# Decentralization and political career concerns ☆

Jiahua Che[a], Kim-Sau Chung[b,*], Yang K. Lu[c]

[a]Department of Economics, Fudan University, China
[b]Department of Economics, Chinese University of Hong Kong, Hong Kong
[c]Department of Economics, Hong Kong University of Science and Technology, Hong Kong

## ARTICLE INFO

## ABSTRACT

Politicians' career paths often start at some subnational governments and end at the national one. Allocation of authorities among national and subnational governments affects (*i*) how tempting the prospects of taking national offices are, and hence how strong bureaucrats' political career concerns are, and (*ii*) whether the incentives generated by these political career concerns can be put into productive use at subnational governments. We illustrate this tradeoff in determining the optimal degree of decentralization using China as a case study. We also compare the equilibrium degree of decentralization in autocracy and in democracy.

© 2016 Elsevier B.V. All rights reserved.

## 1. Introduction

The literature on federalism has evolved through what Qian and Weingast (1997) and Weingast (2009) refer to as two generations of development. The first generation treats each governmental unit as a benevolent social planner, and examines the costs and benefits of decentralization in terms of scale economies, inter-regional spillovers, heterogeneity across regions, etc. The second generation recognizes the incentive problems of politicians at different governmental units, and examines how the degree and the form of decentralization affect their incentives.

Despite these developments, much of the second-generation literature on federalism assumes that politicians at subnational governments are surrogates of their corresponding regions, and represent their regions (however imperfectly due to incentive problems) when they interact with other politicians to make collective decisions at the national government (Besley and Coate, 2003; Coate and Knight, 2007; Knight, 2004; Luelfesmann et al., 2015). Such an

assumption is appropriate when we study legislators in a democracy (such as senators in the U.S.), but less so when the politicians concerned are administrators, whose political career paths typically start at some subnational governments and end (hopefully) at the national one. For such politicians, their political career concerns are part and parcel of their incentives (Myerson, 2006). In an autocracy like China, where politicians at subnational governments are not subject to electoral checks and balances, such political career concerns are arguably even the dominant, if not the only, forces that incentivize these politicians to perform (Li and Zhou, 2005; Xu, 2011). The omission of political career concerns hence renders the second-generation literature on federalism especially inadequate in studying the degree of decentralization in an autocracy.

Recognizing the importance of political career concerns points us to the following new tradeoff in determining the optimal degree of decentralization. A lower degree of decentralization shifts more responsibilities from subnational governments to the national one. Along with this shift in responsibilities are shifts in various kinds of authorities,[1] which make the prospects of taking national offices more tempting. This strengthens the political career concerns of politicians at subnational governments. However, the resulting

---

---

[1] Throughout this paper, we shall use *authorities* as a catch-all term for all kinds of political resources, power, and authorities that accompany political responsibilities. They, however, should not be confused with people who wield these authorities.

stronger incentives to perform may not be put into productive use when too little decentralization leaves too few authorities for subnational governments. Therefore, too much and too little decentralization can both be counter-productive, and the optimal degree of decentralization lies somewhere in between.

In this paper we provide a simple model to formalize this tradeoff. In our model, politicians at subnational governments are called bureaucrats. In responding to their political career concerns, bureaucrats have at least three options to choose from, one is to genuinely work hard, the other is to cook the books in order to increase the chance of promotion,[2] and the third is to engage in rent-seeking by abusing their authorities. When bureaucrats have these options at their disposal, the degree of decentralization would affect their optimal mix. Too little decentralization leaves too few authorities for bureaucrats to abuse or to put into productive use, and hence they spend most of their time cooking the books. A small increase from this low degree of decentralization would only shift authorities from the more productive national government to these do-nothing subnational governments, and result in further decrease in welfare. A moderate increase in the degree of decentralization, however, can tilt the bureaucrats' optimal mix from cooking the books to doing genuine work. Too much decentralization would backfire again. As the prospects of taking national offices become relatively less tempting, bureaucrats no longer bother to cook the books or to prove themselves through doing work, but would instead spend most of their time engaging in rent-seeking by abusing the enormous authorities entrusted to them.

Political career concerns, like career concerns in the marketplace (Dewatripont et al., 1999; Holmstrom, 1999), incentivizes bureaucrats to perform because performing is a way to prove competence. A corollary is that the degree of decentralization, by affecting the strength of bureaucrats' political career concerns, also affects their incentives to prove their competence and hence the society's ability to select competent politicians into the national government. As such, *today's* degree of decentralization has a first-order effect on the productivity of *tomorrow's* (national) government. This kind of dynamic externality is reflected in our model as well, and does not have a natural counterpart in much of the second-generation literature on federalism.[3]

As we dig deeper into the interplay between decentralization and political career concerns, it becomes clear that there are actually two different degrees of decentralization that jointly interact with bureaucrats' political career concerns. While *today's* degree of decentralization determines whether bureaucrats' incentives to perform can be put into productive use, it is *tomorrow's* degree of decentralization that determines how tempting the prospects of taking national offices are, and hence how strong bureaucrat's political career concerns are. Of course, in a steady state equilibrium, tomorrow's degree of decentralization will be the same as today's. But it is still helpful to conceptually distinguish the two, as doing so helps reveal yet a second kind of dynamic externality. Today's degree of decentralization imposes an externality not only on tomorrow's government, as explained in the previous paragraph, but on *yesterday's* government as well. If there is any exogenous factor that affects today's degree of decentralization, the anticipation of that would

affect yesterday's bureaucrats' political career concerns, and hence their incentives to work. Again, this kind of dynamic externality is reflected in our model, and does not have a natural counterpart in the second-generation literature on federalism.

Perhaps the easiest way to appreciate both kinds of dynamic externalities is to compare two different political regimes, one with the politicians at the national government (called the leaders) choosing today's degree of decentralization, the other with citizen-voters choosing it. We shall call the first regime autocracy, and the second democracy, without pretending that real-life autocracies and democracies differ only in this single aspect. Since leaders have a natural self-serving reason to choose a lower degree of decentralization than preferred by citizen-voters, the difference between the two regimes provides an exogenous factor that affects today's degree of decentralization. We show that, if each generation of citizen-voters care only about themselves and do not internalize externalities imposed on the generations before and after, they would tend to choose too high a degree of decentralization. Autocratic leaders, on the other hand, by self-servingly choosing a lower degree of decentralization than preferred by citizen-voters, would paradoxically partially correct for this problem.

Since political career concerns are an especially important component of bureaucrats' incentives in an autocracy, we build our baseline model with an autocracy like China in mind, and proceed to compare it with a democracy subsequently. To model an autocracy like China, we follow Che et al. (2013, 2014) and build our analysis on an overlapping principal-agent model. The main feature of an overlapping principal-agent model is that today's principal was promoted from among yesterday's agents, and will promote one of today's agents as tomorrow's principal. Che et al. (2013, 2014) argue that such a model captures many important features of an autocracy like China that other models, such as those with an infinitely-lived dictator, cannot.[4]

Our paper bridges two previously disjoint literatures. The first literature concerns how high-power incentives can backfire by distorting the composition of efforts (Acemoglu et al., 2008; Holmstrom and Milgrom, 1991; Milgrom, 1988). This insight has inspired a vast literature, but to our best knowledge it has not been used to study the optimal degree of decentralization.

The second literature concerns normative and positive analyses of (different forms of) federalism. On the normative side, this literature highlights various costs and benefits of decentralization,[5] and suggests how the balance between these costs and benefits determines the optimal degree of decentralization.[6] Our paper introduces a new angle to this literature. We observe that career paths of politicians often start at some subnational governments and end at the national one, and hence decentralization, by affecting the distribution of

---

[2] That cooking the books being a viable option should not be new to those familiar with the history of China's Great Leap Forward, where bureaucrats engaged in massive efforts to fake grain production many times higher than the actual figures, squandering much on the way, and misleading the national government into procuring more grain than peasants could spare based on the false figures, hence triggering the worst man-made famine in history.

[3] A notable exception is Boffa et al. (2016), who study how decentralization affects citizen-voters' ability to monitor politicians. In their model, like ours, today's degree of decentralization also affects the productivity of tomorrow's government by directly affecting how citizen-voters' information is aggregated.

[4] Che et al. (2013) focuses are the bigger variations in economic performance both across autocracies and within individual autocracies vis-à-vis democracies. Che et al. (2014) focuses are incentives at the top of the government, where political career concerns are absent. Neither of them explores the interaction between decentralization and political career concerns.

[5] Among the costs of decentralization are: (i) externalities among subnational governments may lead to suboptimal policies (Break, 1967; Cumberland, 1981; Rivlin, 1992); and (ii) national governments may not be strong enough to protect the market (Blanchard and Shleifer, 2001; Cai and Treisman, 2004, 2005). Among the benefits of decentralization are: (i) subnational governments have informational advantage in providing local public goods (Hayek, 1945); (ii) inter-jurisdictional competition can better match citizens' heterogeneous tastes with menus of local public goods (Tiebout, 1956); (iii) inter-jurisdictional competition can provide market-preserving incentives for subnational governments (Weingast, 1995); and (iv) subnational governments, vis-à-vis national governments, may be even poorer advocates for local interests, especially in developing countries (Bardhan, 2002).

[6] Inman and Rubinfeld (1997) and Oates (1999) provide comprehensive surveys of this literature.

authorities among national and subnational governments, affects the incentives generated by bureaucrats' political career concerns.[7]

On the positive-analysis side, this literature documents many empirical regularities regarding decentralization (Arzaghi and Henderson, 2005; Oates, 1972; Oates and Wallis, 1988; Panizza, 1999). Among them is the finding that the level of democracy is positively correlated with the degree of decentralization, which our analysis confirms.

Finally, our paper is naturally related to the emerging literature on political career concerns in an autocracy like China. Inspired by contract-theoretic models of career concerns (Holmstrom, 1999) and (Dewatripont et al., 1999), many empirical studies test and estimate the extent to which bureaucrats in China are evaluated based on their performances (Chen et al., 2005; Li and Zhou, 2005; Li, 2011; Sheng, 2009; Shi and Zhou, 2007). Our paper takes the existence of political career concerns as given, and instead asks how decentralization affects the incentives generated by these concerns.

The rest of the paper is organized as follows. Section 2 introduces the baseline model, which is built with an autocracy like China in mind. Section 3 studies how decentralization affects the incentives generated by bureaucrats' political career concerns. We also use China as a case study to illustrate our main results. Section 4 solves the equilibrium degree of decentralization in an autocratic regime where leaders choose the degree of decentralization, as well as in a democratic regime where citizen-voters choose the degree of decentralization. We then use the comparison between the two regimes to illustrate the two kinds of dynamic externalities explained in this Introduction. Section 5 concludes.

## 2. The baseline model

This is an overlapping-generations model. Every generation lives for two periods: when they are young and old, respectively. There are many citizens in each generation. Each citizen's type is drawn independently from an identical atomless distribution with mean $\bar{z}$ and variance $\sigma_z^2$. The distribution is commonly known, but the type drawn is unknown to anyone, including the citizen himself.

### 2.1. Technology

There is one government, consisting of a leader and two bureaucrats. Bureaucrats are randomly drawn from young citizens. When they turn old, one of the two bureaucrats will be promoted to be the next leader. The one who does not get promoted will become one of the ordinary citizens. A leader is hence necessarily old.

We can think of the leader as managing the national government, while each bureaucrat managing a separate subnational government. Decentralization is measured by how authorities are distributed between the national and the subnational governments. For concreteness, we imagine that there is a unit measure of authorities to be distributed between these different governments in each period. Let $x$ be the (same) measure of authorities entrusted to each subnational government, and $1 - 2x$ is the corresponding measure to the national government.

Each bureaucrat is endowed with one unit of time. He divides that unit into three different activities: he can spend time to manage his subnational government (*production*), he can spend time to cook the books (*manipulation*), or he can spend time to engage in rent-seeking (*embezzlement*). Let $p_i$, $m_i$, and $e_i$ stand for the amounts of time bureaucrat $i$ spends in these three activities, respectively, with $p_i + m_i + e_i \leq 1$.

A subnational government either performs or not. The probability that subnational government $i$ performs depends on how much time, $p_i$, bureaucrat $i$ spends in managing it, as well as his competence, $z_i$. Specifically, it performs with probability $p_i z_i$, and does not perform with probability $1 - p_i z_i$.

When subnational government $i$ performs, the amount of public goods it provides, $g_i$, is proportional to the measure of authorities entrusted to it, $x$. For simplicity, we let $g_i = bx$, where $b > 0$ is a productivity parameter. When it does not perform, $g_i = 0$.

Even when a subnational government does not perform, it may look as if it performs if the corresponding bureaucrat cooks the books. This was what happened during China's Great Leap Forward. To capture this idea, we imagine that the actual performance of subnational government $i$ is not (yet) observable at the time when bureaucrat $i$ is up for promotion. All what the society can observe is a noisy signal, $s_i$, of the actual performance that is subject to manipulation. Specifically, subnational government $i$ is *perceived* to have performed ($s_i = 1$) if it actually has performed, but is still perceived to have performed with some probability even if it actually has not. We assume that the resulting marginal distribution of $s_i$ is $s_i = 1$ with probability $p_i z_i + m_i$, and $s_i = 0$ with probability $1 - (p_i z_i + m_i)$. If bureaucrat $i$ does not spend any time in manipulation (i.e., if $m_i = 0$), the signal would be accurate.

Finally, by spending time engaging in rent-seeking, a bureaucrat can generate some private benefit. The private benefit generated is increasing in the measure of authorities entrusted to him. Specifically, we assume that the private benefit generated for bureaucrat $i$ is $xe_i$.

At the end of a period, the bureaucrat whose subnational government has a higher *perceived*, but not actual, performance will be promoted as the next leader.[8] Ties are broken randomly. At the beginning of a period, the newly promoted leader chooses the degree of decentralization, or equivalently the measure of authorities, $x$, entrusted to each subnational government.

The leader's only action in this model is to choose $x$. We do not explicitly study his incentives to produce/embezzle; after all, he is already at the top of the governmental hierarchy, and his behavior is no longer driven by political career concerns. We hence treat the national government as a black box, which mechanically generates certain amount of public goods for the society and rent for the leader. Specifically, we assume that it generates $G = \beta(1 - 2x) + \lambda z^L$ units of public goods, where $z^L$ is the competence of the leader, and $\beta, \lambda > 0$ are productivity parameters, with $\beta$ being the leader-analog of $b$. We choose the additive functional form to ease our analysis.[9] We also assume that the national government generates $(1 - 2x)\epsilon$ units of private benefit for the leader, with the exogenous parameter $\epsilon > 0$ being the leader-analog of $e_i$. Note that both the public goods and the private benefit generated are increasing in the measure of authorities entrusted to the national government, $(1 - 2x)$.

### 2.2. Payoffs and welfare

In any given period, a generic citizen's payoff is simply the sum of all public goods generated:

$$W(z^L) = (g_i + g_j + G)\mu,$$

---

[7]  We, however, do not explain why career paths often start at the subnational level. Future researches may seek to explain this by identifying special features associated with national and subnational governments, respectively.

[8]  In this model, the sitting leader occupies the top office until he dies, and hence does not have a strict preference over who to promote. We can think of the promotion rule in the text as the sitting leader breaking ties in the interests of the society, who would prefer a merit-based promotion rule given that the next leader's competence affects the national government's public good provision in the next period.

[9]  Specifically, the additive functional form ensures that a leader's equilibrium choice of $x$ would not depend on what he has learned about his own competence, $z^L$, in the previous period. This dramatically simplifies our analysis.

where $\mu > 0$ is the marginal utility of public goods, $g_i$ and $g_j$ are the public goods generated by the two subnational governments, and $G$ is the public goods generated by the national government. We wrote $W$ as a function of $z^L$ to emphasize its dependence on, among other things, the current leader's competence. Naturally, $W$ is also our measurement of welfare.

The leader and bureaucrats are also citizens, and hence also consume public goods. On top of that, they also consume rent. Therefore, the leader's payoff is

$$V(z^L) = W(z^L) + (1 - 2x)\epsilon,$$

and bureaucrat $i$'s current-period payoff is

$$U_i(z^L) = W(z^L) + xe_i - \frac{1}{2}p_i^2,$$

where the last item is the disutility of spending time doing real work.

If $\alpha_i$ is bureaucrat $i$'s probability to get promoted, then his lifetime payoff is $U_i(z^L) + \alpha_i V(z_i) + (1 - \alpha_i)W(z_j)$, where $V(z_i)$ is his next-period payoff as a leader, and $W(z_j)$ is his next-period payoff as a citizen while the other bureaucrat with competence $z_j$ serves as the leader. Note that, for simplicity, we assume no discounting.

For the reader's easy reference, we summarize the notation used in this model in Table 1.

### 2.3. Timing

We summarize the timing in each period as follows:

1. The new leader chooses the decentralization level, $x$, for this period.
2. Two bureaucrats, $i$ and $j$, are randomly drawn from the young citizens. Upon seeing $x$, they simultaneously choose $(p_i, m_i, e_i)$ and $(p_j, m_j, e_j)$.
3. Signals $s_i$ and $s_j$ are observed. The bureaucrat with the higher signal is promoted to be the next leader. Ties are broken randomly.
4. Public goods $g_i$, $g_j$, and $G$ are generated. Every citizen consumes public goods, leader and bureaucrats also consume rent.

### 3. Static analysis

In this section, we temporarily treat the degree of decentralization level as exogenous, and investigate how a bureaucrat would

**Table 1**
Notations.

| Notations | Meanings |
|---|---|
| $b$ | A bureaucrat's ability in translating authorities into public goods |
| $e$ | A bureaucrat's time spent in embezzlement |
| $g$ | Public goods generated by a subnational government |
| $i, j$ | Subnational governments and corresponding bureaucrats |
| $m$ | A bureaucrat's time spent in manipulation |
| $p$ | A bureaucrat's time spent in production |
| $s$ | Perceived performance of a subnational government |
| $x$ | Degree of decentralization |
| $z$ | Competence type |
| $\alpha$ | A bureaucrat's probability to get promoted |
| $\beta$ | Leader-analog of $b$ |
| $\epsilon$ | Leader-analog of $e$ |
| $\lambda$ | Productivity parameter associated with leader's competence |
| $\mu$ | Marginal utility of public goods |
| $G$ | Public goods generated by the national government |
| $U$ | A bureaucrat's payoff |
| $V$ | The leader's payoff |
| $W$ | A generic citizen's payoff |

divide his time between production, manipulation, and embezzlement. This analysis would shed light on how decentralization and political career concerns interact in shaping a bureaucrat's incentives. Throughout this section, we use $x$ and $x'$ to denote the current- and the next-period degrees of decentralization, respectively, which the bureaucrat treats as exogenously given.

### 3.1. Promotion probability

In equilibrium, bureaucrat $i$ knows the strategy $(p_j, m_j, e_j)$ of his opponent bureaucrat $j$, as well as the strategies of all future bureaucrats. This allows him to calculate the probability that his opponent's subnational government being perceived as non-performing, conditional on his opponent's competence being $z_j$:

$$\Pr[s_j = 0 | z_j] = 1 - (p_j z_j + m_j).$$

Bureaucrat $i$'s probability of getting promoted, conditional on the unknown competence $z_i$ and $z_j$, is hence

$$\begin{aligned}
\alpha(z_i, z_j) =& \Pr[s_i = 1 | z_i] \left( \Pr[s_j = 0 | z_j] + \frac{1}{2} \Pr[s_j = 1 | z_j] \right) \\
& + \Pr[s_i = 0 | z_i] \left( \frac{1}{2} \Pr[s_j = 0 | z_j] \right) \\
=& \frac{1}{2} \Pr[s_i = 1 | z_i] + \frac{1}{2} \Pr[s_j = 0 | z_j] \\
=& \frac{1}{2}(p_i z_i + m_i) + \frac{1}{2} \Pr[s_j = 0 | z_j].
\end{aligned}$$

Note that, due to the simplicity of our model, the marginal impact of $p_i$ and $m_i$ on $\alpha(z_i, z_j)$ does not depend on $\Pr[s_j = 0 | z_j]$, and hence does not depend on the opponent's strategy at all. We shall make heavy use of this property below.

### 3.2. Bureaucrat $i$'s problem

Bureaucrat $i$'s problem is

$$\max_{p_i, m_i, e_i} \quad E\left[ U(z^L) + \alpha(z_i, z_j)V(z_i) + (1 - \alpha(z_i, z_j))W(z_j) \right]$$

subject to $\quad p_i, m_i, e_i \geq 0, \quad p_i + m_i + e_i \leq 1,$

where the expectation is taken over $(z_i, z_j)$.

The marginal benefits of $p_i$, $m_i$, and $e_i$ are, respectively,

$$\begin{aligned}
R_p(x|x') &= E\left[ \frac{\partial U(z^L)}{\partial p_i} + \frac{\partial \alpha(z_i, z_j)}{\partial p_i}(V(z_i) - W(z_j)) \right] \\
&= \bar{z}bx\mu - p_i + \frac{1}{2}\left[ \lambda\sigma_z^2\mu + \bar{z}(1 - 2x')\epsilon \right] \\
&\leq \bar{z}bx\mu + \frac{1}{2}\left[ \lambda\sigma_z^2\mu + \bar{z}(1 - 2x')\epsilon \right] =: R_{p=0}(x|x'), \\
R_m(x|x') &= E\left[ \frac{\partial \alpha(z_i, z_j)}{\partial m_i}(V(z_i) - W(z_j)) \right] = \frac{1}{2}(1 - 2x')\epsilon,
\end{aligned}$$

and $\quad R_e(x|x') = E\left[ \frac{\partial U(z^L)}{\partial e_i} \right] = x.$

Note that the marginal benefit of manipulation is a constant independent of the current-period degree of decentralization, $x$, because the whole purpose of manipulation is to increase one's chance to get promoted, and hence its marginal benefit depends on the next-period degree of decentralization, $x'$, only. The marginal benefits of production and of embezzlement, on the other hand, are increasing in $x$, because more authorities enable the bureaucrat to both produce more public goods (which he also enjoys as a citizen) and obtain more rent. Therefore, there exists some threshold $x_1$ (respectively,

$\hat{x}$), which depends on $x'$, such that when $x < x_1$ (respectively, when $x < \hat{x}$), there are too few authorities entrusted to this bureaucrat, so much so that he would rather manipulate than produce (respectively, rather manipulate than embezzle). For larger $x$, how the bureaucrat divides his time between production and embezzlement depends on their respective marginal benefits. Assumption 1 below guarantees that the marginal benefit of embezzlement increases in $x$ faster than that of production, and hence when there are too many authorities entrusted to this bureaucrat, he would rather abuse these authorities than put them into productive use.

Formally, define thresholds $\hat{x}(x')$, $x_1(x')$, and $x_2(x')$ implicitly by

$$R_m(\hat{x}(x')|x') = R_e(\hat{x}(x')|x'),$$
$$R_m(x_1(x')|x') = R_{p=0}(x_1(x')|x'),$$
and $\quad R_{p=0}(x_2(x')|x') = R_e(x_2(x')|x'),$

respectively. Simple algebra shows that these thresholds are

$$\hat{x}(x') = \frac{1}{2}(1 - 2x')\epsilon, \tag{1}$$

$$x_1(x') = \frac{(1 - \bar{z})(1 - 2x')\epsilon - \lambda\sigma_z^2\mu}{2\bar{z}b\mu}, \tag{2}$$

$$\text{and} \quad x_2(x') = \frac{\bar{z}(1 - 2x')\epsilon + \lambda\sigma_z^2\mu}{2(1 - \bar{z}b\mu)}, \tag{3}$$

respectively. The following assumption guarantees that there exists some moderate degree of decentralization such that bureaucrat $i$ will find it optimal to spend some time in production,[10] which we shall maintain throughout this paper without further mentioning.

**Assumption 1.** $\bar{z}b\mu < 1 < \bar{z}b\mu + \bar{z}.$

Given Assumption 1, bureaucrat $i$ should set $e_i = 0$ when $x < \hat{x}(x')$, set $m_i = 0$ when $x > \hat{x}(x')$, and set $p_i = 0$ when either $x < x_1(x')$ or $x > x_2(x')$. When $x \in [x_1(x'), \hat{x}(x')]$, bureaucrat $i$ should set $p_i = min\{p_1(x|x'), 1\}$, where

$$p_1(x|x') := \bar{z}bx\mu + \frac{1}{2}\left[\lambda\sigma_z^2\mu - (1 - \bar{z})(1 - 2x')\epsilon\right]$$

equates $R_p$ and $R_m$, and is linearly increasing in $x$. When $x \in [\hat{x}(x'), x_2(x')]$, bureaucrat $i$ should set $p_i = min\{p_2(x|x'), 1\}$, where

$$p_2(x|x') := \frac{1}{2}\left[\lambda\sigma_z^2\mu + \bar{z}(1 - 2x')\epsilon\right] - x(1 - \bar{z}b\mu) \tag{4}$$

equates $R_p$ and $R_e$, and is linearly decreasing in $x$.

We shall make two further assumptions to ease our dynamic analysis in Section 4; both in effect bound $\sigma_z^2$ from above. The first is to guarantee that bureaucrat $i$ never spends all of his time producing.[11] The second is to guarantee that bureaucrat $i$ spends all of his time embezzling when there is complete decentralization.[12] We shall maintain these assumptions throughout this paper without further mentioning.

**Assumption 2.**

1 $\lambda\sigma_z^2\mu < 2 - (\bar{z}(1 + b\mu) - 1)\epsilon;$
2 $\lambda\sigma_z^2\mu < 1 - \bar{z}(b\mu + \epsilon).$

We summarize bureaucrat $i$'s optimal strategy given Assumptions 1 and 2 as follows.

**Proposition 1.** *Bureaucrat $i$'s optimal strategy is*[13]

$$(p_i, m_i, e_i) = \begin{cases} (0, 1, 0) & \text{when } x \in [0, x_1(x')) \\ (p_1(x|x'), 1 - p_1(x|x'), 0) & \text{when } x \in [max\{x_1(x'), 0\}, \hat{x}(x')) \\ (p_2(x|x'), 0, 1 - p_2(x|x')) & \text{when } x \in [\hat{x}(x'), x_2(x')) \\ (0, 0, 1) & \text{when } x \in [x_2(x'), 1/2] \end{cases}.$$

Bureaucrat $i$'s optimal strategy is illustrated in Fig. 1. When very few authorities are entrusted to his subnational government, there is not much room for rent-seeking, and there is also not much good his subnational government can do even if it is managed well. By spending time to manage his subnational government, he can boost his promotion probability, but even that can be more easily achieved by cooking the books. His optimal strategy is hence to spend all of his time cooking the books. On the other extreme, when lots of authorities are entrusted to his subnational government, the gain from spending time engaging in rent-seeking becomes too big, and he ends up spending all of his time engaging in rent-seeking. Only when the degree of decentralization level is moderate would neither perverse incentive prevail. Given any $x' \in [0, 1/2]$, the time bureaucrat $i$ spends on managing his subnational government is single-peaked in $x$, and Assumption 1 guarantees that there exists some $x$ such that he spends positive amount of time doing so.

Finally, from Eqs. (1)–(3), we can also see that the thresholds $x_1$, $\hat{x}$, and $x_2$ in Fig. 1 are all increasing in $(1 - 2x')$. This is quite intuitive. Recall that $x'$ is the next-period degree of decentralization, and hence $(1 - 2x')$ is the measure of authorities entrusted to the next-period leader. The larger is $(1 - 2x')$, the more private benefit, $(1 - 2x')\epsilon$, the next-period leader can collect, and hence the stronger the political career concerns of the current-period bureaucrats are. When political career concerns are stronger, bureaucrats allocate more time working vis-à-vis engaging in rent-seeking (because working help increase their probabilities to get promoted), and allocate more time cooking the books vis-à-vis working.

### 3.3. A case study: China

Although our analysis has been one with a partial-equilibrium nature so far, it already sheds light on why political career concerns have been accused as a culprit of the worst man-made famine in history (Kung, 2014; Kung and Chen, 2011), while at the same time also being credited as the engine of the most spectacular growth miracle (Li and Zhou, 2005; Xu, 2011).

Throughout the 65-year reign of the Chinese Communist Party, China has experienced two of the most extreme historical events. The Chinese great famine (1959–1961), caused by the extreme policy of Great Leap Forward, was estimated to have claimed millions of victims. Mortality estimates range from 16.5 million (Coale, 1981) to 30 million (Banister, 1987) to 45 million (Dikotter, 2010). Then, barely 20 years later, the same political institution unleashed 30 consecutive years of breathtaking economic growth. According to the National Bureau of Statistics of China, annual GDP growth rate in China averages around 10% from 1979 to 2010. Both extreme historical events have been attributed by different scholars to the political career concerns that pre-dominantly shape the incentives of Chinese bureaucrats. For example, Kung and Chen (2011), find that radical policies motivated by political career concerns can explain 16.83% of the excess death rate during the Chinese great famine.

---

[10] Technically, the assumption guarantees that $x_1(x') < \hat{x}(x') < x_2(x')$ for any $x' \in [0, 1/2]$ and $\sigma_z^2 > 0$.
[11] Technically, the assumption guarantees that, at $x = \hat{x}(x')$, $p_1(x|x')$ (which is the same as $p_2(x|x')$) remains smaller than 1 for any $x' \in [0, 1/2]$.
[12] Technically, the assumption guarantees that $x_2(x') < 1/2$ for any $x' \in [0, 1/2]$.

[13] For any $x' \in [0, 1/2]$, $\hat{x}(x')$ must be nonnegative, and $x_2(x')$ is positive given Assumption 1. However, $x_1(x')$ may be negative.
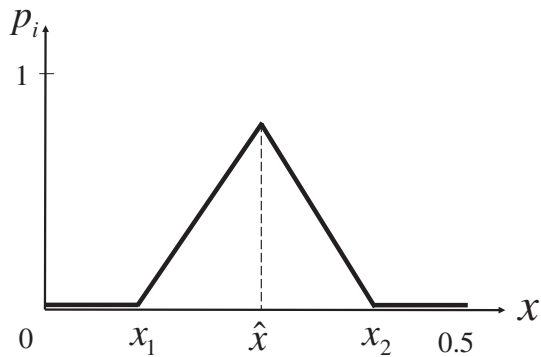
**Fig. 1.** Bureaucrat $i$'s optimal $p_i$ as a function of $x$, given a fixed $x'$.
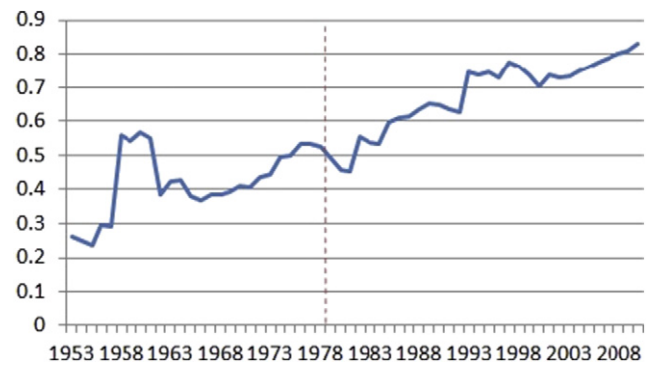


**Fig. 2.** Total subnational government expenditures as a share of total government expenditures in China from 1953 to 2008. Vertical dotted line indicates the year of economic reform.
Source: Source: National Bureau of Statistics of China

Findings of Kung (2014) further support this conclusion. On the other hand, Li and Zhou (2005), for example, provide empirical evidence that political career concerns help motivated bureaucrats to maximize local economic performances.[14] Apparently, the same political career concerns can generate vastly different kinds of incentives for bureaucrats under different environments.

Fig. 2 plots total subnational government expenditures as a share of total government expenditures, an admitted crude approximation of the degree of decentralization, in China from 1953 to 2008. We can see that there was a steady increase in the degree of decentralization, except for the abrupt jump that signifies the Great Leap Forward, which caused the Chinese great famine.

We can think of the policy of Great Leap Forward as being introduced against the backdrop of an extremely low degree of decentralization—an $x$ that is closer to 0 than to $x_1$ in Fig. 1. When $x$ is at such a low level, bureaucrats spent most of their time cooking the books rather than producing or embezzling. There was almost no complaint about corruption during that time, but at the same time few local public goods were produced. When the Great Leap Forward came, the degree of decentralization jumped up, but was still not as high as the average degree in the post-1978 era. This brought $x$ closer to $x_1$, but did not change the fact that bureaucrats were spending most of their time cooking the books. The small increase in the degree of decentralization only shifted authorities from the more productive national government to the do-nothing subnational governments, and resulted in further decrease in welfare.

As the degree of decentralization steady increased to a level even beyond that achieved during the Great Leap Forward (the right half of Fig. 2), $x$ was shifted further rightward, perhaps into the range in between $\hat{x}$ and $x_2$ in Fig. 1. In this range of $x$, bureaucrats no longer spend their time cooking the books, but instead splitting their time doing genuine work and engaging in rent-seeking. There were many more complaints about corruption during the post-1978 era, while at the same time bureaucrats were also credited for doing much good work in promoting economic growth.

This account of the history is of course ultra-simplistic, focusing on a single dimension and ignoring many other differences between pre- and post-1978 China. However, to our best knowledge, this is the first unified theory that explains why the same political career concerns might be responsible for the vastly different kinds of incentives for bureaucrats in pre- and post-1978 China.

## 4. Decentralization in autocracy and in democracy

In the last section, we study a bureaucrat's incentives to work vis-à-vis to cook the books or to engage in rent-seeking, holding the current-period and the next-period degrees of decentralization, $x$ and $x'$, as exogenously given. In any steady-state equilibrium, $x'$ will of course be the same as $x$, and both will be determined endogenously. In this section, we shall compare the endogenously determined equilibrium $x$ in autocracy and in democracy. The main result of this comparison is that the equilibrium degree of decentralization is higher in democracy than in autocracy. More surprisingly, bureaucrats work harder in autocracy than in democracy, and average competence of an autocratic leader is higher than that of a democratic leader. However, the welfare comparison between autocracy and democracy remains ambiguous.

One purpose of this exercise is to show that our model can deliver theoretical predictions that match the empirical findings in the literature on federalism (see Subsection 4.1 below). More importantly, this exercise brings to light two kinds of dynamic externalities that we think are natural in any model studying the interaction between decentralization and political career concerns. When citizens-voters choose the current-period degree of decentralization, $x$, they generate externalities for both next-period and last-period citizens. The higher $x$ they choose lowers the current-period bureaucrats' incentives to work and hence lowers the expected competence of the next-period leader, generating a negative externality for next-period citizens. More subtly, the rational expectations of this higher $x$ also weaken the political career concerns of last-period bureaucrats, and hence weakens their incentives to work, generating a negative externality for last-period citizens as well. In an autocracy, leaders' selfish choice of a lower $x$ (due to their rent-seeking motives) partially corrects these negative externalities.

One caveat of our exercise is that our comparison between autocracy and democracy makes sense only when a steady-state equilibrium indeed exists, which however is not guaranteed. Fortunately, the simplicity of our model allows us to characterize precisely the conditions for existence of a steady state equilibrium. Our comparison results will be stated under these conditions.

Since our baseline model is built with an autocracy like China in mind, we start our analysis of autocracy in this section, and continue with an analysis of democracy by slightly modifying our baseline model.

In an autocracy, the current-period leader chooses the current-period degree of decentralization, $x$, given the common knowledge (among the leader and his two bureaucrats) that the next-period

---

[14] Other related studies include Maskin et al. (2000), Cai and Treisman (2006), and Landry (2008).

degree of decentralization will be $x'$. Ignoring terms that do not depend on $x$, the leader's problem is

$$\max_{x \in [0,1/2]} \quad \bar{V}(x|x') := [2p(x|x')\bar{z}bx + \beta(1-2x)]\mu + (1-2x)\epsilon.$$

In the leader's objective function, the second term is his rent, which increases linearly in the measure of authorities entrusted to him, $1-2x$. The term in the square brackets is the expected amount of total public goods generated, which contains two parts. The first part is the expected amount of public goods generated by his two bureaucrats; in particular, the term $p(x|x')$ is the time each bureaucrat spends on production, and has the expression of

$$p(x|x') = \begin{cases} 0 & \text{when} \quad x \in [0, x_1(x')) \\ p_1(x|x') & \text{when} \quad x \in [\max\{x_1(x'), 0\}, \hat{x}(x')) \\ p_2(x|x') & \text{when} \quad x \in [\hat{x}(x'), x_2(x')) \\ 0 & \text{when} \quad x \in [x_2(x'), 1/2] \end{cases} \quad (5)$$

according to Proposition 1. The second part is the amount of public goods generated by the leader himself, which also increases linearly in the measure of authorities entrusted to him.

Fig. 3 illustrates the typical shape of $\bar{V}(\cdot|x')$. Given any $x' \in [0, 1/2]$, there are as many as four regions. In both the leftmost and the rightmost regions, $\bar{V}(\cdot|x')$ is linearly decreasing in $x$. Those are the regions where his bureaucrats spend no time on production. $\bar{V}(\cdot|x')$ is convex in the middle-left region; this is the region where an increase in the degree of decentralization increases both a bureaucrat's productivity as well as the time he spends in production. $\bar{V}(\cdot|x')$ is concave in the middle-right region; this is the region where an increase in the degree of decentralization increases a bureaucrat's productivity but decreases the time he spends in production. Apparently, the leader's optimal choice of $x$ must either equal to 0 or lie in the middle-right region.

Imagine that we extend the concave part of $\bar{V}(\cdot|x')$ beyond the middle-right region. Specifically, consider the following fictitious payoff function:

$$\hat{V}(x|x') := [2p_2(x|x')\bar{z}bx + \beta(1-2x)]\mu + (1-2x)\epsilon,$$

which is defined over all $x \in \mathbb{R}$ instead of only over $x \in [\hat{x}(x'), x_2(x'))$. Since $p_2(\cdot|x')$ is linear, $\hat{V}(\cdot|x')$ is quadratic. Denote the unique maximum of $\hat{V}(\cdot|x')$ by $x_{IL}(x')$.[15] It can also be shown that $\partial \hat{V}(x_2(x')|x')/\partial x < 0$ and hence $x_{IL}(x') < x_2(x')$ for any $x' \in [0, 1/2]$. The leader's optimal choice of $x$ is hence either 0 or $\max\{\hat{x}(x'), x_{IL}(x')\}$.

**Proposition 2.** *There exists a threshold $\tilde{x}_L$ for next period's degree of decentralization, $x'$, such that the current period's optimal degree of decentralization, $x_L^*$, is strictly positive (i.e., "interior") only if $x'$ is lower than the threshold $\tilde{x}_L$. Specifically, define*

$$\tilde{x}_L := \frac{1}{2}\left(1 - \frac{2(\beta + \epsilon/\mu) - \lambda\sigma_z^2 \bar{z}b\mu}{(\bar{z}b\mu + \bar{z} - 1)\epsilon \bar{z}b}\right). \quad (6)$$

*The leader's optimal choice of $x$ is*

$$x_L^*(x') = \begin{cases} \max\{\hat{x}(x'), x_{IL}(x')\} & \text{if} \quad x' < \tilde{x}_L \\ \{0, \max\{\hat{x}(x'), x_{IL}(x')\}\} & \text{if} \quad x' = \tilde{x}_L \ . \\ 0 & \text{if} \quad x' > \tilde{x}_L \end{cases}$$



**Fig. 3.** The leader's payoff as a function of $x$, given a fixed $x'$.

*Moreover, $x_L^*(x')$ is continuous and strictly decreasing in $x'$ for $x' < \tilde{x}_L$.[16],[17]*

The proof of Proposition 2 is in the Online Appendix. Here we sketch the intuition. Being in his last period of life, the leader does not directly care about next period's degree of decentralization, $x'$. He cares about $x'$ only to the extent that it affects his bureaucrats' incentives to work. The larger $x'$ is, the lower is the gain from being promoted as the next-period leader, and hence the weaker are the bureaucrats' political career concerns. As the bureaucrats become less motivated to work, the leader finds it less desirable to entrust authorities to the subnational governments. Therefore, the leader's optimal decentralization level, $x_L^*$ is strictly decreasing in $x'$. When $x'$ is sufficiently large, so much so that it surpasses a certain threshold, $x_L^*$ drops discontinuously down to 0. This discontinuous drop is due to the non-monotonicity of the leader's payoff function (see Fig. 3), which renders an interval of $x$ immediately above 0 especially unpalatable to the leader. If the leader ever chooses an $x$ in that interval, his bureaucrats' incentives would become so perverse that they would spend most of their time cooking the books instead of working.

Since $x_L^*(x')$ is weakly decreasing in $x'$, and strictly so when it is strictly positive, its graph can cross the 45-degree line at most once. But it may not intersect with the 45-degree line at all because of the discontinuity. Fig. 4 depicts the case when the graph of $x_L^*(x')$ does cross the 45-degree line. The point of crossing will be the steady-state equilibrium degree of decentralization.

**Definition 1.** A rational expectation equilibrium (or simply an equilibrium) is a sequence, $(x_t)_{t=0}^\infty$, of degrees of decentralization such that, $\forall t \geq 0$, $x_t = x_L^*(x_{t+1})$. A steady-state equilibrium is an equilibrium with $x_0 = x_1 = x_2 = \cdots$. A non-zero steady-state is a steady-state equilibrium with $x_0 > 0$.

Our comparison between autocracy and democracy makes sense only when a non-zero steady-state equilibrium exists. The following proposition precisely characterizes when a non-zero steady-state equilibrium exists in autocracy.

**Proposition 3.** *A non-zero steady-state equilibrium exists in autocracy if and only if*

$$\tilde{x}_L \geq \frac{\epsilon}{2(1+\epsilon)},$$

---

[15] For an explicit expression of $x_{IL}(x')$, see Eq.(8) in the Online Appendix. The subscript $L$ reminds us that it is the leader who chooses the degree of decentralization, $x$. (We shall study shortly the case where the citizen-voters instead choose $x$.) The subscript $I$ reminds us that $x_{IL}$ is the "interior" candidate for the optimal $x$. The optimal $x$ needs not be interior, but if it is interior then it coincides with $x_{IL}$ (provided that $x_{IL} \geq \hat{x}$).
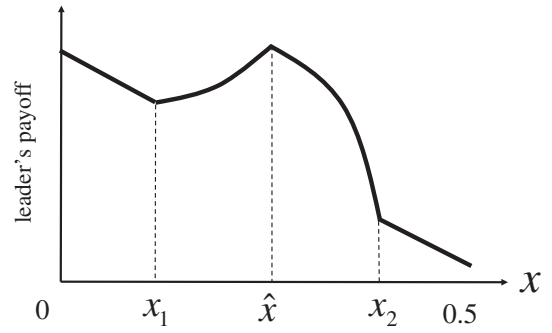
[16] Note that $x_L^*(\cdot)$ is strictly speaking a correspondence. We abuse notation and write it as a function whenever it is single-valued.

[17] Note that $\tilde{x}_L$ as defined in Eq. (6) can lie beyond the interval $[0, 1/2]$, in which case $x_L^*(x')$ is single-valued for every $x' \in [0, 1/2]$.
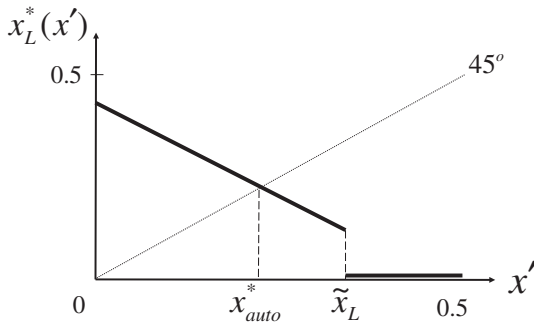
**Fig. 4.** Optimal degree of decentralization, $x_L^*$ as a function next period's degree of decentralization, $x'$.

where $\tilde{x}_L$ is given by Eq. (6). When a non-zero steady-state equilibrium exists, it is the only steady-state equilibrium.

According to Eq. (6), $\tilde{x}_L$ is a function of seven parameters (including $\epsilon$). Therefore, the inequality in Proposition 3 is a joint condition on seven parameters. In the Online Appendix, we provide an intuitive explanation of why and how each of these seven parameters affects the existence of a non-zero steady state equilibrium. We refer the interested reader to the Online Appendix for details.

We now turn to democracy. One of the simplest ways to modify our model to study the equilibrium degree of decentralization in democracy is to have citizen-voters, instead of the leader, choose the degree of decentralization, $x$, at the beginning of a period.[18] In our model, the sizes of young and old citizen-voters are always the same, and hence the identity of the median voter is ambiguous. For all practical purposes, however, it is more reasonable to assume that the median voter is an old citizen, because it is a robust empirical fact that old citizens are more active voters than young citizens are.

The subsequent analysis is exactly the same as that in the case of autocracy. Let $x'$ be the next-period degree of decentralization, which is common knowledge in the current period. Ignoring terms that do not depend on $x$, the citizen-voters' problem is

$$\max_{x\in[0,1/2]} \quad \bar{W}(x|x') := [2p(x|x')\bar{z}bx + \beta(1-2x)]\mu,$$

where $p(x|x')$ is the time each bureaucrat spends in production, and has the same expression as in Eq. (5). The typical shape of $\bar{W}(\cdot|x')$ is similar to that of $\bar{V}(\cdot|x')$ in Fig. 3, except that the slope is less negative or more positive due to the missing term $(1-2x)\epsilon$. Define the fictitious payoff function:

$$\hat{W}(x|x') := [2p_2(x|x')\bar{z}bx + \beta(1-2x)]\mu,$$

and denote the unique maximum of $\hat{W}(\cdot|x')$ by $x_{IC}(x')$.[19] We then have the following analogs of Propositions 2 and 3. Since the proofs are exactly the same as before, they are omitted.

---

[18] There are, of course, many other differences between a democracy and an autocracy that we abstract away. Most prominent among them is that citizen-voters, instead of the current-period leader, choose the next-period leader in a democracy. Our model is not rich enough to allow such a difference to have any significant consequence, because we have already assumed that an autocratic leader breaks ties in favor of the citizens when it comes to promoting the next-period leader.

[19] For an explicit expression of $x_{IC}(x')$, see Eq.(9) in the Online Appendix. The subscript C reminds us that it is citizen-voters who choose the degree of decentralization, $x$. The subscript I reminds us that $x_{IC}$ is the "interior" candidate for the optimal $x$. The optimal $x$ needs not be interior, but if it is interior then it coincides with $x_{IC}$ (provided that $x_{IC} \geq \hat{x}$).

**Proposition 4.** There exists a threshold $\tilde{x}_C$ for next period's degree of decentralization, $x'$, such that the current period's optimal degree of decentralization, $x_C^*$, is strictly positive (i.e., "interior") only if $x'$ is lower than the threshold $\tilde{x}_C$. Specifically, define

$$\tilde{x}_C := \frac{1}{2}\left(1 - \frac{2\beta - \lambda\sigma_z^2\bar{z}b\mu}{(\bar{z}b\mu + \bar{z} - 1)\epsilon\bar{z}b}\right). \tag{7}$$

Citizen-voters' optimal choice of $x$ is

$$x_C^*(x') = \begin{cases} \max\{\hat{x}(x'), x_{IC}(x')\} & \text{if } x' < \tilde{x}_C \\ \{0, \max\{\hat{x}(x'), x_{IC}(x')\}\} & \text{if } x' = \tilde{x}_C \\ 0 & \text{if } x' > \tilde{x}_C \end{cases}.$$

Moreover, $x_C^*(x')$ is continuous and strictly decreasing in $x'$ for $x' < \tilde{x}_C$.

**Proposition 5.** A non-zero steady-state equilibrium exists in democracy if and only if

$$\tilde{x}_C \geq \frac{\epsilon}{2(1+\epsilon)},$$

where $\tilde{x}_C$ is given by Eq. (7). When a non-zero steady-state equilibrium exists, it is the only steady-state equilibrium.

A simple comparison between Eqs. (6) and (7) reveals that $\tilde{x}_L \geq \epsilon/2(1+\epsilon)$ is the more stringent condition, in the sense that a non-zero steady-state equilibrium exists in both autocracy and democracy if and only if $\tilde{x}_L \geq \epsilon/2(1+\epsilon)$. Our comparison between the equilibrium degree of decentralization in autocracy and in democracy will therefore be conducted under this condition.

The following proposition says that the steady-state equilibrium degree of decentralization is higher in a democracy than in an autocracy. The intuition is very simple: autocratic leaders have a natural self-serving reason to choose a lower degree of decentralization than preferred by citizen-voters, because reserving more authorities at the national government increases the private benefits, $(1-2x)\epsilon$, they can collect. Since the proof is very simple, we include it in the main text.

**Proposition 6.** For any $x' \in [0, 1/2]$, $x_C^*(x') \geq x_L^*(x')$.[20] Among other things, this implies that, for any configuration of parameters, if both political regimes have a steady-state equilibrium, the steady-state equilibrium degree of decentralization is weakly higher in a democracy.

**Proof.** The first statement is an immediate result of the fact that the slope of citizen-voters' payoff function, $\bar{W}(\cdot|x')$, is less negative or more positive due to the missing term $(1-2x)\epsilon$:

$$\bar{V}(x_L^*(x')|x') \geq \bar{V}(x_C^*(x')|x')$$
$$\iff \bar{W}(x_L^*(x')|x') + (1-2x_L^*(x'))\epsilon \geq \bar{W}(x_C^*(x')|x') + (1-2x_C^*(x'))\epsilon$$
$$\Rightarrow \bar{W}(x_L^*(x')|x') + (1-2x_L^*(x'))\epsilon \geq \bar{W}(x_L^*(x')|x') + (1-2x_C^*(x'))\epsilon$$
$$\iff x_C^*(x') \geq x_L^*(x'),$$

where the first (respectively, the third) line follows from the fact that $x_L^*(x')$ (respectively, $x_C^*(x')$) maximizes $\bar{V}(\cdot|x')$ (respectively, $\bar{W}(\cdot|x')$).

The second statement follows from the fact that both $x_C^*(x')$ and $x_L^*(x')$ are weakly decreasing in $x'$, and hence if their graphs ever cross the 45-degree line, they cross from above. ■

---

[20] If $x_C^*(x')$ or $x_L^*(x')$ is doubleton-valued, then $x_C^*(x') \geq x_L^*(x')$ means every element of the former is weakly bigger than every element of the latter.

Suppose $\tilde{x}_L \geq \epsilon/2(1 + \epsilon)$ and hence a non-zero steady-state equilibrium exists in both autocracy and democracy. Denote their equilibrium degree of decentralization as $x^*_{auto}$ and $x^*_{demo}$, respectively. The amounts of time bureaucrats spend in production under the two political regimes are hence

$$p^*_{auto} := p\left(x^*_{auto}|x^*_{auto}\right)$$

and $\quad p^*_{demo} := p\left(x^*_{demo}|x^*_{demo}\right),$

respectively.

**Proposition 7.** *Suppose $\tilde{x}_L \geq \epsilon/2(1 + \epsilon)$ and hence a non-zero steady-state equilibrium exists in both autocracy and democracy. Then bureaucrats spend weakly less time in production in a democracy; i.e., $p^*_{demo} \leq p^*_{auto}$. Among other things, this implies that leaders in a democracy have weakly lower expected competence.*

The proof of Proposition 7 is in the Appendix. Here we sketch the intuition. Democracy hurts bureaucrats' incentives to work for two reasons. First, the higher current-period degree of decentralization, $x$, encourages bureaucrats to spend more time in embezzlement. Second, the higher next-period degree of decentralization, $x'$, weakens bureaucrats' political career concerns. Both go in the same direction of reducing bureaucrat's time spent in production. This, among other things, results in less information generated through the political tournament, and hence average competence of the next-period leader through which he is selected.

Since welfare is measured by expected public goods provision, it is affected by (*i*) productivities of the national and the subnational governments, and (*ii*) resources allocation among these different levels of government. While democracy is stronger in terms of (*ii*) because it is citizen-voters who choose $x$, it is weaker in terms of (*i*) because of Proposition 7. It is hence possible to find numerical examples where a democracy results in lower welfare.

### 4.1. Relating the model to empirical observations

One of the key results in this section is that the steady-state equilibrium degree of decentralization is higher in a democracy than in an autocracy. This prediction is supported by empirical studies in the literature on federalism (Arzaghi and Henderson, 2005; Landry, 2008; Panizza, 1999). For example, using cross-country data from 1975, 1980, and 1985, Panizza (1999) finds that the positive correlation between a country's level of democracy and its degree of decentralization is both statistically and economically significant. Similarly, using a sample of 48 countries, Arzaghi and Henderson (2005) find that the correlation between the degree of decentralization and a national democracy index is 0.63 for five-year intervals over the period of 1960–1995, and the correlation between changes in those indices is 0.39 over the same period.

In this section, we have been comparing autocracy and democracy under the assumption that a steady-state equilibrium exists. A steady-state equilibrium, however, does not always exist. In the Online Appendix, we fully characterize all other possible equilibria, including equilibria that feature cycles. Among other things, this allows us to offer more testable implications of our model. For example, our model predicts that the degree of decentralization is more volatile in an autocracy than in a democracy. It is well-documented in the literature that there is a negative correlation between the level of democracy and the volatility in economic performance(Mobarak, 2005; Quinn and Woolley, 2001; Rodrik, 2000). Our model thus suggests for this negative correlation one explanation that has so far been overlooked by the literature

## 5. Concluding remarks

Using an overlapping principal-agent model, we have explored how political career concerns generate a new tradeoff in determining the optimal degree of decentralization. While too much decentralization weakens bureaucrats' political career concerns, and hence weakens their incentives to work, too little decentralization leaves too few authorities for these incentives to be turned into productive use. Today's degree of decentralization imposes externality on both tomorrow's and yesterday's governments. It affects tomorrow's government because it affects how hard today's bureaucrats work, and hence how well competent ones are identified and promoted as tomorrow's leaders. It affects yesterday's government because it affects yesterday's bureaucrats' political career concerns, and hence how hard they work. The channels through which political career concerns affect the optimal degree of decentralization are hence novel and do not have natural counterparts in the previous literature on federalism.

We close this paper with a few remarks on certain modeling choices we made in our analysis. First, we have made quite specific assumptions on a bureaucrat's costs of production, manipulation, and embezzlement. In particular, production incurs a direct utility cost of $\frac{1}{2}p_i^2$, while manipulation and embezzlement incur only indirect costs through the constraint $p_i + m_i + e_i \leq 1$. In fact, many alternative cost functions can similarly deliver a hump-shaped $p_i$ function (as a function of $x$, holding fixed $x'$) like the one depicted in Fig. 1, and hence either too much or too little decentralization would result in the bureaucrat spending little time in production. For example, an alternative assumption is that all three actions of the bureaucrat jointly result in a direct utility cost of $\frac{1}{2}(p_i + m_i + e_i)^2$, and there is no more indirect cost through constraints like $p_i + m_i + e_i \leq 1$. This alternative assumption would result in a graph of $p_i$ that looks like a trapezoid (instead of a triangle as in Fig. 1) in the range $x \in [x_1, x_2]$.[21] We slightly prefer our assumptions to this alternative assumption, because our assumptions generate a range of $x$ where production and embezzlement co-exist, and hence allow us to tell a slightly more realistic story in our case study of China (where we suggest that post-1978 China may correspond to some $x$ in the range $(\hat{x}, x_2)$ where production and embezzlement co-exist).

In general, costs functions that deliver a hump-shaped $p_i$ function need to have the property that either too much temptation to manipulate or too much temptation to embezzle (due to too little and too much decentralization, respectively) would crowd out production. Either a constraint like $p_i + m_i + e_i \leq 1$ or a joint cost function like $\frac{1}{2}(p_i + m_i + e_i)^2$ would satisfy this property. An example that does not satisfy this property is to assume that production incurs a direct utility cost of $\frac{1}{2}p_i^2$, while manipulation and embezzlement jointly result in utility cost of $\frac{1}{2}(m_i + e_i)^2$. In this example, the time the bureaucrat spends in manipulation and embezzlement does not affect his cost of production, and hence production is crowded out neither for small $x$ nor for big $x$.[22]

Second, we have equated welfare to the payoff of a *generic* citizen, which in turn is the sum of all public goods generated: $W = (g_i + g_j + G)\mu$. This stands in sharp contrast with standard models in the previous literature on federalism, where citizens are usually differentiated according to the regions they reside, with payoffs depending more on the local public goods provided in their regions, and less so on those provided in neighboring ones. For example, in Besley and Coate (2003), citizens' payoffs are weighted sums of own-region's public goods and neighboring region's public goods, with weights $(1 - \kappa)$

---

[21] More precisely, under this alternative assumption, $p_i = 0$ for $x < x_1$ and $x > x_2$ (with the same expressions for $x_1$ and $x_2$ as in Eqs. (2) and (3), respectively), and $p_1$ strictly positive at $x = x_1$, and increasing linearly in $x$ for $x \in [x_1, x_2]$.

[22] More precisely, for the cost functions in this example, $p_i$ will be strictly positive at $x = 0$, and increases linearly in $x$ throughout the range $x \in [0, 1/2]$.

and $\kappa$, respectively, where $\kappa \leq 1/2$. Our model hence looks like a very restrictive special case of Besley and Coate (2003) with $\kappa = 1/2$.

However, allowing for more general values of $\kappa$ would not change any of our results. That is, we could have alternatively equated welfare to the average of the payoffs of citizens residing in regions $i$ and $j$: $W = \frac{1}{2}\left(g_i + \frac{\kappa}{1-\kappa}g_j + G\right)\mu + \frac{1}{2}\left(g_j + \frac{\kappa}{1-\kappa}g_i + G\right)\mu$, and modified a bureaucrat's payoff accordingly as $U_i = \left(g_i + \frac{\kappa}{1-\kappa}g_j + G\right)\mu + xe_i - \frac{1}{2}p_i^2$. Such a change would not affect a bureaucrat's incentives to produce, to manipulate, and to embezzle. To see this, note that bureaucrat $i$'s political career concerns are measured by the difference between $V(z_i)$ (the next-period payoff as a leader) and $W(z_j)$ (the next-period payoff as a citizen while the other bureaucrat with competence $z_j$ serves as the leader), which remains $\lambda(z_i - z_j) + (1 - 2x)\epsilon$ in both our and the modified settings (provided that he holds the same belief on the behavior of other agents). Similarly, in both settings, his marginal gain in momentary payoff remains $\mu$ when he increases $p_i$. Hence his optimal mix of $(p_i, m_i, e_i)$ remains the same in both settings.

Intuitively, political career concerns affect the optimal degree of decentralization through brand new channels, and hence it is not surprising that standard factors that determine the optimal degree of decentralization in the previous literature need not play any role in our new theory. In much of the previous literature, politicians at subnational governments represent their regions (however imperfectly due to incentive problems) when they interact with other politicians to make collective decisions at the national government, and hence an important component of their incentives is how the interests of their regions diverge from those of the neighboring regions, which in turn depends sensitively on, say, spillover parameter $\kappa$. In our theory, however, bureaucrats work to prove their competence in order to get promoted, and hence whether and how the interests of their regions diverge from those of neighboring regions have little impact on their incentives.

Of course, in future studies, one may construct hybrid models where both standard and new components loom large in shaping bureaucrats' incentives. However, as a first-pass exercise, it is helpful to shut down standard and well-understood factors in order to build a clean understanding of how political career concerns alone shape bureaucrat's incentives and hence affect the optimal degree of decentralization. For this same reason, we have also shut down many other factors that are considered important determinants of the optimal degree of decentralization in the previous literature. For example, there is no preference heterogeneity across regions, nor within a single region; bureaucrats do not have informational advantage over the leader regarding local conditions; and there is no inter-regional mobility in labor and capital. It will be interesting future research to introduce these factors back into our model to study how they interact with political career concerns in determining the optimal degree of decentralization.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jpubeco.2016.11.017.

## References

Acemoglu, D., Kremer, M., Mian, A., 2008. Incentives in markets, firms, and governments. J. Law Econ. Org. 24 (2), 273–306.
Arzaghi, M., Henderson, J., 2005. Why countries are fiscally decentralizing. J. Public Econ. 89 (7), 1157–1189.
Banister, J., 1987. China's Changing Population. Stanford University Press.
Bardhan, P., 2002. Decentralization of governance and development. J. Econ. Perspect. 16 (4), 185–205.
Besley, T., Coate, S., 2003. Centralized versus decentralized provision of local public goods: a political economy approach. J. Public Econ. 87, 2611–2637.
Blanchard, O., Shleifer, A., 2001. Federalism with and without political centralization: China versus Russia. IMF Staff. Pap. 48, 171–179.
Boffa, F., Piolatto, A., Ponzetto, G.A.M., 2016. Political centralization and government accountability. Q. J. Econ. 131 (1), 381–422.
Break, G.F., 1967. Intergovernmental Fiscal Relations in the United States. Brookings Institution, Washington, DC.
Cai, H., Treisman, D., 2004. State corroding federalism. J. Public Econ. 88 (3-4), 819–843.
Cai, H., Treisman, D., 2005. Does competition for capital discipline governments? Decentralization, globalization, and public policy. Am. Econ. Rev. 95 (3), 817–830.
Cai, H., Treisman, D., 2006. Did government decentralization cause China's economic miracle? World Politics 58 (4), 505–535.
Che, J., Chung, K.-S., Qiao, X., 2013. The good, the bad, and the civil society. J. Public Econ. 106, 68–76.
Che, J., Chung, K.-S., Qiao, X., 2014. Career Concerns, Beijing Style. Mimeo, Chinese University of Hong Kong and Tsinghua University.
Chen, Y. e., Li, H., Zhou, L.-A., 2005. Relative performance evaluation and the turnover of provincial leaders in China. Econ. Lett. 88 (3), 421–425.
Coale, A.J., 1981. Population trends, population policy, and population studies in China. Popul. Dev. Rev. 85–97.
Coate, S., Knight, B., 2007. Socially optimal districting: a theoretical and empirical exploration. Q. J. Econ. 122 (4), 1409–1471.
Cumberland, J.H., 1981. Efficiency and equity in interregional environmental management. Rev. Reg. Stud. 2, 1–9.
Dewatripont, M., Jewitt, I., Tirole, J., 1999. The economics of career concerns, parts i and II. Rev. Econ. Stud. 66 (1), 183–217.
Dikotter, F., 2010. Mao's Great Famine: The History of China's Most Devastating Catastrophe. Walker., pp. 1958–1962.
Hayek, F.A., 1945. The use of knowledge in society. Am. Econ. Rev. 35, 519–530.
Holmstrom, B., 1999. Managerial incentive problems: a dynamic perspective. Rev. Econ. Stud. 66 (1), 169–182.
Holmstrom, B., Milgrom, P.R., 1991. Multitask principal-agent analyses: incentive contracts, asset ownership, and job design. J. Law Econ. Org. 7, 24–52.
Inman, R.P., Rubinfeld, D.L., 1997. Rethinking federalism. J. Econ. Perspect. 11 (4), 43–64.
Knight, B., 2004. Parochial interests and the centralized provision of local public goods: evidence from congressional voting on transportation projects. J. Public Econ. 88, 845–866.
Kung, J., 2014. The emperor strikes back: political status, career incentives and grain procurement during China's great leap famine. Polit. Sci. Res. Methods 1–33. http://dx.doi.org/10.1017/psrm.2014.6.
Kung, J., Chen, S., 2011. The tragedy of the nomenklatura: career incentives and political radicalism during China's great leap famine. Am. Polit. Sci. Rev. 105 (1), 27–45.
Landry, P.F., 2008. Decentralized Authoritarianism in China: The Communist Party's Control of Local Elites in the Post-Mao Era. Cambridge University Press, New York.
Li, H., Zhou, L.-A., 2005. Political turnover and economic performance: the incentive role of personnel control in China. J. Public Econ. 89, 1743–1762.
Li, L., 2011. The incentive role of creating 'Cities' in China. China Econ. Rev. 22 (1), 172–181.
Luelfesmann, C., Kessler, A., Myers, G.M., 2015. The architecture of federations: constitutions, bargaining, and moral hazard. J. Public Econ. 124, 18–29.
Maskin, E., Qian, Y., Xu, C., 2000. Incentives, information, and organizational form. Rev. Econ. Stud. 67, 359–378.
Milgrom, P.R., 1988. Employment contracts, influence activities, and efficient organization design. J. Polit. Econ. 96 (1), 42–60.
Mobarak, A.M., 2005. Democracy, volatility, and economic development. Rev. Econ. Stud. 87 (2), 348–361.
Myerson, R.B., 2006. Federalism and incentives for success of democracy. Q. J. Polit. Sci. 1, 3–23.
Oates, W.E., 1972. Fiscal Federalism. Harcourt Brace Jovanovich Academic Press, New York.
Oates, W.E., 1999. An essay on fiscal federalism. J. Econ. Lit. 37, 1120–1149.
Oates, W.E., Wallis, J.J., 1988. Decentralization in the public sector: an empirical study and local governments. In: Rosen, H.S. (Ed.), Fiscal Decentralization: Quantitative Studies, NBER Project Report. University of Chicago Press, Chicago and London.
Panizza, U., 1999. On the determinants of fiscal centralization: theory and evidence. J. Public Econ. 74, 97–139.
Qian, Y., Weingast, B.R., 1997. Federalism as a commitment to preserving market incentives. J. Econ. Perspect. 11 (4), 83–92.
Quinn, D.P., Woolley, J.T., 2001. Democracy and national economic performance: the preference for stability. Am. J. Polit. Sci. 45 (3), 634–657.
Rivlin, A., 1992. Reviving the American Dream. The Brookings Institution, Washington, DC.
Rodrik, D., 2000. Participatory politics, social cooperation, and economic stability. Am. Econ. Rev. 90 (2), 140–144.
Sheng, Y., 2009. Career Incentives and Political Control under Authoritarianism: Explaining the Political Fortunes of Subnational Leaders in China. Mimeo, Wayne State University.
Shi, Y., Zhou, L.-A., 2007. Regional decentralization and economic efficiency: evidence from separate-planning cities in China. Econ. Res. J. 42 (1), 17–28. (in Chinese, with English summary).
Tiebout, C., 1956. A pure theory of local expenditures. J. Polit. Econ. 64, 416–424.
Weingast, B.R., 1995. The economic role of political institutions: market-preserving federalism and economic development. J. Law Econ. Org. 11, 1–31.
Weingast, B.R., 2009. Second generation fiscal federalism: the implications of fiscal incentives. J. Urban Econ. 65, 279–293.
Xu, C., 2011. The fundamental institutions of China's reforms and development. J. Econ. Lit. 49 (4), 1076–1151.