years, it would be impossible for me to complete my study.

1.0 When Big Brother Meets Big Profit

1.1 Introduction

The tragic death of an overworked content moderator at Bilibili, one of China's largest video streaming platforms, during the Lunar New Year holiday in 2022 startled the public. People were angry for many reasons. Bilibili's first reaction was to deny that the employee had been overworked by deleting his employee profile and attendance record in its internal system. After the news was leaked, the discussion of his death immediately became one of those "sensitive" topics that are prohibited both inside the company and on Bilibili.com. The provoked public indignation by the death of the overworked censor only made Bilibili censors even busier, censoring one more topic. This unexpected incident also made the public aware of the pervasiveness of content moderation on Social Network Service (SNS) platforms in China. Bilibili finally released a public announcement, saying that 1,000 more content moderators would be hired in 2022 to ease the workload of the content moderation team. There were over 2,400 employees in this team, which was roughly 30% of the total employees at Bilibili. Still, Bilibili is not the largest job provider for content moderators in China: By the end of 2020, ByteDance has more than 20,000 content moderators out of its 100,000 employees.

Why do SNS companies hire so many censors? Besides scrubbing content associated with violence, pornography, and fraud, among other things, they also need to comply with orders of the Chinese government (the "Big Brother") to censor content that is deemed politically sensitive or offensive. The platforms have little to negotiate with the government regarding the scope of censorship. However, they have a degree of flexibility regarding how fast and well they respond to such orders (Knockel et al., 2015). Empirical findings show that different platforms have different response delays after the same event (Liu, 2020). Disobedience by social media companies is quite common: 16% of government directives are disobeyed by Weibo because of concerns about censoring more strictly than its competitors (Miller, 2018). This paper explains the choice of compliance level of an SNS platform using a dynamic model of two strategic players: the government and the SNS platform. The government sets a tax rate, the platform observes the tax rate and then chooses the level of censorship, and finally, the government decides whether to punish the platform by shutting it down. The government cares about its tax revenue but wants the platform to remove content it dislikes. The platform avoids censoring too much because it is costly but does not want to be punished by the government.

In the model, choosing a higher level of compliance is costly for the platform. Censorship by algorithm requires regular maintenance. The set of keywords the government wants to block is changing daily. Hiring machine learning experts who can write efficient programs and a large group of censors who manually check suspicious content is expensive. On top of that, SNS companies do not want to censor too much because most censorship is done via friction, in other words, by increasing users' costs to communicate (Roberts, 2018). Internet users are highly impatient: 57 percent of users will abandon a site if it takes more than three seconds for the webpage to load, of which 80 percent will not return (Nagy, 2013). Although censorship via friction is not done by simply increasing the loading time of a website, it often increases the time needed for users to successfully post content and decipher encrypted content posted by other users. Standard practices of censorship on SNS platforms include, but are not limited to, suspending an account temporarily, deactivating an account permanently, hiding a post in the timeline with or without notifying the poster, disabling the reply/forward function for a post, adding a misinformation warning tag to a post. The increased communication cost drives active users away, possibly toward a competitor's platform. Hence it leads to a loss in profit for the platform.

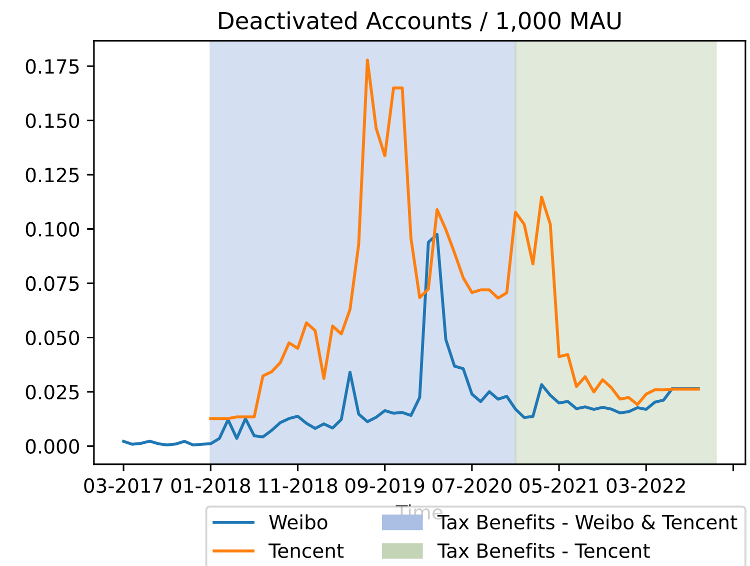
In the model, the government can shut down the platform as punishment for censoring too little. This simplifying assumption captures the idea that the government punishes the platform by hurting its profitability: In reality, the government can punish platforms for disobedience in various ways, for example, by giving them a warning, collecting fines, or shutting them down, either temporarily or permanently. Toutiao ("Today's Headlines"), one of the most popular news platforms in China and a core product of ByteDance, was temporarily shut down for 24 hours in Dec. 2017 by the Cyberspace Administration of China (CAC). Four months later, it permanently shut down its popular joke app Neihan

Duanzi at the request of the State Administration of Radio and Television (SART). At the same time, its CEO Yiming Zhang openly apologized to regulators and the public for allowing the app to "lose its way" and announced an increase of content moderators from 6,000 to 10,000. In 2021 Weibo paid 44 fines totaling $2.25mn (in USD), and Douban incurred 22 fines totaling $1.42mn. These fines themselves are not colossal relative to these companies' revenues. However, these public punishments can trigger a more significant loss. On Dec. 14th, 2021, Weibo stock hit multi-year lows after CAC fined Weibo three million yuan ($470,000), the maximum amount of such fines, for violating cybersecurity laws and publishing illegal information repeatedly. Within one day, Weibo's stock price fell by 10% in Hong Kong and 4.8% in the U.S. stock market. The massive drop in market value reflects the reduced confidence level of the public about the company's future profitability.

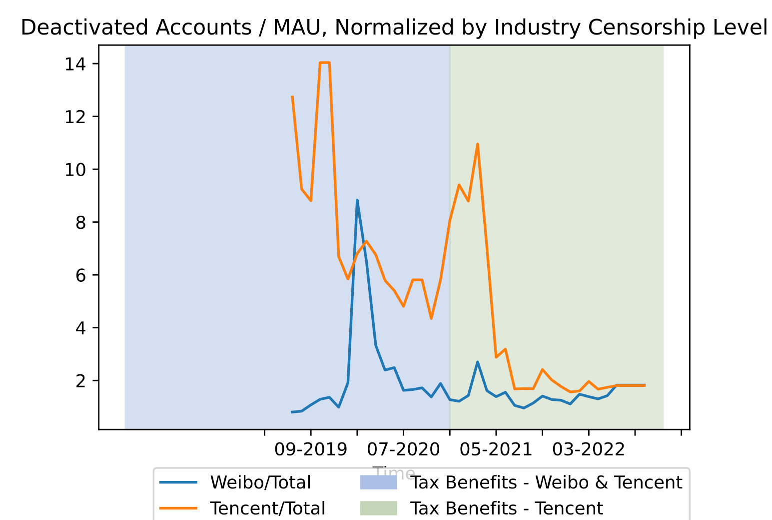
These punishments hurt the "Big Profit" made by SNS companies and the government's tax revenue. Take 2019 as an example: China's total tax revenue was $2,300 billion. Weibo paid around $100 million in tax, while Tencent paid $2.1 billion. The top 500 private companies jointly contributed $200 billion in tax. The general Corporate Income Tax rate in Mainland China is fixed at 25%. While the *de facto* tax rate, which is the ratio of tax expense divided by pretax income, varies across companies and time because some preferential tax treatments apply to some companies. For example, companies that the government approves as "High and New Technology Enterprises" or "Key Software Enterprises" are subject to a preferential corporate income tax rate of 15% and 10%, respectively (PRC State Taxation Administration, 2017, PRC State Taxation Administration, 2022). Both Weibo and Tencent enjoyed these preferential tax treatments for the tax year 2017-2019 (recognized in the year 2018-2020). Weibo lost these benefits while Tencent managed to plow on for one more year in 2021. The model starts with an initial node where the government sets a tax rate for the platform.

1.2 Data

Weibo and Tencent report their censorship performance monthly on their official accounts or service centers. They are two of the largest social media service providers in China: As of Q2 2022, Weibo has 582 million monthly active users (MAU), while the two core social media services provided by Tencent, Tencent QQ and WeChat, have 569 and 1299 million MAU, respectively. The data is collected from Weibo Administrator (微博管理员), which has 135 million followers on Weibo, and Tencent Guard (腾讯卫士), which serves more than one billion WeChat users and 800 million QQ users. Figure 1a shows the self-reported number of deactivated accounts due to politically sensitive or offensive content on the two platforms. Since the two platforms are of different sizes, the numbers are normalized by each platform's Monthly Active Users (MAU) to make the comparison meaningful. The shaded areas indicate when preferential tax benefits are available. Figure 1b displays the ratio of the censorship level on the two platforms to the industry level of censorship. The industry level of censorship is measured by the number of cases reported to the CAC Reporting Center (中央网信办举报中心) by all websites each month, normalized by the number of internet users in China. The data exhibits two noticeable trends: (i) The level of censorship by SNS companies is higher when their *de facto* tax rate is lower. (ii) Bigger SNS companies censor more frequently than smaller ones. The intuition is that the tax rate the government determines in advance affects its future incentives to punish SNS companies when the level of censorship is unsatisfying. If the government cancels some tax benefits for an SNS company, then the *de facto* tax rate is higher, implying that the government cares more about the profit made by the company. Hence the government has weaker incentives to punish the company in case of disobedience because it also hurts its revenue. Therefore, the company also has weaker incentives to comply with the government to censor content. The difference in censorship levels on SNS platforms of different sizes can be explained by stronger incentives of the government to stop circulating sensitive information on bigger SNS platforms.



(a) Frequency of Censorship by Weibo and Tencent



(b) Frequency of Censorship by Weibo and Tencent, Compared to Industry Average

Figure 1: Frequency of Censorship on SNS

Note: The blue area represents the years in which both Weibo and Tencent enjoyed the tax benefits. The green area represents the years in which only Tencent enjoyed the tax benefits.

1.3 The Model

There are two players: The government and the platform. In the first period, the government chooses and commits to a tax rate $\tau \in [0, \bar{\tau}]$ that applies to the platform in the next period. The upper bound $\bar{\tau} \in (0, 1]$ represents the highest tax rate that the government can choose. In China, $\bar{\tau} = 25\%$. The tax rate is public information. Non-strategic users of the platform spend a fixed amount of time on it and hence generate a fixed amount of profit for the platform, which is normalized as 1. The fraction of sensitive content that the users spread, denoted by α, is a random variable that follows a uniform distribution on $[0, 1]$. The distribution is public information, but the realized α is unknown to the platform. Only the government observes α in the final period. This captures the changing and private preferences of the government on political topics. In the second period, the platform chooses its compliance level, $\beta \in [0, 1]$, which measures the fraction of sensitive content the platform successfully catches. Censorship is costly: choosing a higher β incurs a higher cost for the platform. In the final period, the government decides whether to shut down the platform after observing α and β. Let $d \in \{0, 1\}$ represent the government's decision. If $d = 1$ (the government shuts down the platform), both the platform and the government get zero. If $d = 0$ (the government does not shut down the platform), the government taxes the platform's revenue at the chosen tax rate τ. Notice that, in reality, the government punishes SNS platforms for disobedience in various ways. No matter what form the punishment takes, there is one thing in common: It harms the revenue of the SNS company and the government's tax revenue. This model captures it by assuming that the government chooses whether to shut down the platform. Figure 2 summarizes the timeline of the game.

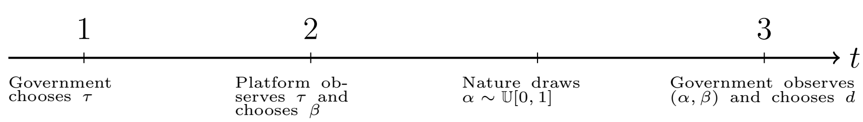


Figure 2: Timeline

If the platform is not shut down, it profits from active users and pays tax and the cost of censorship. The platform's payoff is given by

$$\Pi(\beta) = (1-d)\left(1-\tau-\kappa_1\beta\right),$$

where κ_1 is the coefficient of censorship cost. Notice that the censorship cost includes the hiring cost of content moderators, the reputation cost of censoring, and the loss of active users for censoring more.

Assumption 1.1. *$\kappa_1 \in (0,1)$.*

Censorship is costly, so κ_1 must be positive. It is straightforward that the platform never censors if the cost of censorship is higher than the total revenue. This paper only focuses on the case where $\kappa_1 < 1$.

The government's payoff when $d = 0$ is given by its tax revenue minus dis-utility from the existence of sensitive content on the platform:

$$V(\tau, d) = (1-d)\left[\tau - \kappa_2\alpha(1-\beta)\right],$$

where κ_2 represents how much the government hates sensitive content being spread on the platform. Since the platform size is normalized, κ_2 measures the government's dis-utility from sensitive content per user. For platforms with more active users, κ_2 is higher because every post has a bigger pool of potential audience.

1.4 Results

1.4.1 Equilibrium

The solution concept is Perfect Bayesian Equilibrium.

Proposition 1.1. *There exists a unique equilibrium. There are four cases:*

1. *If $(1-\kappa_1)\kappa_2 < \bar{\tau}^2$, then $\tau^* = \bar{\tau}, \beta^* = 0$, and $\mathbb{E}(d^*) \geq 0$ at equilibrium. There is no tax benefit or censorship. Furthermore, $\mathbb{E}(d^*) > 0$ if and only if $\bar{\tau} < \kappa_2$. A shutdown is possible if κ_2 is big enough.*

2. *If* $(1-\kappa_1)\kappa_2 > \bar{\tau}^2$, $1-\kappa_1 > \bar{\tau}$ *and* $\kappa_2 < \bar{\tau}$, *then* $\tau^* = \bar{\tau}, \beta^* = 0$, *and* $\mathbb{E}(d^*) = 0$ *at equilibrium. There is no tax benefit or censorship. The shutdown never happens.*
3. *If* $1-\kappa_1 > \bar{\tau}$ *and* $\kappa_2 > \bar{\tau}$, *then* $\tau^* = \bar{\tau}, \beta^* = 1 - \frac{\bar{\tau}}{\kappa_2} > 0$, *and* $\mathbb{E}(d^*) = 0$ *at equilibrium. There is no tax benefit. The platform censors just enough to make the government indifferent between shutting it down or not in the worst case (when* $\alpha = 1$*). The shutdown never happens.*
4. *If* $(1-\kappa_1)\kappa_2 > \bar{\tau}^2$ *and* $1-\kappa_1 < \bar{\tau}$, *then* $\tau^* = 1-\kappa_1, \beta^* = 1 - \frac{1-\kappa_1}{\kappa_2} > 0$, *and* $d^* = 0$ *at equilibrium. Tax benefits are offered. The platform censors just enough to make the government indifferent between shutting it down or not in the worst case (when* $\alpha = 1$*). The shutdown never happens.*

Figure 3 locates the four cases in the $(1-\kappa_1) \times \kappa_2$ plane. Only in Case 4 (if censorship is costly and the government's dis-utility from sensitive content is high) the government offers tax benefits to the platform. The platform censors frequently enough to keep the government indifferent between shutting it down or not in the worst case (when $\alpha = 1$). There is always censorship on the platform, and shutdown never happens. In the other three cases the government offers no tax benefit. The platform only censors in Case 3 where the cost of censorship is low and the government is sensitive. Shutdown happens with positive probability only in Case 1 if the government's dis-utility is high enough.

The complete proof can be found in the Appendix. The sketch of the proof is presented below. In the final period, the government's optimal strategy is to shut down the platform whenever α is too high:

$$d^*(\tau, \alpha, \beta) = \mathbb{1}(\kappa_2 \alpha (1-\beta) > \tau) = \mathbb{1}\left(\alpha > \frac{\tau}{\kappa_2(1-\beta)}\right).$$

Given this shutting-down strategy, the platform's expectation is then

$$1 - \mathbb{E}(d^*) = Prob\left(\alpha < \frac{\tau}{\kappa_2(1-\beta)}\right) = \min\left\{\frac{\tau}{\kappa_2(1-\beta)}, 1\right\}.$$

Notice that if $\tau \geq \kappa_2$, then $d^* = 0$ always and hence $\beta^* = 0$ in equilibrium. The best response function of the platform is then

$$\beta^* = \begin{cases} \tilde{\beta}, & \tau \leq 1-\kappa_1 \text{ and } \tau \leq \kappa_2 \\ 0, & \text{otherwise,} \end{cases}$$

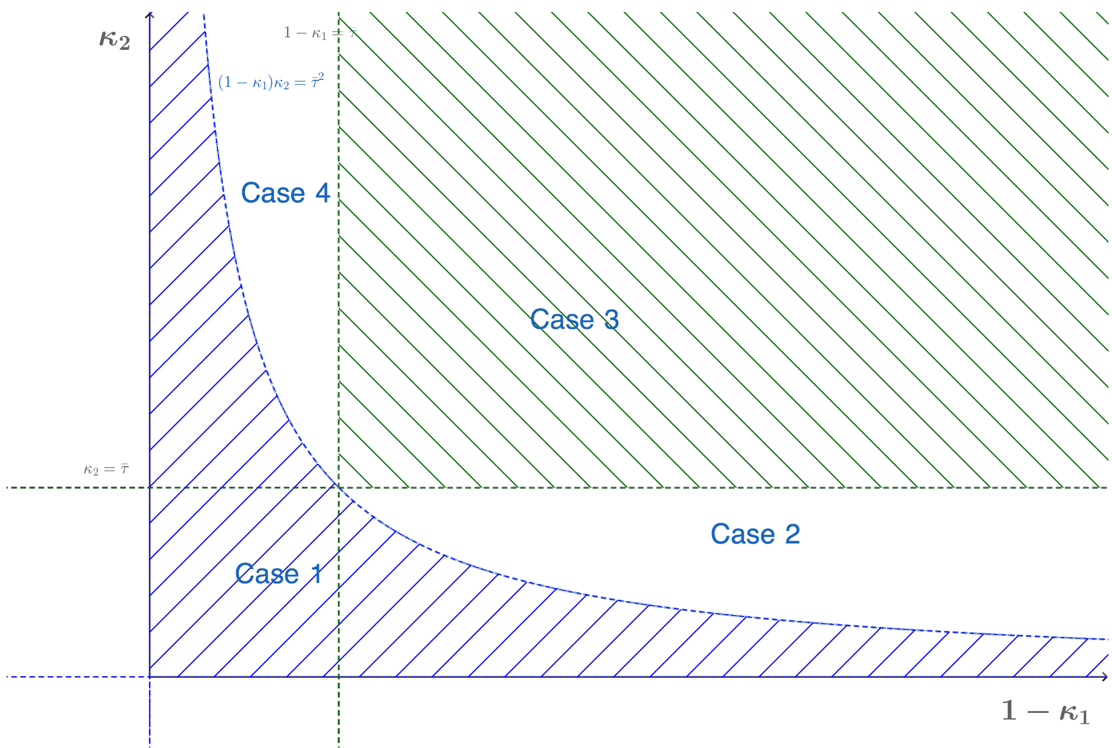


Figure 3: Unique Equilibrium: Four Cases

where

$$\tilde{\beta} = 1 - \frac{\tau}{\kappa_2}.$$

Observe that the platform censors only when the cost of censorship, κ_1, and the tax rate, τ, are low enough. Notice that β^* is decreasing in τ: If the tax rate increases, the platform keeps a smaller share of its revenue while the government gets a bigger share. Hence the government has weaker incentives to shut down the platform, and the platform has weaker incentives to censor. The interior solution $\tilde{\beta}$ makes the government indifferent between shutting down the platform or not in the worst case (when $\alpha = 1$). The platform censors so that the government will not shut them down in the worst case. This explains the first trend observed above.

Then the government's optimal choice of τ can be found by looking at its expected payoff in the first period:

$$\tau^* = \begin{cases} 1 - \kappa_1, & \frac{\bar{\tau}^2}{\kappa_2} < 1 - \kappa_1 < \bar{\tau} \\ \bar{\tau}, & \text{otherwise} \end{cases}$$

From here, it is straightforward to derive all possible cases of equilibrium.

1.4.2 Comparative Statics

Proposition 1.2. *The equilibrium tax rate τ^* strictly decreases in κ_1 if and only if $(1 - \kappa_1)\kappa_2 > \bar{\tau}^2$ and $1 - \kappa_1 < \bar{\tau}$, and remains constant otherwise. The equilibrium censorship level β^* strictly increases in κ_1 if and only if $(1 - \kappa_1)\kappa_2 > \bar{\tau}^2$ and $1 - \kappa_1 < \bar{\tau}$, and remains constant otherwise. The equilibrium censorship level β^* strictly increases in κ_2 if and only if $(1 - \kappa_1)\kappa_2 > \bar{\tau}^2$ and $\kappa_2 > \bar{\tau}$, and remains constant otherwise.*

Notice that, by the platform's best response function, the cost of censorship, κ_1, only affects the censorship level indirectly: Depending on high the cost of censorship is relative to the tax rate, the platform either censors so that the government is indifferent between shutting it down when $\alpha = 1$ or does not censor at all. However, κ_1 is irrelevant to the government's shutting down strategy. In Cases 1 and 2, there is no tax benefit or censorship at equilibrium. An increase in κ_1 has no effect on β^* and makes it harder for the government to incentivize censorship. Hence an increase in κ_1 leads to no change in the equilibrium τ^*

and β^*. In Case 3, there is no tax benefit but censorship at equilibrium. A marginal increase in κ_1 does not change the platform's behavior: It still wants to censor to keep the government indifferent. Therefore, an increase in κ_1 leads to no change in the equilibrium τ^* and β^*. However, in Case 4, the platform is also on the edge of censoring or not. If κ_1 increases, the platform would stop censoring if nothing else is changed. Hence the government sets a lower tax rate to incentivize censorship by the platform. However, a lower tax rate amplifies the government's incentives to punish the platform as well. Therefore, it leads to an increase in the level of censorship required to keep the government indifferent. The equilibrium level of censorship increases while the equilibrium tax rate decreases. In Case 3 and 4, if the government's dis-utility from sensitive content, κ_2, increases, the platform is forced to censor more so that the government is still indifferent between shutting it down or not in the worst case.

The difference in censorship levels on SNS platforms of different sizes (the second trend) can be explained by the positive correlation between β^* and κ_2: The same piece of sensitive information receives more attention on a bigger platform. Hence the government dislikes sensitive content on SNS platforms with more users. So κ_2 is higher when the SNS platform has more users; hence, the equilibrium censorship level is higher.

In July 2022, the Russian internet regulator fined Google \$370 million for promoting the dissemination of false content on YouTube. A week later, Russia fined Google \$34 million under the pretext of breaching competition rules in the video hosting market. These punishments can be interpreted as the result of the sudden increase in α and κ_2 after the Russian invasion of Ukraine. The equilibrium censorship level β^* increases, whereas Google failed to meet the Russian government's expectations.

A possible explanation for the cessation of tax benefits for Weibo but not Tencent in 2021 is a northwestern movement for both companies in the $(1-\kappa_1) \times \kappa_2$ plane. Their κ_2 increases due to continuing growth in their active users. There is a huge increase in κ_1 after the pandemic: More Chinese internet users utilize more complicated methods to escape censorship after experiencing multiple lockdowns in real life and online. Weibo has moved to the Case 1 area, while Tencent remains inside the Case 4 area because Tencent has a bigger κ_2 than Weibo.

1.5 Conclusion

This paper provides a simple dynamic game theory model that illustrates the interaction between social media companies and the government that wants the SNS companies to remove politically sensitive content. The key trade-off for the government is between tax revenue and censorship. The tax rate that the government approves in advance affects how the government can manage censorship by SNS companies in the future: If the tax rate is high, the threat to punish the SNS platform in case of insufficient censorship is less convincing because it also hurts the government's tax revenue more. As a result, the platform has weaker incentives to comply with the government's order to censor. The model also explains the different censorship levels on SNS platforms of different sizes. It is consistent with the observation that the smaller platform, Weibo, loses tax benefits earlier than the bigger platform, Tencent.

2.0 When To Censor

2.1 Introduction

A censor examines contents and suppresses any parts that are considered unacceptable. Although it is nearly impossible to censor every piece of information on the Internet while it barely costs anything to duplicate some digital content, Internet censorship is pervasive worldwide. One apparent reason is that social networking service providers have to comply with the laws of the countries where their users are. For example, Twitter may remove content deemed illegal or objectionable after receiving complaints or requests from third parties, including government officials, companies, and other outside parties. What contents need to be removed? The answer varies across countries since the criterion for removal of contents is often defined by the government: In France, it could be antisemitic content or hate speech; In India, it could be parodies of the ruling elite; In South Korea, it could be any tweet by North Korea government; In China, it could be the social network itself or SOS posts by desperate Weibo users in the sudden lock-down of Wuhan in January 2020. Besides censorship under overt pressure, censorship by social networking service providers also exists. For example, thousands of ISIS-related Twitter accounts were suspended in 2015. Social media platforms in China hire thousands of inspectors to manually remove offensive content to avoid being punished by the government.

However, censorship may backfire if people know it exists. Noticing that a keyword is being blocked often makes people wonder why it is being blocked. For example, when a Chinese netizen notices that searching a place name on social media platforms returns only limited or no results, the immediate reaction is that something terrible must have happened there. It provides a reasonable way to fact-check a rumor in China: If searching the related keywords returns no result, the rumor is probably true. One recent example is the protest on a bridge in Beijing: On October 13th, 2022, a protestor burnt tires and hung banners against Xi's zero-Covid policy on Sitong Bridge, a bridge above a main transit route in central Beijing. Soon people noticed that the keywords "北京天桥" (bridge in Beijing) and

"横幅" (banner) were banned on all social media sites. People then started searching for and circulating videos and images of this protest via private channels and websites outside the Great Firewall, such as Twitter. Chinese people greeted that day by asking each other, "Do you know why the bridge in Beijing cannot be talked about?".

This paper explains the difference among the censorship strategies used by different governments when the choice of censorship strategies can be informative. Gratton et al. (2018) explains *October Surprises* by focusing on how Sender strategically chooses the starting time of the learning process. In their model, Sender has private information about his type and private information that would trigger a public learning process. This paper looks at the opposite situation: If a learning process has started, when does the censor want to stop it? The censor privately knows his type but has no control over how and when the public learning process starts. Instead, they can stop a learning process after it starts. For example, the outbreak of a scandal can trigger public discussion of a new topic, toward which the public does not yet know the government's attitude. At the beginning of 2022, a video of a chained woman, mom of seven boys and a girl, and "wife" of a local man in the countryside of Xuzhou went viral on Douyin (TikTok) in China. Netizens soon discovered that she was likely a victim of human trafficking decades ago. The local propaganda department in Xuzhou first claimed that their investigation showed that the marriage was legal and that there was no human trafficking involved. The public was not convinced at all and decided to investigate themselves. The government then set up another investigation team and, at the same time, started to obstruct voluntary investigators. Two female volunteers were detained by local police while investigating the case and trying to help the chained woman. Online discussions about contradictions among different reports by the authorities were strictly censored. Chinese citizens, especially Chinese women, started to question whether the Chinese Communist Party (CCP) cared about the well-being of Chinese women after observing the constant disappearance of discussions about the chained woman. After all, the slogan "妇女能顶半边天", translated as "women hold up half the sky", was an essential part of the propaganda since the late 1950s in China. Most Chinese women were still under the impression that the CCP was not against feminism until they found out the government did not allow any more discussion of this chained woman.

The literature on censorship mainly consists of empirical analysis of how censorship is implemented in authoritarian countries (King et al., 2013; King et al., 2017; Roberts, 2018; Gallagher and Miller, 2019). Chen and Xu (2017) argue that there are two possible ways an authoritarian government can benefit from allowing the public exchange of information among citizens: The government can learn about the dissatisfaction level of the public and amend policies accordingly; Citizens will be discouraged from protesting if they learn from public communication that other citizens have different opinions. This paper shows that, besides these reasons listed above, the authoritarian government mimics a government of good type via randomization of censoring and allowing the public exchange of information.

2.2 Model

Time horizon is discrete and finite, denoted by $t \in \{1, \dots, T, T+1\}$. There are two players: a Censor and a Receiver. Censor privately knows his type, $\theta \in \{G, B\}$ (good type or bad type). Receiver's prior belief is denoted by $Pr(\theta = G) = \pi$, where $\pi \in (0, 1)$ is common knowledge. Censor wants to maximize the receiver's posterior belief that $\theta = G$ at a given date $t = T$, denoted by s. Notice that the time horizon includes a period $t = T + 1$ that captures all future periods after the exogenous deadline T. Depending on the situation, the exogenous deadline can be interpreted differently. For example, it could represent the week of the National Congress of the Chinese Communist Party in China or Election Day in the United States.

A post exogenously arrives in the first period, $t = 1$. The model focuses on what different types of censors would do after the post arrives. [1] A learning process for the receiver is then triggered by the post. It can be interpreted as new information brought in by discussion and debates in the post's comment section. Without loss of generality, it is assumed that a sequence of signals is generated following a stochastic process

$$\mathcal{L} = \{L_\theta(t) | 1 \leq t \leq T\}.$$

[1] Since the arrival time of the post is normalized to 1, the actual meaning of T is the number of periods between the arrival of the post and the exogenous deadline.

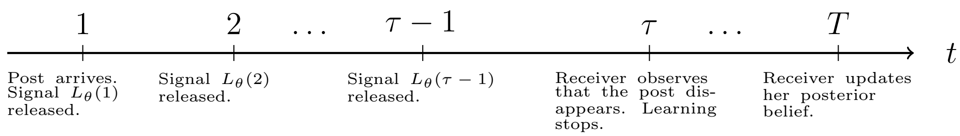


Figure 4: Timeline

The receiver stops receiving signals from the learning process if the censor or a technical shock removes the post happens. In every period, an exogenous technical shock happens with probability $\varepsilon \in [0, 1)$. The censor can choose whether to remove the post in each period unless a technical shock has happened.

Figure 4 shows the timeline if the post disappears at $t = \tau$, either because of a technical shock or because the censor removes it: The receiver combines two pieces of information to form her posterior belief α at T: (i) η, the interim belief that $\theta = G$ based on the the fact that the post disappears at τ, and (ii) signals she received from the learning process from period 1 to $(\tau - 1)$. These signals follow the stochastic process $\mathcal{L}_\tau = \{L_\theta(t) | 1 \leq t \leq \tau - 1\}$.[2]

Denote by $H(\cdot|\tau, \eta)$ the distribution over Receiver's posterior belief, α, generated by $\mathcal{L}_\tau$ conditional on her interim belief η. Similarly, denote by $H_\theta(\cdot|\tau, \eta)$ the distribution over Receiver's posterior belief, α, generated by $\mathcal{L}_\tau$ conditional on her interim belief, η, and Censor's type, θ. Notice that $H(\cdot|1, \eta)$ and $H_\theta(\cdot|1, \eta)$ are degenerate distributions that assign probability one to $\alpha = \eta$.

In equilibrium, the receiver updates her belief from π to η based on the equilibrium strategy of the censor. If the good and bad censors use different strategies, the timing of disappearance, τ, is then informative. In other words, in non-raveling equilibria, censor signals about his type via the timing of removal.

Assumption 2.1. *(i) Every signal generated by $\mathcal{L}$ is informative (in Blackwell's sense). Removing the post sooner reveals strictly less information about the censor's type: For any $\tau, \tau' \in \{1, \ldots, T, T+1\}$ with $\tau < \tau'$, $H(\cdot|\tau', \eta)$ is a strict mean-preserving spread of $H(\cdot|\tau, \eta)$. (ii) The learning process never fully reveals censor's type: The support of $H(\cdot|T, \eta)$ is a subset*

[2] Note that $\mathcal{L}_1 = \emptyset$, and $\mathcal{L}_{T+1} = \mathcal{L}$.

of $(0,1)$ *for any* η.

2.3 Main Result

The solution concept is Perfect Bayesian Equilibrium. A typical strategy of the censor is denoted by $s_\theta \in [0,1]^T$, where $s_\theta(t)$ represents the probability of removing the post if the post still exists in period t. Let P_θ and p_θ be the c.d.f. and density function of the disappearing time of the post given censor's type θ and strategy in equilibrium. Note that as long as $\varepsilon \neq 0$, in every period t either $p_\theta(t) > 0$, or $p_\theta(t) = 0$ and $s_\theta(\hat{t}) = 1$ for some $\hat{t} < t$.

In any equilibrium, the receiver's interim belief if the post disappears in period τ is given by $\eta = \mu(\tau)$ with

$$\mu(\tau) = \frac{\pi p_G(\tau)}{\pi p_G(\tau) + (1-\pi)p_B(\tau)}.$$

Lemma 2.1. *Let* $\mathbb{E}[\alpha|\tau,\eta,\theta]$ *denote the expectation of Receiver's posterior belief* α *conditional on the disappearing time* τ*, her interim belief* η*, and Censor's type* θ*. For any* $\tau < \tau' \in \{1,\ldots,T,T+1\}$ *and any* $\eta < \eta'$*,*

1. $\mathbb{E}[\alpha|\tau,\eta,\theta] < \mathbb{E}[\alpha|\tau,\eta',\theta]$ *for* $\theta \in \{G,B\}$*;*
2. $\mathbb{E}[\alpha|\tau,\eta,G] < \mathbb{E}[\alpha|\tau',\eta,G]$*;*
3. $\mathbb{E}[\alpha|\tau,\eta,B] > \mathbb{E}[\alpha|\tau',\eta,B]$*.*

All omitted proofs are in the Appendix. Censor always benefits from a higher interim belief, regardless of his type. For a fixed interim belief, more information for the receiver is detrimental for the bad censor while beneficial for the good censor.

Lemma 2.2. *In any equilibrium:*

- *For all* $\tau < \tau' \in \{1,\ldots,T,T+1\}$*, if good Censor weakly prefers that the post disappears at* τ *than at* τ'*, then* $\mu(\tau) > \mu(\tau')$ *and bad Censor strictly prefers that the post disappears at* τ *than* τ'*.*

- *For all $t \leq T$, $s_G(t) = 0$, and hence*

$$p_G(t) = \begin{cases} (1-\varepsilon)^{t-1}\varepsilon & t \leq T \\ (1-\varepsilon)^T & t = T+1 \end{cases}$$

Suppose a good censor weakly prefers that the post disappears at τ than at $\tau' > \tau$. It must be because the receiver's belief is more tilted toward $\theta = G$ upon observing that the post disappears at τ than at τ'. Moreover, this gain in credibility must dominate the loss from less learning for the good censor. Hence a bad censor should strictly prefer that the post disappears at τ than τ' because, unlike the good censor, the bad censor benefits from less learning.

The second result is a corollary of the first one. Bad censor must have already removed the post at τ if good censor assigns positive probability to remove the post at τ, i.e., for any $\tau \leq T$, $s_G(\tau) > 0$ implies $s_B(\tau) = 1$. Suppose $s_G(\tau) > 0$ for some $\tau \leq T$ in equilibrium, then $p_B(\tau') = 0$ for all $\tau' > \tau$ and hence Receiver can conclude that $\theta = G$ upon observing that the post disappears in any period $\tau' > \tau$. Therefore a good censor should strictly prefer that the post disappears at τ' than τ, contradicting that $s_G(\tau) > 0$. As a result, a good censor never removes the post in any period $t \leq T$ in any equilibrium.

Lemma 2.3. *In any equilibrium:*

- *If $\varepsilon = 0$, bad Censor never removes the post.*
- *If $\varepsilon > 0$, $s_B(t) < 1$ for any $t \leq T$ and $p_B(T+1) > 0$.*
- *If $\varepsilon > 0$, $s_B(t) > 0$ for any $t \leq T$.*

If the exogenous technical shock does not exist, given that the good censor never removes the post, the receiver can conclude that the censor is of the bad type upon observing the removal of the post. Hence the bad censor should strictly prefer to mimic the good censor and never removes the post.

To understand the second result, suppose $s_B(t) = 1$ for some $t \leq T$ in equilibrium. Then the post never disappears after period t. In other words, for any $t' > t$, $p_B(t') = 0$. Since the good censor never removes the post, all information sets are on-path. Therefore, the receiver believes that $\theta = G$ if the post disappears after t. The bad censor is strictly better off by choosing $s_B(t) = 0$ instead. Therefore, in any PBE, $s_B(t) < 1$ for any $t \leq T$.

The third result is by induction on time. Since $s_B < 1$, with some positive probability, the post never disappears, i.e., $p_B(T+1) > 0$. It thus implies that the bad censor weakly prefers that the post disappears at $T+1$ than at period T because otherwise, he would choose $s_B(T) = 1$ instead. Recall that more learning hurts the bad censor, so the gain in credibility at $T+1$ must dominate the loss from more learning. Hence $s_B(T) > \frac{(1-\varepsilon)}{(1-\varepsilon)+1} > 0$. Suppose $s_B(t) > 0$ for some $t \leq T$. Then the bad censor weakly prefers that the post disappears at t than at $(t-1)$. Again, the gain in credibility must dominate the loss from more learning, so $s_B(t-1) > \frac{(1-\varepsilon)s_B(t)}{(1-\varepsilon)s_B(t)+1} > 0$. By induction, $s_B > 0$.

Lemma 2.4. *In any equilibrium, for any $\tau \geq 1$,*

$$\int \alpha dH_B(\alpha|\tau, \mu(\tau)) = \mu(1), \tag{1}$$

and

$$\sum_{\tau=1}^{T+1} \frac{1-\mu(\tau)}{\mu(\tau)} p_G(\tau) = \frac{1-\pi}{\pi} \tag{2}$$

These results are essentially the Martingale property: They follow from that the receiver updates her belief by Bayes's rule and the censor's equilibrium strategy. The expectation of posterior beliefs must not change.

Proposition 2.1. *There exists a divine equilibrium in which*

$$p_G(\tau) = \begin{cases} (1-\varepsilon)^{\tau-1}\varepsilon & \tau \leq T \\ (1-\varepsilon)^T & \tau = T+1 \end{cases},$$

and for all $\tau \in [1, T+1]$

$$\frac{\mu(\tau)}{1-\mu(\tau)} p_B(\tau) = \frac{\pi}{1-\pi} p_G(\tau),$$

where $\mu(\tau) \in (0,1)$ is uniquely determined by (1) and (2) .

In this equilibrium, the good type never censors, while the bad type randomly censors. The bad censor is indifferent about when the post disappears.

2.4 Conclusion

The censor wants to prevent the receiver from learning more about his private type. However, it may backfire if the receiver notices the existence of censorship. The receiver can learn from the existence and, more specifically, the timing of censorship as long as good and bad censors use different strategies. The censor, therefore, faces the trade-off between allowing the receiver to learn more about his true type and revealing that the censor is more likely a good type when he chooses to remove the post later. In the divine equilibrium, a good censor never removes the post, while a bad censor randomly removes the post to mimic the good censor whenever a technical shock exists. The positive probability of technical shock in each period is critical: If the receiver knows that there is no technical shock at all, she believes the censor is of bad type as soon as she observes the disappearance of a post. Hence it is impossible for a bad censor to disguise censorship.

3.0 Governmental Structure: Vertical or Horizontal?

3.1 Introduction

To facilitate the economic reform in the late 1970s, sub-national 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 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 in figure 5.

(a) Horizontal Structure (b) Vertical Structure

Figure 5: 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 sub-national 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.

3.2 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 summarized in Figure 6.

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 tree diagram in Figure 7. The top node represents a province, position 1 and 2 represent two prefectures

[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.

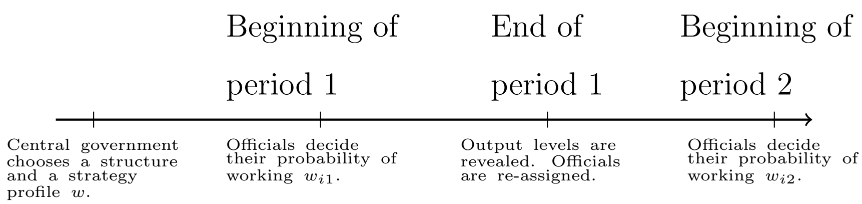


Figure 6: Timeline

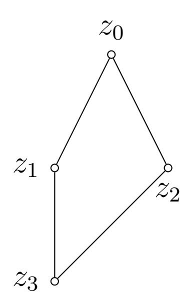
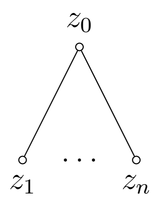
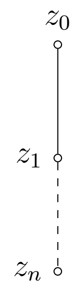


Figure 7: Example: Four-Official Three-Layer Structure



(a) horizontal structure h_n



(b) vertical structure v_n

Figure 8: Graphic illustration

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, as depicted in Figure 8. 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.

3.3 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 3.1. *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.

3.3.1 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 3.1. *Given any n and P, the optimal horizontal structure $h^*_{n,P}$ implements an* *all-working strategy* $w_n^1 = (1, \dots, 1)$ *with the following distribution of resources:*

$$z_0 = \frac{(1-\delta(1-P))\left(1-\delta\left(1-\frac{P}{n}\right)\right)}{(1-\delta(1-P))\left(1-\delta\left(1-\frac{P}{n}\right)\right)+\delta P}, \tag{3}$$

and

$$z_1 = \ldots = z_n = \frac{\delta\frac{P}{n}}{(1-\delta(1-P))\left(1-\delta\left(1-\frac{P}{n}\right)\right)+\delta P}. \tag{4}$$

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* $Y^*_{n,P}(w_n^1) = \sum_{i=1} nz_i = 1 - z_0$ *is increasing in* n*.*

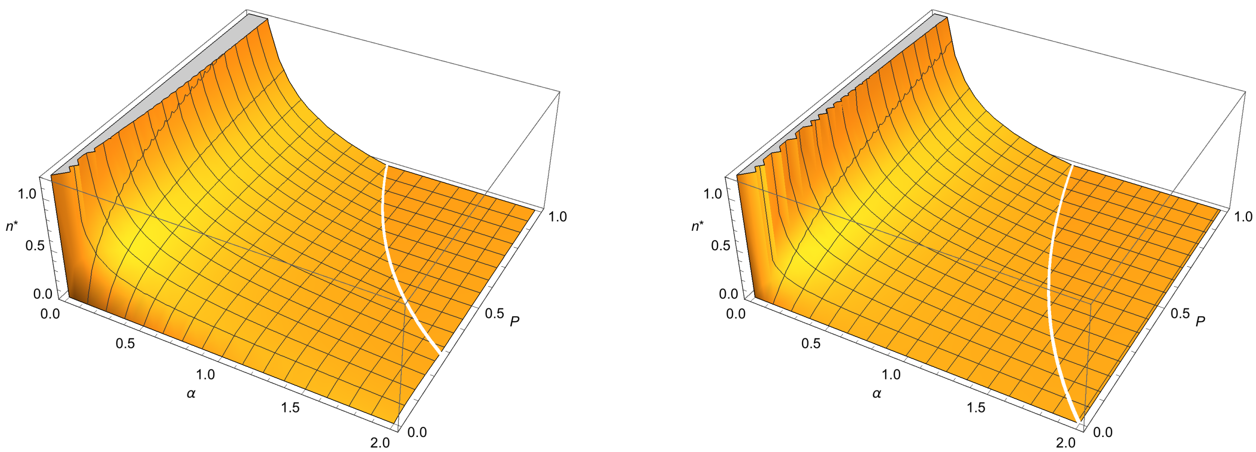
It is natural to ask what is the optimal horizontal structure if the central government is free to choose n, the number of local offices, and P, the Leader's probability of getting fired. Proposition 3.1 shows that given any P, a bigger n leads to a higher total output. But the increase in the total 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 3.2. *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*.*



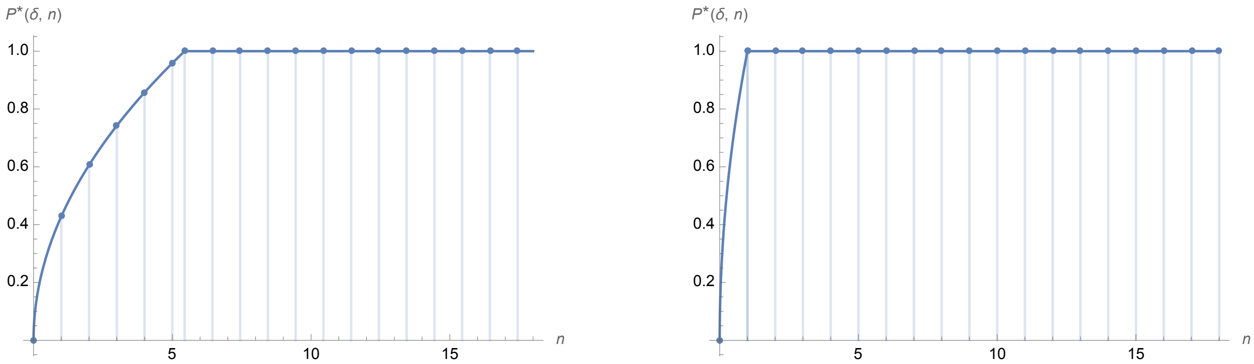
(a) $n^*(\alpha, 0.7, P)$ (b) $n^*(\alpha, 0.5, P)$

Figure 9: Example of $n^*(\alpha, \delta, P)$

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 $n^*(\alpha, \delta, P) > 0$.

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, $n^*(\alpha, \delta, P)$. Figure 9 shows examples of optimal n^* as a function of α and P given a fixed value of δ.

What is the optimal value of P when n is fixed? The office cost is fixed given that



(a) $P^*(0.7, n)$ (b) $P^*(0.5, n)$

Figure 10: Example of $P^*(\delta, n)$

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 10 shows examples of optimal P^* as a function of n given a fixed value of δ.

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.

3.3.2 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.

3.3.3 Main Result

Proposition 3.3. *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 3.4. *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.

3.4 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.

4.0 When Would an Autocrat Open the Emigration Floodgates?

4.1 Introduction

In an autocracy, the autocrat chooses what he/she wants. Emigration policy is a crucial part of the choice set since it can be used to manipulate the domestic demographic characteristics and hence citizens' behavior. History does witness a lot of variations of emigration policies in autocracies both across time and region. The Soviet government strictly limited emigration from the late 1920s to the 1970s. The Berlin Wall existed from 1961 to 1989. China has also experienced changes in emigration regulations since the formulation of the People's Republic of China in 1949. North Korea, in contrast, sticks to the no-migration policy and is infamous for heavy penalties for defection. This paper provides a simple model that integrates the government's choice, citizens' emigration decision, and revolution decision to analyze the costs and benefits of opening the floodgates. The following scrutinisation of emigration policy changes in China motivates this model.

Before 1949 most Chinese emigrants were low-skilled labor force and refugees to countries in Asia and the Americas. People were prevented from leaving the country during the planned economy era from 1949 to the late 1970s.[1] Most people could not apply for a passport. Even moving within China was strictly restricted. The ban on migration was gradually lifted following the Opening Up reform in 1978. Many people were able to reunite with their families overseas due to liberalized emigration policies. China started to export a large amount of labor (mainly low-skilled) to the world. Thousands of students have been sponsored by the government to study abroad yearly since the early 1980s. The slogan was "努力学习，学成回国，报效祖国", translated as "study hard (abroad), come back after graduation to repay your motherland." The wave of self-financed student emigrants came with the boost in the Chinese economy and the adoption of a more flexible policy on studying abroad in 1984. The Chinese Communist Party's new slogan in 1992 was "支持留学，鼓励回国，来去自由", meaning that "students should be supported to study abroad, encouraged

[1] Despite this regulation, more than two million people illegally emigrated to Hong Kong in this period.

to return to China but free to choose whether to return". The outflow of students grew while the return rate decreased during this period.

Why did the government's policy change so much in such a short time? One thing that must not be ignored is the 1989 Tiananmen Square protests. Students led hunger strikes and sit-ins in Tiananmen Square in Beijing starting in April, calling for democracy, freedom of speech and the press, and constitutional due process. Demonstrations soon spread to about 400 cities nationwide by May. The protests in Beijing were suppressed on June 4th by troops with rifles and tanks under the orders of the government. Mass arrests were carried out afterward. The Beijing Public Security Bureau identified 21 students as leaders of the protests. Seven fled to Hong Kong, Taiwan, and the United States, while the rest were arrested and spent a few years in prison. After being released, eight imprisoned students went to top universities abroad and eventually settled overseas.

Why were some student leaders on the list never arrested? Was it because they ran away fast or because the government turned a blind eye? Why did the Chinese government issue passports to these imprisoned student leaders after they were released? One possible explanation is that the government let these students abroad to avoid potential disturbance. It may also be true that the government did not arrest all 21 students in 1989 simply due to incapability. However, what happened decades later when the government was able to imprison them is worth investigating: Several students who ran away in 1989 successfully entered China decades later, even though Beijing has never retracted the most-wanted list. Wu'erkaixi tried several times to return to mainland China to turn himself in to the police after 2004 but was deported every time. The way the government dealt with these student leaders, paired with the relaxation of study abroad regulations since 1992, could be considered a way to reduce threats from students. The contemporary migration trends in China are summarized as follows. The four most popular ways of emigration are: (i) The rich invest in the destination country to obtain permanent residence. (ii) Highly skilled labor and students move to countries such as the United States, the United Kingdom, Canada, and Australia (the brain drain). Several policies encouraging returns of this group, either temporarily or permanently, have been introduced. But there is no exit control generally. Returning is a choice but not a requirement, except for less than five percent of government-sponsored students. (iii) Family members apply for immigrant visas as dependents of the rich and the highly skilled. (iv) The unskilled emigrate through project-tied collective labor deployment and individual overseas employment. However, the application fee is high, and the workers must return upon the completion of a job contract. [2] Meanwhile, the number of permanent immigrants to China is extremely small and negligible compared to that of emigrants. There are about five million more emigrants than immigrants each year (Xiang, 2016).

Constantly losing the young and educated labor force harms productivity and hence the GDP. So there must be some benefits of opening the floodgates that dominate this loss in productivity under certain conditions. This model combines revolution and emigration and shows that under certain conditions, the gains of relaxing emigration policies dominate the loss in GDP. Two important assumptions are used to make the model tractable. First, it is assumed that only one representative strategic citizen chooses to stay in the country (and not protest), (stay and) protest, or emigrate. All other citizens are assumed to be non-strategic. This representative strategic citizen captures the small group of citizens who can and believe that they can potentially take over the autocracy and, if they cannot have their way, would emigrate if possible. They are not "cogs in the wheels" of the system who are playing by the rules or the social contract of the system. Typically they are young, educated, and able to think outside the confines of the system. In China's case, the student leaders of the 1989 Tiananmen Square protests are considered strategic citizens. It was not uncommon in the history of other autocracies that the well-educated became revolutionary forces and posed threats to the government. Ho Chi-Minh, a founding member of the French Communist Party, the first leader of Vietnam's nationalist movement, and the symbol of Vietnamese liberation, received his political education in France from 1919 to 1923. The Gwangju Uprising in South Korea in May 1980 was started by university students. They led demonstrations for an end to martial law and democratization. Lots of people who took over the former Soviet in the early 1990s were educated mathematicians and physicists and not those educated in Marxist political economy. This assumption greatly simplifies the analysis of the revolution decision by setting aside the coordination and free-riding problems that

[2] The common practice is to collect a deposit before they leave China and only return it after they board the airplane back home when the contract ends.

are common in protest models (Morris and Shin, 2001; Bueno De Mesquita, 2010), making it possible to focus on the interaction of revolution and emigration decisions. Second, it is assumed that the unknown true state of the world determines how hard it is to take over and that there are only two possible states. In one state, the government is overthrown if and only if the strategic citizen protests; In another, the government never loses office. The government faces a lower risk of losing office by allowing the strategic citizen to leave of his/her own free will.

4.2 Literature Review

Much attention has been given to the reasons behind the rising interest in emigration in autocracies. Recent surveys show that around half of wealthy individuals in China have considered emigrating and sending their children abroad (Hurun, 2016). Three possible motivations are on the table: higher quality of education abroad, higher quality of life abroad, and a broader concern about the political, social, and economic stability in China (Xiang, 2016; Huang, 2017).

However, little of the literature on immigration sheds light on the other side of the story. Knowing what incentives and strategy space the game players (citizens) face, it is natural to take a further step to analyze the choices of the game designer (the autocrat). Since the autocrat has *de facto* power to modify the emigration policy to their taste, why do they write the emigration policy such that emigration is possible for some citizens but almost impossible for the rest? To answer this question, this model combines the government's choice of emigration policies, the citizens' emigration choice, and revolutionary action.

Some of the literature on Arab Spring also discuss the relationship between political movements and population movements but are mainly empirical and case studies (Boubakri, 2013; Fargues, 2017). Sellars (2019) models the interaction between migration and collective action with empirical evidence from Mexico and Japan but assumes in her model that individuals' choice space is exogenously given. Her model shows how emigration can reduce pressure on the government; This model allows the government to manipulate the cost of emigration and therefore alters the set of feasible choices available to citizens. It looks at how this interaction impacts the government's choices as it is torn between fear of losing talents and liking for stability.

4.3 Model

Consider a representative strategic citizen who maximizes her expected income by choosing from the action space $\{Stay, Emigrate, Protest\}$.[3] The citizen earns wage $(1-\tau)w$ after tax if she stays in the country, where τ is the fixed income tax rate. Alternatively, she can pay an emigration cost c to emigrate and earn αw abroad (also after tax) with $\alpha > 1 - \tau$. Notice that the citizen's wage abroad is assumed to be proportional to her domestic wage. It captures the idea that the citizen's wage reflects her ability and that countries differ in average wage levels. Alternatively, she can work in her home country and, at the same time, pay an upfront cost, κ, to participate in an underground protest against the government.

Other citizens in this country are assumed to be non-strategic: They never move to another country or participate in a protest. For example, loyalty to the ruling party could be a reason why some citizens do not behave strategically. Since the incumbent party cares only about the behavior of the representative strategic citizen, non-strategic citizens' actions are irrelevant. Denote the total income of non-strategic citizens before tax by γw.

If the protest succeeds, the incumbent party is overthrown, and the strategic citizen receives $\beta + (1 - \tau)w - \kappa$, where β represents the benefit of taking office; Otherwise, the incumbent party remains in office and collects the tax, and the strategic citizen gets only $(1 - \tau)w - \kappa$.

The unknown true state of the world $T \in \{B, G\}$ determines how hard it is to overthrow the incumbent party. A bad state $T = B$ means the protest succeeds if and only if the strategic citizen protests, while a good state $T = G$ means a protest can never succeed. A

[3] I use one representative strategic citizen to avoid the free-ride problem and the coordination problem in the revolution game.

		Result of the protest	
		Succeed	Fail
Citizen	Stay	$(1-\tau)w$	$(1-\tau)w$
	Emigrate	$\alpha w - c$	$\alpha w - c$
	Protest	$\beta + (1-\tau)w - \kappa$	$(1-\tau)w - \kappa$
Government		-L	τ*GDP

Table 1: Return to the citizen and the government

common prior on the distribution is assumed:

$$P(T = B) = \eta, P(T = G) = 1 - \eta.$$

The risk-neutral incumbent party collects the flat-rate income tax if the protest fails and incurs a loss of L if the protest succeeds. The loss L represents the dis-utility from being overthrown. The objective function of the government is given by

$$\pi = \tau Prob(\text{protest fails})\mathbb{E}(GDP|\text{protest fails}) - Prob(\text{protest succeeds})L.$$

The first term is the expected amount of tax the government collects. Note that "tax" in this model represents what the ruling party takes away from its citizens and puts into its' pocket. Corruption, for example, is considered part of this tax. Public spending, however, is not included in this tax.

Table 1 summarizes the return to the citizen and the government under different results of the protest.

Potentially, the government can manipulate emigration cost c by modifying emigration policies, the participation cost κ by adjusting regulations regarding protests, as well as the tax rate τ. The following sections discuss what happens if the government is allowed to choose c only, τ only, or both (c, τ). In each case, the representative citizen makes her choice after observing the government's choice without knowing the true state. Finally, the true state is realized, and everyone gets their return.

4.4 Main Result

4.4.1 Benchmark I: The Government Chooses c

The citizen chooses the action that gives her the highest return:

$$\max\{(1-\tau)w, \alpha w - c, \eta\beta + (1-\tau)w - \kappa\}.$$

Let $(1-e-p, e, p) \in \Delta(\{Stay, Emigrate, Protest\})$ be the strategic citizen's equilibrium strategy. The same notation is also used in the rest of the paper. Recall that the protest succeeds if and only if $T = B$ and the citizen chooses *Protest*. Note that $p > 0$ requires

$$\eta\beta + (1-\tau)w - \kappa \geq \max\{(1-\tau)w, \alpha w - c\}.$$

Proposition 4.1. *If $\eta\beta > \kappa$ and $\tau w(1-\eta(1+\gamma)) \leq \eta L$ the government chooses any $c < (\alpha+\tau-1)w + \kappa - \eta\beta$ such that $e = 1$ and $\pi = \tau\gamma w$.*

If $\eta\beta > \kappa$ and $\tau w(1-\eta(1+\gamma)) \geq \eta L$ the government chooses any $c > (\alpha+\tau-1)w+\kappa-\eta\beta$ such that $p = 1$ and $\pi = (1-\eta)\tau(1+\gamma)w - \eta L$.

Proof. If $\eta\beta > \kappa$ then *Protest* strictly dominates *Stay* and hence $1-e-p$ must be zero. If $c > (\alpha+\tau-1)w+\kappa-\eta\beta$ then $p = 1$ and $\pi = (1-\eta)\tau(1+\gamma)w-\eta L$. If $c < (\alpha+\tau-1)w+\kappa-\eta\beta$ then $e = 1$ and $\pi = \tau\gamma w$. Therefore, $c < (\alpha+\tau-1)w+\kappa-\eta\beta$ is optimal for the government if $(1-\eta)\tau(1+\gamma)w - \eta L \leq \tau\gamma w$, which is equivalent to $\tau w(1-\eta(1+\gamma)) \leq \eta L$; And $c > (\alpha+\tau-1)w+\kappa-\eta\beta$ is optimal for the government if $\tau w(1-\eta(1+\gamma)) \geq \eta L$. □

If *Stay* is never the citizen's best choice, a low emigration cost may be optimal for the government. More specifically, when the dis-utility of losing office is big enough ($\tau w(1-\eta(1+\gamma)) \leq \eta L$), making the citizen leave the country is optimal for the government.

An example of this case where emigration is an affordable choice while the government is worried about the stability of society is the 1980s in China. During this period, the economy of China is relatively undeveloped, making the difference between domestic wage and foreign wage huge (α is big), while η is relatively high.

In 1972, would-be emigrants of the USSR had to pay a fee as high as twenty times an annual salary for the higher education they received in the USSR. The choice of a big c

was intended to alleviate the brain drain but was soon revoked as it caused international condemnations. The Dymshits‘‘Kuznetsov hijacking affair in 1970, an attempt by 16 Soviet refuseniks to steal a civilian aircraft and escape to the West, was followed by the Soviet Union aliyah, the mass emigration of Soviet Jews to Israel. The Soviet government could interpret the hijacking affair as a signal that η increased.

Proposition 4.2. *If* $\eta\beta < \kappa$ *the government chooses any* $c > (\alpha+\tau-1)w$ *such that* $e = p = 0$ *and* $\pi = \tau(1+\gamma)w$*.*

Proof. If $\eta\beta < \kappa$ then *Stay* strictly dominates *Protest* and hence p must be zero. If $c < (\alpha+\tau-1)w$ then $e = 1$ and $\pi = \tau\gamma w$. So the government chooses $c > (\alpha+\tau-1)w$ such that the strategic citizen chooses *Stay* ($e = p = 0$) and gets $\pi = \tau(1+\gamma)w > \tau\gamma w$. □

If *Protest* is never the citizen's best choice, keeping the citizen in the country is always optimal for the government.

An example where emigration is almost impossible and revolutionary activities are savagely suppressed is the period from 1949 to the late 1970s in China. Although α is at its peak in this period, citizens being unable to apply for a passport can be interpreted as c approaching infinity. Other examples include North Korea and the Eastern Bloc. State censorship and restrictions on emigration are at their highest levels. It is almost impossible for citizens to protest or to emigrate.

The non-generic case of $\eta\beta = \kappa$ is omitted throughout the paper.

4.4.2 Benchmark II: The Government Chooses τ

Consider another benchmark model where the emigration cost c is fixed and the tax rate $\tau \in [0, \hat{\tau}]$ is chosen by the government, where $\hat{\tau} \in (0, 1)$ represents the upper bound of the government's taxing ability.

Proposition 4.3. *If* $\eta\beta > \kappa$ *and* $\hat{\tau} > \max\{1-\alpha+\frac{c+\eta\beta-\kappa}{w}, (1-\eta)\frac{1+\gamma}{\gamma}(1-\alpha+\frac{c+\eta\beta-\kappa}{w}) - \frac{\eta L}{\gamma w}\}$*, the government chooses* $\tau = \hat{\tau}$ *such that* $e = 1$ *and* $\pi = \hat{\tau}\gamma w$*.*

If $\eta\beta > \kappa$ *and* $\hat{\tau} < 1-\alpha+\frac{c+\eta\beta-\kappa}{w}$*, the government chooses* $\tau = \hat{\tau}$ *such that* $p = 1$ *and* $\pi = (1-\eta)\hat{\tau}(1+\gamma)w - \eta L$*.*

If $\eta\beta > \kappa$ and if $1-\alpha+\frac{c+\eta\beta-\kappa}{w} < \hat{\tau} < (1-\eta)\frac{1+\gamma}{\gamma}(1-\alpha+\frac{c+\eta\beta-\kappa}{w}) - \frac{\eta L}{\gamma w}$ the government chooses $\tau = 1-\alpha+\frac{c+\eta\beta-\kappa}{w}$ such that $p=1$ and $\pi = (1-\eta)(1+\gamma)\left((1-\alpha)w + c + \eta\beta - \kappa\right) - \eta L$.

Proof. If $\eta\beta > \kappa$ then *Protest* strictly dominates *Stay* and hence $1-e-p$ must be zero. If $\tau \leq 1-\alpha+\frac{c+\eta\beta-\kappa}{w}$ then $p=1$ is an equilibrium and $\pi = (1-\eta)\tau(1+\gamma)w - \eta L$. If $\tau \geq 1-\alpha+\frac{c+\eta\beta-\kappa}{w}$ then $e=1$ is an equilibrium and $\pi = \tau\gamma w$.

The discussion is further divided into two cases because of the existence of the upper bound on τ:

- If $\hat{\tau} < 1-\alpha+\frac{c+\eta\beta-\kappa}{w}$ then $e=1$ can never be implemented. Hence $\tau=\hat{\tau}$ is optimal and $p=1$, $\pi = (1-\eta)\hat{\tau}(1+\gamma)w - \eta L$ in equilibrium.
- If $\hat{\tau} > 1-\alpha+\frac{c+\eta\beta-\kappa}{w}$ then the government gets at most $(1-\eta)(1+\gamma)\left((1-\alpha)w + c + \eta\beta - \kappa\right) - \eta L$ by choosing $\tau = 1-\alpha+\frac{c+\eta\beta-\kappa}{w}$, and at most $\hat{\tau}\gamma w$ by choosing $\tau=\hat{\tau}$. Therefore, if $\hat{\tau} > \max\{1-\alpha+\frac{c+\eta\beta-\kappa}{w}, (1-\eta)\frac{1+\gamma}{\gamma}(1-\alpha+\frac{c+\eta\beta-\kappa}{w}) - \frac{\eta L}{\gamma w}\}$, $\tau=\hat{\tau}$ is optimal for the government and $e=1$, $\pi=\hat{\tau}\gamma w$ in equilibrium; If $1-\alpha+\frac{c+\eta\beta-\kappa}{w} < \hat{\tau} < (1-\eta)\frac{1+\gamma}{\gamma}(1-\alpha+\frac{c+\eta\beta-\kappa}{w}) - \frac{\eta L}{\gamma w}$ then $\tau = 1-\alpha+\frac{c+\eta\beta-\kappa}{w}$ is optimal for the government and $p=1$, $\pi = (1-\eta)(1+\gamma)\left((1-\alpha)w + c + \eta\beta - \kappa\right) - \eta L$ in equilibrium.

□

If *Stay* is never the citizen's best choice and emigration is profitable enough for the strategic citizen (i.e., if α and w are big enough and c is small enough), the government uses a high tax rate to force the strategic citizen out of the country.

Proposition 4.4. *If $\eta\beta < \kappa$ the government chooses any $c > (\alpha+\tau-1)w$ such that $e=p=0$ and $\pi = \tau(1+\gamma)w$.*

Proof. If $\eta\beta < \kappa$ then *Stay* strictly dominates *Protest* and hence p must be zero. If $\tau \geq 1-\alpha+\frac{c}{w}$ then $e=1$ is an equilibrium and $\pi=\tau\gamma w$. If $\tau \leq 1-\alpha+\frac{c}{w}$ then $e=p=0$ is an equilibrium and $\pi=\tau(1+\gamma)w$. The optimal choice of τ depends on the value of $\hat{\tau}$:

- If $\hat{\tau} \geq (1+\gamma)(1-\alpha+\frac{c}{w})$ then $\tau=\hat{\tau}$ is optimal and in equilibrium $e=1$ and $\pi=\hat{\tau}\gamma w$.
- If $1-\alpha+\frac{c}{w} < \hat{\tau} < (1+\gamma)(1-\alpha+\frac{c}{w})$ then $\tau = 1-\alpha+\frac{c}{w}$ is optimal and in equilibrium $e=p=0$ and $\pi = (1-\alpha)w+c)(1+\gamma)$.

- If $\hat{\tau} < 1 - \alpha + \frac{c}{w}$ then $\tau = \hat{\tau}$ is optimal and in equilibrium $e = p = 0$ and $\pi = \hat{\tau}(1 + \gamma)w$.

□

If *Protest* is never the citizen's best choice, the government either chooses a small enough τ and collects tax from all citizens or a big enough τ and collects tax only from non-strategic citizens. Making the strategic citizen leave is optimal if $\hat{\tau}$ is big enough and γ is small enough.

4.4.3 The Government Chooses (c, τ)

Now assume the government can choose both the tax rate $\tau \in [0, \hat{\tau}]$ and the emigration cost c, where $\hat{\tau} \in (0, 1)$ is the upper bound of the government's taxing ability.

Proposition 4.5. *If $\eta\beta > \kappa$ and $(1 - \eta(1 + \gamma))\hat{\tau}w \leq \eta L$ the government chooses any $(c, \hat{\tau})$ that satisfies $c < (\alpha + \hat{\tau} - 1)w + \kappa - \eta\beta$. In equilibrium $e = 1$ and $\pi = \hat{\tau}\gamma w$.*

If $\eta\beta > \kappa$ and $(1 - \eta(1 + \gamma))\hat{\tau}w \geq \eta L$ the government chooses $(c, \hat{\tau})$ with $c > (\alpha + \hat{\tau} - 1)w + \kappa - \eta\beta$. In equilibrium $p = 1$ and $\pi = (1 - \eta)\hat{\tau}(1 + \gamma)w - \eta L$.

Proof. If $\eta\beta > \kappa$ then *Protest* strictly dominates *Stay* and hence $1 - e - p$ must be zero. If $c > (\alpha + \tau - 1)w + \kappa - \eta\beta$ then $p = 1$ and $\pi = (1 - \eta)\tau(1 + \gamma)w - \eta L \leq (1 - \eta)\hat{\tau}(1 + \gamma)w - \eta L$, with equality if and only if $\tau = \hat{\tau}$. If $c < (\alpha + \tau - 1)w + \kappa - \eta\beta$ the citizen chooses $e = 1$ and then $\pi = \tau\gamma w \leq \hat{\tau}\gamma w$ with equality if and only if $\tau = \hat{\tau}$.

Therefore, a small c is optimal for the government if $(1 - \eta)\hat{\tau}(1 + \gamma)w - \eta L \leq \hat{\tau}\gamma w$, which is equivalent to $(1 - \eta(1 + \gamma))\hat{\tau}w \leq \eta L$. However, if $(1 - \eta(1 + \gamma))\hat{\tau}w \geq \eta L$, it is optimal for the government to choose $(c, \hat{\tau})$ with $c > (\alpha + \hat{\tau} - 1)w + \kappa - \eta\beta$. □

The intuition is the same as in the first benchmark model. If *Stay* is never the citizen's best choice and the dis-utility of losing office is big enough ($(1 - \eta)\hat{\tau}w \leq \eta L$), the government prefers to make the strategic citizen leave the country. Also observe that whenever the choice of the tax rate is relevant, the government should choose the highest tax rate possible, $\hat{\tau}$, because the "cost" of increasing the tax rate (providing the citizen a stronger incentive to emigrate) can be eliminated by increasing the emigration cost accordingly. A higher emigration cost does not directly affect the government's payoff, but a higher tax rate does.

So the tax rate is a better tool than the emigration cost for the government to manipulate citizens' choices.

Proposition 4.6. *If $\eta\beta < \kappa$ the government chooses $(c, \hat{\tau})$ with $c > (\alpha + \hat{\tau} - 1)w$ such that $e = p = 0$ and $\pi = \hat{\tau}w$.*

Proof. If $\eta\beta < \kappa$ then *Stay* strictly dominates *Protest* and hence p must be zero. If $c < (\alpha + \tau - 1)w$ then $e = 1$ and $\pi = 0$. So the government chooses $c > (\alpha + \tau - 1)w$ such that the citizen chooses *Stay* ($e = p = 0$) and gets $\pi = \tau w$. In this case π is maximized when $\tau = \hat{\tau}$. □

If *Protest* is never the citizen's best choice, keeping the citizen in the country is better for the government. Also, the government should choose the highest tax rate possible for the same reason as above.

4.5 Concluding Remarks

In this model, people take into consideration their income levels in the home country and overseas as well as the result of the protest against the government when making emigration decisions. The government, the game designer who determines the emigration cost and the tax rate, then faces the balance of stability and economic performance: If emigration is an attractive option for the strategic citizen, then it is more likely that the autocrat can remain in power, but the GDP is low; Otherwise, the GDP is high, but it is more likely that the autocrat loses office.

With other parameters fixed, when *Protest* dominates *Stay* for the citizen and when the government is not able to extract very much from the citizens ($\hat{\tau}$ is small) while losing office is fatal to the incumbent party (L is large), the government wants to open the floodgates to avoid the risk of losing office since losing GDP has little impact on the amount of tax it can collect; In contrast, when the government can take away a big fraction of the GDP ($\hat{\tau}$ is big) and the loss L is small, the government wants all citizens to stay in the country, even though they are going to protest because losing the GDP is too costly while losing office is bearable.

This model can also be extended to allow the government to choose κ, the upfront cost of participating in the protest. However, choosing a bigger κ should be costly for the government. Otherwise, the government can prevent the protest from happening at no cost.

Appendix A Appendix to Chapter 1

A.1 Proof of Proposition 1.1

Proof. In the final period, the government solves

$$\max_{d\in\{0,1\}} V(\tau, d) = (1-d)\left[\tau - \kappa_2\alpha(1-\beta)\right].$$

Therefore, the optimal shutting down strategy is given by

$$d^*(\tau,\alpha,\beta) = \nVdash(\kappa_2\alpha(1-\beta) > \tau) = \nVdash\left(\alpha > \frac{\tau}{\kappa_2(1-\beta)}\right). \tag{A.1.1}$$

Hence the platform's expectation is

$$1-\mathbb{E}(d^*) = Prob\left(\alpha < \frac{\tau}{\kappa_2(1-\beta)}\right) = \min\left\{\frac{\tau}{\kappa_2(1-\beta)}, 1\right\}.$$

In the second period, the platform solves

$$\begin{aligned}\max_{\beta\in[0,1]} \mathbb{E}\left(\Pi(\beta)\right) &= \left(1-\mathbb{E}(d^*)\right)(1-\tau-\kappa_1\beta)\\ &= \begin{cases}\frac{\tau(1-\tau-\kappa_1\beta)}{\kappa_2(1-\beta)}, & \tau \le \kappa_2(1-\beta)\\ 1-\tau-\kappa_1\beta, & \tau \ge \kappa_2(1-\beta)\end{cases}\end{aligned}$$

If $\tau > \kappa_2$, the objective function decreases with β and hence $\beta^* = 0$. Otherwise, the objective function decreases with β if $\beta \ge \tilde{\beta} = 1 - \frac{\tau}{\kappa_2} > 0$. For $\beta < \tilde{\beta}$, the objective function is monotone in β. It increases with β if and only if $\tau \le 1-\kappa_1$. Therefore, the platform's best response function is

$$\beta^* = \begin{cases}\tilde{\beta}, & \tau \le 1-\kappa_1 \text{ and } \tau \le \kappa_2\\ 0, & \text{otherwise.}\end{cases} \tag{A.1.2}$$

The government's expected payoff in the first period from choosing τ is

$$\mathbb{E}V(\tau, d^*) = \begin{cases} \int_0^1 [\tau - \alpha\tau]\, d\alpha = \frac{\tau}{2}, & \tau < 1-\kappa_1 \text{ and } \tau < \kappa_2 \\ \int_0^{\frac{\tau}{\kappa_2}} [\tau - \kappa_2\alpha]\, d\alpha = \frac{\tau^2}{2\kappa_2}, & 1-\kappa_1 < \tau < \kappa_2 \\ \int_0^1 [\tau - \kappa_2\alpha]\, d\alpha = \tau - \frac{\kappa_2}{2}, & \tau > \kappa_2. \end{cases}$$

Therefore the optimal τ is

$$\tau^* = \begin{cases} 1-\kappa_1, & \frac{\bar{\tau}^2}{\kappa_2} < 1-\kappa_1 < \bar{\tau} \\ \bar{\tau}, & \text{otherwise} \end{cases} \tag{A.1.3}$$

The last step is to go through all possible cases of the parameters, $(\kappa_1, \kappa_2, \bar{\tau})$, and find the corresponding τ^*, β^* and $\mathbb{E}(d^*)$ by checking Equation (A.1.3), (A.1.2), and (A.1.1).

- Case 1: If $(1-\kappa_1)\kappa_2 < \bar{\tau}^2$, then $\tau^* = \bar{\tau}$ by Equation (A.1.3). Therefore, by Equation (A.1.2), $\beta^* = 0$. It can be derived from Equation (A.1.1) that $\mathbb{E}V(d^*) = \max\{1-\frac{\bar{\tau}}{\kappa_2}, 0\} \geq 0$. Furthermore, $\mathbb{E}V(d^*) > 0$ if and only if $\bar{\tau} < \kappa_2$.
- Case 2: If $(1-\kappa_1)\kappa_2 > \bar{\tau}^2$, $1-\kappa_1 > \bar{\tau}$ and $\kappa_2 < \bar{\tau}$, then, similarly, $\tau^* = \bar{\tau}$, $\beta^* = 0$, and $\mathbb{E}(d^*) = 0$.
- Case 3: If $1-\kappa_1 > \bar{\tau}$ and $\kappa_2 > \bar{\tau}$, then $\tau^* = \bar{\tau}$, $\beta^* = 1 - \frac{\bar{\tau}}{\kappa_2} > 0$, and $\mathbb{E}(d^*) = 0$ at equilibrium.
- Case 4: If $(1-\kappa_1)\kappa_2 > \bar{\tau}^2$ and $1-\kappa_1 < \bar{\tau}$, then $\tau^* = 1-\kappa_1, \beta^* = 1 - \frac{1-\kappa_1}{\kappa_2} > 0$, and $d^* = 0$.

□

Appendix B Appendix to Chapter 2

B.1 Proof of Lemma 2.1

Proof. By Assumption 2.1, if $\tau > 1$, releasing signals generated by the learning process $\mathcal{L}_\tau$ is the same as releasing an informative signal y. By Bayes' rule, posterior α is given by

$$\alpha = \frac{\eta q(y|G)}{\eta q(y|G) + (1-\eta)q(y|B)},$$

where $q(y|\theta)$ is the density of signal y given Censor's type θ. Therefore,

$$\frac{q(y|G)}{q(y|B)} = \frac{1-\eta}{\eta}\frac{\alpha}{1-\alpha}. \tag{B.1.1}$$

Notice that the left hand side of (B.1.1) is constant as the signal is fixed. Hence any interim belief η and the corresponding posterior belief α must satisfy that $\frac{1-\eta}{\eta}\frac{\alpha}{1-\alpha}$ is constant. If η increases, the corresponding α should increase. If $\tau = 1$, the posterior equals the interior: $\alpha = \eta$. Hence part 1 follows.

Again, by Assumption 2.1, given a fixed interior belief η for any $\tau < \tau'$, the post disappearing at τ' is the same as the post disappearing at τ and then an informative signal y with density $q(y|\theta)$ is released. Then

$$\begin{aligned}
\mathbb{E}_\alpha[\alpha|\tau', \eta, G] &= \mathbb{E}_{\alpha,y}[\frac{\alpha q(y|G)}{\alpha q(y|G) + (1-\alpha)q(y|B)}|\tau, \eta, G] \\
&= \mathbb{E}_\alpha[\mathbb{E}_y[\frac{\alpha q(y|G)}{\alpha q(y|G) + (1-\alpha)q(y|B)}|\tau, \alpha, G]|\tau, \eta, G] \\
&= \mathbb{E}_\alpha[\mathbb{E}_y[\frac{\alpha}{\alpha + (1-\alpha)\frac{q(y|B)}{q(y|G)}}|\tau, \alpha, G]|\tau, \eta, G] \\
&> \mathbb{E}_\alpha[\frac{\alpha}{\alpha + (1-\alpha)\mathbb{E}_y[\frac{q(y|B)}{q(y|G)}|\tau, \alpha, G]}|\tau, \eta, G] \\
&= \mathbb{E}_\alpha[\alpha|\tau, \eta, G],
\end{aligned}$$

where the first line follows from Bayes' rule, the second line from the law of iterated expectations, and the fourth from Jensen's inequality applied to the strictly convex function

$f(z) = \alpha/(\alpha + (1-\alpha)z)$, and the last from the definition of expectations. This proves part 2.

Analogously, part 3 holds because

$$\begin{aligned}
\mathbb{E}_\alpha[\alpha|\tau', \eta, B] &= \mathbb{E}_{\alpha,y}[\frac{\alpha q(y|G)}{\alpha q(y|G) + (1-\eta)q(y|B)}|\tau, \eta, B] \\
&= \mathbb{E}_\alpha[\mathbb{E}_y[\frac{\alpha q(y|G)}{\alpha q(y|G) + (1-\eta)q(y|B)}|\tau, \alpha, B]|\tau, \eta, B] \\
&= \mathbb{E}_\alpha[\mathbb{E}_y[\frac{\alpha \frac{q(y|G)}{q(y|B)}}{\alpha \frac{q(y|G)}{q(y|B)} + (1-\eta)}|\tau, \alpha, B]|\tau, \eta, B] \\
&< \mathbb{E}_\alpha[\frac{\alpha \mathbb{E}_y[\frac{q(y|G)}{q(y|B)}|\tau, \alpha, B]}{\alpha \mathbb{E}_y[\frac{q(y|G)}{q(y|B)}|\tau, \alpha, B] + (1-\eta)}|\tau, \eta, B] \\
&= \mathbb{E}_\alpha[\alpha|\tau, \eta, B],
\end{aligned}$$

where the fourth line holds by applying Jensen's inequality to the strictly concave function $f(z) = \alpha z/(\alpha z + 1 - \alpha)$. □

Appendix C Appendix to Chapter 3

C.1 Proof of Lemma 3.1

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. □

C.2 Proof of Proposition 3.1

A few lemmas are needed to prove Proposition 3.1:

Lemma C.1. *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 C.2. *Given any $w = (w_1, \dots, w_n) \in (0, 1]^n$, in the unique horizontal structure $h^*_{n,P}(w) = (z_0, z_1, \dots, z_n)$, $\forall i \neq j$,*

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

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.

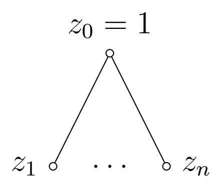
- If $w_i > w_j$, then $Q_i(w_{-i}) > Q_j(w_{-j})$. Because, except for themselves, i and j have the same group of opponents, $\{1, \dots, n\} \setminus \{i, j\}$; While i's another opponent, j, works with a probability lower than j's opponent, i. Thus, $Q_i(w_{-i}) > Q_j(w_{-j})$.
- If $Q_i(w_{-i})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

The three statements complete the circle. □

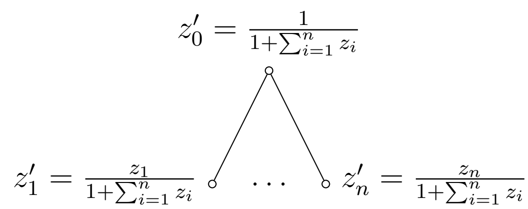
Lemma C.3. *For any $w = (w_1, \ldots, w_n) \in (0,1]^n$, let $m = |\{1 \le i \le n | w_i = w_n\}$ and denote by $\underline{w} = w_{n-m+1} = \ldots = w_n$. The equilibrium total output of the proper horizontal structure corresponding to w, $Y^*_{n,P}(w_1, \ldots, w_{n-m}, \underline{w}, \ldots, \underline{w})$ 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,

$$z_i = \frac{\delta PQ_iV_L}{1 - \delta(1 - PQ_i)}. \tag{C.2.1}$$

Differentiating z_i with respect to Q_i twice gives

$$\frac{\partial z_i}{\partial Q_i} = \frac{(1-\delta)\delta PV_L}{(1 - \delta(1 - PQ_i))^2} > 0 \tag{C.2.2}$$

and

$$\frac{\partial^2 z_i}{\partial Q_i^2} < 0. \tag{C.2.3}$$

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 $i = 1, \ldots, n$,

$$Q_i = \frac{\prod_{j\neq i} w_j}{n} + \frac{\sum_{k\neq i}(1-w_k)\prod_{j\neq i,k} w_j}{n-1} + \dots$$
$$+ \frac{\sum_{j\neq i} w_j \prod_{k\neq i,j}(1-w_k)}{2} + \prod_{k\neq i}(1-w_k)$$

This generates some useful observations about derivatives:

$$\frac{\partial Q_i}{\partial w_j} < 0,\ \forall i \neq j \in \{1,\dots,n\}. \tag{C.2.4}$$

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 $w_1 \geq \dots \geq w_{n-1} > w_n$ (i.e. $m = 1$) first. Notice that

$$\begin{aligned} \frac{\partial Q_n}{\partial w_{n-1}} &= \frac{\partial Q_{n-1}}{\partial w_n} \\ &= -\frac{\prod_{j=1}^{n-2} w_j}{n(n-1)} - \frac{\sum_{k=1}^{n-2}(1-w_k)\prod_{j\neq n,n-1,k} w_j}{(n-1)(n-2)} - \dots - \frac{\prod_{k=1}^{n-2}(1-w_k)}{2\times 1}, \end{aligned} \tag{C.2.5}$$

and for any $i \leq n-2$,

$$\begin{aligned} \frac{\partial^2 Q_i}{\partial w_n \partial w_{n-1}} =& \frac{2\prod_{j\neq n,n-1,i} w_j}{n(n-1)(n-2)} + \frac{2\sum_{k\neq n,n-1,i}(1-w_k)\prod_{j\neq n,n-1,i,k} w_j}{(n-1)(n-2)(n-3)} \\ &+ \dots + \frac{2\prod_{k\neq n,n-1,i}(1-w_k)}{3\times 2\times 1} > 0. \end{aligned} \tag{C.2.6}$$

Equilibrium total output is $Y_n(w_1,\dots,w_n) = \sum_{i=1}^n w_i z_i' = \frac{\sum_{i=1}^n w_i z_i}{1+\sum_{i=1}^n z_i}$. By (C.2.2) and (C.2.4), $\forall i \neq j \in \{1,\dots,n\}$, the denominator decreases in w_n. It then suffices to show the numerator, $F_n(w_1,\dots,w_n) := \sum_{i=1}^n w_i z_i$, 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 (C.2.3) and (C.2.4). 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 (C.2.5) and (C.2.6), it can be verified that the first $(n-2)$ terms on the right-hand side of equation (C.2.5) 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 numerator 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,

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

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

$$Q=\sum_{i=0}^{n-1}\frac{1}{n-i}\binom{n-1}{i}(1-w)^iw^{n-1-i}, \tag{C.2.7}$$

and $\frac{\partial Q}{\partial w}<0$. Equation (C.2.1) and equation (C.2.2) 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,

$$\frac{\partial nwz}{\partial w}=n(z+w\frac{\partial z}{\partial Q}\frac{\partial Q}{\partial w})>n\frac{z}{Q}\left(Q+w\frac{\partial Q}{\partial w}\right). \tag{C.2.8}$$

Hence it suffices to show that $Q + w\frac{\partial Q}{\partial w} > 0$. From equation (C.2.7) it follows that

$$\begin{aligned}\frac{\partial Q}{\partial w} =& \sum_{i=0}^{n-1} \frac{1}{n-i}\binom{n-1}{i}[-i(1-w)^{i-1}w^{n-1-i} + (n-1-i)(1-w)^{i}w^{n-2-i}] \\ =& \sum_{i=0}^{n-1} \frac{1}{n-i}\binom{n-1}{i}(1-w)^{i-1}w^{n-2-i}[-iw + (n-1-i)(1-w)].\end{aligned} \tag{C.2.9}$$

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 (C.2.8) it then follows that $F_n = nwz$ is increasing in w, which finishes the proof. $\square$

Proof of Proposition 3.1. Given any equilibrium strategy profile $w = (w_1, \dots, w_n) \in (0, 1]^n$ and its corresponding *proper* horizontal structure, assuming without loss of generality that $w_1 \geq w_2 \geq \dots w_n$, go through the following process:

Step 1. Substitute w with $(w_1, \dots, w_{n-1}, w_{n-1})$, 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 $w = (w_1, \dots, w_1)$; 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 $w_n^1 := (1, \dots, 1)$ is optimal for the central government. □

C.3 Proof of Proposition 3.2

Proof. I follow the notations used in proof of Lemma C.3. Notice that if $n = 0$, $\Pi = -\alpha$. Now consider the case when $n > 0$. Since optimal horizontal structures are symmetric, for simplicity, let $z(n) = z_1 = \dots = z_n = \frac{\delta P \frac{1}{n} V_L}{1-\delta(1-P\frac{1}{n})}$, 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

$$\max_{n \geq 0} \Pi = \frac{nz(n)}{1 + nz(n)} - \alpha(n+1). \tag{C.3.1}$$

First order condition:

$$\frac{\partial \Pi}{\partial n} = \frac{z(n) + nz'(n)}{(1 + nz(n))^2} - \alpha = 0. \tag{C.3.2}$$

Solving (C.3.2) leads to

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

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

$$\frac{\partial}{\partial P}\left(\frac{nz(P)}{1+nz(P)} - \alpha(n+1\right) = \frac{nz'(P)}{(1+nz(P))^2} = 0. \quad \text{(C.3.3)}$$

Solving (C.3.3) gives

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

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

$$\begin{cases} \hat{n} = n^*(\alpha,\delta,\hat{P}) > 0 \\ \hat{P} = P^*(\delta,\hat{n}) \end{cases}. \quad \text{(C.3.4)}$$

It is easy to verify that if and only if $\alpha \le 1$ and $\alpha(1-\delta+\delta^2)^2 \le (1-\delta)^4$, 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.3.4) implies

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

and

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

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,

$$\frac{d\hat{x}}{d\alpha}=-\frac{[(1-\delta)(\hat{x}+1)^2+\hat{x}^2]^2}{g'(\hat{x})}<0, \tag{C.3.5}$$

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.3.5), $\frac{d\hat{n}}{d\alpha},\frac{d\hat{P}}{d\alpha}<0$. 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 $\tilde{g}(P)$:

$$\tilde{g}(P) = \alpha[\frac{(1-\delta(1-P))^2}{1-\delta} + (\frac{\delta}{1-\delta})^2 P^2]^2 - (1-\delta(1-P)). \tag{C.3.6}$$

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 $\delta \in (0, 1)$. □

C.4 Proof of Proposition 3.3

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

$$z_{0,h} = \frac{1}{1 + n \times \frac{1}{1-\delta(1-P)} \times \frac{\delta\frac{P}{n}}{1-\delta(1-\frac{P}{n})}} \tag{C.4.1}$$

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

The total output of $v^*_{n,P}$ with equilibrium strategy profile $w^1_n = (1, \ldots, 1) \in \mathbb{R}^n$ 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 $i = 1, \ldots, n$, 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

$$z_{0,v}[1+(\frac{\delta P}{1-\delta(1-P)}+\big(\frac{\delta P}{1-\delta(1-P)})^2+\ldots+(\frac{\delta P}{1-\delta(1-P)})^n\big)\frac{1}{1-\delta(1-P)}] = 1. \tag{C.4.2}$$

To show $Y_{V,n} \geq Y_{H,n}$ is equivalent to show $\frac{z_{0,h}}{z_{0,v}} \geq 1$. Plug in (C.4.1) and (C.4.2), it is then equivalent to show

$$\frac{1}{1-\delta(1-P)}(\frac{\delta P}{1-\delta(1-P)} + \ldots + (\frac{\delta P}{1-\delta(1-P)})^n) \geq \frac{\delta P}{1-\delta(1-P)} \frac{1}{1-\delta(1-\frac{P}{n})}$$

$$\Leftrightarrow \frac{1-(\frac{\delta P}{1-\delta(1-P)})^n}{1-\delta} \geq \frac{1}{1-\delta(1-\frac{P}{n})}$$

$$\Leftrightarrow ((1-\delta)+\delta P)^n \geq (\delta P)^{n-1}(n(1-\delta)+\delta P) = (\delta P)^n + n(1-\delta)(\delta P)^{n-1}.$$

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. □

C.5 Proof of Proposition 3.4

Proof. By the previous lemma, all Bureaucrats must be indifferent between working and shirking. Let $g^*_{(n_1,\ldots,n_k),P}$ 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 $n_1 + \ldots + n_k = n$ 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.

□

Bibliography

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.

Boubakri, H. (2013). Revolution and international migration in tunisia. *MPC Research Report No. 4*.

Bueno De Mesquita, E. (2010). Regime change and revolutionary entrepreneurs. *American Political Science Review*, *104*(3), 446–466.

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

Chen, J., & Xu, Y. (2017). Why do authoritarian regimes allow citizens to voice opinions publicly? *The Journal of Politics*, *79*(3), 792–803.

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

Fargues, P. (2017). Mass migration and uprisings in arab countries: An analytical framework. *International Development Policy— Revue internationale de politique de d'veloppement*, *7*(7).

Gallagher, M. E., & Miller, B. (2019). Who not what: The logic of china''s information control strategy. *University of Michigan*.

Gratton, G., Holden, R., & Kolotilin, A. (2018). When to drop a bombshell. *The Review of Economic Studies*, *85*(4), 2139–2172.

Huang, H. (2017). Who wants to leave china? *Journal of East Asian Studies*, *17*(2), 191–213.

Hurun, I. (2016). Hurun report: Chinese investment immigration white paper 2016.

King, G., Pan, J., & Roberts, M. E. (2013). How censorship in china allows government criticism but silences collective expression. *American Political Science Review*, 326–343.

King, G., Pan, J., & Roberts, M. E. (2017). How the chinese government fabricates social media posts for strategic distraction, not engaged argument. *American political science review*, *111*(3), 484–501.

Knockel, J., Crete-Nishihata, M., Ng, J. Q., Senft, A., & Crandall, J. R. (2015). Every rose has its thorn: Censorship and surveillance on social video platforms in china. *5th USENIX Workshop on Free and Open Communications on the Internet (FOCI 15).*

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

Liu, Z. J. (2020). Impact of market structure on regulatory compliance: Evidence from online censorship in china. *Johns Hopkins Carey Business School Research Paper*, (20-13).

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

Miller, B. A. P. (2018). The limits of commercialized censorship in china.

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

Morris, S., & Shin, H. S. (2001). Global games: Theory and applications.

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

Nagy, Z. (2013). Improved speed on intelligent web sites. *Recent Advances in Computer Science*, *1*(14), 215–220.

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

PRC State Taxation Administration. (2017, June 18). *Guojia shuiwu zongju guanyu shishi gaoxin jishu qiye suodeshui youhui zhengce youguan wenti de gonggao* 国家税务总局关于实施高新技术企业所得税优惠政策有关问题的公告*[bulletin 24 issued by state taxation administration on tax incentives for high and new technology enterprises]*. Retrieved October 23, 2022, from http://www.chinatax.gov.cn/chinatax/n362/c6078025/content.html

PRC State Taxation Administration. (2022, May 20). *Ruanjian qiye he jicheng dianlu qiye shuifei youhui zhengce zhiyin* 软件企业和集成电路企业税费优惠政策指引*[guidance* *on tax incentives for software enterprises and integrated circuit enterprises]*. Retrieved October 23, 2022, from http://www.chinatax.gov.cn/chinatax/n810341/n810825/c101434/c5175486/content.html

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.

Roberts, M. E. (2018). *Censored*. Princeton University Press.

Sellars, E. A. (2019). Emigration and collective action. *The Journal of Politics*, *Forthcoming*.

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

Xiang, B. (2016). *Emigration trends and policies in china: Movement of the wealthy and highly skilled*. Migration Policy Institute, Transatlantic Council on Migration.

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