DISCUSSION PAPER SERIES

No. 8855

A MODEL OF EQUILIBRIUM INSTITUTIONS

Bernardo Guimarães and Kevin D. Sheedy

INTERNATIONAL MACROECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP8855.asp**

ISSN 0265-8003

A MODEL OF EQUILIBRIUM INSTITUTIONS

Bernardo Guimarães, Sao Paulo School of Economics (FGV) and CEPR
Kevin D. Sheedy, London School of Economics

Discussion Paper No. 8855
February 2012

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 8855

February 2012

ABSTRACT

A Model of Equilibrium Institutions*

Institutions that serve the interests of an elite are often cited as an important reason for poor economic performance. This paper builds a model of institutions that allocate resources and power to maximize the payoff of an elite, but where any group that exerts sufficient fighting effort can launch a rebellion that destroys the existing institutions. The rebels are then able to establish new institutions as a new elite, which will similarly face threats of rebellion. The paper analyses the economic consequences of the institutions that emerge as the equilibrium of this struggle for power. High levels of economic activity depend on protecting private property from expropriation, but the model predicts this can only be achieved if power is not as concentrated as the elite would like it to be, ex post. Power sharing endogenously enables the elite to act as a government committed to property rights, which would otherwise be time inconsistent. But sharing power entails sharing rents, so in equilibrium power is too concentrated, leading to inefficiently low investment.

JEL Classification: E02, O43 and P48
Keywords: institutions, political economy, power struggle, property rights and time inconsistency

Bernardo Guimarães
São Paulo School of Economics
Fundacao Getulio Vargas
Rua Itapeva, 474
Sao Paulo
BRAZIL
Email: bernardo.guimaraes@fgv.br

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=160881

Kevin D. Sheedy
Department of Economics
London School of Economics and Political Science
Houghton Street
London
Email: k.d.sheedy@lse.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=168764

* We thank Bruno Decreuse, Erik Eyster, Emmanuel Farhi, Carlos Eduardo Goncalves, Ethan Ilzetzki, Peter Jensen, Per Krusell, Dirk Niepelt, Andrea Prat, Ronny Razin, Alwyn Young, and seminar participants at U. Amsterdam, the Anglo-French-Italian Macroeconomic workshop, U. Bonn, U. Carlos III, Central European U., Columbia U., CREI conference "The political economy of economic development", U. Cyprus, EEA 2011, EIEF, U. Erlangen-Nuremberg, ESSIM 2010, FGV (Rio), U. Glasgow, Harvard U., U. Helsinki, ISI Delhi Economic Growth and Development conference 2010, Institute for International Economic Studies, the Joint French Macroeconomic Workshop, London School of Economics, Paris School of Economics, PUC (Rio), RES 2011, U. Sao Paulo, Sao Paulo School of Economics (FGV), SED 2011, U. Surrey, U. St. Gallen, and Yale U. for helpful comments.

Submitted 07 February 2012

Princes who want to make themselves despotic have always begun by uniting all magistracies in their person.

Montesquieu (1748), *De l'esprit des lois*

1 Introduction

Economic activity is influenced by the rules that prevail in a society, whether those rules be decreeing taxes and transfers, regulating markets, enforcing private property, or limiting the arbitrary exercise of government power. These rules depend on a country's institutions, but what in turn explains which institutions will arise in a given place and time?

This question is important because many researchers have posited institutions as a significant determinant of economic development and the large differences found in the cross-country distribution of income.[1] Inefficient rules are often attributed to institutions serving the interests of an elite rather than the interests of society as a whole. However, why does elite control of a country's institutions have more than just distributional consequences? Why would an elite have incentives to set up institutions that shrink the total pie?

The institutions designed by elites are likely to reflect the constraints imposed by their desire to remain in power. Thus, to analyse the economic consequences of control of institutions by an elite, it is necessary to think about how groups come to power and what they do to remain there. This paper proposes a model of institutions that emerge as the equilibrium of this struggle for power, in a world of ex-ante identical individuals. Institutions serve the interests of those in power, but can be overthrown by any group willing to exert enough fighting effort to defeat those defending them. The power struggle is represented by a simple "conflict technology" that encompasses all different types of "rebellions" from "popular uprisings" to "coups d'état", with no exogenous restrictions placed on the groups that can fight for power.

The exogenous "technology" available to create institutions is basic. In the model, institutions control the allocation of resources and power among the individuals in a society. This means that institutions can mandate any transfers of goods between different individuals. "Power" is a primitive that gives its holder an advantage in the conflict that would occur if a group decides to rebel against the current institutions. Elites are free to design institutions stipulating the allocation of resources and power without any restrictions other than surviving the power struggle, that is, avoiding rebellions. In this environment, is it possible for rules such as private property to emerge and survive, without directly assuming they are on the menu of institutional choices? And would institutions supporting such rules be chosen by the group in power given the constraints endogenously imposed by the power struggle?

The challenge is that many rules necessary for economic efficiency require restrictions on the untrammelled exercise of the elite's power. For example, evidence suggests institutions guaranteeing

[1] For example, see North (1990), North and Weingast (1989), Engerman and Sokoloff (1997), Hall and Jones (1999) and Acemoglu, Johnson and Robinson (2005).

private property against expropriation are especially important in supporting high levels of economic activity (Acemoglu and Johnson, 2005). A system of private property requires that transfers such as expropriation are not made, even if they are physically feasible and in the interests of those in power at the time. Elites will always want to ensure they are sufficiently strong to survive the power struggle, but there needs to be a means of preventing them also using their power to sweep aside existing rules and reshaping institutions to conform to their ex post interests. The problem is essentially the classic question of "who will guard the guardians?"

This question might call to mind such notions as "independent judiciaries", "the rule of law", "political representation", and the like. While it would be possible to introduce such devices into the model by assumption, this would be as a *deus ex machina* that simply overrides the ability of members of the elite to exploit their own power. Introducing an exogenous "higher-level" institutional technology that directly allows for the protection of property rights (perhaps at some cost) would not do justice to the question at hand if the aim is to understand why or why not such features of institutions arise in equilibrium. Importantly, as it turns out, institutions that do protect property rights may arise endogenously *without* adding any extra assumptions to the model that explicitly allow for this.

The analysis of the model starts from a point where new institutions are being established, which will determine the group in power and the taxes and transfers that will be imposed. The new institutions are chosen to maximize the payoff of the group now in power. Once they have been created, there are opportunities for rebellions, to which all individuals have access on the same terms. Rebellion is the only way of changing institutions, so if no rebellions occur, economic activity takes place and the allocation of resources laid down by the prevailing institutions is implemented. If a successful rebellion does occur, the current institutions are destroyed and the model is back at the stage of establishing new ones, which will now reflect the interests of the group that emerged victorious from the foregoing conflict.

There is complete information about incentives to rebel, and the conflict technology is non-stochastic, so the goal of any elite is designing institutions in its own interests subject to no group of individuals having an incentive to launch a viable rebellion. By doing this, no rebellions will occur in equilibrium, hence no inefficiencies will result from actual conflict. Incentives to rebel depend on the payoffs the rebels expect to receive as the new elite when new institutions are subsequently established. Hence the equilibrium institutions are the fixed point of the constrained maximization problem of the elite in power subject to the threat of rebellion, where subsequent elites would be similarly constrained by equivalent threats of rebellion. As in George Orwell's *Animal Farm*, there is no essential difference between the "pigs" and the "men" they replace, but in equilibrium, some individuals will be "more equal" than others.

In an endowment economy, the model gives rise to a simple theory of distribution. The distribution of resources is uniquely determined and tied to the distribution of power. Those with equal power receive the same payoff, and those with more power receive a higher payoff. The intuition is that in comparing two individuals of equal power, the one with the lower payoff has more to gain from rebellion and is therefore willing to exert more fighting effort; while comparing two individuals with the same payoff, the one with more power poses a greater danger if he supports a rebellion because of his superior fighting strength. Since any rebellion will be launched by a subset of the population, the elite would like to minimize the fighting strength of the rebellion comprising the group of individuals with the greatest incentive to fight. This means rewarding the powerful to keep them on side, while otherwise equalizing payoffs to avoid concentrating disenchantment with the institutions. Sharing power thus always entails sharing rents.

There is a basic trade-off in this environment that characterizes the equilibrium institutions. On the one hand, the greater the number of individuals sharing power inside the elite, the greater their ability to defend the institutions they establish against rebellions, which allows them to levy higher taxes on those outside the elite. On the other hand, the proceeds must then be divided more thinly among more individuals. The equilibrium elite size maximizes the payoff of an elite member by striking a balance between these two effects.

Are institutions that are designed to maximize the payoff of those in power ever consistent with efficient outcomes? In some cases, the answer turns out to be yes. For example, suppose there is a technology that transforms rivalrous consumption goods into a public good that benefits everyone, and which has no impact on any other aspect of the environment. Such a public good will be optimally provided by the elite in equilibrium, as if its provision were chosen by a benevolent government. The intuition is that the elite must consider the impact of its choices on everyone because of the threat of rebellion. Provision of public goods reduces the incentive to rebel, just as higher taxes increase it. Since institutions can specify any transfers between individuals, the elite has incentives to choose institutions supporting a Pareto-improving "deal" to provide public goods in exchange for higher taxes, and no group has an incentive to rebel against this. Here, the contestability of institutions through rebellions leads to an efficient outcome, even though the gains may not be distributed equally.

To explore whether the equilibrium institutions support economic efficiency in other settings, the model is extended so that individuals have access to an investment technology. Individuals who invest incur an immediate effort cost, while the fruits of their investment are realized only after a lag. During this time, there is the ever-present opportunity for any group of individuals to launch a rebellion against the prevailing institutions. Were a rebellion to occur after investments have been made, the group in power following the rebellion would have incentives to expropriate fully investors' capital because the effort cost of investing is now sunk.

To provide appropriate incentives for individuals to invest, the institutions established prior to investment decisions must offer investors a higher payoff, and importantly, those institutions must survive rebellions so that what they prescribe is actually put into practice. In an endowment economy, the elite's principal concern is in avoiding a "popular uprising", a rebellion of outsiders. When offering incentives to investors, the danger of this type of rebellion increases, but it becomes essential also to avoid a "coup d'état", a rebellion launched by insiders. The higher payoff enjoyed by investors is only in the interests of the elite ex ante, so ex post there is a time-inconsistency problem. In other words, the members of the elite themselves want to rebel against the existing institutions so they can rewrite the rules protecting investors' private property.

It is therefore necessary to reduce the incentive to rebel for those inside and outside the elite simultaneously. This can only be done by expanding the size of the elite — the problem cannot be solved with any system of transfers. If higher payoffs were offered to some to reduce their willingness to rebel, resources must be taken away from others, increasing their incentive to rebel. Fundamentally, transfers can only redistribute disgruntlement with the institutions. In contrast, enlarging the elite reduces the attractiveness of all types of rebellions by increasing the number of individuals who will lose power if a rebellion succeeds. This increases the size of the group willing to defend the current institutions.

Simply adding the possibility of investment to the model therefore gives rise to an equilibrium with power sharing among a larger elite. Sharing power is an *endogenous* commitment mechanism that allows the elite to act as a government bound to a set of policies that would otherwise be time inconsistent. The analysis thus highlights the importance of sharing power as a way of guaranteeing the stability of institutions, allowing in particular for incentives to invest. This resonates with Montesquieu's doctrine of the separation of powers, now accepted and followed in well-functioning systems of government. Note that power is not shared here with those individuals who are actually investing. The additional individuals in the elite in no sense represent nor care about those who invest — but they do care about their own rents under the status quo. By this means, a group of self-interested individuals is able to act a government that commits to protection of property rights.

Although it is possible to sustain protection against expropriation by sharing power among a sufficiently large group, in equilibrium, there is too little power sharing and thus too little protection of property rights. In other words, capital taxation is too high, and investment is inefficiently low. Total output available for consumption could be increased by having a larger group in power to reduce the proportion of investors' returns that is expropriated. But the equilibrium institutions fail to generate this efficient outcome. The intuition comes from the inseparability of power and rents, which follows from the threat posed by powerful individuals were conflict to occur. It is not possible in equilibrium for the elite to share power with more individuals yet not grant them the same payoff as their equally strong peers. This places an endogenous and binding limit on the set of possible transfers, so Pareto-improving deals remain unfulfilled.

While the model is quite abstract, it is congruent with a number of historical examples, two of which are discussed later in the paper: the disappearance of private corporations (the *societas publicanorum*) when power was concentrated under the Roman emperors, and the tenacious resistance of the Stuart kings of England to sharing power with parliament.

The plan of the paper is as follows. Section 2 first discusses how the paper relates to and differs from other contributions in the literature. Section 3 presents the basic model of power and distribution. The benchmark case of public-good provision is briefly studied in section 4, after which private investment is analysed in section 5. Finally, section 6 draws some conclusions.

2 Comparison with the existing literature

Since Downs (1957) emphasized the importance of studying governments composed of self-interested agents, a vast literature on political economy has flourished (see, for example, Persson and Tabellini, 2000). Most of this literature focuses on democracies, so institutions are not themselves explained in terms of the decisions of self-interested agents. But in much of the developing world and during most of human history, political regimes have differed greatly from democracies.

Recently, some models have been developed with the aim of understanding institutions themselves. Greif (2006) combines a rich historical analysis of trade and institutions in medieval times with economic modelling, part of which focuses on the form of government and political institutions that emerged in Genoa. Acemoglu and Robinson (2006, 2008) analyse the conditions leading to democracy or dictatorship in an environment where an elite is trying to maintain its power, while citizens prefer a more egalitarian state. In Besley and Persson (2009a,b, 2010), society comprises two groups of agents that alternate in power, and make investments in two technologies that respectively allow the state to tax and to enforce contracts. Differently from these contributions, this paper aims to make primitive assumptions only on the mechanisms through which institutions are created and destroyed, while imposing no other restrictions on agents' choices. Hence there is a more basic institutional technology and no ex-ante differences among individuals, and there are no relevant constraints on the choice of institutions besides the threat of rebellions. In our view, this makes this paper more suited to studying the constraints on institutions that arise.

This paper shares some features of the literature on coalition formation (see Ray, 2007).[2] As in that literature, the process of establishing rules is non-cooperative, but it is assumed in the absence of rebellion that such rules are followed. Moreover, the modelling of rebellions here is related to the idea of blocking in coalitions (Ray, 2007, part III) in the sense that there is no explicit game-form. What distinguishes this paper is the actual modelling of power, rebellions, and conflict.

The model assumes that if the institutions established by the elite survive the power struggle then they do indeed determine the allocation of resources once production has taken place. But how would these institutions manage to control the allocation of goods ex post? As pointed out by Basu (2000) and Mailath, Morris and Postlewaite (2001), laws and institutions do not change the physical nature of the game, all they can do is affect how agents coordinate on some pattern of behaviour. But in reality, laws and institutions are seen to have a strong impact on behaviour, and this feature must be present in any model for those institutions that do not trigger rebellions.

One possible interpretation of the approach in this paper is similar to the application put forward by Myerson (2009) of Schelling's (1960) notion of focal points in the organization of society. The "rules of the game" are self enforcing as long as society coordinates on punishing whomever deviates from the rules — and whomever deviates from punishing the deviator. Following this, theorizing about institutions is theorizing about (i) how rules (or focal points) are chosen, and (ii) how rules can change. For example, Myerson (2004) explores the idea of justice as a focal point influencing the

[2] Baron and Ferejohn (1989) analyse bargaining in legislatures using this approach, while Levy (2004) studies political parties as coalitions. Other recent contributions include Acemoglu, Egorov and Sonin (2008) and Piccione and Razin (2009).

allocation of resources in society. This paper takes a more cynical view of our fellow human beings: those in power choose the laws and institutions to maximize their own payoffs, and institutions can only be destroyed by rebellions. There is no modelling of the post-production game.

This paper is also related to the literature on social conflict and predation, surveyed by Garfinkel and Skaperdas (2007).[3] It is easy to envisage how conflict could be important in a state of nature: individuals could devote their time to fighting and stealing from others. However, when there are fights, there are deadweight losses. Thus, it would be efficient if individuals could agree on transfers to avoid conflict. This paper supposes such deals are possible: individuals pay taxes to the group in power, which allocates resources according to some predetermined rules. Here, differently from the literature on conflict, individuals fight to be part of the group that sets the rules, not directly over what has been produced. Moreover, they fight in groups, not as isolated individuals.

Acemoglu (2003) raises the question of why institutions can be inefficient (a breakdown of what is termed the "political Coase theorem") and highlights the importance of commitment problems. Commitment is indeed the key issue in the main application of the model of this paper, but the question here is the means by which commitment can arise endogenously, and whether the group in power has sufficient incentives to make use of the endogenous commitment mechanism. There are other theoretical models focused on political issues that lead to inefficiencies in protecting property rights. Examples include Glaeser, Scheinkman and Shleifer (2003), Acemoglu (2008), Guriev and Sonin (2009), and Myerson (2010). Here, the risk of expropriation of capital and the consequent need to protect property rights is just a natural consequence of the possibility of investment and the "rebellion technology" that allows institutions to be destroyed and replaced.

Lastly, it is possible to draw an analogy between this paper and models of democracy (for example, in Persson and Tabellini, 2000) in the sense that the "election technology" there is replaced by a "rebellion technology" here.

3 The model of power and distribution

This section presents an analysis of the equilibrium institutions in a simple endowment economy. Subsequent sections apply the model developed here to economies with added production technologies.

3.1 Preferences and technologies

There is an area containing a measure-one population of ex-ante identical individuals indexed by $\imath \in \Omega$. Individuals receive utility $\mathcal{U}$ from their own consumption C of a homogeneous good and disutility if they exert *fighting effort* F:

$$\mathcal{U} = u(C) - F, \qquad [3.1]$$

[3]For instance, see Grossman and Kim (1995) and Hirshleifer (1995).

where $u(\cdot)$ is a strictly increasing and weakly concave function. Individuals who become *workers* receive an exogenous endowment of q units of goods per worker.

3.2 Institutions

Institutions stipulate the distribution of resources and power. They specify the set $\mathcal{W}$ of workers, and the set $\mathcal{P}$ of individuals currently in *power*, referred to as the *elite*. Each position of power confers an equal advantage on its holder in the event of any conflict, as described below. The elite $\mathcal{P}$ can have any size between 0% and 50% of the population. Those individuals in power cannot simultaneously be workers.[4]

Institutions specify the transfers of goods that are made between individuals. A worker $\imath \in \mathcal{W}$ facing an individual-specific tax $\tau(\imath)$ (or transfer if negative) consumes

$$C_{\mathrm{w}}(\imath) = \mathrm{q} - \tau(\imath). \tag{3.2}$$

Tax revenue is used to finance the consumption of the elite. If $C_{\mathrm{p}}(\imath)$ is the individual-specific consumption of a member of the elite $\imath \in \mathcal{P}$ then the overall budget constraint is

$$\int_{\mathcal{P}} C_{\mathrm{p}}(\imath)\mathrm{d}\imath = \int_{\mathcal{W}} \tau(\imath)\mathrm{d}\imath. \tag{3.3}$$

Any system of transfers is physically feasible subject only to this budget constraint and non-negativity constraints on the consumption of each individual.

Formally, institutions are a collection $\mathscr{I} = \{\mathcal{P}, \mathcal{W}, \tau(\imath), C_{\mathrm{p}}(\imath)\}$, where the sets $\mathcal{P}$ and $\mathcal{W}$ satisfy $\mathcal{P} \cup \mathcal{W} = \Omega$, $\mathcal{P} \cap \mathcal{W} = \emptyset$, and $|\mathcal{P}| < 1/2$ (with $|\cdot|$ denoting the measure of a set), and where the tax and consumption distributions $\tau(\imath)$ and $C_{\mathrm{p}}(\imath)$ are consistent with the budget constraint [3.3] and non-negativity constraints.

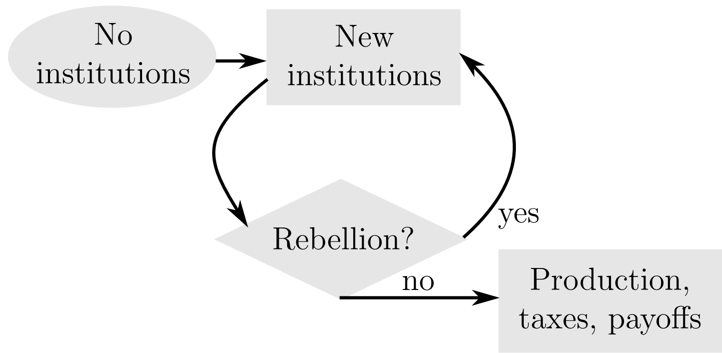
Once institutions exist, the allocation of resources and power they decree prevails unless a successful *rebellion* occurs. Rebellions are the only means of changing institutions, and are explained in detail below.

The sequence of events is depicted in Figure 1. Institutions are established. There are then opportunities for rebellions. If a successful rebellion occurs, new institutions are established, followed by more opportunities for rebellions. When no rebellions occur, workers receive their endowments, the rules laid down by the prevailing institutions are implemented, and payoffs are received.

New institutions are created as soon as a rebellion has destroyed existing ones, or if no institutions previously existed. The fundamental assumption is that institutions are set up to maximize the average of the payoffs $\mathcal{U}_{\mathrm{p}}(\imath)$ of those who will be in the elite ($\imath \in \mathcal{P}$). The choice variables are the taxes $\tau(\imath)$ levied on workers ($\imath \in \mathcal{W}$), the consumption $C_{\mathrm{p}}(\imath)$ of members of the elite, and the status of each individual, that is, an assignment of each $\imath \in \Omega$ to one of the sets $\mathcal{P}$ or $\mathcal{W}$.

[4] The assumption that those in power do not receive the endowments of workers is not essential for the main results. It does mean that there is an opportunity cost of individuals being in power, so strictly speaking, this is not a pure endowment economy. However, the resulting "guns versus butter" inefficiency is *not* the focus of this paper.

Figure 1: *Sequence of events*



The maximization problem for new institutions is subject only to the following constraints. The basic feasibility conditions (budget constraint, non-negativity constraints, and maximum number of individuals in power) must be respected. There must be no profitable opportunity for any group to rebel, as described below. An *elite selection function* $\mathscr{E}(\cdot)$ must be respected, which determines the identities of those who are to be in the elite, *conditional on* the elite having size p.

Formally, the elite selection function $\mathscr{E}(\cdot)$ is a mapping $\mathscr{E} : [0, 1/2) \to \mathscr{P}$, with $\mathscr{P}$ denoting the set of (measurable) subsets of Ω. It has two properties: $|\mathscr{E}(p)| = p$ for all $p \in [0, 1/2)$, and $\mathscr{E}(p_1) \subseteq \mathscr{E}(p_2)$ if $p_1 \leq p_2$. If p is the size (measure) of the elite, the constraint imposed by the elite selection function is $\mathcal{P} = \mathscr{E}(p)$. The function $\mathscr{E}(\cdot)$ itself is a state variable that was set down at the time of the previous rebellion, or in the case of no previous institutions, was randomly chosen by nature. The elite selection function is such that any larger elite would necessarily include those already selected to join a smaller elite, so it can be thought of as specifying individuals' priority in receiving positions of power, in other words, the "pecking order" in society.

Finally, in cases where a non-degenerate distribution of taxes $\tau(\imath)$ or elite consumption $C_{\mathrm{p}}(\imath)$ is chosen, each individual's tax and consumption level must be a random draw from the corresponding distribution. Every individual's draw becomes known as soon as the institutions are set up.

3.3 Rebellions

A successful *rebellion* destroys the existing institutions and defines a new elite selection function, which will be a state variable when new institutions are subsequently formed. If a rebellion occurs, all individuals know the new elite selection function that will be put in place if it succeeds.

A rebellion is successful if the fighting strength of its *rebel army* $\mathcal{R}$ is sufficiently large. The rebel army is the set of all individuals satisfying two conditions. First, members must anticipate receiving a payoff under the post-rebellion institutions at least as high as what they would receive under the current institutions. Second, members must anticipate receiving a place in the post-rebellion elite.

Formally, a particular rebellion can be entirely characterized by the new elite selection function $\mathscr{E}'(\cdot)$ it stipulates. In what follows, the notation $'$ indicates the post-rebellion value of a variable.

Consider a particular elite selection function $\mathscr{E}'(\cdot)$. The conditions defining the membership of the rebel army depend on beliefs about the new institutions $\mathscr{I}' = \{\mathcal{P}', \mathcal{W}', \tau'(\imath), C'_{\mathrm{p}}(\imath)\}$ that would be created after the rebellion. Let p'^{e} denote beliefs about the post-rebellion elite size $p' = |\mathcal{P}'|$. If the rebellion succeeds, individuals in the set $\mathscr{E}'(p'^{\mathrm{e}})$ anticipate a place in the new elite. Let $\mathcal{U}'^{\mathrm{e}}_{\mathrm{p}}$ denote the expected utility of such individuals once the new institutions are in place, and $\mathcal{U}'^{\mathrm{e}}_{\mathrm{w}}$ the expected utility of those individuals who anticipate being workers under the new institutions:

$$\mathcal{U}'^{\mathrm{e}}_{\mathrm{p}} = \frac{1}{p'^{\mathrm{e}}} \int_{\mathscr{E}'(p'^{\mathrm{e}})} u(C'^{\mathrm{e}}_{\mathrm{p}}(\jmath))\mathrm{d}\jmath, \quad \text{and } \mathcal{U}'^{\mathrm{e}}_{\mathrm{w}} = \frac{1}{1-p'^{\mathrm{e}}} \int_{\Omega \backslash \mathscr{E}'(p'^{\mathrm{e}})} u(\mathrm{q} - \tau'^{\mathrm{e}}(\jmath))\mathrm{d}\jmath. \qquad [3.4]$$

Now take an individual $\imath \in \mathscr{E}'(p'^{\mathrm{e}})$ who anticipates a place in the post-rebellion elite. The utility of this individual under the prevailing institutions $\mathscr{I} = \{\mathcal{P}, \mathcal{W}, \tau(\imath), C_{\mathrm{p}}(\imath)\}$ is denoted by $\mathcal{U}(\imath)$, noting that this individual could be inside or outside the current elite. Such an individual (weakly) gains from the success of the rebellion if $\mathcal{U}'^{\mathrm{e}}_{\mathrm{p}} \geq \mathcal{U}(\imath)$. The set $\mathcal{F}$ comprises all individuals for whom this inequality holds, and the rebel army $\mathcal{R}$ raised by the rebellion with new elite selection function $\mathscr{E}'(\cdot)$ is thus

$$\mathcal{R} = \mathcal{F} \cap \mathscr{E}'(p'^{\mathrm{e}}), \quad \text{with } \mathcal{F} = \{\imath \in \Omega \mid \mathcal{U}'^{\mathrm{e}}_{\mathrm{p}} \geq \mathcal{U}(\imath)\}. \qquad [3.5]$$

In the event of a rebellion occurring, the survival of the current institutions rests on the *incumbent army* $\mathcal{A}$, the set of individuals defending the institutions. An individual belongs to the incumbent army if three conditions are met. First, the individual is in power ($\imath \in \mathcal{P}$) under the current institutions. Second, the individual does not join the rebel army ($\imath \notin \mathcal{R}$) according to the criteria set out above. Third, given beliefs about the post-rebellion institutions, the individual expects a lower payoff than under the current institutions. The incumbent army is thus given by

$$\mathcal{A} = \mathcal{D} \backslash \mathcal{R}, \quad \text{with } \mathcal{D} = \{\imath \in \mathcal{P} \mid \mathcal{U}_{\mathrm{p}}(\imath) > \mathcal{U}'^{\mathrm{e}}_{\mathrm{w}}\}, \qquad [3.6]$$

where $\mathcal{D}$ denotes the set of elite members who stand to gain by defending the current institutions.[5]

The outcome of a rebellion is determined by which of the rebel and incumbent armies has the greater *fighting strength*. Each army's fighting strength is the integral of the fighting strengths of its members. The fighting strength of individual $\imath \in \mathcal{R}$ in the rebel army is the amount of fighting effort $F(\imath)$ he exerts, which is assumed to equal to the maximum the individual would be willing to put in to change the institutions. This is the difference between the individual's payoff under the current institutions $\mathcal{U}(\imath)$ and his anticipated payoff under the post-rebellion institutions $\mathcal{U}'^{\mathrm{e}}_{\mathrm{p}}$. Each individual $\imath \in \mathcal{A}$ in the incumbent army has fighting strength measured by the parameter δ, which is obtained at no utility cost to these individuals. The rebellion succeeds if and only if

$$\int_{\mathcal{R}} F(\imath)\mathrm{d}\imath > \int_{\mathcal{A}} \delta \mathrm{d}\imath, \quad \text{where } F(\imath) = \mathcal{U}'^{\mathrm{e}}_{\mathrm{p}} - \mathcal{U}(\imath). \qquad [3.7]$$

All new elite selection functions are feasible when launching a rebellion, so if [3.7] holds for *any* $\mathscr{E}'(\cdot)$ then the current institutions will be destroyed. If [3.7] holds for multiple $\mathscr{E}'(\cdot)$ then one of

[5] This definition assumes without loss of generality that those in power who do not join the rebel army anticipate becoming workers if the rebellion succeeds.

these new elite selection functions is randomly chosen.

Although the term *rebellion* has been used to describe the process of destroying the current institutions, the formal definition encompasses "popular uprisings" (the rebel army comprises only workers), "coups d'état" (the rebel army is a subset of the current elite), "suspensions of the constitution" (the rebel army includes all members of the current elite), and "revolutions" that receive the backing of some insiders from the current regime (the rebel army includes a mixture of workers and elite members).

3.4 Establishing institutions

The constrained maximization problem for the new institutions $\mathscr{I} = \{\mathcal{P}, \mathcal{W}, \tau(\imath), C_{\mathrm{p}}(\imath)\}$ reduces to a choice of the extent of power sharing (the elite size, denoted by p), and distributions of taxes and elite consumption, denoted by $\tau(\cdot)$ and $C_{\mathrm{p}}(\cdot)$:

$$\max_{p,\tau(\cdot),C_{\mathrm{p}}(\cdot)} \frac{1}{p}\int_{\mathscr{E}(p)} \mathcal{U}_{\mathrm{p}}(\imath)\mathrm{d}\imath \;\; \text{subject to the budget constraint [3.3]}, \; C(\imath) \geq 0, \; p < 1/2,$$
$$\text{and for all } \mathscr{E}'(\cdot): \int_{\mathcal{R}} F(\imath)\mathrm{d}\imath \leq \int_{\mathcal{A}} \delta \mathrm{d}\imath, \qquad \text{where } F(\imath), \mathcal{R} \text{ and } \mathcal{A} \text{ are given by [3.4]–[3.7].} \tag{3.8}$$

Given the solution $\{p, \tau(\cdot), C_{\mathrm{p}}(\cdot)\}$, the current predetermined elite selection function $\mathscr{E}(\cdot)$ determines $\mathcal{P} = \mathscr{E}(p)$ and $\mathcal{W} = \Omega \backslash \mathscr{E}(p)$, and each individual's $\tau(\imath)$ or $C_{\mathrm{p}}(\imath)$ is drawn from the distributions $\tau(\cdot)$ and $C_{\mathrm{p}}(\cdot)$ if these are non-degenerate. The new institutions solving this constrained maximization problem will not trigger any rebellions, hence no further fighting effort $F(\imath)$ is exerted by any individual. Since any past fighting effort is sunk, the utility function [3.1] implies that payoffs of elite members and workers are $\mathcal{U}_{\mathrm{p}}(\imath) = u(C_{\mathrm{p}}(\imath))$ and $\mathcal{U}_{\mathrm{w}}(\imath) = u(\mathrm{q} - \tau(\imath))$ respectively.

In the absence of any uncertainty, beliefs about subsequent institutions must coincide with outcomes, that is, $p'^{\mathrm{e}} = p'$, $\mathcal{U}_{\mathrm{p}}'^{\mathrm{e}} = \mathcal{U}_{\mathrm{p}}'$, and $\mathcal{U}_{\mathrm{w}}'^{\mathrm{e}} = \mathcal{U}_{\mathrm{w}}'$. Determining the equilibrium institutions thus requires knowledge of the equilibrium that would prevail when new institutions are set up following a rebellion. But finding the equilibrium institutions at that subsequent stage means solving a constrained maximization problem of the same form as [3.8] in that any subsequent elite would also be subject to threats of rebellion. These threats mean that the post-post-rebellion equilibrium p'', $\mathcal{U}_{\mathrm{p}}''$, and $\mathcal{U}_{\mathrm{w}}''$ needs to be known to determine the post-rebellion outcomes p', $\mathcal{U}_{\mathrm{p}}'$, and $\mathcal{U}_{\mathrm{w}}'$, and so on recursively, *ad infinitum*.

At every point at which new institutions can be formed, attention is restricted to institutions that depend only on fundamental state variables. Formally, *Markovian institutions* are a collection $\{p, \tau(\cdot), C_{\mathrm{p}}(\cdot)\}$ that is a function only of those state variables actually affecting which institutional choices are physically feasible now or in the future, rather than of the entire history at that point. A *Markovian equilibrium* is a set of Markovian institutions $\{p^*, \tau^*(\cdot), C_{\mathrm{p}}^*(\cdot)\}$ solving the constrained maximization problem [3.8], where the institutions that would be set up following any subsequent rebellions are expected to be Markovian solutions of the corresponding constrained maximization problems.

In the simple endowment economy model, the only state variable is the elite selection function $\mathscr{E}(\cdot)$. While this state variable is relevant for determining the identities of the elite, the maximum attainable average elite payoff in the constrained maximization problem [3.8] is unaffected by the particular elite selection function in force because all individuals are ex ante identical. Thus, there are in fact *no* relevant state variables in the problem of finding the optimal $\{p, \tau(\cdot), C_{\mathrm{p}}(\cdot)\}$. The Markovian equilibrium $\{p^*, \tau^*(\cdot), C_{\mathrm{p}}^*(\cdot)\}$ is then found in two steps. First, the solution of [3.8] is obtained taking as given a particular set of subsequent institutions $\{p', \tau'(\cdot), C_{\mathrm{p}}'(\cdot)\}$. Second, the equilibrium conditions of an identical elite size $p' = p^*$ and identical distributions of taxes $\tau'(\cdot) = \tau^*(\cdot)$ and elite consumption $C_{\mathrm{p}}'(\cdot) = C_{\mathrm{p}}^*(\cdot)$ are imposed, where the latter conditions can be stated precisely as $\mathbb{P}[\tau^*(\imath) \leq \mathrm{\tau}] = \mathbb{P}[\tau'(\imath) \leq \mathrm{\tau}]$ for all $\mathrm{\tau}$, and $\mathbb{P}[C_{\mathrm{p}}^*(\imath) \leq \mathrm{C}] = \mathbb{P}[C_{\mathrm{p}}'(\imath) \leq \mathrm{C}]$ for all C. This fixed-point problem can be solved without reference to the elite selection function, which is then used only to pin down the identities of the elite.

3.5 Equilibrium

Proposition 1 *Any Markovian equilibrium must have the following properties:*

(i) Equalization of workers' payoffs: $\mathcal{U}_{\mathrm{w}}^*(\imath) = \mathcal{U}_{\mathrm{w}}^*$ *for all* $\imath \in \mathcal{W}$ *(with measure one).*

(ii) Sharing power implies sharing rents: $\mathcal{U}_{\mathrm{p}}^*(\imath) = \mathcal{U}_{\mathrm{p}}^*$ *for all* $\imath \in \mathcal{P}$ *(with measure one).*

(iii) Power determines rents: $\mathcal{U}_{\mathrm{p}}^* - \mathcal{U}_{\mathrm{w}}^* = \delta$.

(iv) The equilibrium institutions can be characterized by solving the maximization problem [3.8] subject only to a single "no-rebellion" constraint:

$$\mathcal{U}_{\mathrm{p}}' - \mathcal{U}_{\mathrm{w}} \leq \delta \frac{p}{p'}. \tag{3.9}$$

(v) A Markovian equilibrium always exists and is unique. The equilibrium elite size p^* *satisfies* $0 < p^* \leq 2 - \varphi$, *where* $\varphi \equiv (1 + \sqrt{5})/2 \approx 1.618$.[6]

Proof See appendix A.1. ■

The first two parts of the proposition demonstrate that the elite has a strong incentive to avoid inequality except where it is justified by differences in power. These results hold even when the utility function is linear in consumption, and so do not depend on strict concavity.

The intuition for the payoff-equalization results is that the most dangerous composition of the rebel army is the one including those individuals with the greatest incentive to fight. A rebel army will always be a subset of the whole population. As a consequence, if there were payoff inequality among workers, the most dangerous rebel army would not include those workers who

[6] A necessary and sufficient condition for all non-negativity constraints on consumption to be slack in equilibrium is $\delta/u'(0)\mathrm{q} < 1 + \mathrm{q}/u^{-1}(u(0)+\delta)$, with $\delta/u'(0)\mathrm{q} < 1$ being sufficient for this. The condition $\delta/u'(0)\mathrm{q} > \varphi$ is sufficient for an equilibrium in which the non-negativity constraint on workers' consumption is binding. The term $\varphi \equiv (1+\sqrt{5})/2$ is known as the *Golden ratio* or *mean of Phidias*. The constraint $p < 1/2$ is always slack in equilibrium.

receive a relatively high payoff. The elite could then reduce the effectiveness of this rebel army by redistributing from relatively well-off workers to the worse off. This slackens the set of no-rebellion constraints, allowing the elite to achieve a higher payoff. The elite's tax policy exploits these gains to the maximum possible extent when all workers' payoffs are equalized.[7] Similarly, inequality in elite payoffs is undesirable because elite members receiving a relatively low payoff can defect and join a rebel army. Equalizing elite payoffs by redistributing consumption does not directly lower average elite utility when the utility function is weakly concave, while it has the advantage of weakening the most dangerous rebel army. Since defections from the elite weaken the incumbent army, there is no version of this argument that calls for equalization of payoffs *between* workers and the elite.

Given that workers receive equal endowments q, payoff equalization implies all workers pay the same tax τ. Payoff equalization among the elite implies that all members receive the same consumption level C_{p}, which can be found using the budget constraint [3.3]. Thus:

$$\mathcal{U}_{\mathrm{w}} = u(C_{\mathrm{w}}) = u(\mathrm{q} - \tau), \quad \text{and } \mathcal{U}_{\mathrm{p}} = u(C_{\mathrm{p}}) = u\left(\frac{(1-p)\tau}{p}\right). \tag{3.10}$$

As a consequence of the payoff-equalization results, all that matters for the composition of a rebel army are the fractions σ_{w} and σ_{p} of its total numbers drawn from workers and from the current elite. The equilibrium institutions solving [3.8] are then the solution of the simpler problem

$$\max_{p,\tau} \mathcal{U}_{\mathrm{p}} \text{ s.t. } \sigma_{\mathrm{w}} \max\{\mathcal{U}'_{\mathrm{p}} - \mathcal{U}_{\mathrm{w}}, 0\} + \sigma_{\mathrm{p}}(\mathcal{U}'_{\mathrm{p}} - \mathcal{U}_{\mathrm{p}} + \delta)\mathbb{1}[\mathcal{U}'_{\mathrm{p}} \geq \mathcal{U}_{\mathrm{p}}] \leq \delta\frac{p}{p'}, \tag{3.11}$$

for all non-negative proportions σ_{w} and σ_{p} that are feasible given the size of the rebel army p' and the sizes of the groups of workers and elite members under the current institutions, that is, $\sigma_{\mathrm{w}} \leq (1-p)/p'$, $\sigma_{\mathrm{p}} \leq p/p'$, and $\sigma_{\mathrm{w}} + \sigma_{\mathrm{p}} = 1$. The general form of the no-rebellion constraints stated in [3.11] is derived from the participation constraints on membership of the rebel and incumbent armies described in equations [3.4]–[3.7],[8] with $\mathbb{1}[\cdot]$ denoting the indicator function.

The fourth claim in Proposition 1 states that the equilibrium institutions can be characterized by a single "no-rebellion" constraint [3.9], which is equivalent to setting $\sigma_{\mathrm{w}} = 1$ and $\sigma_{\mathrm{p}} = 0$ in the general constraints of [3.11] (and noting that $\mathcal{U}'_{\mathrm{p}}$ will exceed $\mathcal{U}_{\mathrm{w}}$ in a Markovian equilibrium). Satisfaction of [3.9] is clearly necessary, but the proposition shows that this single constraint is also *sufficient* to characterize the Markovian equilibrium institutions.[9] The constrained maximization problem thus reduces to

$$\max_{p,\tau} \mathcal{U}_{\mathrm{p}} \text{ s.t. } \mathcal{U}'_{\mathrm{p}} - \mathcal{U}_{\mathrm{w}} \leq \delta\frac{p}{p'}, \tag{3.12}$$

[7]This result is different from those found in some models of electoral competition such as Myerson (1993). In the equilibrium of that model, politicians offer different payoffs to different agents. But there is a similarity with the model here because in neither case will agents' payoffs depend on their initial endowments.

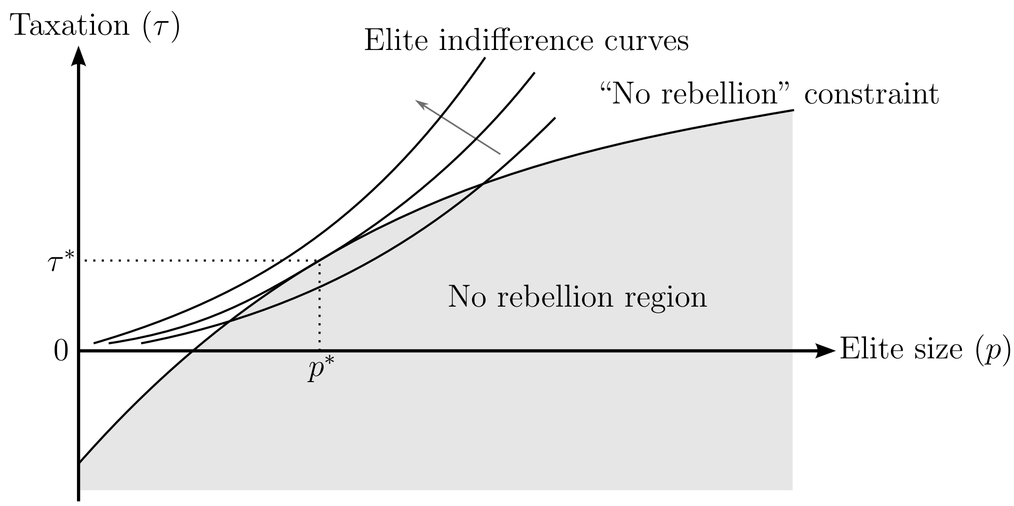
[8]The requirement $\mathcal{U}_{\mathrm{p}} > \mathcal{U}'_{\mathrm{w}}$ that elite members who do not join a rebel army are willing to defend the current institutions is always satisfied in equilibrium.

[9]This finding is specific to the simple endowment economy model of this section; subsequent extensions of the model will find that other no-rebellion constraints become binding. Which composition or compositions of the rebel army are associated with binding no-rebellion constraints in equilibrium is endogenous and will depend on the situation being analysed.

where $\mathcal{U}_{\mathrm{w}}$ and $\mathcal{U}_{\mathrm{p}}$ are as in equation [3.10], with beliefs p' and $\mathcal{U}'_{\mathrm{p}}$ taken as given. In a Markovian equilibrium, these beliefs are equal to the corresponding values p^* and $\mathcal{U}^*_{\mathrm{p}}$ that solve the maximization problem.

As can be seen from [3.10], the payoff of each member of the elite is increasing in the tax τ, and decreasing in the size of the elite p because the total tax revenue must be distributed more widely (and because there are fewer workers available to tax). The indifference curves of the elite over τ and p are plotted in Figure 2.

Figure 2: *Trade-off between elite size and taxation*



An increase in τ reduces the payoffs of workers, making them more willing to fight in a rebellion, while an increase in the size of the elite boosts the fighting strength of the incumbent army, making rebellion less attractive to workers. The elite thus has two margins to ensure that it avoids rebellions. It can reduce taxes τ (the "carrot"), or increase its size p (the "stick"). This corresponds to the upward-sloping "no-rebellion" constraint depicted in Figure 2.[10]

After taking into account the binding no-rebellion constraint, the key decision the elite must make is how widely to share power. It faces a fundamental trade-off in determining its optimal size: a larger size will strengthen the elite and allow higher taxes to be extracted from workers, but will also spread the revenue from these taxes more thinly among a larger number of individuals (and also reduce the tax base). Proposition 1 shows that it is not possible in equilibrium to add extra individuals to the elite without offering these individuals the same high payoff received by other members. Thus, sharing power entails sharing rents. The elite then shares power with an extra individual if and only if this allows it to increase its average payoff. The allocation of power therefore reflects the interests of the elite, rather than the interests of society. In a Markovian equilibrium, the utility value of the rents received by those in the elite depends only on the exogenous power parameter δ.

[10]When utility is linear in consumption, the no-rebellion constraint is a straight line. The diagram shows the general case where utility is strictly concave in consumption, resulting in the constraint being a concave function.

3.6 Example: linear utility in consumption

There are three exogenous parameters in the model: the power parameter δ, the endowment q of a worker, and the utility function $u(\cdot)$ in consumption. This section illustrates the workings of the model with a linear utility function $u(C) = C$. The maximization problem [3.12] of the elite becomes

$$\max_{p,\tau} \frac{(1-p)\tau}{p} \quad \text{s.t.} \quad C'_{\mathrm{p}} - (\mathrm{q} - \tau) \leq \delta \frac{p}{p'}, \tag{3.13}$$

after substituting the expressions for $\mathcal{U}_{\mathrm{p}}$ and $\mathcal{U}_{\mathrm{w}}$ from [3.10]. The single no-rebellion constraint is binding, and can be used to solve explicitly for the tax $\tau = \mathrm{q} - C'_{\mathrm{p}} + \delta p/p'$. By substituting this level of taxes into the objective function, consumption of the elite is given by

$$C_{\mathrm{p}} = \frac{1-p}{p}\left(\mathrm{q} - C'_{\mathrm{p}} + \delta \frac{p}{p'}\right). \tag{3.14a}$$

The problem is now an unconstrained choice of elite size p to maximize the expression for elite consumption, with beliefs p' and C'_{p} taken as given. The first-order condition is

$$\frac{C^*_{\mathrm{p}}}{1-p^*} = (1-p^*)\frac{\delta}{p'}. \tag{3.14b}$$

Now the Markovian equilibrium conditions ($p^* = p'$, $C^*_{\mathrm{p}} = C'_{\mathrm{p}}$) are imposed in [3.14a], which leads to $C^*_{\mathrm{p}} = (\mathrm{q}+\delta)(1-p^*)$. Combining this with equation [3.14b] (and using $p^* = p'$ again) yields the Markovian equilibrium:[11]

$$p^* = \frac{\delta}{\mathrm{q}+2\delta}, \quad C^*_{\mathrm{p}} = \frac{(\mathrm{q}+\delta)^2}{\mathrm{q}+2\delta}, \quad \text{and} \quad C^*_{\mathrm{w}} = \frac{(\mathrm{q}+\delta)^2}{\mathrm{q}+2\delta} - \delta. \tag{3.15}$$

Notice in this case that the size of the elite is a function of the ratio δ/q.[12]

3.7 Discussion

The model described above is designed to capture in a simple manner the "power struggle" for control of institutions determining the allocation of resources. Institutions can be overthrown by rebellions and replaced by new ones. The success of a rebellion is settled by a basic "conflict technology" that avoids going into the punches and sword thrusts of battle. Everyone has access on the same terms

[11] The restriction $\delta/\mathrm{q} \leq \varphi$ is assumed, where φ is the Golden ratio (see footnote 6). When the utility function is linear, this restriction is necessary and sufficient for an equilibrium in which the non-negativity constraint on workers' consumption is not binding.

[12] The power parameter δ affects the equilibrium in three ways. First, an increase in δ makes the elite stronger because the rebels have to exert greater fighting effort to defeat the incumbent army. This "income effect" leads to an increase in τ and a decrease in p. Second, the payoff that the rebels would receive once in power increases as their position will also be stronger once they have supplanted the current elite, making rebellion more attractive. This offsetting "income effect" decreases τ and increases p. Third, an increase in δ raises the effectiveness of the marginal fighter in the incumbent army, leading to a "substitution effect" whereby the elite increases its size in order to extract higher taxes. As long as the non-negativity constraint on workers' consumption is not binding, the third effect dominates and the size of the elite is increasing in δ.

to opportunities for rebellion, irrespective of their current status. Several assumptions are made for simplicity. The combatants' strengths are linear in the strengths of the individuals that make up the rebel and incumbent armies. Members of the incumbent army have a predetermined fighting strength (the parameter δ), so this is inelastic with respect to current fighting effort. Given fighting strengths, there is no uncertainty about which side will emerge victorious.

In modelling conflict, it is necessary to introduce some asymmetry between the rebel and incumbent armies for the notion of being "in power" to be meaningful. The parameter δ represents the advantage incumbents derive from their entrenched position. It is fighting strength that is obtained at no *current* effort cost, while the rebels can only obtain fighting strength from current effort. The inelasticity of incumbent-army members' fighting strength with respect to current effort can be seen as an inessential simplification.[13] When the current elite designs institutions, it has several margins it can adjust to ensure it avoids rebellions, such as varying the number of individuals in power, or offering transfers to those who might join a rebel army. The ability to adjust δ at some cost in addition to these does not fundamentally change the problem. The rebel army, however, lacks these alternative margins, so it is essential that its fighting strength is responsive to the effort put in by its members.

One interpretation of the parameter δ is that the individuals currently in power possess some defensive fortifications, such as a castle, which place them at an immediate fighting advantage over any rebels who must breach these from outside. A broader interpretation is that δ represents the more severe coordination problems faced by rebels from outside the current elite. Authority depends on a chain of command, where individuals follow instructions in expectation of punishment from others if they disobey. The rebels confront the challenge of persuading enough individuals that they should fear punishment for disobeying them rather than the incumbents people are accustomed to.[14]

The assumptions of the model allow for coordination among individuals in launching rebellions, but subject to some restrictions. These restrictions are intended as a simple representation of the plausible constraints that ought to be placed on the set of possible deals or "contracts" among the rebels. The fundamental contracting problem is the issue of enforceability in a world with no exogenous commitment technologies. The rebellion mechanism is intended to be as flexible as possible in allowing for "deals" subject to the limits of enforceability.

The rebellion "contract" implicit in the model requires a prescribed amount of fighting effort from each rebel in return for a position sufficiently high up the pecking order to yield a place in the new elite. Formally, this deal is represented by the new elite selection function together with the restriction that the rebel army is limited to those who expect a place in the subsequent elite. This rebellion contract raises two questions. Why are other forms of contract ruled out? And what suggests a contract of this form is not susceptible to enforceability concerns?

[13] The inelasticity assumption can also be viewed as a requirement of internal consistency given the deterministic nature of the conflict technology. If the incumbents know they will be defeated, it is difficult to rationalize their exerting any current fighting effort.

[14] Under this interpretation, the successful rebellion condition in [3.7] can be seen as the rebels' effort requirement to demonstrate that they have the strength and the organization to overcome these problems. Once the incumbents see this tipping point is reached, they surrender and no actual fighting takes place.

Consider a group of individuals with incentives to come together and agree to fight, enabling new institutions to be established that offer each of them a better payoff than what they currently receive. The maximum fighting effort each would be willing to agree to is equal to each's expected utility gains. There are two facets of the enforceability problem for this "deal". First, after the success of the rebellion, will there be incentives to create the particular institutions that were agreed beforehand? Second, will individual rebels honour their agreed levels of fighting effort?

A rebellion contract prescribing exactly what institutions are to be set up faces formidable enforceability problems. Once the current institutions are overthrown, there is no higher authority (no "meta-institutions") that can compel the group now in power to act against its interests ex post. This type of contract is therefore ruled out. The new institutions must maximize elite payoffs[15] starting afresh from the world as they now find it, unconstrained by history except for fundamental state variables.[16] Any bygones will be bygones. In particular, this rules out the trigger strategies of repeated games as commitment devices.[17]

As Proposition 1 shows, the elite has a strict preference to avoid payoff inequality except where it is matched by differences in power. It follows that the new elite would have incentives not to honour contracts that specified either transfers to those who contributed fighting effort during the rebellion or fines for those who did not, hence such contracts cannot be written ex ante.

Now consider the second aspect of the enforceability problem. Taking as given the payoff improvement an individual expects if a rebellion succeeds (subject to the restrictions on what can be agreed in advance regarding the new institutions), will it be possible to enforce the agreed amount of fighting effort from an individual who is a party to a rebellion contract? The basic problem is that each atomistic individual (correctly) does not perceive himself as pivotal in determining whether the rebellion succeeds. Thus, left to his own devices, he would have an incentive to shirk and free-ride on others' fighting effort. To a large extent, rebel armies may be able to control individual members through internal discipline, but a non-negligible enforceability problem remains when some necessary fighting is done at an individual's discretion.

To ensure that all agreed fighting effort is actually exerted, there needs to be a credible punishment that can be imposed on shirkers after the fact. However, according to Proposition 1, only differences in payoffs that reflect differences in power are in the interests of the elite ex post. This suggests that the offer of a position of power conditional on the requisite amount of fighting effort can provide a credible incentive not to shirk.[18] While the rebellion contract cannot determine the

[15] In the model, institutions being in the interests of the elite is interpreted to mean maximizing the average elite payoff. Moving away from this simplifying assumption would require modelling the hierarchy and power relations within the elite. See Myerson (2008) for a model which addresses that question.

[16] For simplicity, the occurrence of conflict does not itself affect any fundamental state variables. This implicitly assumes members of the rebel army can be demobilized costlessly once the fighting is over if, off the equilibrium path, there were more rebels than places in the new elite. Adding a cost of demobilization would make the size of the previous rebellion a state variable at the stage new institutions are created, which would add a significant complication to the model without obviously delivering any new insights.

[17] Models that allow trigger strategies face the problem of multiple equilibria because there is always a range of possible punishments consistent with equilibrium.

[18] Suppose that a fraction ξ of an individual's agreed fighting effort (associated with some necessary tasks) cannot be directly enforced at the time through the rebel army's own discipline mechanisms, but that the full amount of fighting effort F must be exerted otherwise the individual does not obtain fighting strength F. Suppose also that

total number of positions of power under the new institutions, it can control the identities of those who will receive these positions. Formally, this is the elite selection function in the model.[19] But for those who anticipate becoming workers, even if they gain from the success of the rebellion, there is no worse position they can be credibly assigned if they fail to put in the agreed level of fighting effort. There is nothing to prevent such individuals from shirking,[20] which leads to the restriction that all rebels must expect a place in the subsequent elite.

4 Public goods

In the model of section 3 there was no scope for the elite to do what governments are customarily supposed to do, such as provide public goods. This section introduces a technology that allows for production of public goods. The elite can set up institutions that decree spending on public goods and make use of tax revenue to pay for this. It is then natural to ask whether resources will be efficiently allocated to public-good production.

The new technology converts units of output into public goods. If g units of goods per person are converted then everyone receives an extra $\Gamma(g)$ units of the consumption good. The function $\Gamma(\cdot)$ is strictly increasing, strictly concave, and satisfies the usual Inada conditions. The definition of institutions $\mathscr{I}$ from section 3.2 is augmented to specify public-good provision g, hence $\mathscr{I} = \{\mathcal{P}, \mathcal{W}, \tau(\imath), C_{\mathrm{p}}(\imath), g\}$. All individuals observe the choice of g and take it into account — along with all other aspects of the institutions — when deciding whether to participate in a rebellion, and if so, the amount of fighting effort to exert. The model is otherwise identical to that of section 3.

The utility of an individual is now

$$\mathcal{U} = u(C) - F, \quad \text{with } C = c + \Gamma(g), \qquad [4.1]$$

where C denotes the individual's overall consumption, comprising private consumption c and the consumption $\Gamma(g)$ each person obtains from the public good. Given the overall budget constraint,

each individual's total fighting effort is verifiable after the rebellion. After agreeing to the rebellion contract, the individual is directly compelled to exert effort $(1-\xi)F$. If he exerts the remaining effort ξF then he receives his place in the subsequent elite with continuation payoff $\mathcal{U}'_{\mathrm{p}}$. If he shirks and exerts no further effort, he is demoted to worker status and receives payoff $\mathcal{U}'_{\mathrm{w}}$. Therefore, for individual $\imath$ to join the rebel army, incentive compatibility requires $\mathcal{U}'_{\mathrm{p}} - \mathcal{U}'_{\mathrm{w}} \geq \xi F(\imath)$, while the maximum-effort participation constraint is $\mathcal{U}'_{\mathrm{p}} - \mathcal{U}(\imath) \geq F(\imath)$. If ξ is positive but sufficiently small then the incentive compatibility constraint is satisfied but not binding, while the participation constraint binds, as was implicitly assumed in the description of the rebellion mechanism.

[19] The particular elite selection function in place has no effect on the maximum attainable elite payoff, so carrying out the punishment does not affect the others in the elite. Intuitively, since all individuals are ex ante identical, individuals in power do not care about the identities of those with whom they share power, only the total number of such people. Without a device such as an elite selection function, there would be a fundamental indeterminacy regarding the identities of the elite, which would need to be resolved in some other way.

[20] The incentive compatibility constraint discussed in footnote 18 would be violated for these individuals with any positive ξ, no matter how small. One alternative approach that could incentivize more individuals not to shirk offers a lottery in return for an agreed amount of fighting effort, where the prize is a place in the elite. While this mechanism could induce fighting effort from more individuals, the amount of effort each individual would agree to is lower because the lottery is less valuable than a place in the elite with certainty.

consumption per person is

$$pC_{\mathrm{p}} + (1-p)C_{\mathrm{w}} = (1-p)\mathrm{q} - g + \Gamma(g). \qquad [4.2]$$

A benevolent social planner would choose the first-best level of public-good provision $g = \hat{g}$, determined by $\Gamma_g(\hat{g}) = 1$, which maximizes the total amount of goods available for consumption.

The payoff-equalization insights of Proposition 1 continue to apply to this new environment, hence it is possible without loss of generality to focus on institutions specifying the size of the elite p (with identities determined by an elite selection function), the tax τ levied on all workers, and the necessarily common public-good provision g. The budget constraint [4.2] can be used to find the consumption levels of a worker and a member of the elite under a particular set of institutions:

$$C_{\mathrm{w}} = \mathrm{q} - \tau + \Gamma(g), \quad \text{and} \quad C_{\mathrm{p}} = \frac{(1-p)\tau - g}{p} + \Gamma(g). \qquad [4.3]$$

The argument of Proposition 1 that the equilibrium institutions can be characterized by imposing only the no-rebellion constraint for a rebel army composed entirely of workers also carries over to this new environment. Thus, the equilibrium institutions are the solution of

$$\max_{p,\tau,g} u\left(\frac{(1-p)\tau - g}{p} + \Gamma(g)\right) \text{ s.t. } \mathcal{U}'_{\mathrm{p}} - u\left(\mathrm{q} - \tau + \Gamma(g)\right) \leq \delta \frac{p}{p'}, \qquad [4.4]$$

with beliefs p' and $\mathcal{U}'_{\mathrm{p}}$ taken as given, but with $p' = p^*$, $\tau' = \tau^*$ and $g' = g^*$ in equilibrium. Setting up the Lagrangian for this problem with multiplier χ on the no-rebellion constraint yields first-order conditions for τ and g:

$$\frac{u_C(C^*_{\mathrm{p}})}{\chi^* u_C(C^*_{\mathrm{w}})}\left(\frac{1-p^*}{p^*}\right) = 1, \quad \text{and} \quad \frac{u_C(C^*_{\mathrm{p}})}{\chi^* u_C(C^*_{\mathrm{w}})}\left(\frac{1}{p^*} - \Gamma_g(g^*)\right) = \Gamma_g(g^*). \qquad [4.5]$$

By eliminating the term $u_C(C^*_{\mathrm{p}})/\chi^* u_C(C^*_{\mathrm{w}})$ from the equations above, public-good provision g^* under the equilibrium institutions is determined by $\Gamma_g(g^*) = 1$. This is identical to the condition for the provision $\hat{g}$ by a benevolent social planner, so $g^* = \hat{g}$. The equilibrium institutions are therefore economically efficient in respect of public-good production.[21]

To understand this result, observe that the no-rebellion constraint implies the elite cannot disregard the interests of the workers, even though it does not care about them directly. Provision of the public good slackens the no-rebellion constraint, while the taxes raised to finance it tighten the constraint. By optimally trading off the benefits of the public good against the cost of production, the elite effectively maximizes the size of the pie, making use of transfers to ensure everyone is indifferent between rebelling or not. The efficiency result can be seen as a "political" analogue of the Coase theorem,[22] where the contestability of institutions through rebellions plays the role of legal

[21]The distribution of total output between workers and elite members depends on the other parameters of the model, including the utility function $u(\cdot)$. In equilibrium, all individuals will receive a higher overall payoff as a result of the public-good technology becoming available, though in general, the benefits will not be shared equally.

[22]See Acemoglu (2003) for a discussion of this analogy.

property rights. The ability of institutions to make transfers between individuals is crucial to this finding. The institutional technology permits any transfers, but more importantly, the constraints imposed by the power struggle do not interfere with the transfers needed here.

The analysis shows that although the elite is extracting rents from workers, this does not preclude it from acting as if it were benevolent in other contexts. Hence, the overall welfare of workers might be larger or smaller compared to a world in which no-one can compel others to act against their will. This reflects the ambivalent effects on ordinary people of having a ruling elite.[23]

The result found in this section is far from surprising and can be obtained in several other settings, as discussed by Persson and Tabellini (2000) in the context of voting and elections. Here the result provides a benchmark case where the equilibrium institutions are economically efficient.

5 Investment

This section adds the possibility of investment to the analysis of equilibrium institutions. Individuals can now exert effort to obtain a greater quantity of goods, but there is a time lag between the effort being made and the fruits of the investment being received. During this span of time, there are opportunities for rebellion against the prevailing institutions. The model is otherwise identical to that of section 3. In particular, there are no changes to the mechanism through which institutions are created and destroyed. However, if investment occurs then this changes incentives for rebellion, and thus affects the elite's design of institutions. The following analysis considers how the equilibrium institutions will provide incentives for individuals to invest, and to what extent it will be done — in particular, whether these institutions will achieve economic efficiency.

5.1 Environment

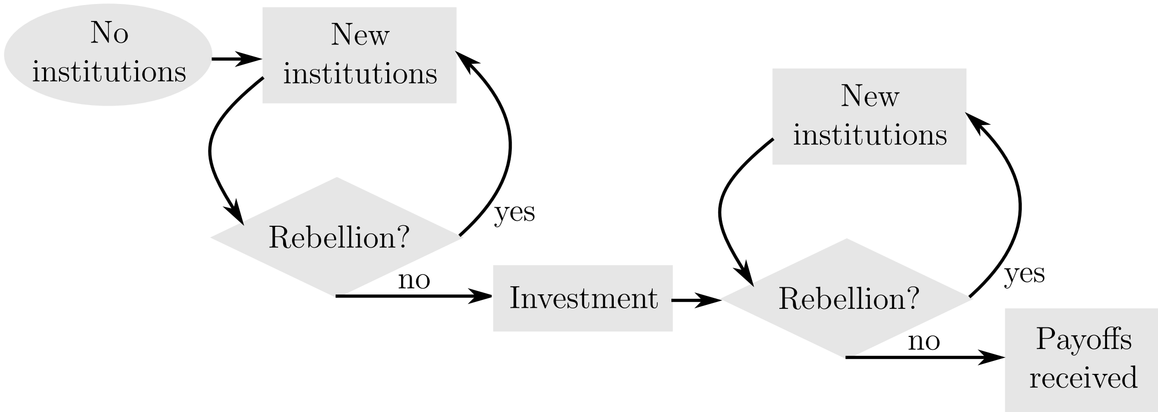
The sequence of events is depicted in Figure 3. Before any investment decisions are made, institutions are first established through a process identical to that described in section 3 (compare Figure 1). Institutions can now specify transfers contingent on whether individuals are holding capital (i.e. taxes on capital), as well as determine the size of the group in power and make other transfers. Once institutions are established that do not immediately trigger rebellion, there are opportunities to invest. After investment decisions are made, there is another round of opportunities for rebellion, with new institutions established if a rebellion occurs. When the prevailing institutions do not lead to any further rebellions, what those institutions specify is implemented and payoffs are received.

Individuals who are in power (members of the elite, denoted by $\imath \in \mathcal{P}$) have fighting strength δ in defence, as in the model of section 3. *Economically active* individuals (those *not* in the elite, denoted by $\imath \in \mathcal{N}$) at the post-investment stage receive an endowment of q units of goods.

There are μ *investment opportunities.* An investment opportunity is the option to produce κ units of *capital* in the future in return for incurring a present effort cost θ (in utility units), which is

[23] This trade-off is mentioned in the Bible (1 Samuel 8:10–20). The people want a king to provide public goods, despite being warned by the prophet Samuel that the king would use his power in his own interests. Many centuries later, in far too many cases, the warnings of Samuel remain as relevant as ever.

Figure 3: *Sequence of events with investment*



sunk by the time the capital is produced. Capital here simply means more units of the consumption good. All investment opportunities yield the same amount of capital κ, but each features an effort cost that is an independent draw from the distribution

$$\theta \sim \text{Uniform}\left[\psi, \kappa\right], \qquad [5.1]$$

where $0 < \psi < \kappa$, with ψ being the minimum effort cost.[24] An individual's receipt of an investment opportunity, the required effort cost θ, and whether the opportunity is taken, are private information, while possession of capital is common knowledge.[25] It is further assumed that investment opportunities are received only by those economically active individuals outside the elite,[26] and that investment opportunities are randomly assigned at the investment stage with no-one having prior knowledge of whether he will receive one, nor the required effort cost θ if so.[27]

An individual's utility $\mathcal{U}$ is now

$$\mathcal{U} = u(C) - \theta I - F, \qquad [5.2]$$

where $I \in \{0, 1\}$ is an indicator variable for whether an investment opportunity is received and

[24]The uniform distribution is chosen for simplicity. The choice of distribution does not affect the qualitative results.

[25]Since investment opportunities are private information when taken, it is not feasible to have institutions specifying a "command economy" where individuals perform investments by decree, nor is it feasible to make holding power contingent on receiving and taking an investment opportunity.

[26]Allowing the elite to invest adds extra complications to the model. It might be thought important to have investors inside the elite to provide appropriate incentives. As will be seen, this is not the case, and moreover, the advantage of having investors inside the elite most likely applies to the case where they are brought into the elite *before* they decide whether to take investment opportunities. Given the information restrictions, which represent the not implausible difficulties of identifying talented investors in advance, bringing them in at that stage is not feasible.

[27]This modelling device places individuals behind a "veil of ignorance" about their talents as investors when the pre-investment stage institutions are determined. Doing this avoids having to track whether talented investors are disproportionately inside or outside the elite, which would add a (relevant) state variable to the problem of determining the pre-investment stage institutions, significantly complicating the analysis. However, it will turn out that the no-rebellion constraints are slack at the pre-investment stage, so this assumption need not significantly affect the results.

taken, and F denotes any fighting effort, as in the model of section 3.

The following parameter restrictions are imposed:

$$\frac{\delta}{q} \leq \varphi \equiv \frac{1+\sqrt{5}}{2}, \quad \mu \leq \frac{q}{2(q+2\delta)}, \quad \kappa < \delta, \quad \text{and} \;\; u(C) = C. \qquad [5.3]$$

The first restriction is the bound from section 3.6 needed to ensure non-negativity constraints are always slack in equilibrium. The second restriction states that the measure μ of individuals who receive an investment opportunity is not too large, which ensures that capitalists are not the predominant group by numbers.[28] The third restriction places a physical limit on the economy's maximum capital stock. Finally, individuals' preferences are assumed to be linear in consumption for analytical tractability. This allows for a simple closed-form solution, but it is expected that similar results would be found for the general class of concave utility functions.

5.2 Equilibrium institutions

Characterizing the equilibrium institutions requires working backwards, determining the new equilibrium institutions if a rebellion were to occur at the post-investment stage, and then analysing what institutions will be chosen at the pre-investment stage. The elite at the pre-investment stage will want to choose institutions that survive rebellion at all points.

5.2.1 Post-investment stage institutions after a rebellion

Suppose a rebellion occurs at some point after investment decisions have been made. Let $K(\imath)$ denote the capital currently held by individual $\imath$. The effort cost θ of investing is now sunk, so the continuation value of utility $\mathcal{U} = C - F$ is the same for both an expropriated capitalist and an individual who never possessed any capital to begin with. Since holdings of capital are common knowledge, an argument similar to Proposition 1 shows that the institutions chosen by the elite in the unique Markovian equilibrium would equalize continuation payoffs for economically active individuals outside the elite ($\imath \in \mathcal{N}$). This means that any notional claims to capital will be set aside and individuals' payoffs will be determined according to their power, with capital redistributed accordingly under a new set of institutional rules.

Let τ_q denote the net tax paid by an economically active individual (one receiving the endowment q) independent of the individual's holdings of capital, and $\tau_\kappa(\imath)$ the tax on capital paid by individual $\imath$. For a general distribution of capital, payoff equalization requires $\tau_\kappa(\imath) = K(\imath)$, that is, a 100% tax on capital. The budget constraint faced by the elite is then

$$pC_p = (1-p)\tau_q + \int_\Omega \tau_\kappa(\imath)\mathrm{d}\imath = (1-p)\tau_q + K, \quad \text{with} \;\; K = \int_\Omega K(\imath)\mathrm{d}\imath,$$

[28]If capitalists were to become the predominant group then the nature of the binding constraints might change and the problem would become significantly more algebraically convoluted. While this analysis could in principle add some twists to the results, it would not affect any of the conclusions in this paper, so it is left for future research.

where K denotes the total capital stock, and C_{p} is the consumption of each member of the elite (the equilibrium features payoff equalization among the elite, using the argument of Proposition 1). Given the consumption C_{n} the $1-p$ economically active individuals outside the elite, the consumption C_{p} of the p elite members can be found using the budget constraint. The levels of utility are:

$$\mathcal{U}_{\mathrm{n}} = C_{\mathrm{n}} = (\mathrm{q} - \tau_{\mathrm{q}}) + (K(\imath) - \tau_{\kappa}(\imath)) = \mathrm{q} - \tau_{\mathrm{q}}, \quad \text{and } \mathcal{U}_{\mathrm{p}} = C_{\mathrm{p}} = \frac{(1-p)(\mathrm{q} - \mathcal{U}_{\mathrm{n}}) + K}{p}. \qquad [5.4]$$

Using the argument of Proposition 1, the institutions in the Markovian equilibrium following a rebellion at the post-investment stage can be characterized by maximizing the elite payoff in [5.4] subject to a single no-rebellion constraint

$$\mathcal{U}_{\mathrm{n}} \leq \mathcal{U}_{\mathrm{p}}'(K) - \delta \frac{p}{p'(K)},$$

where $p'(K)$ and $\mathcal{U}_{\mathrm{p}}'(K)$ are the beliefs about the subsequent institutions following a further rebellion.[29] In a Markovian equilibrium, these beliefs may be functions of the total capital stock K, which is a relevant state variable here (the model of section 3 featured no relevant state variables). Given payoff equalization, the choice variables are the lump-sum tax τ_{q} and the elite size p (with identities of the elite determined by the elite selection function given the elite size p).

Solving this constrained maximization problem and then imposing the Markovian equilibrium conditions $p(K) = p'(K)$ and $\mathcal{U}_{\mathrm{p}}(K) = \mathcal{U}_{\mathrm{p}}'(K)$ leads to a unique equilibrium. The equilibrium values of each variable are as follows, denoted by a † superscript:

$$p^{\dagger} = \frac{\delta}{\mathrm{q} + 2\delta}, \quad \mathcal{U}_{\mathrm{p}}^{\dagger}(K) = \frac{(\mathrm{q}+\delta)^2}{\mathrm{q}+2\delta} + K, \quad \text{and } \mathcal{U}_{\mathrm{w}}^{\dagger}(K) = \frac{(\mathrm{q}+\delta)^2}{\mathrm{q}+2\delta} - \delta + K. \qquad [5.5]$$

The equilibrium elite size $p^{\dagger}$ is independent of K and is the same as that found in the endowment economy model with linear utility from section 3.6.[30] The results show that were a rebellion to occur at the post-investment stage, the entire capital stock would be expropriated and equally distributed among the whole population. The elite chooses to redistribute the capital among all individuals because the presence of the capital increases incentives for further rebellions.

5.2.2 Pre-investment stage institutions

Institutions chosen at the pre-investment stage specify the set of individuals in power, denoted by $\mathcal{P}$, and the set of economically active individuals outside the elite, denoted by $\mathcal{N}$. Depending on the institutions and the arrival of investment opportunities, some of the individuals in $\mathcal{N}$ may become *investors*, denoted by $\mathcal{I}$. Those economically active individuals who do not become investors are referred to as *workers*, denoted by $\mathcal{W}$.

It is shown formally in Proposition 2 below that the elite's payoff is strictly lower when there

[29] Under the parameter restrictions in [5.3], all non-negativity constraints on consumption are slack, the constraint $p < 1/2$ is slack, and the participation constraint $\mathcal{U}_{\mathrm{p}}^{\dagger}(K) > \mathcal{U}_{\mathrm{w}}^{\dagger}(K)$ for the incumbent army is slack.

[30] This analytically convenient finding is owing to the linearity of utility in consumption.

is inequality in payoffs among workers or among members of the elite (but differences in power justify inequality between these groups). Moreover, there is no loss of generality in considering institutions that give all investors the same level of consumption.[31] It follows that attention can be restricted to institutions specifying a constant tax τ_κ on those holding capital, a constant tax τ_q on all economically active individuals (workers and investors), and an elite size p.[32]

In equilibrium, institutions will survive rebellion at all points, so the incentive to take an investment opportunity is assessed under the hypothesis that the capital tax τ_κ and all other rules specified by the current institutions will be implemented. If an investment opportunity is not received or is not taken then an individual becomes a worker. Any economically active individual $\imath$ holding capital $K(\imath) = \kappa$ also receives the endowment q, but now pays total tax $\tau_q + \tau_\kappa$. Thus, all investors have consumption $C_i = q + \kappa - \tau_q - \tau_\kappa$. The utilities of a worker ($\imath \in \mathcal{W}$) and an investor ($\imath \in \mathcal{I}$) who takes an opportunity with effort cost θ are:

$$\mathcal{U}_w = C_w = q - \tau_q, \quad \text{and } \mathcal{U}_i(\theta) = C_i - \theta = (q - \tau_q) + (\kappa - \tau_\kappa) - \theta. \qquad [5.6]$$

An individual with an investment opportunity takes it if $\mathcal{U}_i(\theta) \geq \mathcal{U}_w$, which is equivalent to the effort cost θ being not more than a threshold $\tilde{\theta}$, where $\mathcal{U}_i(\tilde{\theta}) = \mathcal{U}_w$. The proportion of those receiving an investment opportunity who take it is denoted by s, which is calculated given $\tilde{\theta}$ using the uniform distribution for θ specified in [5.1]. Expressions for these variables and the total measure i of investors and the total capital stock K are:

$$\tilde{\theta} = \kappa - \tau_\kappa, \quad s = \mathbb{P}_\theta[\theta \leq \tilde{\theta}] = \frac{\tilde{\theta} - \psi}{\kappa - \psi}, \quad i = \mu s, \quad \text{and } K = i\kappa. \qquad [5.7]$$

The elite aims to design institutions that survive opportunities for rebellion both before and after investment decisions are made. Potential rebellions at these stages need to be considered separately owing to the change in the environment as a result of the presence of capital and the revelation of information after the investment stage.

At the post-investment stage, there are three groups of individuals: workers ($\mathcal{W}$), those in power ($\mathcal{P}$), and investors ($\mathcal{I}$) who have already incurred a sunk effort cost, and who are now referred to as *capitalists*. Workers receive the payoff $\mathcal{U}_w$ given in equation [5.6]. All capitalists receive the same level of consumption C_i and so have the same continuation payoff $\mathcal{U}_k$. The budget constraint can then be used to find the consumption of those in the elite. The utilities of capitalists and elite

[31]The arguments relating to the distribution of consumption among investors are different from those for workers or elite members. As will be seen, providing incentives to investors implies that their continuation payoffs at the post-investment stage must be higher than those of workers. Since these individuals have identical power (investors are outside the elite), investors will not be included in the rebel army with the greatest fighting strength. As a result, when the utility function $u(\cdot)$ is linear, there is no equivalent result stating that ex-post inequality among investors' payoffs reduces the payoff of the elite. However, given that capital κ is common to all investors and the effort cost θ is private information, there is no loss of utility to the elite from choosing a constant capital tax τ_κ to be paid by all investors, nor indeed would this restriction have consequences for any aggregate variables. Given linearity of $u(\cdot)$, institutions specifying lotteries of capital taxes for investors could also maximize the elite's payoff if the dispersion of taxes were sufficiently small, but this would not change any of the results regarding the aggregate amount of investment in equilibrium.

[32]The identities of the individuals in the elite are determined by the elite selection function given the elite size p.

members are:

$$\mathcal{U}_{\text{k}} = C_{\text{i}} = (\text{q} - \tau_{\text{q}}) + (\kappa - \tau_{\kappa}), \quad \text{and } \mathcal{U}_{\text{p}} = C_{\text{p}} = \frac{(1-p)\tau_{\text{q}} + i\tau_{\kappa}}{p}. \qquad [5.8]$$

In principle, rebel armies could comprise any of these groups of individuals, or any mixture of them.

At the pre-investment stage, there are only two groups of distinct individuals: those in power ($\mathcal{P}$), and those economically active individuals ($\mathcal{N}$) outside the elite (who do not know yet whether they will become workers or investors). Let $\mathcal{U}_{\text{n}}$ denote the expected utility of an economically active individual under the current institutions. If α denotes the probability that such an individual will receive an investment opportunity, his expected payoff is given by:[33]

$$\mathcal{U}_{\text{n}} = (1-\alpha)\mathcal{U}_{\text{w}} + \alpha \mathbb{E}_{\theta} \max\{\mathcal{U}_{\text{i}}(\theta), \mathcal{U}_{\text{w}}\}, \quad \text{where } \alpha = \frac{\mu}{1-p}. \qquad [5.9]$$

This payoff can be written in terms of the expected surplus $\mathcal{S}_{\text{i}}(\tilde{\theta})$ of those receiving an investment opportunity:

$$\mathcal{U}_{\text{n}} = (\text{q} - \tau_{\text{q}}) + \alpha \mathcal{S}_{\text{i}}(\tilde{\theta}), \quad \text{where } \mathcal{S}_{\text{i}}(\tilde{\theta}) \equiv \mathbb{E}_{\theta} \max\{\tilde{\theta} - \theta, 0\}. \qquad [5.10]$$

Incentives to rebel depend on what payoffs the rebels expect to receive under the post-rebellion institutions. For a rebellion at the post-investment stage, the unique Markovian equilibrium institutions that would be set up after the rebellion have already been characterized as a function of the aggregate capital stock K. The subsequent elite would be of size $p^{\dagger}$ and receive utility $\mathcal{U}_{\text{p}}^{\dagger}(K)$ as given in equation [5.5], with the capital stock K predetermined according to [5.7]. At the pre-investment stage, there will be beliefs p' and $\mathcal{U}_{\text{p}}'$ about what elite size and elite payoff would prevail under the institutions formed after a rebellion at that stage. These beliefs will be determined using the restriction to Markovian choices of institutions.

The problem of choosing institutions is thus

$$\max_{p,\tau_{\text{q}},\tau_{\kappa}} \mathcal{U}_{\text{p}} \;\text{ s.t. }\; \sigma_{\text{n}} \max\{\mathcal{U}_{\text{p}}' - \mathcal{U}_{\text{n}}, 0\} + \sigma_{\text{p}}(\mathcal{U}_{\text{p}}' - \mathcal{U}_{\text{p}} + \delta)\mathbb{1}[\mathcal{U}_{\text{p}}' \geq \mathcal{U}_{\text{p}}] \leq \delta \frac{p}{p'} \;\text{ and } \qquad [5.11]$$

$$\sigma_{\text{w}}^{\dagger} \max\{\mathcal{U}_{\text{p}}^{\dagger}(K) - \mathcal{U}_{\text{n}}, 0\} + \sigma_{\text{p}}^{\dagger}(\mathcal{U}_{\text{p}}^{\dagger}(K) - \mathcal{U}_{\text{p}} + \delta)\mathbb{1}[\mathcal{U}_{\text{p}}^{\dagger}(K) \geq \mathcal{U}_{\text{p}}] + \sigma_{\text{i}}^{\dagger} \max\{\mathcal{U}_{\text{p}}^{\dagger}(K) - \mathcal{U}_{\text{k}}, 0\} \leq \delta \frac{p}{p^{\dagger}},$$

for all possible σ_{n}, σ_{p}, $\sigma_{\text{w}}^{\dagger}$, $\sigma_{\text{p}}^{\dagger}$, and $\sigma_{\text{i}}^{\dagger}$, where these coefficients indicate, respectively, the proportions of economically active and elite members in a pre-investment stage rebel army, and the proportions of workers, elite members, and investors in a post-investment stage rebel army. These non-negative coefficients must satisfy the natural restrictions $\sigma_n \leq (1-p)/p'$, $\sigma_p \leq p/p'$, and $\sigma_{\text{n}} + \sigma_{\text{p}} = 1$, together with $\sigma_{\text{w}}^{\dagger} \leq w/p^{\dagger}$, $\sigma_{\text{p}}^{\dagger} \leq p/p^{\dagger}$, $\sigma_{\text{i}}^{\dagger} \leq i/p^{\dagger}$, and $\sigma_{\text{w}}^{\dagger} + \sigma_{\text{p}}^{\dagger} + \sigma_{\text{i}}^{\dagger} = 1$.

The following result confirms the payoff-equalization claims made earlier and characterizes which of the many possible no-rebellion constraints are binding in equilibrium.[34]

[33] Note that the parameter restrictions in [5.3] imply $\mu < 1/2$, and since $p < 1/2$, the number of economically active individuals is always more than 50%, and hence more than the number of investment opportunities. This justifies the formula for i in [5.7] and the formula for α in [5.9].

[34] It is verified later that the non-negativity constraint $C_{\text{w}} \geq 0$, the constraint $p < 1/2$, and the participation

Proposition 2 *Consider an arbitrary choice of the capital tax τ_κ, which determines $\tilde{\theta}$ and s according to* [5.7]*. Any Markovian equilibrium with $s > 0$ must have the following features:*

(i) Payoff equalization among all workers, and payoff equalization among all elite members.

(ii) All no-rebellion constraints at the pre-investment stage are slack. Two no-rebellion constraints at the post-investment stage are binding, namely $(\sigma^\dagger_{\mathrm{w}}, \sigma^\dagger_{\mathrm{p}}, \sigma^\dagger_{\mathrm{i}}) = (1, 0, 0)$ and $(\sigma^\dagger_{\mathrm{w}}, \sigma^\dagger_{\mathrm{p}}, \sigma^\dagger_{\mathrm{i}}) = (0, 1, 0)$ (as well as any constraints where $\sigma^\dagger_{\mathrm{w}} + \sigma^\dagger_{\mathrm{p}} = 1$, which are satisfied as a consequence of those two).

(iii) The binding no-rebellion constraints imply that

$$p = p^\dagger + \frac{\mu\tilde{\theta}s}{\delta}, \qquad [5.12a]$$

so incentives for investment require a larger elite size p at the pre-investment stage than the size $p^\dagger$ that would be optimal at the post-investment stage.

(iv) Given the elite size from [5.12a]*, the payoff of a member of the elite is*

$$\mathcal{U}_{\mathrm{p}} = \frac{(\mathrm{q}+\delta)^2}{\mathrm{q}+2\delta} + \mu\left(\kappa - \left(\frac{\mathrm{q}+2\delta}{\delta}\right)\tilde{\theta}\right)s. \qquad [5.12b]$$

PROOF See appendix A.2. ■

The proposition shows that no investor belongs to a rebel army associated with a binding no-rebellion constraint. The basic reason is that providing incentives to investors means granting them higher consumption than workers, and thus higher utility ex post once the sunk effort cost of investing has already been incurred (ex ante, the marginal investor has the same utility as a worker). This can be seen by combining equations [5.6], [5.7] and [5.8] to obtain $\mathcal{U}_{\mathrm{k}} = \mathcal{U}_{\mathrm{w}} + \tilde{\theta}$, where $\tilde{\theta} > 0$ when $s > 0$. The analysis of section 5.2.1 shows that the institutions following a rebellion at the post-investment stage will not respect individual holdings of capital prior to the rebellion. Thus, what investors stand to receive following a rebellion (net of fighting costs) is no different from that of workers who rebel (their power is identical), while what they lose is superior. Accordingly, they are less willing to fight to replace the current institutions. This implies the distribution of income needed to provide incentives to invest is not one that investors themselves could enforce by a credible threat to participate in rebellions.

The fundamental problem is that the distribution of income needed to encourage investment diverges from that consistent with the distribution of power, and so there are incentives for groups to rebel against institutions prescribing property rights. As usual, the no-rebellion constraint is binding for workers since the elite gains by extracting as much as possible from them. What is novel here is that discouraging rebellion by workers is no longer sufficient in the presence of investment opportunities: the elite must also worry about rebellion from within its own ranks. The elite would

constraints $\mathcal{U}_{\mathrm{p}} > \mathcal{U}'_{\mathrm{w}}$ and $\mathcal{U}_{\mathrm{p}} > \mathcal{U}^\dagger_{\mathrm{w}}(K)$ for the incumbent army are all slack.

like to design institutions encouraging investment by not taxing capital too heavily, but there is also the temptation ex post for elite members to participate in a rebellion that will allow them to create new institutions permitting full expropriation of capital. The fact that the effort cost of investment is sunk gives rise to a time-inconsistency problem, which is reflected in the threat of rebellion coming from inside as well as outside the elite.[35]

Given this time-inconsistency problem, it might be thought impossible to sustain any investment in equilibrium because individuals cannot commit not to rebel. Since the defence of the current institutions relies on those in power, a rebellion backed by all members of the elite succeeds without requiring any fighting effort. If the elite size p were equal to $p^\dagger$, those in power would be able to destroy the current institutions through a costless "suspension of the constitution", allowing new institutions to be created while leaving the current elite members in power, essentially granting them full discretion ex post to rewrite the rules completely. However, if $p > p^\dagger$, costless suspension of the constitution is not possible. The equilibrium elite size after the rebellion is smaller than beforehand, so some members of the existing elite must lose their positions. The rebellion launched by insiders is now necessarily a "coup d'état" that shrinks the elite. Conflict with those elite members who lose their positions makes this a costly course of action.

The formal analysis in Proposition 2 confirms that satisfaction of the no-rebellion constraints for workers and elite members is equivalent to ensuring the pre-investment elite size p is large in relation to $p^\dagger$, the equilibrium elite size after a rebellion at the post-investment stage. As the proportion s of investors rises, the required elite size p increases. The choice of capital tax τ_κ (which determines s via equation [5.7]) can be interpreted broadly as revealing the extent of protection of private property against expropriation (whether directly, or indirectly through taxes). The elite size p can be interpreted as how widely members of the elite choose to share power. The claim in [5.12a] is that credible limits on expropriation require more power sharing than what would be optimal for an elite member after investment decisions have actually been made.

The proposition shows that not only is this increase in power sharing *sufficient* for credible protection of property rights; it is also *necessary*. There is no other design of institutions that can both establish credible incentives for investors and survive the power struggle. In particular, it might be thought possible to solve the problem through some system of taxes and transfers. But discouraging rebellion by workers would require *lower* taxes, while discouraging rebellion by members of the elite would require *higher* taxes. Further taxes on investors would of course destroy the very incentives that must be preserved. The only way to discourage rebellion from both inside and outside the elite simultaneously is an increase in the elite size. Fundamentally, transfers are a zero-sum game, and can only redistribute disgruntlement with the current institutions.[36]

Sharing power among a wider group of individuals therefore allows the elite to act as a government committed to policies that would otherwise be time inconsistent. Even though all individuals act

[35] The no-rebellion constraint for the elite places a lower bound on the elite payoff $\mathcal{U}_\mathrm{p}$ even though the institutions are set up to maximize $\mathcal{U}_\mathrm{p}$ ex ante. The constraint then represents the absence of incentives to deviate from the initial institutions through rebellion ex post.

[36] On the other hand, the notion of being in power is essentially an ability to impose costs on others when fighting occurs at a lower cost to oneself.

with discretion, overcoming the time-inconsistency problem is feasible. Sharing power thus emerges endogenously as a commitment device. It provides a solution to the classic problem of "who will guard the guardians?": institutions can be protected from those who hold power when some of them fear losing their privileged status if the institutions are destroyed from within.[37]

This analysis rests on two key assumptions made earlier about institutions and the power struggle. First, institutions do lay down the allocation of resources and power that is respected unless they are overthrown by a rebellion. For institutions to function in supporting cooperation among individuals who have chosen to stop fighting, it is essential that particular individuals or groups cannot arbitrarily modify aspects of those institutions while managing at the same time to avoid calling into question the whole structure. This assumption is necessary for the sheer existence of an environment where rules are followed. It captures the idea that elites can in principle create institutions that specify known taxes to which individuals acquiesce when they decide not to rebel.

Second, there is no means of enforcing commitments made at the time of rebellion to take actions that are not optimal after the rebellion has succeeded. The underlying idea is that institutions can implicitly support deals between individuals, but there are no "meta-institutions" to enforce deals concerning the choice of institutions themselves. This means that following the destruction of a set of institutions, there must be optimization over *all* possible dimensions of the new institutions. In particular, following a rebellion after investment decisions have been made, the degree of power sharing is reoptimized in addition to the tax system.

Once individuals have incurred the sunk effort costs of investing, those in power would like to be able to sign a "rebellion contract" where they agree to overthrow the institutions and expropriate capital, but bind themselves not to change the number of positions of power. However, each has an incentive to reduce the extent of power sharing (imposing the loss of status on others within the former elite), so this contract could only be enforced by some exogenous higher authority. In the absence of such a thing, those in power may rebel against the existing institutions, but cannot commit to what they will then do next.[38]

5.3 The equilibrium and efficient choices of capital taxation

Economic development ultimately requires rewarding the productive rather than just the strong, and for this to happen, institutions must credibly protect the property rights of investors. It is an endogenous feature of the model that institutions with broader power sharing can achieve that goal, but does the elite have an incentive to build institutions conducive to investment?

Given payoff equalization within groups, the elite chooses three institutional variables: p, τ_{q} and τ_{κ}. Proposition 2 identifies two binding no-rebellion constraints at the post-investment stage, which

[37] The extra members of the elite are in no way intrinsically different from existing members and do not have access to any special technology directly protecting property rights.

[38] If it were possible for the rebels to commit to restrict reoptimization to certain areas following a rebellion, paradoxically this makes it harder for institutions to sustain credible commitments. For example, suppose the composition of the elite is defined on the first page of the constitution, and limits on expropriation of private property are specified on the second page. If it were feasible somehow to prevent a successful rebellion from rewriting page one of the constitution then this would annihilate the credibility of page two.

can be used to eliminate two of these variables. The proposition also shows that no pre-investment-stage constraints are binding: groups would always have an incentive to defer rebellion until after investments have been made. This leaves one degree of freedom to maximize the elite payoff $\mathcal{U}_{\mathrm{p}}$. Equation [5.12b] gives an expression for $\mathcal{U}_{\mathrm{p}}$ as a function of s and $\tilde{\theta}$, with p and τ_{q} eliminated using the binding constraints. Noting the relationship between $\tilde{\theta}$ and s from [5.7], the expression in [5.12b] can be differentiated to obtain the equilibrium value of s^*:

$$s^* = \max\left\{0, \frac{\delta\kappa - (\mathrm{q}+2\delta)\psi}{2(\mathrm{q}+2\delta)(\kappa-\psi)}\right\}. \qquad [5.13]$$

As confirmed below in Proposition 3, this is the unique Markovian equilibrium of the model. Does it correspond to the efficient level of investment?

Since θ lies in the range $[\psi, \kappa]$, it is easy to see from [5.10] that the surplus $\mathcal{S}_{\mathrm{i}}(\tilde{\theta})$ from investment opportunities is maximized when $\hat{\tilde{\theta}} = \kappa$. This is the first-best level of investment. From [5.7] it follows that $\hat{\tau}_\kappa = 0$ and $\hat{s} = 1$, so capital taxes should be zero. From a public finance perspective, in a world where the government needs to raise a particular amount of revenue, this is also the capital tax that would be chosen optimally because lump-sum taxes and transfers are available. Equation [5.13] clearly shows that $s^* < 1$, and hence $\tau_\kappa^* > 0$, so equilibrium investment always falls short of the first-best level as capital taxes are positive.

However, the first best is not the most interesting welfare benchmark. The model demonstrates power sharing is required for protection of property rights, but a larger elite diverts more individuals from directly productive occupations (individuals in the elite do not receive the endowment q). This means there is an opportunity cost of increasing the elite size. Does the finding that $s^* < 1$ then simply reflect the resource cost of adding more individuals to the elite, compared to the hypothetical first-best world where there is no need to provide credible incentives?

To address this question, consider the following notion of *constrained efficiency*. Suppose that it were possible exogenously to impose some choice of capital tax τ_κ on the institutions selected by any elite at the pre-investment stage. All other aspects of institutions would be chosen as before by elites to maximize their own payoffs subject to avoiding rebellions. The constrained efficient level of the capital tax is what would then be chosen by a benevolent agent who takes into account the other institutional choices made by elites to survive the power struggle. The benevolent agent would appreciate that more investment requires greater protection of property rights, and thus more power sharing. The concept of constrained efficiency requires setting the benefit of more investment against the resource cost of the larger elite.[39] In the public-good application considered in section 4, a benevolent agent could not improve upon the efficiency of the institutions by imposing a level of public-good provision different from what prevails in equilibrium. Here, the issue is whether the equilibrium amount of investment coincides with its constrained efficient level.

The benevolent agent maximizes the average ex ante utility $\bar{\mathcal{U}}$ of all individuals. All institutional

[39]If there were no resource cost of increasing the size of the elite then the notion of constrained efficiency would coincide with first best. Note that in the model with the public-good technology from section 4, the first-best and constrained efficient outcomes are the same.

variables other than τ_κ are determined in equilibrium as before, and notice that the claims in Proposition 2 regarding payoff equalization and which constraints will bind continue to hold even with the value of s corresponding to the benevolent agent's choice of τ_κ. The elite size p and proportion s of individuals accepting investment opportunities determine the number of investors $i = \mu s$ (all of whom have $\theta \leq \tilde{\theta}$) and the number of workers $w = 1 - p - i$. Writing average utility $\bar{\mathcal{U}}$ in terms of the payoffs of each of these groups:

$$\bar{\mathcal{U}} \equiv \int_\Omega \mathcal{U}(\imath)\mathrm{d}\imath = p\mathcal{U}_\mathrm{p} + (1 - p - \mu s)\mathcal{U}_\mathrm{w} + \mu s \mathbb{E}_\theta[\mathcal{U}_\mathrm{i}(\theta)|\theta \leq \tilde{\theta}]. \quad [5.14]$$

The condition $\mathcal{U}_\mathrm{i}(\tilde{\theta}) = \mathcal{U}_\mathrm{w}$ holds for the marginal investor with effort cost $\tilde{\theta}$. Average utility can then be rewritten in terms of the investors' surplus $\mathcal{S}_\mathrm{i}(\tilde{\theta})$ from [5.10]:

$$\bar{\mathcal{U}} = p\mathcal{U}_\mathrm{p} + (1 - p)\mathcal{U}_\mathrm{w} + \mu\mathcal{S}_\mathrm{i}(\tilde{\theta}).$$

Proposition 2 implies that the only two binding no-rebellion constraints are for workers and members of the elite at the post-investment stage. As a consequence, it can be seen from [5.11] that worker and elite payoffs are tied together by $\mathcal{U}_\mathrm{w} = \mathcal{U}_\mathrm{p} - \delta$. Since the benevolent agent takes such constraints into account, this relationship is substituted into the expression for $\bar{\mathcal{U}}$:

$$\bar{\mathcal{U}} = \mathcal{U}_\mathrm{p} - \delta(1 - p) + \mu\mathcal{S}_\mathrm{i}(\tilde{\theta}). \quad [5.15]$$

There are thus two differences between the expressions for average utility $\bar{\mathcal{U}}$ and the elite's objective $\mathcal{U}_\mathrm{p}$. The second term on the right-hand side is related to the distribution of resources between different individuals, and the third term reflects the investors' surplus.

Proposition 3 *(i) The unique Markovian equilibrium s^* is given by the expression in* [5.13].

(ii) The constrained efficient value of s (denoted by $s^\diamond$) that maximizes $\bar{\mathcal{U}}$ in [5.15] *is given by the following expression (if the non-negativity constraint on workers' consumption does not bind):*

$$s^\diamond = \max\left\{0, \frac{\delta\kappa - (\mathrm{q} + \delta)\psi}{(2\mathrm{q} + \delta)(\kappa - \psi)}\right\}. \quad [5.16]$$

(iii) s^* is positive when $\kappa/\psi - 1 > 1 + \mathrm{q}/\delta$, while $\kappa/\psi - 1 > \mathrm{q}/\delta$ is necessary for $s^\diamond > 0$. Whenever $s^\diamond > 0$, it must be the case that $s^* < s^\diamond$.

Proof See appendix A.3. ■

There are two reasons why the constrained-efficient choice of capital taxes is lower — and thus associated with more investment — than what prevails in equilibrium. The first (and more interesting) distortion follows from the distributional consequences of protection against expropriation (the second term in equation [5.15]). Strong property rights require sharing power, which in turn requires sharing rents because members of the elite are more powerful than other individuals. Those in the elite can implement the constrained-efficient allocation, which requires a larger p, but the need to avoid rebellions implies that they would have to offer rents to the extra elite members. Therefore, the cost to elite members of expanding their number is not simply the lost output from diverting individuals away from directly productive activities.

In contrast to the environment in section 4, the power struggle imposes an endogenous and binding constraint on the set of possible transfers among individuals. This leads to a breakdown of the political analogue of the Coase theorem. A larger elite gives rise to the conditions necessary for individuals to undertake investment projects that increase output, but the association between power and rents places a lower bound on the consumption of each elite member. The impossibility of sharing power without sharing rents thus drives a wedge between maximizing total output and maximizing the elite's own payoff.

The second distortion is that the equilibrium choice of capital taxes does not take account of investors' surplus (the third term in equation [5.15]). Since investors' effort costs θ are not public information, it is impossible for the elite to make transfers conditional on this information. The no-rebellion constraints for rebel armies including such individuals will be slack, so no benefit accrues to the elite from marginal increases in investors' payoffs. This increases the wedge between total output and the elite's payoff.[40]

The inefficiently high capital taxes can be interpreted as inadequate protection against government expropriation of private property. Acemoglu, Johnson and Robinson (2005) present evidence that institutional failures in this area are especially damaging to economic performance. But why should property rights be so susceptible to political failures compared to other aspects of institutions? The model here sheds light on this question.[41]

5.4 Historical analogies

The results highlight the importance of sharing power in enabling governments to overcome time-inconsistency problems, for example, creating an environment in which individuals can make investments without fearing expropriation of a large portion of their return. The results also show that rulers will not share power as much as would be desirable for economic efficiency. Although the model is too abstract to match any given historical episode precisely, these conclusions resonate with a number of examples.

Broadly speaking, the extra individuals in power required to sustain property rights (in the model, the difference between p^* and $p^\dagger$, as given in Proposition 2) might be interpreted as a "parliament" or any other group of people with the power to resist attempts to change institutions coming especially from others in power. Members of parliaments are usually thought of as repre-

[40] The first and second distortions correspond respectively to the second and third terms in the expression for $\bar{\mathcal{U}}$ in [5.15]. The effect of each of them on first-order condition determining s^* is thus seen to be independent of the other distortion.

[41] The welfare implications of the parameter δ are non-trivial in this application of the model. In the public-good application of section 4, a larger δ can only be harmful to workers because it allows higher taxes to be sustained, resulting in a more unequal distribution of income. In contrast, in an economy with a very small value of δ, there would not be any investment in equilibrium. A larger δ parameter makes it easier for the elite to remain in power, which directly benefits elite members, but might also allow them to offer some protection of property rights.

senting those who elect them, but so are democratically elected presidents, and the rationale for parliaments with large numbers of members must go beyond simply defending minorities. Power sharing makes institutions more stable because it makes it costlier for some members of the elite to replace the current institutions with new ones — with potentially different rules on how resources and power are distributed. Once power is too concentrated, institutions become subject to the whims of those in power, as noted by Montesquieu.

In seventeenth-century England, the Glorious Revolution led to power sharing between king and parliament. By accepting the Bill of Rights, King William III conceded that power would be shared. North and Weingast (1989) argue that the Glorious Revolution began an era of secure property rights and put an end to confiscatory government. As a result, the English government was able to borrow much more, and at substantially lower rates. This was certainly in the interests of the king, yet the earlier Stuart kings had staunchly resisted sharing power with parliament. According to the model, secure property rights require just such power sharing to make it costly for the king to rewrite the rules ex post. However, the existence of a parliament with real power implies that rents have to be shared, so even if the total pie becomes larger, with a smaller share, the amount received by the king might end up being lower.

Malmendier (2009) studies the Roman *societas publicanorum*, which were groups (precursors of the modern business corporation) to which the government contracted functions such as tax collection and public works. Their demise occurred with the transition from the Roman republic to the Roman empire. Why? According to Malmendier (2009), one possible explanation is that "the Roman Republic was a system of checks and balances. But the emperors centralized power and could, in principle, bend law and enforcement in their favor". In other words, while power was decentralized, it was possible to have rules that guaranteed the government's contract with the *societas publicanorum* and their property rights, presumably because changing the rules would result in some of the individuals in power coming into conflict with their peers, which would be costly. Once power was centralized, protection against expropriation was not possible any longer.

6 Concluding remarks

Research in economics has frequently progressed by focusing on the behaviour of individuals subject to some fundamental constraints or frictions and deriving the resulting implications for the economy. For example, it is often claimed that unemployment, credit rationing, and missing markets ought not to be directly assumed, but instead derived from the likes of search frictions, limited pledgeability, or asymmetric information.

This is arguably a far cry from the state of the art in research on social conflict and institutions. This literature typically assumes the existence of exogenous groups, imposes ad-hoc limits on what those in and out of power can do, on what happens when another group takes power, and makes a variety of different assumptions for the many dimensions of the power struggle. Furthermore, the workings of political institutions such as courts, constitutions, and representative bodies are often exogenously assumed. While many of these assumptions might be matched by their counterparts in reality, one is left to wonder how different the implications would be of a model where those features were not imposed, but emerged endogenously as a result of some more elementary frictions. Moreover, one consequence of the existing approach is that assumptions end up being very specific to the particular problem being analysed. A unified framework that can be used for analysis of different questions related to institutions is lacking — which is not surprising once we start to think about the challenge of integrating social conflict, governance, and individual choice into an economic model. Yet we believe that building such a framework could yield substantial gains in understanding institutions and their economic consequences.

The model of this paper attempts a step in that direction. Institutions are assumed to maximize the payoffs of those in power subject only to the power struggle, with no arbitrary constraints on transfers or policies. The power struggle is captured by a single rebellion mechanism that allows individuals to form groups and fight for power. Those in power have an advantage in defending the current institutions, but the option of rebelling is open to everyone on the same terms.

Our goal is to understand what features of political institutions arise in equilibrium starting from the basics of preferences, technologies, and the power struggle. In this paper, the general framework was used to study a situation where investment is possible but can be expropriated. We do not assume that property rights can be protected through some explicit and exogenous institutional technology. Instead, we derive the means by which this can be done endogenously starting from the primitives of the environment, and ask whether such protection will be efficiently provided by the equilibrium institutions. In order to generate commitment to rules that would otherwise be time inconsistent, a larger elite is endogenously formed. But the same conflict mechanism that explains how power sharing can overcome the commitment problem also implies that sharing power cannot be done without sharing rents. This imposes endogenous limits on the set of possible transfers and precludes Pareto-improving deals. In equilibrium, there is too little power sharing, and thus not enough institutional stability to offer investors the strong property rights required for economic efficiency.

References

Acemoglu, D. (2003), "Why not a political Coase theorem? Social conflict, commitment, and politics", *Journal of Comparative Economics*, **31(4)**:620–652. 6, 18

——— (2008), "Oligarchic versus democratic societies", *Journal of the European Economic Association*, **6(1)**:1–44 (March). 6

Acemoglu, D., Egorov, G. and Sonin, K. (2008), "Coalition formation in non-democracies", *Review of Economic Studies*, **75(4)**:987–1009 (October). 5

Acemoglu, D. and Johnson, S. (2005), "Unbundling institutions", *Journal of Political Economy*, **113(5)**:949–995. 2

Acemoglu, D., Johnson, S. and Robinson, J. A. (2005), "Institutions as a fundamental cause of long-run growth", in P. Aghion and S. N. Durlauf (eds.), *Handbook of Economic Growth*, North-Holland, chap. 6, pp. 386–472. 1, 30

Acemoglu, D. and Robinson, J. A. (2006), *Economic Origins of Dictatorship and Democracy*, Cambridge University Press. 5

——— (2008), "Persistence of power, elites, and institutions", *American Economic Review*, **98(1)**:267–293 (March). 5

Baron, D. P. and Ferejohn, J. A. (1989), "Bargaining in legislatures", *American Political Science Review*, **83(4)**:1181–1206 (December). 5

Basu, K. (2000), *Prelude to Political Economy*, Oxford University Press. 5

Besley, T. and Persson, T. (2009a), "The origins of state capacity: Property rights, taxation, and politics", *American Economic Review*, **99(4)**:1218–1244 (September). 5

——— (2009b), "Repression or civil war?", *American Economic Review*, **99(2)**:292–297 (May). 5

——— (2010), "State capacity, conflict and development", *Econometrica*, forthcoming. 5

Downs, A. (1957), "An economic theory of political action in a democracy", *Journal of Political Economy*, **65(2)**:135–150 (April). 5

Engerman, S. L. and Sokoloff, K. L. (1997), "Factor endowments, institutions, and differential paths of growth among New World economies: A view from economic historians of the United States", in S. H. Haber (ed.), *How Latin America Fell Behind*, Stanford University Press, chap. 10, pp. 260–304. 1

Garfinkel, M. R. and Skaperdas, S. (2007), "Economics of conflict: An overview", in T. Sandler and K. Hartley (eds.), *Handbook of Defense Economics*, vol. 2, North-Holland, chap. 22, pp. 649–709. 6

Glaeser, E., Scheinkman, J. and Shleifer, A. (2003), "The injustice of inequality", *Journal of Monetary Economics*, **50(1)**:199–222 (January). 6

Greif, A. (2006), *Institutions and the Path to the Modern Economy*, Cambridge University Press. 5

Grossman, H. I. and Kim, M. (1995), "Swords or plowshares? A theory of the security of claims to property", *Journal of Political Economy*, **103(6)**:1275–1288 (December). 6

Guriev, S. and Sonin, K. (2009), "Dictators and oligarchs: A dynamic theory of contested property rights", *Journal of Public Economics*, **93(1–2)**:1–13 (February). 6

Hall, R. E. and Jones, C. I. (1999), "Why do some countries produce so much more output per worker than others?", *Quarterly Journal of Economics*, **114(1)**:83–116 (February). 1

Hirshleifer, J. (1995), "Anarchy and its breakdown", *Journal of Political Economy*, **103(1)**:26–52 (February). 6

Levy, G. (2004), "A model of political parties", *Journal of Economic Theory*, **115(2)**:250–277 (April). 5

Mailath, G., Morris, S. and Postlewaite, A. (2001), "Laws and authority", working paper, Princeton University. 5

Malmendier, U. (2009), "Law and finance at the origin", *Journal of Economic Literature*, **47(4)**:1076–1108. 31

Myerson, R. B. (1993), "Incentives to cultivate favored minorities under alternative electoral rules", *American Political Science Review*, **87(4)**:856–869 (December). 12

——— (2004), "Justice, institutions and multiple equilibria", *Chicago Journal of International Law*, **5(1)**:91–107. 5

——— (2008), "The autocrat's credibility problem and foundations of the constitutional state", *American Political Science Review*, **102(1)**:125–139. 16

——— (2009), "Learning from Schelling's *Strategy of Conflict*", *Journal of Economic Literature*, **47(4)**:1109–1125. 5

——— (2010), "Capitalist investment and political liberalization", *Theoretical Economics*, **5(1)**:73–91. 6

North, D. C. (1990), *Institutions, Institutional Change, and Economic Performance*, Cambridge University Press. 1

North, D. C. and Weingast, B. R. (1989), "Constitutions and commitment: The evolution of institutions governing public choice in seventeenth-century England", *Journal of Economic History*, **49(4)**:803–832 (December). 1, 31

Persson, T. and Tabellini, G. (2000), *Political Economics: Explaining Economic Policy*, MIT Press. 5, 6, 19

Piccione, M. and Razin, R. (2009), "Coalition formation under power relations", *Theoretical Economics*, **4(1)**:1–15 (March). 5

Ray, D. (2007), *A Game-Theoretic Perspective on Coalition Formation*, Oxford University Press. 5

Schelling, T. C. (1960), *The Strategy of Conflict*, Harvard University Press. 5

A Technical appendix

A.1 Proof of Proposition 1

Consider a set of institutions $\mathscr{I} = \{\mathcal{P}, \mathcal{W}, \tau(\imath), C_{\mathrm{p}}(\imath)\}$ that constitute a Markovian equilibrium. Let $\bar{\mathcal{U}}_{\mathrm{w}}$ and $\bar{\mathcal{U}}_{\mathrm{p}}$ denote the average payoffs of workers and elite members under these institutions:

$$\bar{\mathcal{U}}_{\mathrm{w}} \equiv \frac{1}{|\mathcal{W}|}\int_{\mathcal{W}} \mathcal{U}_{\mathrm{w}}(\imath)\mathrm{d}\imath, \quad \text{and } \bar{\mathcal{U}}_{\mathrm{p}} \equiv \frac{1}{|\mathcal{P}|}\int_{\mathcal{P}} \mathcal{U}_{\mathrm{p}}(\imath)\mathrm{d}\imath. \tag{A.1.1}$$

To be an equilibrium, the institutions must maximize $\bar{\mathcal{U}}_{\mathrm{p}}$ subject to the general no-rebellion constraint [3.7]. Using the equations [3.5] and [3.6] that define the rebel and incumbent armies for a particular new elite selection function $\mathscr{E}'(\cdot)$, the general no-rebellion constraint can be stated as follows:

$$\begin{aligned}\int_{\mathcal{W}\cap\mathscr{E}'(p')} \max\{\mathcal{U}'_{\mathrm{p}} - \mathcal{U}_{\mathrm{w}}(\imath), 0\}\mathrm{d}\imath + \int_{\mathcal{P}\cap\mathscr{E}'(p')} \max\{\mathcal{U}'_{\mathrm{p}} - \mathcal{U}_{\mathrm{p}}(\imath), 0\}\mathrm{d}\imath \\ \leq \int_{\mathcal{P}\cap\mathscr{E}'(p')} \delta\mathbb{1}[\mathcal{U}_{\mathrm{p}}(\imath) > \mathcal{U}'_{\mathrm{p}}]\mathrm{d}\imath + \int_{\mathcal{P}\setminus\mathscr{E}'(p')} \delta\mathbb{1}[\mathcal{U}_{\mathrm{p}}(\imath) > \mathcal{U}'_{\mathrm{w}}]\mathrm{d}\imath \ \text{ for all } \mathscr{E}'(\cdot).\end{aligned} \tag{A.1.2}$$

To be a Markovian equilibrium, the institutions must also satisfy the conditions $p = p'$, $\bar{\mathcal{U}}_{\mathrm{p}} = \mathcal{U}'_{\mathrm{p}}$, and $\bar{\mathcal{U}}_{\mathrm{w}} = \mathcal{U}'_{\mathrm{w}}$.

Notice that if $\mathcal{U}_{\mathrm{w}}(\imath) > \mathcal{U}'_{\mathrm{p}}$ for a positive measure of workers $\imath \in \mathcal{W}$ then taxes $\tau(\imath)$ on those workers can be raised without increasing their fighting effort $\max\{\mathcal{U}'_{\mathrm{p}} - \mathcal{U}_{\mathrm{w}}(\imath), 0\}$. If the extra tax revenue is distributed among the elite then $\bar{\mathcal{U}}_{\mathrm{p}}$ is strictly increased, while no $\mathcal{U}_{\mathrm{p}}(\imath)$ is lower. It follows that if [A.1.2] was initially satisfied then it continues to be so after this deviation. Since this deviation is both feasible and raises the elite's objective function, the original institutions featuring $\mathcal{U}_{\mathrm{w}}(\imath) > \mathcal{U}'_{\mathrm{p}}$ cannot be an equilibrium. Therefore, attention can be restricted to institutions satisfying $\mathcal{U}_{\mathrm{w}}(\imath) \leq \mathcal{U}'_{\mathrm{p}}$ for all $\imath \in \mathcal{W}$. This implies $\bar{\mathcal{U}}_{\mathrm{w}} \leq \mathcal{U}'_{\mathrm{p}}$, so all Markovian equilibria must be such that $\bar{\mathcal{U}}_{\mathrm{w}} \leq \bar{\mathcal{U}}_{\mathrm{p}}$.

Next, consider the case where $\mathcal{U}_{\mathrm{p}}(\imath) < \mathcal{U}'_{\mathrm{w}}$ for a positive measure of elite members $\imath \in \mathcal{P}$. Since this implies $\mathcal{U}_{\mathrm{p}}(\imath) < \bar{\mathcal{U}}_{\mathrm{p}}$ in a Markovian equilibrium, expulsion of these individuals from the elite (and shrinking the elite size so they are not replaced) would strictly increase $\bar{\mathcal{U}}_{\mathrm{p}}$. Assume that an alternative set of institutions grant these expelled individuals the same consumption as when they were elite members. Since those outside the elite receive the endowment q, there are also extra resources to distribute, and assume these are distributed among elite members. This means that $\mathcal{U}_{\mathrm{p}}(\imath)$ is no lower for anyone who remains within the elite under these alternative institutions. Consequently, the left-hand side of [A.1.2] is no higher for any $\mathscr{E}'(\cdot)$, and since the expelled individuals had $\mathcal{U}_{\mathrm{p}}(\imath) < \mathcal{U}'_{\mathrm{w}} \leq \mathcal{U}'_{\mathrm{p}}$, the right-hand side of [A.1.2] is unaffected for all $\mathscr{E}'(\cdot)$. The deviation raising $\bar{\mathcal{U}}_{\mathrm{p}}$ is therefore feasible, which shows that all Markovian equilibria must be such that $\mathcal{U}_{\mathrm{p}}(\imath) \geq \mathcal{U}'_{\mathrm{w}}$ for all $\imath \in \mathcal{P}$.

Restricting attention to cases where $\mathcal{U}_{\mathrm{w}}(\imath) \leq \mathcal{U}'_{\mathrm{p}}$ and $\mathcal{U}_{\mathrm{p}}(\imath) \geq \mathcal{U}'_{\mathrm{w}}$ for all $\imath$, the general no-rebellion constraint [A.1.2] is equivalent to

$$\int_{\mathcal{W}\cap\mathscr{E}'(p')} (\mathcal{U}'_{\mathrm{p}} - \mathcal{U}_{\mathrm{w}}(\imath))\mathrm{d}\imath + \int_{\mathcal{P}\cap\mathscr{E}'(p')} \mathbb{1}[\mathcal{U}_{\mathrm{p}}(\imath) \leq \mathcal{U}'_{\mathrm{p}}](\mathcal{U}'_{\mathrm{p}} - \mathcal{U}_{\mathrm{p}}(\imath) + \delta)\mathrm{d}\imath \leq \delta p \ \text{ for all } \mathscr{E}'(\cdot), \tag{A.1.3}$$

where $p = |\mathcal{P}|$ is the measure of the set $\mathcal{P}$ of individuals in power. Given a particular new elite selection function $\mathscr{E}'(\cdot)$, the subset of workers offered a place in the new elite is $\mathcal{E}_{\mathrm{w}} = \mathcal{W} \cap \mathscr{E}'(p')$, and the subset of current elite members who receive a place in the subsequent elite is $\mathcal{E}_{\mathrm{p}} = \mathcal{P} \cap \mathscr{E}'(p')$. Conversely, for any sets $\mathcal{E}_{\mathrm{w}} \subseteq \mathcal{W}$ and $\mathcal{E}_{\mathrm{p}} \subseteq \mathcal{P}$ such that $|\mathcal{E}_{\mathrm{w}} \cup \mathcal{E}_{\mathrm{p}}| = p'$, there exists a new elite selection function $\mathscr{E}'(\cdot)$ generating these sets. Therefore, the general no-rebellion constraint [A.1.3] can be stated equivalently as

$$\int_{\mathcal{E}_{\mathrm{w}}} (\mathcal{U}'_{\mathrm{p}} - \mathcal{U}_{\mathrm{w}}(\imath))\mathrm{d}\imath + \int_{\mathcal{E}_{\mathrm{p}}} \mathbb{1}[\mathcal{U}_{\mathrm{p}}(\imath) \leq \mathcal{U}'_{\mathrm{p}}](\mathcal{U}'_{\mathrm{p}} - \mathcal{U}_{\mathrm{p}}(\imath) + \delta) \leq \delta p \ \text{ for all } \mathcal{E}_{\mathrm{w}} \subseteq \mathcal{W}, \mathcal{E}_{\mathrm{p}} \subseteq \mathcal{P} \ \text{ with } |\mathcal{E}_{\mathrm{w}} \cup \mathcal{E}_{\mathrm{p}}| = p'. \tag{A.1.4}$$

In what follows, let σ denote the fraction of places in the post-rebellion elite that would be filled by those who are workers under the current institutions. With this definition, note that $\sigma = |\mathcal{E}_{\mathrm{w}}|/p'$ and $1-\sigma = |\mathcal{E}_{\mathrm{p}}|/p'$. Given p and p', and hence $|\mathcal{W}| = 1-p$, there are limits on the range of possible σ values associated with sets $\mathcal{E}_{\mathrm{w}} \subseteq \mathcal{W}$, $\mathcal{E}_{\mathrm{p}} \subseteq \mathcal{P}$ with $|\mathcal{E}_{\mathrm{w}} \cup \mathcal{E}_{\mathrm{p}}| = p'$. In particular, σ must lie between $\underline{\sigma}$ and $\overline{\sigma}$:

$$\underline{\sigma} \equiv \max\left\{0, \frac{p'-p}{p'}\right\}, \quad \text{and} \;\; \overline{\sigma} \equiv \min\left\{\frac{1-p}{p'}, 1\right\}. \qquad \text{[A.1.5]}$$

Now define the following functions $\mathscr{F}_{\mathrm{w}}(\sigma)$ and $\mathscr{F}_{\mathrm{p}}(\sigma)$ of $\sigma \in [\underline{\sigma}, \overline{\sigma}]$:

$$\mathscr{F}_{\mathrm{w}}(\sigma) \equiv \max_{\substack{\mathcal{E}_{\mathrm{w}} \subseteq \mathcal{W} \\ |\mathcal{E}_{\mathrm{w}}| = \sigma p'}} \int_{\mathcal{E}_{\mathrm{w}}} (\mathcal{U}_{\mathrm{p}}' - \mathcal{U}_{\mathrm{w}}(\imath))\mathrm{d}\imath, \quad \text{and} \;\; \mathscr{F}_{\mathrm{p}}(\sigma) \equiv \max_{\substack{\mathcal{E}_{\mathrm{p}} \subseteq \mathcal{P} \\ |\mathcal{E}_{\mathrm{p}}| = (1-\sigma)p'}} \int_{\mathcal{E}_{\mathrm{p}}} \mathbb{1}[\mathcal{U}_{\mathrm{p}}(\imath) \leq \mathcal{U}_{\mathrm{p}}'](\mathcal{U}_{\mathrm{p}}' - \mathcal{U}_{\mathrm{p}}(\imath) + \delta)\mathrm{d}\imath. \qquad \text{[A.1.6]}$$

These functions represent the maximum contributions to the net strength of the rebel army from those who are currently workers and those who are currently elite members. Maximization is over the compositions of the subsequent elite from among current workers and current elite members respectively, taking as given the fraction σ of places in the subsequent elite assigned to current workers. The general no-rebellion constraint [A.1.4] imposes an upper bound on the sum of the effective fighting strengths of workers and current elite members in the rebel army for all new elite selection functions, that is, for all compositions of the subsequent elite. This constraint therefore holds if and only if it holds for the maximized contributions of workers and elite members for all possible values of σ. Therefore, [A.1.4] is equivalent to

$$\mathscr{F}(\sigma) \equiv \mathscr{F}_{\mathrm{w}}(\sigma) + \mathscr{F}_{\mathrm{p}}(\sigma) \leq \delta p \;\; \text{for all} \;\; \sigma \in [\underline{\sigma}, \overline{\sigma}], \qquad \text{[A.1.7]}$$

with bounds $\underline{\sigma}$ and $\overline{\sigma}$ from [A.1.5]. Another equivalent form of the constraint is to state it in terms of the maximized value of $\mathscr{F}(\sigma)$, the sum of $\mathscr{F}_{\mathrm{w}}(\sigma)$ and $\mathscr{F}_{\mathrm{p}}(\sigma)$, over all values of σ:

$$\mathscr{F}^{\ddagger} \equiv \max_{\sigma \in [\underline{\sigma}, \overline{\sigma}]} \mathscr{F}(\sigma) \leq \delta p. \qquad \text{[A.1.8]}$$

Therefore, equilibrium institutions maximize $\bar{\mathcal{U}}_{\mathrm{p}}$ subject to the above constraint, the non-negativity constraints, $p < 1/2$, and the elite selection function $\mathcal{P} = \mathscr{E}(p)$.

Note that the constraint [A.1.8] must be binding. Suppose first the non-negativity constraints are slack. If it were the case that $\mathscr{F}^{\ddagger} < \delta p$ then taxes on all workers could be increased by some strictly positive amount, with the proceeds distributed to members of the elite. It is clear from the definition of $\mathscr{F}_{\mathrm{w}}(\sigma)$ in [A.1.6] that $\mathscr{F}_{\mathrm{w}}(\sigma)$ is continuous with respect to this tax change. Furthermore, with $\mathcal{U}_{\mathrm{p}}(\imath)$ no lower for any $\imath \in \mathcal{P}$, the definition of $\mathscr{F}_{\mathrm{p}}(\sigma)$ implies that $\mathscr{F}_{\mathrm{p}}(\sigma)$ is no larger for any $\sigma \in [\underline{\sigma}, \overline{\sigma}]$, while $\bar{\mathcal{U}}_{\mathrm{p}}$ is strictly increased. Therefore, there is some positive tax increase that ensures $\mathscr{F}^{\ddagger}$ remains below δp, and is thus feasible.

Now consider a case where some non-negativity constraints are binding. These cannot be binding on all individuals. Since any Markovian equilibrium must feature $\mathcal{U}_{\mathrm{p}}(\imath) \geq \bar{\mathcal{U}}_{\mathrm{w}}$, non-negativity constraints could be binding only for workers and possibly those elite members receiving the minimum in the distribution of elite payoffs. It follows that there exists a positive measure of elite members at or below the average elite payoff for whom non-negativity constraints are not binding. Consider a deviation whereby some positive measure of these individuals are expelled from the elite and now face taxes that give them a payoff lower by some strictly positive amount than what they previously received. Suppose the extra resources following this deviation are distributed among the remaining elite members. This strictly increases the average elite payoff, and continues to satisfy the no-rebellion constraints for sufficiently small changes since these are continuous. There is thus a profitable deviation in both cases, so the constraint $\mathscr{F}^{\ddagger} \leq \delta p$ must be binding.

Now consider a Markovian equilibrium. Since $p' = p = p^*$, and $p^* < 1/2$, it follows from [A.1.5] that $\underline{\sigma} = 0$ and $\overline{\sigma} = 1$. The Markovian equilibrium institutions must therefore maximize $\bar{\mathcal{U}}_{\mathrm{p}}$ subject to

$$\mathscr{F}^{*\ddagger} \equiv \max_{\sigma \in [0,1]} \mathscr{F}^*(\sigma) \leq \delta p^*, \qquad \text{[A.1.9]}$$

where $\mathscr{F}^*(\sigma)$, $\mathscr{F}_{\mathrm{w}}^*(\sigma)$ and $\mathscr{F}_{\mathrm{p}}^*(\sigma)$ denote the functions in [A.1.7] and [A.1.8] evaluated using the Markovian equilibrium payoffs.

Payoff equalization among workers

Conjecture that there is a Markovian equilibrium in which a positive measure of workers receive payoffs $\mathcal{U}^*_{\mathrm{w}}(\imath)$ different from the average $\bar{\mathcal{U}}^*_{\mathrm{w}}$.

Suppose first that $\mathscr{F}^*(0) < \mathscr{F}^{*\ddagger}$. Each worker's payoff is $\mathcal{U}^*_{\mathrm{w}}(\imath) = u(C^*_{\mathrm{w}}(\imath))$, where $C^*_{\mathrm{w}}(\imath) = \mathrm{q} - \tau^*(\imath)\}$ are the consumption levels implied by the distribution of taxes specified by the Markovian equilibrium institutions. Let $\bar{C}_{\mathrm{w}}$ denote the level of consumption required to give a worker the average utility of all workers in the Markovian equilibrium, that is, $u(\bar{C}_{\mathrm{w}}) = \bar{\mathcal{U}}^*_{\mathrm{w}}$. Since the utility function $u(\cdot)$ is weakly concave, Jensen's inequality implies

$$u\left(\frac{1}{|\mathcal{W}|}\int_{\mathcal{W}} C^*_{\mathrm{w}}(\imath)\mathrm{d}\imath\right) \geq \frac{1}{|\mathcal{W}|}\int_{\mathcal{W}} u(C^*_{\mathrm{w}}(\imath))\mathrm{d}\imath = \bar{\mathcal{U}}^*_{\mathrm{w}} = u(\bar{C}_{\mathrm{w}}).$$

The utility function $u(\cdot)$ is strictly increasing, so this means that

$$\bar{C}_{\mathrm{w}} \leq \frac{1}{|\mathcal{W}|}\int_{\mathcal{W}} C^*_{\mathrm{w}}(\imath)\mathrm{d}\imath.$$

Consider the following deviation from the Markovian equilibrium institutions. Each worker faces the same tax, and the level of this tax implies that all workers receive consumption $\bar{C}_{\mathrm{w}}$. The equation above shows that this deviation does not reduce the resources available to finance the elite's consumption. Any extra resources are distributed among the elite, so that they now receive consumption $C_{\mathrm{p}}(\imath) \geq C^*_{\mathrm{p}}(\imath)$. Let $\mathcal{U}_{\mathrm{p}}(\imath)$ denote the payoffs of elite members after the deviation, which must satisfy $\mathcal{U}_{\mathrm{p}}(\imath) \geq \mathcal{U}^*_{\mathrm{p}}(\imath)$ for all $\imath \in \mathcal{P}$. Let $\mathscr{F}_{\mathrm{p}}(\sigma)$ be the function from [A.1.6] calculated using the new distribution of elite payoffs, while $\mathscr{F}^*_{\mathrm{p}}(\sigma)$ is the corresponding function calculated using the payoffs in the conjectured Markovian equilibrium. The definition in [A.1.6] implies that $\mathscr{F}_{\mathrm{p}}(\sigma) \leq \mathscr{F}^*_{\mathrm{p}}(\sigma)$ for all $\sigma \in [0,1]$.

After the deviation, all workers $\imath \in \mathcal{W}$ receive the same payoff $\mathcal{U}_{\mathrm{w}}(\imath) = \bar{\mathcal{U}}_{\mathrm{w}}$, which by construction is equal to $\bar{\mathcal{U}}^*_{\mathrm{w}}$. The definition of $\mathscr{F}_{\mathrm{w}}(\sigma)$ in [A.1.6] and $\mathcal{U}'_{\mathrm{p}} = \bar{\mathcal{U}}^*_{p}$ then imply that $\mathscr{F}_{\mathrm{w}}(\sigma) = \sigma p^*(\bar{\mathcal{U}}^*_{\mathrm{p}} - \bar{\mathcal{U}}^*_{\mathrm{w}})$ for all $\sigma \in [0,1]$.

Observe that in a Markovian equilibrium, $p' = p^* < 1/2$, so it follows that $\sigma p^* < 1 - p^* = |\mathcal{W}|$ for all $\sigma \in [0,1]$. This means that for any $\sigma \in [0,1]$, the set of workers included in the subsequent elite following a rebellion always leaves out some positive measure of workers from the rebel army. Hence, with payoff inequality among a positive measure of workers in the Markovian equilibrium, it must be the case that the smallest possible integral of the workers' payoffs in the rebel army is less than the average worker payoff multiplied by the number of workers in the rebel army (when the measure of workers included is positive):

$$\min_{\substack{\mathcal{E}_{\mathrm{w}}\subseteq\mathcal{W}\\ |\mathcal{E}_{\mathrm{w}}|=\sigma p^*}} \int_{\mathcal{E}_{\mathrm{w}}} \mathcal{U}^*_{\mathrm{w}}(\imath)\mathrm{d}\imath < \sigma p^* \bar{\mathcal{U}}^*_{\mathrm{w}}, \quad \text{for all } \sigma \in (0,1].$$

Therefore, noting the definition in [A.1.6]:

$$\mathscr{F}^*_{\mathrm{w}}(\sigma) = \max_{\substack{\mathcal{E}_{\mathrm{w}}\subseteq\mathcal{W}\\ |\mathcal{E}_{\mathrm{w}}|=\sigma p^*}} \int_{\mathcal{E}_{\mathrm{w}}} (\bar{\mathcal{U}}^*_{\mathrm{p}} - \mathcal{U}^*_{\mathrm{w}}(\imath))\mathrm{d}\imath > \sigma p^*(\bar{\mathcal{U}}^*_{\mathrm{p}} - \bar{\mathcal{U}}^*_{\mathrm{w}}) = \mathscr{F}_{\mathrm{w}}(\sigma), \quad \text{for all } \sigma \in (0,1].$$

Together with the earlier result $\mathscr{F}_{\mathrm{p}}(\sigma) \leq \mathscr{F}^*_{\mathrm{p}}(\sigma)$ for all $\sigma \in [0,1]$, the equation above establishes that $\mathscr{F}(\sigma) < \mathscr{F}^*(\sigma)$ for all $\sigma \in (0,1]$. Now take any $\sigma^\ddagger \in [0,1]$ that solves the maximization problem from [A.1.9], that is, $\mathscr{F}^*(\sigma^\ddagger) = \mathscr{F}^{*\ddagger}$. Since $\mathscr{F}^*(0) < \mathscr{F}^{*\ddagger}$ in the case under consideration, this value must satisfy $\sigma^\ddagger > 0$. The analysis above has shown $\mathscr{F}(\sigma^\ddagger) < \mathscr{F}^*(\sigma^\ddagger)$ for any such $\sigma^\ddagger$, and therefore $\mathscr{F}^\ddagger < \mathscr{F}^{*\ddagger}$, where $\mathscr{F}^{*\ddagger}$ and $\mathscr{F}^\ddagger$ are respectively the maximized values of the function in [A.1.9] before and after the deviation. Since $\mathscr{F}^{*\ddagger} = \delta p^*$, it follows that $\mathscr{F}^\ddagger < \delta p^*$, so the no-rebellion constraint is slack after the deviation. It has been seen how to increase $\bar{\mathcal{U}}_{\mathrm{p}}$ starting from this point by raising taxes, while continuing to satisfy the no-rebellion constraint [A.1.9].

The remaining case to consider is $\mathscr{F}^*(0) \geq \mathscr{F}^{*\ddagger}$. The definition of $\mathscr{F}^{*\ddagger}$ in [A.1.9] implies $\mathscr{F}^*(0) \leq \mathscr{F}^{*\ddagger}$, so this case is equivalent to $\mathscr{F}(0) = \mathscr{F}^{*\ddagger}$.

The definition of $\mathscr{F}_{\mathrm{w}}(\sigma)$ in [A.1.6] implies that as the share σ of current workers offered places in the subsequent elite increases, those joining the strongest rebel army would have current payoffs that are no lower than those who have already joined. Thus, the average payoff of those workers in the strongest rebel army for a given σ cannot decrease as σ increases. Consequently, $\mathscr{F}_{\mathrm{w}}(\sigma)/\sigma$ must be weakly decreasing in σ. Together with $\mathscr{F}_{\mathrm{w}}(0) = 0$, this implies the function $\mathscr{F}_{\mathrm{w}}(\sigma)$ is weakly concave in σ. A similar argument applies to $\mathscr{F}_{\mathrm{p}}(\sigma)$, so $\mathscr{F}_{\mathrm{p}}(\sigma)/(1-\sigma)$ is weakly increasing in σ (the average is weakly decreasing in the proportion $1-\sigma$ of places in the subsequent elite offered to current elite members). Together with $\mathscr{F}_{\mathrm{p}}(1) = 0$, this means that $(\mathscr{F}_{\mathrm{p}}(\sigma) - \mathscr{F}_{\mathrm{p}}(1))/(\sigma - 1)$ is weakly decreasing in σ, hence $\mathscr{F}_{\mathrm{p}}(\sigma)$ is weakly concave. It follows that the sum $\mathscr{F}(\sigma) = \mathscr{F}_{\mathrm{w}}(\sigma) + \mathscr{F}_{\mathrm{p}}(\sigma)$ is also a weakly concave function of σ.

Note that equation [A.1.9] together with $\mathscr{F}^*(0) = \mathscr{F}^{*\ddagger}$ implies $\mathscr{F}^*(\sigma) \leq \mathscr{F}^*(0)$ for all $\sigma \in [0,1]$, which means that $\mathscr{F}^*(\sigma)$ cannot have a positive slope at $\sigma = 0$. Since $\mathscr{F}^*(\sigma)$ is weakly concave for all $\sigma \in [0,1]$, the slope of $\mathscr{F}^*(\sigma)$ cannot increase with σ, so it must be the case that $\mathscr{F}^*(\sigma)$ is weakly decreasing for all $\sigma \in [0,1]$. Observe that when $\sigma = 0$, all members of the current elite can be offered places in the subsequent elite because $p' = p^*$ in a Markovian equilibrium. If $\mathscr{F}^*_{\mathrm{w}}(\sigma) + \mathscr{F}^*_{\mathrm{p}}(\sigma)$ cannot increase with σ then this says that the net strength of the rebel army is never increased by increasing the share of positions in the subsequent elite (and hence places in the rebel army) assigned to current workers. This can only happen if the highest possible contribution to the fighting strength of the rebel army from a positive measure of workers is never more than the lowest net contribution by a positive measure of current elite members. Given the definitions of $\mathscr{F}^*_{\mathrm{w}}(\sigma)$ and $\mathscr{F}^*_{\mathrm{p}}(\sigma)$ in [A.1.6], this condition can be stated formally as:

$$\sup\left\{f \,\middle|\, \left|\{\imath \in \mathcal{W} \,|\, \bar{\mathcal{U}}^*_{\mathrm{p}} - \mathcal{U}^*_{\mathrm{w}}(\imath) \leq f\}\right| < 1\right\} \leq \inf\left\{f \,\middle|\, \left|\{\imath \in \mathcal{P} \,|\, \mathbb{1}[\mathcal{U}^*_{\mathrm{p}}(\imath) \leq \bar{\mathcal{U}}^*_{\mathrm{p}}](\bar{\mathcal{U}}^*_{\mathrm{p}} - \mathcal{U}^*_{\mathrm{p}}(\imath) + \delta) \leq f\}\right| > 0\right\}. \tag{A.1.10}$$

The existence of payoff inequality among a positive measure of workers in the Markovian equilibrium implies there must be a positive measure of workers $\imath \in \mathcal{W}$ such that

$$\bar{\mathcal{U}}^*_{\mathrm{p}} - \mathcal{U}^*_{\mathrm{w}}(\imath) \leq \sup\left\{f \,\middle|\, \left|\{\imath \in \mathcal{W} \,|\, \bar{\mathcal{U}}^*_{\mathrm{p}} - \mathcal{U}^*_{\mathrm{w}}(\imath) \leq f\}\right| < 1\right\} - \Delta, \quad \text{for some } \Delta > 0.$$

Consider a new distribution of taxes on workers that increases the tax on all workers satisfying the condition above by a strictly positive amount, but one that does not lower these workers' utilities by more than Δ. All other workers' taxes are unaffected. Let $\mathcal{U}_{\mathrm{w}}(\imath)$ denote workers' payoffs after this deviation from the conjectured Markovian equilibrium. Given the condition above and [A.1.10], the new worker payoffs satisfy the inequality:

$$\sup\left\{f \,\middle|\, \left|\{\imath \in \mathcal{W} \,|\, \bar{\mathcal{U}}^*_{\mathrm{p}} - \mathcal{U}_{\mathrm{w}}(\imath) \leq f\}\right| < 1\right\} \leq \sup\left\{f \,\middle|\, \left|\{\imath \in \mathcal{W} \,|\, \bar{\mathcal{U}}^*_{\mathrm{p}} - \mathcal{U}^*_{\mathrm{w}}(\imath) \leq f\}\right| < 1\right\}$$
$$\leq \inf\left\{f \,\middle|\, \left|\{\imath \in \mathcal{P} \,|\, \mathbb{1}[\mathcal{U}^*_{\mathrm{p}}(\imath) \leq \bar{\mathcal{U}}^*_{\mathrm{p}}](\bar{\mathcal{U}}^*_{\mathrm{p}} - \mathcal{U}^*_{\mathrm{p}}(\imath) + \delta) \leq f\}\right| > 0\right\}.$$

If the function $\mathscr{F}_{\mathrm{w}}(\sigma)$ is defined using the new distribution of worker payoffs then the inequality above demonstrates that the sum $\mathscr{F}_{\mathrm{w}}(\sigma) + \mathscr{F}^*_{\mathrm{p}}(\sigma)$ is weakly decreasing for all $\sigma \in [0,1]$.

The extra tax revenue generated by the deviation considered above is distributed among the elite, ensuring the new elite payoffs satisfy $\mathcal{U}_{\mathrm{p}}(\imath) \geq \mathcal{U}^*_{\mathrm{p}}(\imath)$ for all $\imath \in \mathcal{P}$. The average elite payoff $\bar{\mathcal{U}}_{\mathrm{p}}$ is strictly increased because taxes are increased by a positive amount on a positive measure of workers. The function $\mathscr{F}_{\mathrm{p}}(\sigma)$ is calculated using the new distribution of elite payoffs, and equation [A.1.9] shows this satisfies $\mathscr{F}_{\mathrm{p}}(\sigma) \leq \mathscr{F}^*_{\mathrm{p}}(\sigma)$ for all $\sigma \in [0,1]$.

By putting these results together and noting that $\mathscr{F}_{\mathrm{w}}(0) = \mathscr{F}^*(0) = 0$:

$$\mathscr{F}(\sigma) \equiv \mathscr{F}_{\mathrm{w}}(\sigma) + \mathscr{F}_{\mathrm{p}}(\sigma) \leq \mathscr{F}_{\mathrm{w}}(\sigma) + \mathscr{F}^*_{\mathrm{p}}(\sigma) \leq \mathscr{F}_{\mathrm{w}}(0) + \mathscr{F}^*_{\mathrm{p}}(0) = \mathscr{F}^*_{\mathrm{w}}(0) + \mathscr{F}^*_{\mathrm{p}}(0) = \mathscr{F}^{*\ddagger} = \delta p^*,$$

and hence $\mathscr{F}(\sigma) \leq \delta p^*$ for all $\sigma \in [0,1]$ after the tax change. This shows that the no-rebellion constraint remains satisfied, while the elite objective function is increased.

Therefore, payoff inequality for a positive measure of workers is not consistent with Markovian equilibrium. Any equilibrium must feature $\mathcal{U}_{\mathrm{w}}^{*}(\imath) = \bar{\mathcal{U}}_{\mathrm{w}}^{*}$ for all $\imath \in \mathcal{W}$ (except for measure-zero sets of workers).

Payoff equalization among members of the elite

Since payoff equalization $\mathcal{U}_{\mathrm{w}}^{*}(\imath) = \mathcal{U}_{\mathrm{w}}^{*}$ has been established for all workers (up to measure zero sets) in any Markovian equilibrium, with $\mathcal{U}_{\mathrm{p}}' = \bar{\mathcal{U}}_{\mathrm{p}}^{*}$ the function $\mathscr{F}_{\mathrm{w}}(\sigma)$ from [A.1.6] reduces to

$$\mathscr{F}_{\mathrm{w}}^{*}(\sigma) = \sigma p^{*}(\bar{\mathcal{U}}_{\mathrm{p}}^{*} - \mathcal{U}_{\mathrm{w}}^{*}). \qquad \text{[A.1.11]}$$

The definition of $\mathscr{F}_{\mathrm{p}}^{*}(\sigma)$ from [A.1.6] makes it clear that $\mathscr{F}_{\mathrm{p}}^{*}(1) = 0$, so since any equilibrium institutions must satisfy the no-rebellion constraint [A.1.9], it follows that $\mathscr{F}_{\mathrm{w}}^{*}(1) = \mathscr{F}^{*}(1) \leq \mathscr{F}^{*\ddagger} = \delta p^{*}$. With the equation above, this implies:

$$\bar{\mathcal{U}}_{\mathrm{p}}^{*} - \mathcal{U}_{\mathrm{w}}^{*} \leq \delta. \qquad \text{[A.1.12]}$$

Suppose there is a Markovian equilibrium with $\bar{\mathcal{U}}_{\mathrm{p}}^{*} - \mathcal{U}_{\mathrm{w}}^{*} < \delta$. Starting from such institutions, consider the following deviation. First, redistribute consumption equally among elite members so that all elite members receive the average $\bar{C}_{\mathrm{p}}^{*}$ specified by the conjectured Markovian equilibrium institutions. Let $C_{\mathrm{p}}(\imath) = \bar{C}_{\mathrm{p}}^{*}$ denote the new distribution of elite consumption and $\mathcal{U}_{\mathrm{p}}(\imath) = u(C_{\mathrm{p}}(\imath))$ the elite payoffs after the deviation. Given the weak concavity of the utility function $u(\cdot)$, Jensen's inequality implies

$$\mathcal{U}_{\mathrm{p}}(\imath) = u(\bar{C}_{\mathrm{p}}^{*}) = u\left(\frac{1}{|\mathcal{P}|}\int_{\mathcal{P}} C_{\mathrm{p}}^{*}(\imath)\mathrm{d}\imath\right) \geq \frac{1}{|\mathcal{P}|}\int_{\mathcal{P}} u(C_{\mathrm{p}}^{*}(\imath))\mathrm{d}\imath = \bar{\mathcal{U}}_{\mathrm{p}}^{*}. \qquad \text{[A.1.13]}$$

A second change made to the conjecture Markovian equilibrium institutions is a strictly positive tax increase levied equally on all workers. This is chosen to be sufficiently small so that $\bar{\mathcal{U}}_{\mathrm{p}}^{*} - \mathcal{U}_{\mathrm{w}} \leq \delta$ holds, where $\mathcal{U}_{\mathrm{w}}$ is the common payoff of all workers after the deviation. The proceeds of the tax are distributed among the elite, so the average elite payoff $\bar{\mathcal{U}}_{\mathrm{p}}$ is strictly larger. Furthermore, using the inequality in [A.1.13], it must be the case that $\mathcal{U}_{\mathrm{p}}(\imath) \geq \bar{\mathcal{U}}_{\mathrm{p}}^{*}$ for all $\imath \in \mathcal{P}$ after the deviations.

Since $\mathcal{U}_{\mathrm{p}}' = \bar{\mathcal{U}}_{\mathrm{p}}^{*}$ in a Markovian equilibrium, $\mathcal{U}_{\mathrm{p}}(\imath) \geq \bar{\mathcal{U}}_{\mathrm{p}}^{*}$ implies that for all $\imath \in \mathcal{P}$:

$$\mathbb{1}[\mathcal{U}_{\mathrm{p}}(\imath) \leq \mathcal{U}_{\mathrm{p}}'](\mathcal{U}_{\mathrm{p}}' - \mathcal{U}_{\mathrm{p}}(\imath) + \delta) \leq \delta.$$

As a result, the function $\mathscr{F}_{\mathrm{p}}(\sigma)$ from [A.1.6] (calculated using the elite members' payoffs after the deviation) can be bounded above:

$$\mathscr{F}_{\mathrm{p}}(\sigma) \leq (1-\sigma)p^{*}\delta, \quad \text{for all } \sigma \in [0,1].$$

Following the tax increase, all workers' payoffs remain equalized, so the equivalent of equation [A.1.11] holds for the function $\mathscr{F}_{\mathrm{w}}(\sigma)$ calculated using workers' payoffs after the deviation (and $\mathcal{U}_{\mathrm{p}}' = \bar{\mathcal{U}}_{\mathrm{p}}^{*}$), hence $\mathscr{F}_{\mathrm{w}}(\sigma) = \sigma p^{*}(\bar{\mathcal{U}}_{\mathrm{p}}^{*} - \mathcal{U}_{\mathrm{w}})$ for all $\sigma \in [0,1]$. Given that $\bar{\mathcal{U}}_{\mathrm{p}}^{*} - \mathcal{U}_{\mathrm{w}} \leq \delta$, it must be the case that $\mathscr{F}_{\mathrm{w}}(\sigma) \leq \sigma p^{*}\delta$. Putting the bounds for $\mathscr{F}_{\mathrm{w}}(\sigma)$ and $\mathscr{F}_{\mathrm{p}}(\sigma)$ together yields:

$$\mathscr{F}(\sigma) = \mathscr{F}_{\mathrm{w}}(\sigma) + \mathscr{F}_{\mathrm{p}}(\sigma) \leq \sigma p^{*}\delta + (1-\sigma)p^{*}\delta = \delta p^{*}, \quad \text{for all } \sigma \in [0,1].$$

This establishes that the institutions following the deviation from the conjectured Markovian equilibrium increase $\bar{\mathcal{U}}_{\mathrm{p}}$ while continuing to satisfy the general no-rebellion constraint [A.1.9]. Therefore, any Markovian equilibrium must feature $\bar{\mathcal{U}}_{\mathrm{p}}^{*} - \mathcal{U}_{\mathrm{w}}^{*} \geq \delta$. Combined with [A.1.12], the following condition necessarily holds in any Markovian equilibrium:

$$\bar{\mathcal{U}}_{\mathrm{p}}^{*} - \mathcal{U}_{\mathrm{w}}^{*} = \delta. \qquad \text{[A.1.14]}$$

Now consider the possibility that there is a Markovian equilibrium in which there is inequality among a group of elite members with positive measure. This means there must be a positive measure of elite members with payoffs $\mathcal{U}_{\mathrm{p}}^{*}(\imath)$ strictly below the average $\bar{\mathcal{U}}_{\mathrm{p}}^{*}$. In a Markovian equilibrium, $\mathcal{U}_{\mathrm{p}}' = \bar{\mathcal{U}}_{\mathrm{p}}^{*}$, so it follows that for these members of the elite:

$$\mathbb{1}[\mathcal{U}_{\mathrm{p}}^{*}(\imath) \leq \mathcal{U}_{\mathrm{p}}'](\mathcal{U}_{\mathrm{p}}^{*}(\imath) \leq \mathcal{U}_{\mathrm{p}}' + \delta) > \delta.$$

As the fraction of places in the subsequent elite assigned to current workers approaches $\sigma = 1$, the measure of elite members who can join the rebel army becomes arbitrarily small. Using the definition of $\mathscr{F}_{\mathrm{p}}(\sigma)$ from [A.1.6] and the continuity of the integral, it must be the case for all $\sigma < 1$ sufficiently close to 1 that the following holds:

$$\mathscr{F}^*_{\mathrm{p}}(\sigma) > (1-\sigma)p^*\delta. \qquad \text{[A.1.15]}$$

Since payoff equalization has already been shown to be a property of any Markovian equilibrium, equations [A.1.11] and [A.1.14] imply:

$$\mathscr{F}^*_{\mathrm{w}}(\sigma) = \sigma p^*(\bar{\mathcal{U}}^*_{\mathrm{p}} - \mathcal{U}^*_{\mathrm{w}}) = \sigma p^*\delta.$$

Together with the inequality from [A.1.15], it follows that for some $\sigma \in (0,1)$:

$$\mathscr{F}^*(\sigma) > \sigma p^*\delta + (1-\sigma)p^*\delta = \delta p^*.$$

This demonstrates that $\mathscr{F}^{*\ddagger} > \delta p^*$, violating the general no-rebellion constraint in [A.1.9].

Therefore, payoff inequality among a positive measure of elite members is inconsistent with the necessary properties of any Markovian equilibrium derived earlier. Any equilibrium must feature $\mathcal{U}^*_{\mathrm{p}}(\imath) = \bar{\mathcal{U}}^*_{\mathrm{p}}$ for all $\imath \in \mathcal{P}$ (except for measure-zero sets of elite members).

Power determines rents

With $\mathcal{U}^*_{\mathrm{p}}(\imath) = \bar{\mathcal{U}}^*_{\mathrm{p}}$ for all $\imath \in \mathcal{P}$, this means that $\bar{\mathcal{U}}^*_{\mathrm{p}} = \bar{\mathcal{U}}^*_{\mathrm{p}}$, so equation [A.1.14] implies $\mathcal{U}^*_{\mathrm{p}} = \mathcal{U}^*_{\mathrm{w}} + \delta$.

Reduction to a single no-rebellion constraint

With payoff equalization among all workers and among all elite members, the choice of institutions reduces to a choice of elite size p (with identities determined by the elite selection function) and a common level of taxes τ levied on workers (with elite consumption then determined by the resource constraint). Equilibrium institutions are then the solution to the following maximization problem

$$\max_{p,\tau} \mathcal{U}_{\mathrm{p}} \;\text{ subject to }\; \sigma(\mathcal{U}'_{\mathrm{p}} - \mathcal{U}_{\mathrm{w}}) + (1-\sigma)\mathbb{1}[\mathcal{U}_{\mathrm{p}} \leq \mathcal{U}'_{\mathrm{p}}](\mathcal{U}'_{\mathrm{p}} - \mathcal{U}_{\mathrm{p}} + \delta) \leq \delta\frac{p}{p'} \;\text{ for all }\; \sigma \in [\underline{\sigma}, 1], \qquad \text{[A.1.16]}$$

and subject to the other constraints (non-negative consumption, $p < 1/2$). The general form of the no-rebellion constraint is derived from [A.1.3] after payoff equalization is assumed, with σ denoting the fraction of places in the post-rebellion elite assigned to those who are currently workers. The bound $\underline{\sigma}$ is defined in equation [A.1.5] (note that the constraint $p < 1/2$ implies $\overline{\sigma} = 1$). A Markovian equilibrium is a solution (p^*, τ^*) of the maximization problem [A.1.16] taking p' and $\mathcal{U}'_{\mathrm{p}}$ as given, but with $p' = p^*$ and $\mathcal{U}'_{\mathrm{p}} = \mathcal{U}^*_{\mathrm{p}}$ in equilibrium.

Now consider a simpler maximization problem where the no-rebellion constraint in [A.1.16] is imposed only for $\sigma = 1$:

$$\max_{p,\tau} \mathcal{U}_{\mathrm{p}} \;\text{ subject to }\; \mathcal{U}'_{\mathrm{p}} - \mathcal{U}_{\mathrm{w}} \leq \delta\frac{p}{p'}, \qquad \text{[A.1.17]}$$

and subject to the non-negativity constraints on consumption and $p < 1/2$. A Markovian equilibrium of this alternative problem is defined as a solution $(p^\star, \tau^\star)$ of the maximization problem [A.1.17], taking p' and $\mathcal{U}'_{\mathrm{p}}$ as given, but with $p' = p^\star$ and $\mathcal{U}'_{\mathrm{p}} = \mathcal{U}^\star_{\mathrm{p}}$ in equilibrium.

Start by considering a Markovian equilibrium $(p^\star, \tau^\star)$ of the simpler problem [A.1.17]. Since these institutions must satisfy the no-rebellion constraint from [A.1.17] and $p' = p^\star$ and $\mathcal{U}'_{\mathrm{p}} = \mathcal{U}^\star_{\mathrm{p}}$, it follows that $\mathcal{U}^\star_{\mathrm{p}} - \mathcal{U}^\star_{\mathrm{w}} \leq \delta$. Therefore, for any $\sigma \in [0,1]$:

$$\sigma(\mathcal{U}^\star_{\mathrm{p}} - \mathcal{U}^\star_{\mathrm{w}}) + (1-\sigma)\mathbb{1}[\mathcal{U}^\star_{\mathrm{p}} \leq \mathcal{U}^\star_{\mathrm{p}}](\mathcal{U}^\star_{\mathrm{p}} - \mathcal{U}^\star_{\mathrm{p}} + \delta) = \sigma(\mathcal{U}^\star_{\mathrm{p}} - \mathcal{U}^\star_{\mathrm{w}}) + (1-\sigma)\delta \leq \sigma\delta + (1-\sigma)\delta = \delta = \delta\frac{p^\star}{p^\star}.$$

This demonstrates that $(p^\star, \tau^\star)$ satisfies the no-rebellion constraint of the original problem [A.1.16] when $p' = p^\star$ and $\mathcal{U}'_{\mathrm{p}} = \mathcal{U}^\star_{\mathrm{p}}$. Now take any other institutions (p, τ) not subject to rebellion in the original problem [A.1.16], again when $p' = p^\star$ and $\mathcal{U}'_{\mathrm{p}} = \mathcal{U}^\star_{\mathrm{p}}$. Evaluating the no-rebellion constraint at $\sigma = 1$ yields $\mathcal{U}^\star_{\mathrm{p}} - \mathcal{U}_{\mathrm{w}} \leq \delta p/p^\star$, which shows that these institutions also satisfy the no-rebellion constraint of the simpler problem [A.1.17]. Since $(p^\star, \tau^\star)$ maximizes $\mathcal{U}_{\text{p}}$ over all institutions satisfying the constraint in [A.1.17], it must be the case that $\mathcal{U}_{\text{p}} \leq \mathcal{U}_{\text{p}}^\star$ for any institutions (p, τ) consistent with the constraint in [A.1.16]. Therefore, $(p^\star, \tau^\star)$ is also a Markovian equilibrium of [A.1.16] as well.

Now consider the converse. Take a Markovian equilibrium (p^*, τ^*) of the original problem [A.1.16]. These institutions are clearly consistent with the constraint in [A.1.17] when $p' = p^*$ and $\mathcal{U}_{\text{p}}' = \mathcal{U}_{\text{p}}^*$ because the constraint is a special case of that in [A.1.16] when $\sigma = 1$. Suppose for contradiction that (p^*, τ^*) is not a Markovian equilibrium of the problem [A.1.17]. Since it satisfies the no-rebellion constraint, it must therefore be the case that there exists another set of institutions (p, τ) such that $\mathcal{U}_{\text{p}} > \mathcal{U}_{\text{p}}^*$ satisfying the no-rebellion constraint in [A.1.17] when $p' = p^*$ and $\mathcal{U}_{\text{p}}' = \mathcal{U}_{\text{p}}^*$. Now take any $\sigma \in [\underline{\sigma}, 1]$ and multiply both sides of the inequality in [A.1.17] by this number to obtain:

$$\sigma(\mathcal{U}_{\text{p}}^* - \mathcal{U}_{\text{w}}) \leq \sigma \delta \frac{p}{p^*} \leq \delta \frac{p}{p^*}.$$

Observe that $(1-\sigma)\mathbb{1}[\mathcal{U}_{\text{p}} \leq \mathcal{U}_{\text{p}}^*](\mathcal{U}_{\text{p}}^* - \mathcal{U}_{\text{p}} + \delta) = 0$, so this demonstrates that (p, τ) satisfies [A.1.16] for all $\sigma \in [\underline{\sigma}, 1]$. Since these institutions satisfy the no-rebellion constraint in [A.1.16], the resulting elite payoff cannot be higher than the elite payoff in equilibrium, hence $\mathcal{U}_{\text{p}} \leq \mathcal{U}_{\text{p}}^*$. This contradicts the inequality $\mathcal{U}_{\text{p}} > \mathcal{U}_{\text{p}}^*$ obtained earlier, and thus proves that (p^*, τ^*) must be a Markovian equilibrium of the simpler problem [A.1.17].

In summary, it has been shown that the set of Markovian equilibria of the original problem [A.1.16] is identical to the set of Markovian equilibrium of the simpler problem [A.1.17]. Therefore, there is no loss of generality in imposing [3.9] as the only no-rebellion constraint.

Existence and uniqueness of the Markovian equilibrium

With payoff equalization, the resource constraint implies $C_{\text{p}} = (1-p)\tau/p$. Worker and elite payoffs are given by $\mathcal{U}_{\text{w}} = u(\text{q} - \tau)$ and $\mathcal{U}_{\text{p}} = u((1-p)\tau/p)$ respectively. It has been shown that the Markovian equilibrium institutions can be characterized as a solution of [A.1.17]. Any Markovian equilibrium (p^*, τ^*) is thus a solution of the maximization problem

$$\max_{p,\tau} u\left(\frac{(1-p)\tau}{p}\right) \text{ subject to } \mathcal{U}_{\text{p}}^* - u(\text{q} - \tau) \leq \delta \frac{p}{p^*}, \qquad \text{[A.1.18]}$$

taking p^* and $\mathcal{U}_{\text{p}}^*$ as given, but with $p = p^*$ and $\mathcal{U}_{\text{p}} = \mathcal{U}_{\text{p}}^*$ in equilibrium. The solution of the maximization problem must also respect the constraint $p < 1/2$, and the non-negativity constraints on all individuals' consumption, which are equivalent to $0 \leq \tau \leq \text{q}$ here.

It has already been seen that the no-rebellion constraint must be binding, which was reduced to a single equation in [A.1.18]. This equation can be solved to obtain τ as a function of p for given values of p^* and $\mathcal{U}_{\text{p}}^*$:

$$\tau = \text{q} - u^{-1}\left(\mathcal{U}_{\text{p}}^* - \delta \frac{p}{p^*}\right). \qquad \text{[A.1.19]}$$

Differentiating shows that the constraint implicitly specifies a positive relationship between τ and p:

$$\frac{\partial \tau}{\partial p} = \frac{\delta}{p^*} \frac{1}{u'\left(u^{-1}\left(\mathcal{U}_{\text{p}}^* - \delta \frac{p}{p^*}\right)\right)} > 0. \qquad \text{[A.1.20]}$$

The problem [A.1.18] is equivalent to maximizing $C_{\text{p}} = (1-p)\tau/p$ over values of p after substituting for τ using equation [A.1.19]:

$$\max_{p} \frac{(1-p)}{p}\left(\text{q} - u^{-1}\left(\mathcal{U}_{\text{p}}^* - \delta \frac{p}{p^*}\right)\right), \qquad \text{[A.1.21]}$$

subject to $p < 1/2$ and the value of τ implied by [A.1.19] being such that $0 \leq \tau \leq \text{q}$. Taking the derivative of C_p with respect to p (making use of [A.1.20]):

$$\frac{\partial C_\mathrm{p}}{\partial p} = \frac{1}{p^2}\left(\frac{p}{p^*}\frac{\delta(1-p)}{u'\left(u^{-1}\left(\mathcal{U}_\mathrm{p}^* - \delta\frac{p}{p^*}\right)\right)} - \left(\mathrm{q} - u^{-1}\left(\mathcal{U}_\mathrm{p}^* - \delta\frac{p}{p^*}\right)\right)\right). \qquad \text{[A.1.22]}$$

The second derivative is

$$\frac{\partial^2 C_\mathrm{p}}{\partial p^2} = -\frac{2}{u'\left(u^{-1}\left(\mathcal{U}_\mathrm{p}^* - \delta\frac{p}{p^*}\right)\right)}\frac{\delta p}{p^*} + \frac{u''\left(u^{-1}\left(\mathcal{U}_\mathrm{p}^* - \delta\frac{p}{p^*}\right)\right)}{\left\{u'\left(u^{-1}\left(\mathcal{U}_\mathrm{p}^* - \delta\frac{p}{p^*}\right)\right)\right\}^3} - \frac{2}{p}\frac{\partial C_\mathrm{p}}{\partial p}. \qquad \text{[A.1.23]}$$

In a Markovian equilibrium it is necessary to have $p = p^*$, in which case the binding no-rebellion constraint [A.1.19] reduces to

$$\tau^* = \mathrm{q} - u^{-1}(\mathcal{U}_\mathrm{p}^* - \delta),$$

or equivalently

$$u\left(\frac{(1-p^*)(\mathrm{q} - C_\mathrm{w}^*)}{p^*}\right) = \mathcal{U}_\mathrm{p}^* = u(\mathrm{q} - \tau^*) + \delta = u(C_\mathrm{w}^*) + \delta, \qquad \text{[A.1.24]}$$

where $C_\mathrm{w}^* = \mathrm{q} - \tau^*$ is the consumption of a worker. This equation is in turn equivalent to

$$p^* u^{-1}(u(C_\mathrm{w}^*) + \delta) = (1-p^*)(\mathrm{q} - C_\mathrm{w}^*). \qquad \text{[A.1.25]}$$

Evaluating the derivative of C_p from [A.1.22] at $p = p^*$:

$$\left.\frac{\partial C_\mathrm{p}}{\partial p}\right|_{p=p^*} = \frac{1}{p^{*2}}\left(\frac{\delta(1-p^*)}{u'\left(u^{-1}\left(\mathcal{U}_\mathrm{p}^* - \delta\right)\right)} - \left(\mathrm{q} - u^{-1}\left(\mathcal{U}_\mathrm{p}^* - \delta\right)\right)\right).$$

Since [A.1.24] implies that $u^{-1}(\mathcal{U}_\mathrm{p}^* - \delta) = C_\mathrm{w}^*$, this can be simplified as follows:

$$\left.\frac{\partial C_\mathrm{p}}{\partial p}\right|_{p=p^*} = \frac{1}{p^{*2}}\left(\frac{\delta(1-p^*)}{u'\left(C_\mathrm{w}^*\right)} - (\mathrm{q} - C_\mathrm{w}^*)\right). \qquad \text{[A.1.26]}$$

Now define the following functions $\mathscr{G}(p, \varkappa)$ and $\mathscr{U}(p, \varkappa)$, which will represent respectively (with $\varkappa = C_\mathrm{w}$) the no-rebellion constraint and the first-order condition in a Markovian equilibrium:

$$\mathscr{G}(p, \varkappa) \equiv p u^{-1}(u(\varkappa) + \delta) - (1-p)(\mathrm{q} - \varkappa), \quad \mathscr{U}(p, \varkappa) \equiv \delta(1-p) - (\mathrm{q} - \varkappa)u'(\varkappa). \qquad \text{[A.1.27]}$$

The no-rebellion constraint [A.1.25] for a Markovian equilibrium holds when $\mathscr{G}(p^*, C_\mathrm{w}^*)$ is zero, while the derivative of C_p in [A.1.26] is proportional to $\mathscr{U}(p^*, C_\mathrm{w}^*)$:

$$\mathscr{G}(p^*, C_\mathrm{w}^*) = 0, \quad \text{and} \quad \left.\frac{\partial C_\mathrm{p}}{\partial p}\right|_{p=p^*} = \frac{1}{p^{*2}}\frac{1}{u'(C_\mathrm{w}^*)}\mathscr{U}(p^*, C_\mathrm{w}^*). \qquad \text{[A.1.28]}$$

The partial derivatives of the function $\mathscr{G}(p, \varkappa)$ from [A.1.27] are:

$$\frac{\partial \mathscr{G}}{\partial p} = u^{-1}(u(\varkappa) + \delta) + (\mathrm{q} - \varkappa), \quad \frac{\partial \mathscr{G}}{\partial \varkappa} = p\frac{u'(\varkappa)}{u'\left(u^{-1}(u(\varkappa) + \delta)\right)} + (1-p). \qquad \text{[A.1.29]}$$

Since $u(\cdot)$ is strictly increasing, $u'(\varkappa) > 0$ and $u'(u^{-1}(u(\varkappa) + \delta)) > 0$, and also $u^{-1}(u(\varkappa) + \delta) > \varkappa$. It follows that both of the partial derivatives above are strictly positive for all $0 \leq p < 1/2$ and $0 \leq \varkappa \leq \mathrm{q}$. The partial derivatives of the function $\mathscr{U}(p, \varkappa)$ from [A.1.27] are:

$$\frac{\partial \mathscr{U}}{\partial p} = -\delta, \quad \frac{\partial \mathscr{U}}{\partial \varkappa} = u'(\varkappa) - (\mathrm{q} - \varkappa)u''(\varkappa). \qquad \text{[A.1.30]}$$

The properties of $u(\cdot)$ ensure that $u'(\varkappa) > 0$ and $u''(\varkappa) \leq 0$, so $\mathscr{U}(p,\varkappa)$ is strictly decreasing in p and strictly increasing in $\varkappa$.

Now consider two functions $\mathscr{H}(p)$ and $\mathscr{V}(p)$ defined implicitly by the equations:

$$\mathscr{G}(p,\mathscr{H}(p)) = 0, \quad \text{and} \quad \mathscr{U}(p,\mathscr{V}(p)) = 0. \tag{[A.1.31]}$$

Where these functions are defined, the signs of their partial derivatives can be deduced using [A.1.29] and [A.1.30]:

$$\mathscr{H}'(p) = -\frac{\partial\mathscr{G}}{\partial p}\Big/\frac{\partial\mathscr{G}}{\partial\varkappa} < 0, \quad \text{and} \quad \mathscr{V}'(p) = -\frac{\partial\mathscr{U}}{\partial p}\Big/\frac{\partial\mathscr{U}}{\partial\varkappa} > 0. \tag{[A.1.32]}$$

Observe from [A.1.27] that $\mathscr{U}(1,\varkappa) = -(\mathrm{q}-\varkappa)u'(\varkappa)$, with the definition [A.1.31] then implying $\mathscr{V}(1) = \mathrm{q}$ given $u'(\cdot) > 0$. [A.1.32] shows that $\mathscr{V}(p)$ is strictly increasing in p, so it follows by continuity that either there exists a $\underline{\pi} > 0$ such that $\mathscr{V}(\underline{\pi}) = 0$, or $\mathscr{V}(0) \geq 0$, in which case $\underline{\pi}$ is set to zero. With the resulting $\underline{\pi} \in [0,1)$, define $\underline{\varkappa} \equiv \mathscr{V}(\underline{\pi})$, noting that this satisfies $0 \leq \underline{\varkappa} < \mathrm{q}$ because $\mathscr{V}(\underline{\pi}) \geq 0$ and $\mathscr{V}(1) = \mathrm{q}$. The function $\mathscr{V}(p)$ is then well defined on the interval $[\underline{\pi},1]$ in the sense of returning a value of $\varkappa$ in the interval $[\underline{\varkappa},\mathrm{q}]$.

It can be seen from [A.1.27] that $\mathscr{G}(0,\varkappa) = -(\mathrm{q}-\varkappa)$, so the definition in [A.1.31] implies $\mathscr{H}(0) = \mathrm{q}$. Note that since the utility function $u(\cdot)$ is strictly increasing, so is its inverse $u^{-1}(\cdot)$. It follows that $u^{-1}(u(\varkappa)+\delta) > \varkappa$ and thus for all $p > 0$:

$$\mathscr{G}(p,\varkappa) > p\varkappa - (1-p)(\mathrm{q}-\varkappa) = \varkappa - \mathrm{q}(1-p).$$

Using [A.1.31], this means that $0 = \mathscr{G}(p,\mathscr{H}(p)) > \mathscr{H}(p) - \mathrm{q}(1-p)$, and hence

$$\mathscr{H}(p) < \mathrm{q}(1-p), \quad \text{for all } \; p > 0.$$

[A.1.32] shows that $\mathscr{H}(p)$ is strictly decreasing in p, so given $\mathscr{H}(0) = \mathrm{q}$ and the bound above, it follows by continuity that there exists a $\overline{\pi} \in (0,1)$ such that $\mathscr{H}(\overline{\pi}) = 0$. The function $\mathscr{H}(p)$ is then well defined on the interval $[0,\overline{\pi}]$ in the sense of returning a value of $\varkappa$ in the interval $[0,\mathrm{q}]$.

Let $\mathscr{H}^{-1}(\varkappa)$ denote the inverse function of $\mathscr{H}(p)$, defined on $[0,\mathrm{q}]$. Similarly, $\mathscr{V}^{-1}(\varkappa)$ is the inverse function of $\mathscr{V}(p)$, defined on $[\underline{\varkappa},\mathrm{q}]$, where $0 \leq \underline{\varkappa} \equiv \mathscr{V}(\underline{\pi}) < \mathrm{q}$. Since $\mathscr{H}(p)$ is strictly decreasing and $\mathscr{V}(p)$ is strictly increasing according to [A.1.32], their inverse functions inherit these properties. Now define the following function $\mathscr{A}(\varkappa)$ on $[\underline{\varkappa},\mathrm{q}]$:

$$\mathscr{A}(\varkappa) \equiv \mathscr{V}^{-1}(\varkappa) - \mathscr{H}^{-1}(\varkappa), \tag{[A.1.33]}$$

where the properties of $\mathscr{H}(p)$ and $\mathscr{V}(p)$ imply that $\mathscr{A}(\varkappa)$ is strictly increasing in $\varkappa$. Note also that $\mathscr{A}(\mathrm{q}) = 1 - 0 = 1$ since $\mathscr{H}(0) = \mathrm{q}$ and $\mathscr{V}(1) = \mathrm{q}$.

Consider first the case where $\underline{\pi} < \overline{\pi}$. The definition [A.1.33] and $\mathscr{V}(\underline{\pi}) = \underline{\varkappa}$ imply:

$$\mathscr{A}(\underline{\varkappa}) = \underline{\pi} - \mathscr{H}^{-1}(\underline{\varkappa}).$$

The definition of $\underline{\pi}$ was such that $\underline{\pi} = 0$ if $\underline{\varkappa} > 0$, and so $\mathscr{A}(\underline{\varkappa}) = -\mathscr{H}^{-1}(\underline{\varkappa})$. Given that $\underline{\varkappa} < \mathrm{q}$, $\mathscr{H}(0) = \mathrm{q}$, and $\mathscr{H}(p)$ is strictly decreasing, it follows that $\mathscr{H}^{-1}(\underline{\varkappa}) > 0$, hence $\mathscr{A}(\underline{\varkappa}) < 0$. When $\underline{\varkappa} = 0$, note that $\mathscr{H}^{-1}(\underline{\varkappa}) = \overline{\pi}$ since $\mathscr{H}(\overline{\pi}) = 0$. This implies that $\mathscr{A}(\underline{\varkappa}) = \underline{\pi} - \overline{\pi} < 0$ because $\underline{\pi} < \overline{\pi}$ in the case under consideration. Therefore, it has been shown that $\mathscr{A}(\underline{\varkappa}) < 0$ and $\mathscr{A}(\mathrm{q}) > 0$, thus $\mathscr{A}(\varkappa)$ being continuous and strictly increasing proves there exists a unique value $\varkappa^* \in (0,\mathrm{q})$ such that $\mathscr{A}(\varkappa^*) = 0$.

Consider the remaining case where $\underline{\pi} \geq \overline{\pi}$. It is necessary that $\underline{\pi} > 0$ in this case since $\overline{\pi} > 0$, and the definition of $\underline{\pi}$ then guarantees that $\mathscr{V}(\underline{\pi}) = 0$, and $\underline{\varkappa} = 0$ because $\underline{\varkappa} = \mathscr{V}(\underline{\pi})$. Therefore, $\mathscr{A}(\underline{\varkappa}) = \mathscr{A}(0) = \underline{\pi} - \overline{\pi} \geq 0$. Since $\mathscr{A}(\varkappa)$ is continuous and strictly increasing and $\mathscr{A}(\mathrm{q}) > 0$ as before, it follows that either $\varkappa^* = 0$ is the only possible solution of $\mathscr{A}(\varkappa) = 0$ for $\varkappa \in [0,\mathrm{q}]$, or there is no solution of the equation. Whether or not a solution exists, set $\varkappa^* = 0$ in this case, noting that $\mathscr{A}(\varkappa^*) \geq 0$.

Depending on which case above prevails there is either $\mathscr{A}(\varkappa^*) = 0$ or $\mathscr{A}(\varkappa^*) \geq 0$. In both cases, $\varkappa^* < \mathrm{q}$. Define $p^* = \mathscr{H}^{-1}(\varkappa^*)$, and since $\mathscr{H}(p)$ is strictly decreasing and $\mathscr{H}(0) = \mathrm{q}$, it follows that $p^* > 0$. The definition of $\mathscr{H}(p)$ in [A.1.31] also implies that $\mathscr{G}(p^*, \varkappa^*) = 0$. Therefore, using [A.1.27]:

$$q - \varkappa^* = \frac{p^*}{1-p^*} u^{-1}(u(\varkappa^*) + \delta). \qquad \text{[A.1.34]}$$

Since $\mathscr{A}(\varkappa^*) \geq 0$, equation [A.1.33] implies $\mathscr{V}^{-1}(\varkappa^*) \geq \mathscr{H}^{-1}(\varkappa^*) = p^*$. As has been shown in [A.1.30], $\mathscr{U}(p, \varkappa)$ is decreasing in p. The definition [A.1.31] implies $\mathscr{U}(\mathscr{V}^{-1}(\varkappa), \varkappa) = 0$, hence it then follows that $\mathscr{U}(p^*, \varkappa^*) \geq 0$. Using [A.1.27]:

$$q - \varkappa^* \leq \frac{\delta(1-p^*)}{u'(\varkappa^*)},$$

and combining this with [A.1.34] yields:

$$\frac{p^*}{1-p^*} u^{-1}(u(\varkappa^*) + \delta) \leq \frac{\delta(1-p^*)}{u'(\varkappa^*)}.$$

Therefore, the following inequality must hold:

$$u^{-1}(u(\varkappa^*) + \delta) \leq \frac{(1-p^*)^2}{p^*} \frac{\delta}{u'(C_w^*)}. \qquad \text{[A.1.35]}$$

Now note that since $u(\cdot)$ is a concave function, its inverse $u^{-1}(\cdot)$ is a convex function, so it is bounded below by its tangent at $u(\varkappa^*)$. Together with $\varkappa^* \geq 0$, this leads to:

$$u^{-1}(u(\varkappa^*) + \delta) \geq u^{-1}(u(\varkappa^*)) + \frac{1}{u'(u^{-1}(u(\varkappa^*)))} \delta = \varkappa^* + \frac{\delta}{u'(\varkappa^*)} \geq \frac{\delta}{u'(\varkappa^*)}.$$

By combining this the earlier inequality in [A.1.35]:

$$\frac{\delta}{u'(\varkappa^*)} \leq \frac{(1-p^*)^2}{p^*} \frac{\delta}{u'(\varkappa^*)}, \quad \text{and hence } 1 \leq \frac{(1-p^*)^2}{p^*}.$$

Therefore, the value of p^* must satisfy the quadratic inequality $\mathscr{B}(p^*) \geq 0$ where:

$$\mathscr{B}(p) \equiv (1-p)^2 - p = p^2 - 3p + 1.$$

Since $\mathscr{B}(0) > 0$ and $\mathscr{B}(1) < 0$, the quadratic $\mathscr{B}(p)$ has exactly one root $\bar{p} \in (0, 1)$. The product of the roots is positive, so this must be the smallest root, which can then be obtained using the formula:

$$\bar{p} = \frac{3 - \sqrt{5}}{2} = 2 - \left(\frac{1+\sqrt{5}}{2}\right) = 2 - \varphi, \qquad \text{[A.1.36]}$$

where $\varphi \equiv (1 + \sqrt{5})/2$ is the *Golden ratio* introduced in footnote 6. As $\mathscr{B}(p^*) \geq 0$, it must be the case that $p^* \leq \bar{p}$, and therefore $p^* \leq 2 - \varphi$.

Given the constraint $p < 1/2$, the search for a Markovian equilibrium is restricted to the interval $p \in [0, 1/2]$. The non-negativity constraints on consumption are equivalent to $0 \leq C_w \leq q$, where $C_w = q - \tau$. Equation [A.1.17] shows that the value of τ consistent with the binding no-rebellion constraint is strictly increasing in p. Since the utility function $u(\cdot)$ is strictly increasing and weakly concave, equations [A.1.22] and [A.1.23] imply that any critical point of the objective function C_p must be a local maximum. Therefore, these observations show that the general necessary and sufficient condition for a Markovian equilibrium is:

$$\left.\frac{\partial C_p}{\partial p}\right|_{p=p^*} \begin{cases} \leq 0 & \text{if } p^* = 0 \text{ or } C_w^* = q \\ = 0 & \text{if } 0 < p^* < 1/2 \text{ and } 0 < C_w^* < q \\ \geq 0 & \text{if } p^* = 1/2 \text{ or } C_w^* = 0 \end{cases}. \qquad \text{[A.1.37]}$$

Consider first the possibility of a Markovian equilibrium with $p^* = 0$ or $C_w^* = q$. Equation [A.1.28] and

[A.1.31] imply that $C_\text{w}^* = \mathscr{H}(p^*)$, and since $\mathscr{H}(0) = \text{q}$ it follows that any such equilibrium must feature $p^* = 0$ and $C_\text{w}^* = \text{q}$. Thus, by using [A.1.27], $\mathscr{U}(p^*, C_\text{w}^*) = \delta > 0$. From equation [A.1.28] it follows that $\partial C_\text{p}/\partial p \to \infty$ at $p^* = 0$. But the first-order condition [A.1.37] would require $\partial C_\text{p}/\partial p \leq 0$ for this type of equilibrium. Therefore, there are no Markovian equilibria with either $p^* = 0$ or $C_\text{w}^* = \text{q}$.

Now consider the possibility of a Markovian equilibrium with $p^* = 1/2$. From the relevant first-order condition in [A.1.37] and [A.1.28], such an equilibrium would need to satisfy $\mathscr{U}(p^*, C_\text{w}^*) \geq 0$ and $\mathscr{G}(p^*, C_\text{w}^*) = 0$. But it has been shown that $p^* \leq \overline{p}$ for any value of p^* consistent with these conditions, where $\overline{p}$ is defined in [A.1.36]. It can be seen that $\overline{p} < 1/2$, so there are no Markovian equilibria with $p^* = 1/2$.

Next, consider a case of a Markovian equilibrium with $0 < p^* < 1/2$ and $0 < C_\text{w}^* < \text{q}$. Using [A.1.37] and [A.1.28], the required conditions are $\mathscr{G}(p^*, C_\text{w}^*) = 0$ and $\mathscr{U}(p^*, C_\text{w}^*) = 0$. From [A.1.31], this is seen to be equivalent to $p^* = \mathscr{H}^{-1}(C - \text{w}^*)$ and $p^* = \mathscr{V}^{-1}(C_\text{w}^*)$, and to $\mathscr{A}(C_\text{w}^*) = 0$ using [A.1.33]. In the case where $\underline{\pi} < \overline{\pi}$ such a solution has been shown to exist, and to be unique. There is no solution when $\underline{\pi} \geq \overline{\pi}$. Therefore, a Markovian equilibrium of this type exists (and is unique among those of this type) if and only if $\underline{\pi} < \overline{\pi}$.

Finally, consider the case of a Markovian equilibrium with $C_\text{w}^* = 0$. According to [A.1.28], this must satisfy $\mathscr{G}(p^*, 0) = 0$, and hence $p^* = \mathscr{H}^{-1}(0)$ using [A.1.31]. Using [A.1.37] and [A.1.28], the first-order condition in this case requires $\mathscr{U}(p^*, 0) \geq 0$. The definition in [A.1.31] implies $\mathscr{U}(\mathscr{V}^{-1}(0), 0) = 0$, and since [A.1.30] shows $\mathscr{U}(p, \varkappa)$ is strictly decreasing in p, it follows that $\mathscr{U}(\mathscr{H}^{-1}(0), 0) \geq 0$ if and only if $\mathscr{V}^{-1}(0) \geq \mathscr{H}^{-1}(0)$. This is seen to be equivalent to $\mathscr{A}(\varkappa^*) \geq 0$ using [A.1.33]. The earlier analysis shows this inequality is satisfied if and only if $\underline{\pi} \geq \overline{\pi}$. Hence a unique Markovian equilibrium exists in this case too.

Therefore, irrespective of whether $\underline{\pi} < \overline{\pi}$ or $\underline{\pi} \geq \overline{\pi}$ holds, a unique Markovian equilibrium exists. In the case $\underline{\pi} < \overline{\pi}$, the equilibrium features $0 < C_\text{w}^* < \text{q}$, so all non-negativity constraints are slack. In the case $\underline{\pi} \geq \overline{\pi}$, the equilibrium features $C_\text{w}^* = 0$, so the non-negativity constraint is binding for workers. In all cases, $0 < p^* \leq 2 - \varphi < 1/2$.

Define the following normalized value of the parameter δ:

$$\beta \equiv \frac{\delta}{u'(0)\text{q}}. \tag{A.1.38}$$

Using this definition, observe from [A.1.27] that $\mathscr{U}(0, 0) = u'(0)\text{q}(\beta - 1)$. Hence, if $\beta \geq 1$ then it must be the case that $\mathscr{U}(0, 0) \leq 0$. Note that $\mathscr{U}(0, \mathscr{V}(0)) = 0$ follows from the definition in [A.1.31], which implies $\mathscr{V}(0) \geq 0$ since $\mathscr{U}(p, \varkappa)$ is strictly increasing in $\varkappa$ according to [A.1.30]. The earlier construction of $\underline{\pi}$ is such that $\underline{\pi} = 0$ whenever $\mathscr{V}(0) \geq 0$. As $\overline{\pi} > 0$, it follows that $\underline{\pi} < \overline{\pi}$, and thus that the unique Markovian equilibrium does not have a binding non-negativity constraint for workers.

Now suppose that $\beta > 1$, in which case $\mathscr{U}(0, 0) > 0$ using the argument above. It follows that $\underline{\pi} > 0$, with this variable defined as $\underline{\pi} = \mathscr{V}^{-1}(0)$. Given [A.1.31], this means that $\underline{\pi}$ satisfies $\mathscr{U}(\underline{\pi}, 0) = 0$, and by using [A.1.27]:

$$\delta(1 - \underline{\pi}) - u'(0)\text{q} = 0.$$

Solving this equation explicitly in terms of the parameter β defined in [A.1.38] yields:

$$\underline{\pi} = \frac{\beta - 1}{\beta}, \tag{A.1.39}$$

The term $\overline{\pi}$ is always defined as the solution of the equation $\mathscr{G}(\overline{\pi}, 0) = 0$. Equation [A.1.27] then leads to

$$\overline{\pi} u^{-1}(u(0) + \delta) = (1 - \overline{\pi})\text{q},$$

which can be rearranged as follows:

$$\frac{\overline{\pi}}{1 - \overline{\pi}} = \frac{\text{q}}{u^{-1}(u(0) + \delta)}.$$

Observe from [A.1.39] that $\underline{\pi}/(1 - \underline{\pi}) = \beta - 1$. The inequality $\underline{\pi} < \overline{\pi}$ is equivalent to $\underline{\pi}/(1 - \underline{\pi}) < \overline{\pi}/(1 - \overline{\pi})$.

By using the equation above, $\underline{\pi} < \overline{\pi}$ is also equivalent to:

$$\beta < 1 + \frac{q}{u^{-1}(u(0) + \delta)}. \qquad [A.1.40]$$

Substituting for β from [A.1.38] thus provides the necessary and sufficient condition for an equilibrium with all non-negativity constraints being slack since $\underline{\pi} < \overline{\pi}$ is necessary and sufficient for this.

Finally, note that since the utility function $u(\cdot)$ is weakly concave, its inverse $u^{-1}(\cdot)$ is weakly convex. That implies

$$u^{-1}(u(0) + \delta) \geq u^{-1}(u(0)) + \frac{1}{u'(u^{-1}(u(0)))}\delta = \frac{\delta}{u'(0)},$$

and by using the definition of β from [A.1.38]:

$$\frac{q}{u^{-1}(u(0) + \delta)} \leq \beta^{-1}.$$

Now suppose condition [A.1.40] holds. Putting [A.1.40] together with the inequality above yields:

$$\beta < 1 + \beta^{-1}, \quad \text{or equivalently} \;\; \Phi(\beta) < 0 \;\; \text{where} \;\; \Phi(\beta) \equiv \beta^2 - \beta - 1.$$

The quadratic equation $\Phi(\beta) = 0$ has a positive and a negative root, and since $\Phi(0) < 0$, the function $\Phi(\beta)$ takes negative values when β lies strictly between its two roots. Observe that $\Phi(\varphi) = 0$, so the *Golden ratio* $\varphi \equiv (1 + \sqrt{5})/2$ is the positive root. The condition $\Phi(\beta) < 0$ then requires $\beta < \varphi$, so this is a necessary condition for an equilibrium with no binding non-negativity constraints. It follows that $\beta \geq \varphi$ is then sufficient for an equilibrium with a binding non-negativity constraint for workers since this is the only other possible equilibrium. This completes the proof.

A.2 Proof of Proposition 2

Consider a Markovian equilibrium $\{p^*, \tau^*_q(\imath), \tau^*_\kappa, C^*_p(\imath)\}$ in which a positive fraction s^* of investment opportunities are taken.

There are always sufficient non-elite members or workers to fill a rebel army following a rebellion at any stage

At the pre-investment stage, individuals outside the elite do not yet know whether they will be investors or workers. The number of non-elite members is $n = 1 - p$, so given that $p < 1/2$ and $p' < 1/2$, n will always exceed p', the size of the subsequent elite following a rebellion at the pre-investment stage.

At the post-investment stage, workers are those outside the elite who did not become investors. Their number is $w = 1 - p - i$. At this point, any rebellion would feature a subsequent elite size of $p^\dagger$. Observe that

$$w - p^\dagger = (1 - p - i) - p^\dagger = (1 - p) - \mu + (\mu - i) - p^\dagger = \left((1 - p) - \frac{1}{2}\right) + \left(\frac{1}{2} - p^\dagger - \mu\right) + (1 - s)\mu,$$

where μ is the measure of investment opportunities, and $i = \mu s$, given the definition in [5.7]. The formula for $p^\dagger$ from [5.5] implies

$$\frac{1}{2} - p^\dagger = \frac{q}{2(q + 2\delta)},$$

and hence

$$w - p^\dagger = \left((1 - p) - \frac{1}{2}\right) + \left(\frac{q}{2(q + 2\delta)} - \mu\right) + (1 - s)\mu > 0,$$

for all parameters consistent with [5.3] and all feasible values of p and s. Therefore, $w > p^\dagger$, so there are always sufficient workers to fill a rebel army at the post-investment stage.

Payoff equalization for workers

Suppose a Markovian equilibrium features a non-degenerate distribution of taxes $\{\tau_{\mathrm{q}}^*(\imath)\}$ on economically active individuals, and hence dispersion in worker payoffs $\{\mathcal{U}_{\mathrm{w}}^*(\imath)\}$. Following the argument in Proposition 1, in the case where no workers belong to a rebel army associated with a binding no-rebellion constraint, it would be possible for the elite to increase taxes on a subset of workers without violating any no-rebellion constraint by targeting the tax increases on those receiving higher payoffs (and the increase in elite payoffs cannot increase fighting effort from elite members who join a rebel army). The second case is where some workers belong to a rebel army associated with a binding no-rebellion constraint. Since workers (or non-elite members at the pre-investment stage) are more numerous than the maximum number of places in the rebel army, no rebel army can include all of them. Payoff equalization among workers thus strictly reduces the fighting effort of any rebel army including a positive number of workers. This slackens the binding no-rebellion constraints and allows the elite to raise taxes. Since a higher elite payoff is feasible in both cases, there are no Markovian equilibria with payoff inequality among a set of workers with positive measure.

Therefore, the search for a Markovian equilibrium can be restricted to those that feature payoff equalization for all workers, that is, a single tax on the endowment τ_{q} and a common payoff $\mathcal{U}_{\mathrm{w}}(\imath) = \mathcal{U}_{\mathrm{w}}$ for all $\imath \in \mathcal{W}$. Taxes can always be raised to ensure $\mathcal{U}_{\mathrm{w}} \leq \mathcal{U}_{\mathrm{p}}^{\dagger}(K)$ without any danger of rebellion, and this would raise the elite's payoff, so $\mathcal{U}_{\mathrm{p}}^{\dagger}(K)$ is an upper bound for worker payoffs in equilibrium. Given that there are always sufficient workers at the post-investment stage, $\sigma_{\mathrm{w}}^{\dagger} = 1$ is feasible in the general no-rebellion constraint [5.11] at the post-investment stage. Therefore, the following no-rebellion constraint must always hold:

$$\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \mathcal{U}_{\mathrm{w}} \leq \delta \frac{p}{p^{\dagger}}. \qquad \text{[A.2.1]}$$

No investors in a rebel army with a binding constraint at the post-investment stage

Ex post, investors receive the common payoff $\mathcal{U}_{\mathrm{k}}$ given in [5.8] (given a equal tax distribution $\tau_{\mathrm{q}}(\imath) = \tau_{\mathrm{q}}$ and an equal tax τ_{κ} on capital — the latter is without loss of generality as discussed in footnote 31. Comparison with [5.6] shows that $\mathcal{U}_{\mathrm{k}} = \mathcal{U}_{\mathrm{w}} + (\kappa - \tau_{\kappa})$, and given the incentive compatibility condition for investors from [5.7], this implies:

$$\mathcal{U}_{\mathrm{k}} = \mathcal{U}_{\mathrm{w}} + \tilde{\theta}. \qquad \text{[A.2.2]}$$

Since $0 < \psi \leq \tilde{\theta} \leq \kappa$ according to [5.1], any Markovian equilibrium must feature $\tilde{\theta}^* > 0$. Hence, $\mathcal{U}_{\mathrm{k}}^* = \mathcal{U}_{\mathrm{w}}^* + \tilde{\theta}^* > \mathcal{U}_{\mathrm{w}}^*$. Therefore, in an equilibrium with $s^* > 0$, there is a positive measure of investors, but it can be seen from [5.11] that these would exert strictly less fighting effort than any worker included in the rebel army. Since it has been shown there are enough workers to fill the whole rebel army, investors would never be included in a rebel army associated with a binding no-rebellion constraint. Therefore, in what follows, the search for a Markovian equilibrium can be confined to cases where the restriction $\sigma_{\mathrm{i}}^{\dagger} = 0$ is imposed on the general no-rebellion constraint in [5.11].

The general no-rebellion constraint at the post-investment stage

In what follows, let σ denote the fraction of places in the subsequent elite (and hence positions in the rebel army) that are filled by those who are currently workers if a rebellion occurs at the post-investment stage. Define the function $\mathscr{R}(\sigma)$:

$$\mathscr{R}(\sigma) \equiv \delta \frac{p}{p^{\dagger}} - \frac{1}{p^{\dagger}} \max_{\substack{\mathcal{E}_{\mathrm{p}} \subseteq \mathcal{P} \\ |\mathcal{E}_{\mathrm{p}}| = (1-\sigma)p^{\dagger}}} \int_{\mathcal{E}_{\mathrm{p}}} \mathbb{1}[\mathcal{U}_{\mathrm{p}}(\imath) \leq \mathcal{U}_{\mathrm{p}}^{\dagger}(K)]\{\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \mathcal{U}_{\mathrm{p}}(\imath) + \delta\} \mathrm{d}\imath - \sigma(\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \mathcal{U}_{\mathrm{w}}), \qquad \text{[A.2.3]}$$

where $\mathcal{U}_{\mathrm{p}}^{\dagger}(K)$ is the payoff of an elite member under the institutions that would be formed following a post-investment rebellion, and is $p^{\dagger}$ the subsequent elite size, as given in [5.5]. The function above is defined for a given distribution of elite payoffs $\mathcal{U}_{\mathrm{p}}(\imath)$. The set $\mathcal{E}_{\mathrm{p}}$ denotes the subset of the current elite $\mathcal{P}$ who are offered places in the subsequent elite, which can be varied by considering different elite selection functions $\mathscr{E}'(\cdot)$. However, even if $\imath \in \mathcal{E}_{\mathrm{p}}$, elite member $\imath \in \mathcal{P}$ is only willing to join the rebel army if $\mathcal{U}_{\mathrm{p}}(\imath) \leq \mathcal{U}_{\mathrm{p}}^{\dagger}(K)$. If that condition is not met, this member of the elite will belong to the incumbent army.

Using the argument from Proposition 1, it can be seen that the set of relevant post-investment no-rebellion constraints (given payoff equalization for workers and $\sigma_{\mathrm{i}}^{\dagger}=0$) is equivalent to

$$\mathscr{R}(\sigma) \geq 0 \text{ for all } \sigma \in [\underline{\sigma}, 1], \quad \text{where } \underline{\sigma} \equiv \max\left\{0, \frac{p^{\dagger}-p}{p^{\dagger}}\right\}. \qquad \text{[A.2.4]}$$

The term $\underline{\sigma}$ is the minimum number of places in the subsequent elite that can be offered to those who are currently workers given the sizes of the current and subsequent elites (note that a rebel army with $\sigma=1$ is feasible because there are always enough workers to fill all places in the subsequent elite).

Reduction to single no-rebellion constraint at the pre-investment stage

The argument here follows Proposition 1. The set of Markovian equilibria imposing the full range of pre-investment no-rebellion constraints (and the appropriate post-investment constraints) is the same as the set of Markovian equilibria in which the only constraint imposed at the pre-investment stage is for a rebel army drawn solely from non-elite members (of whom there is always a sufficient number). Therefore, the only relevant pre-investment stage no-rebellion constraint sets $\sigma_{\mathrm{p}}=0$ in [5.11] to obtain $\max\{\mathcal{U}_{\mathrm{p}}'-\mathcal{U}_{\mathrm{n}}, 0\} \leq \delta p/p'$, where an expression for the expected utility $\mathcal{U}_{\mathrm{n}}$ of a non-elite member is given in [5.10]. Since the right-hand side of the inequality is non-negative, it is equivalent to:

$$\mathcal{U}_{\mathrm{p}}'-\mathcal{U}_{\mathrm{n}} \leq \delta \frac{p}{p'}. \qquad \text{[A.2.5]}$$

The elite's objective function and the distribution of elite payoffs

Let $\bar{C}_{\mathrm{p}}$ denote average elite consumption and $\bar{\mathcal{U}}_{\mathrm{p}}$ average elite utility. The utility function is linear in consumption, so $\mathcal{U}_{\mathrm{p}}(\imath)=C_{\mathrm{p}}(\imath)$, and thus $\bar{\mathcal{U}}_{\mathrm{p}}=\bar{C}_{\mathrm{p}}$. Using the elite's budget constraint, the measure of investors $i=\mu s$, and the incentive compatibility condition for investors from [5.7], which implies $\tau_{\kappa}=\kappa-\tilde{\theta}$, average elite utility and consumption are given by:

$$\bar{\mathcal{U}}_{\mathrm{p}}=\bar{C}_{\mathrm{p}}=\frac{(1-p)\tau_{\mathrm{q}}+i\tau_{\kappa}}{p}=\frac{(1-p)\tau_{\mathrm{q}}+\mu(\kappa-\tilde{\theta})s}{p}. \qquad \text{[A.2.6]}$$

This is the elite's objective function that is maximized when new institutions are created. Note that it does not directly depend on the distribution of elite consumption owing to the linearity of the utility function.

Take any distribution of elite consumption $\{C_{\mathrm{p}}(\imath)\}$, and the consequent distribution of elite payoffs $\{\mathcal{U}_{\mathrm{p}}(\imath)\}$. Let λ denote the fraction of elite members receiving a payoff strictly greater than $\mathcal{U}_{\mathrm{p}}^{\dagger}(K)$, and let γ denote the amount on average by which $\mathcal{U}_{\mathrm{p}}(\imath)$ exceeds $\mathcal{U}_{\mathrm{p}}^{\dagger}(K)$ for this group of elite members. Finally, let $\check{\mathcal{U}}_{\mathrm{p}}$ denote the average payoff for the fraction $1-\lambda$ of elite members receiving a payoff less than or equal to $\mathcal{U}_{\mathrm{p}}^{\dagger}(K)$. Given the linearity of utility in consumption, these variables are related to the overall average elite payoff $\bar{\mathcal{U}}_{\mathrm{p}}$ as follows:

$$\bar{\mathcal{U}}_{\mathrm{p}}=\lambda(\mathcal{U}_{\mathrm{p}}^{\dagger}(K)+\gamma)+(1-\lambda)\check{\mathcal{U}}_{\mathrm{p}}.$$

This equation can be solved for $\check{\mathcal{U}}_{\mathrm{p}}$ taking λ and γ as given:

$$\check{\mathcal{U}}_{\mathrm{p}}=\frac{\bar{\mathcal{U}}_{\mathrm{p}}-\lambda(\mathcal{U}_{\mathrm{p}}^{\dagger}(K)+\gamma)}{1-\lambda}, \quad \text{and hence } \mathcal{U}_{\mathrm{p}}^{\dagger}(K)-\check{\mathcal{U}}_{\mathrm{p}}=\frac{\mathcal{U}_{\mathrm{p}}^{\dagger}(K)-\bar{\mathcal{U}}_{\mathrm{p}}+\lambda\gamma}{1-\lambda}. \qquad \text{[A.2.7]}$$

Now define $\omega(\lambda)$ to be the minimum fraction of places in the rebel army filled by workers if the λ fraction of current elite members are not to be offered a place in the subsequent elite:

$$\omega(\lambda) \equiv \max\left\{0, \frac{p^{\dagger}-(1-\lambda)p}{p^{\dagger}}\right\}, \qquad \text{[A.2.8]}$$

with $\omega(\lambda)$ being no less than the physical lower bound $\underline{\sigma}$ defined in [A.2.4]. Given the distribution of elite payoffs, a fraction λ of elite members will not join a rebel army even if included in the subsequent elite because they would receive a higher payoff if the current institutions survive. Such elite members are effectively loyal, in that they cannot be induced to leave the incumbent army and join the rebels. Consequently, for any $\sigma \in [\underline{\sigma}, \omega(\lambda))$, it must be the case that $\mathscr{R}(\sigma) \geq \mathscr{R}(\omega(\lambda))$, so the no-rebellion constraints [A.2.4] are redundant for σ values in this range. It is necessary and sufficient only to verify that $\mathscr{R}(\sigma) \geq 0$ for all $\sigma \in [\omega(\lambda), 1]$.

Now take the general distribution of elite payoffs $\{\mathcal{U}_{\mathrm{p}}(\imath)\}$ and hypothetically suppose that all consumption levels for those with payoffs less than or equal to $\mathcal{U}_{\mathrm{p}}^{\dagger}(K)$ were equalized (with no change to the distribution of consumption for those with a payoff above this threshold). Given linearity of utility, it follows that with this new distribution of elite payoffs, a fraction $1-\lambda$ all receive the same payoff $\check{\mathcal{U}}_{\mathrm{p}}$ from [A.2.7], with the remaining fraction λ all having payoffs above $\mathcal{U}_{\mathrm{p}}^{\dagger}(K)$. Let $\mathscr{C}(\sigma)$ denote the equivalent of the function $\mathscr{R}(\sigma)$ from [A.2.3] defined using this alternative distribution of elite payoffs, where $\sigma \in [\underline{\sigma}, 1]$ is as before the fraction of places in the rebel army assigned to those who are currently workers. Since those elite members with $\mathcal{U}_{\mathrm{p}}(\imath) \leq \mathcal{U}_{\mathrm{p}}^{\dagger}(K)$ all have the same payoff $\check{\mathcal{U}}_{\mathrm{p}}$, the function $\mathscr{C}(\sigma)$ can be written explicitly (making use of the corresponding definition of $\mathscr{R}(\sigma)$ in [A.2.3]) as follows:

$$\mathscr{C}(\sigma) = \delta\frac{p}{p^{\dagger}} - \sigma(\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \mathcal{U}_{\mathrm{w}}) - \begin{cases} (1-\omega(\lambda))(\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \check{\mathcal{U}}_{\mathrm{p}} + \delta) & \text{if } \sigma \in [\underline{\sigma}, \omega(\lambda)); \\ (1-\sigma)(\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \check{\mathcal{U}}_{\mathrm{p}} + \delta) & \text{if } \sigma \in [\omega(\lambda), 1]. \end{cases} \quad \text{[A.2.9]}$$

Using the argument developed in Proposition 1, payoff equalization among those elite members currently receiving no more than $\mathcal{U}_{\mathrm{p}}^{\dagger}(K)$ cannot increase the net fighting strength of the rebel army (but the same need not be true for payoff equalization among all elite members including those with payoffs above $\mathcal{U}_{\mathrm{p}}^{\dagger}(K)$). Since the functions $\mathscr{C}(\sigma)$ and $\mathscr{R}(\sigma)$ represent the net fighting strengths of the incumbent army with and without payoff equalization among the $1-\lambda$ fraction of the elite, it follows that

$$\mathscr{R}(\sigma) \leq \mathscr{C}(\sigma) \ \text{ for all } \ \sigma \in [\underline{\sigma}, 1]. \quad \text{[A.2.10a]}$$

Consider a case where $\sigma \in (\omega(\lambda), 1)$. This means that not all places in the rebel army are taken by workers, while there are more than enough current elite members (the fraction $1-\lambda$) willing to fill the remaining places. Given that only a subset of these individuals can actually be included, this subset will be those most dissatisfied with their current payoffs. Hence, if there is a positive measure of elite members with $\mathcal{U}_{\mathrm{p}}(\imath) \leq \mathcal{U}_{\mathrm{p}}^{\dagger}(K)$ and $\mathcal{U}_{\mathrm{p}}(\imath) \neq \check{\mathcal{U}}_{\mathrm{p}}$, the argument from Proposition 1 shows that

$$\mathscr{R}(\sigma) < \mathscr{C}(\sigma) \ \text{ for all } \ \sigma \in (\omega(\lambda), 1). \quad \text{[A.2.10b]}$$

Now define a function $\mathscr{N}(\lambda) \equiv \mathscr{C}(\omega(\lambda))$ using $\omega(\lambda)$ from [A.2.8] and $\mathscr{C}(\sigma)$ from [A.2.9]. Equation [A.2.9] implies

$$\mathscr{N}(\lambda) = \delta\left(\frac{p - (1-\omega(\lambda))p^{\dagger}}{p^{\dagger}}\right) - (1-\omega(\lambda))(\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \check{\mathcal{U}}_{\mathrm{p}}) - \omega(\lambda)(\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \mathcal{U}_{\mathrm{w}}), \quad \text{[A.2.11]}$$

for all $\lambda \in [0,1]$. In addition, define the threshold $\lambda^{\#}$ where for $\lambda \geq \lambda^{\#}$, a positive measure of workers would be required to fill the rebel army after including all the $1-\lambda$ fraction of current elite members who are willing to join. This threshold is formally defined as follows:

$$\lambda^{\#} \equiv \max\left\{0, 1 - \frac{p^{\dagger}}{p}\right\}, \quad \text{[A.2.12]}$$

observing that $0 \leq \lambda^{\#} < 1$. First, consider the case where $\lambda \in [0, \lambda^{\#})$. If there is a λ in this range then necessarily $\lambda^{\#} > 0$, and hence $1 - \lambda^{\#} = p^{\dagger}/p$. Since $\lambda < \lambda^{\#}$ in this case, it follows that $(1-\lambda)p > p^{\dagger}$, and thus $\omega(\lambda) = 0$ according to the definition in [A.2.8]. Therefore the expression for $\mathscr{N}(\lambda)$ in [A.2.11] reduces to:

$$\mathscr{N}(\lambda) = \delta\left(\frac{p - p^{\dagger}}{p^{\dagger}}\right) - (\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \check{\mathcal{U}}_{\mathrm{p}}) = \delta\left(\frac{p - p^{\dagger}}{p^{\dagger}}\right) - \left(\frac{\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \bar{\mathcal{U}}_{\mathrm{p}} + \lambda\gamma}{1-\lambda}\right), \quad \text{[A.2.13a]}$$

where the second equality uses the relationship between $\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \check{\mathcal{U}}_{\mathrm{p}}$ and $\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \bar{\mathcal{U}}_{\mathrm{p}}$ in [A.2.7].

Next, consider the remaining case where $\lambda \in [\lambda^{\#}, 1]$. Using [A.2.12], it follows that $\lambda \geq 1 - p^{\dagger}/p$, and hence $(1-\lambda)p \leq p^{\dagger}$. Given this inequality, the definition in [A.2.8] implies that

$$\omega(\lambda) = 1 - (1-\lambda)\frac{p}{p^{\dagger}}.$$

By substituting this into the expression for $\mathscr{N}(\lambda)$ from [A.2.11]:

$$\mathscr{N}(\lambda) = \delta\frac{p}{p^{\dagger}} - \delta(1-\lambda)\frac{p}{p^{\dagger}} - (1-\lambda)\frac{p}{p^{\dagger}}(\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \check{\mathcal{U}}_{\mathrm{p}}) - \left(1 - (1-\lambda)\frac{p}{p^{\dagger}}\right)(\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \mathcal{U}_{\mathrm{w}}).$$

Making some simplifications to the above and using [A.2.7] to substitute for $\check{\mathcal{U}}_{\mathrm{p}}$ in terms of $\bar{\mathcal{U}}_{\mathrm{p}}$:

$$\mathscr{N}(\lambda) = \lambda\delta\frac{p}{p^{\dagger}} - \frac{p}{p^{\dagger}}(\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \bar{\mathcal{U}}_{\mathrm{p}}) - \frac{p}{p^{\dagger}}\lambda\gamma - \left(1 - \frac{p}{p^{\dagger}}\right)(\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \mathcal{U}_{\mathrm{w}}) - \lambda\frac{p}{p^{\dagger}}(\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \mathcal{U}_{\mathrm{w}}),$$

and then after collecting terms in λ:

$$\mathscr{N}(\lambda) = \frac{p}{p^{\dagger}}(\delta + \mathcal{U}_{\mathrm{w}} - \mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \gamma)\lambda - \frac{p}{p^{\dagger}}(\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \bar{\mathcal{U}}_{\mathrm{p}}) - \left(1 - \frac{p}{p^{\dagger}}\right)(\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \mathcal{U}_{\mathrm{w}}). \quad \text{[A.2.13b]}$$

Equations [A.2.13a] and [A.2.13b] provide a complete specification of the function $\mathscr{N}(\lambda)$ for all $\lambda \in [0,1]$:

$$\mathscr{N}(\lambda) = \begin{cases} \delta\left(\frac{p-p^{\dagger}}{p^{\dagger}}\right) - \left(\frac{\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \bar{\mathcal{U}}_{\mathrm{p}}}{1-\lambda}\right) - \frac{\lambda}{1-\lambda}\gamma & \text{if } \lambda \in [0, \lambda^{\#}); \\ \frac{p}{p^{\dagger}}(\delta + \mathcal{U}_{\mathrm{w}} - \mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \gamma)\lambda - \frac{p}{p^{\dagger}}(\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \bar{\mathcal{U}}_{\mathrm{p}}) - \left(1 - \frac{p}{p^{\dagger}}\right)(\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \mathcal{U}_{\mathrm{w}}) & \text{if } \lambda \in [\lambda^{\#}, 1]. \end{cases} \quad \text{[A.2.14]}$$

The function $\mathscr{N}(\lambda)$ is seen to be continuous for all $\lambda \in [0,1]$ since the two branches coincide when evaluated at $\lambda = \lambda^{\#}$, as defined in [A.2.12].

Time inconsistency problem for members of the elite

Equation [5.6] giving the payoff of a worker implies $\tau_{\mathrm{q}} = \mathrm{q} - \mathcal{U}_{\mathrm{w}}$, so the average elite payoff [A.2.6] can be written as:

$$\bar{\mathcal{U}}_{\mathrm{p}} = \frac{(1-p)(\mathrm{q} - \mathcal{U}_{\mathrm{w}}) + \mu(\kappa - \tilde{\theta})s}{p}.$$

This expression can be rearranged as follows:

$$\bar{\mathcal{U}}_{\mathrm{p}} = \frac{(1-p)\left(\mathrm{q} + \delta\frac{p}{p^{\dagger}} - \mathcal{U}_{\mathrm{p}}^{\dagger}(K)\right) + \mu(\kappa - \tilde{\theta})s}{p} - \left(\frac{1-p}{p}\right)\left(\mathcal{U}_{\mathrm{w}} - \mathcal{U}_{\mathrm{p}}^{\dagger}(K) + \delta\frac{p}{p^{\dagger}}\right),$$

from which an expression for $\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \bar{\mathcal{U}}_{\mathrm{p}}$ can be obtained:

$$\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \bar{\mathcal{U}}_{\mathrm{p}} = \frac{1}{p}\left(\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \left((1-p)\left(\mathrm{q} + \delta\frac{p}{p^{\dagger}}\right) + \mu(\kappa - \tilde{\theta})s\right)\right) + \left(\frac{1-p}{p}\right)\left(\mathcal{U}_{\mathrm{w}} - \mathcal{U}_{\mathrm{p}}^{\dagger}(K) + \delta\frac{p}{p^{\dagger}}\right). \quad \text{[A.2.15]}$$

Substituting for $\mathcal{U}_{\text{p}}^{\dagger}(K)$ from [5.5] and rearranging the first term in brackets yields:

$$\begin{aligned}\mathcal{U}_{\text{p}}^{\dagger}(K)-\left((1-p)\left(\text{q}+\delta\frac{p}{p^{\dagger}}\right)+\mu(\kappa-\tilde{\theta})s\right) &= \frac{(\text{q}+\delta)^2}{\text{q}+2\delta}+K-(1-p)\left(\text{q}+\delta\frac{p}{p^{\dagger}}\right)-\mu\kappa s+\mu\tilde{\theta}s \\ &= (\text{q}+\delta)(1-p^{\dagger})+\mu\kappa s-\text{q}(1-p)-\frac{\delta}{p^{\dagger}}p+\frac{\delta}{p^{\dagger}}p^2-\mu\kappa s+\mu\tilde{\theta}s \\ &= \frac{\delta}{p^{\dagger}}\left(p^2-p+\frac{\text{q}p^{\dagger}}{\delta}p+\frac{p^{\dagger}}{\delta}\left((\text{q}+\delta)(1-p^{\dagger})-\text{q}\right)\right)+\mu\tilde{\theta}s \\ &= \frac{\delta}{p^{\dagger}}\left(p^2-\left(1-\frac{\text{q}}{\text{q}+2\delta}\right)p+\frac{p^{\dagger}}{\delta}\left(\frac{(\text{q}+\delta)^2-\text{q}(\text{q}+2\delta)}{\text{q}+2\delta}\right)\right)+\mu\tilde{\theta}s \\ &= \frac{\delta}{p^{\dagger}}\left(p^2-2\left(\frac{\delta}{\text{q}+2\delta}\right)p+p^{\dagger}\left(\frac{\delta}{\text{q}+2\delta}\right)\right)+\mu\tilde{\theta}s \\ &= \frac{\delta}{p^{\dagger}}\left(p^2-2p^{\dagger}p+{p^{\dagger}}^2\right)+\mu\tilde{\theta}s=\frac{\delta}{p^{\dagger}}\left(p-p^{\dagger}\right)^2+\mu\tilde{\theta}s.\end{aligned} \quad \text{[A.2.16]}$$

Since $\tilde{\theta}>0$, whenever $s>0$ the final term in the above is strictly positive. The other term is non-negative, hence:

$$\mathcal{U}_{\text{p}}^{\dagger}(K)-\left((1-p)\left(\text{q}+\delta\frac{p}{p^{\dagger}}\right)+\mu(\kappa-\tilde{\theta})s\right)>0. \quad \text{[A.2.17]}$$

The no-rebellion constraint [A.2.1] for workers at the post-investment stage implies

$$\mathcal{U}_{\text{p}}^{\dagger}(K)-\mathcal{U}_{\text{w}}+\delta\frac{p}{p^{\dagger}}\geq 0,$$

and taking this together with [A.2.15] and [A.2.17], it is proved that for any $s>0$:

$$\mathcal{U}_{\text{p}}^{\dagger}(K)>\bar{\mathcal{U}}_{\text{p}}. \quad \text{[A.2.18]}$$

This says that the elite payoff following a rebellion at the post-investment stage is always strictly larger than any feasible average elite payoff with no rebellion. As a consequence, there must be an upper bound less than one for feasible values of λ. Given the non-negativity constraints, $\check{\mathcal{U}}_{\text{p}}\geq 0$, so equation [A.2.7] implies $\lambda\leq\bar{\lambda}$, where

$$\bar{\lambda}\equiv\frac{\bar{\mathcal{U}}_{\text{p}}}{\mathcal{U}_{\text{p}}^{\dagger}(K)+\gamma},$$

with [A.2.18] demonstrating that $0<\bar{\lambda}<1$.

The pre-investment no-rebellion constraint cannot bind on its own

Consider first the possibility of a Markovian equilibrium with $s>0$ in which the pre-investment stage constraint [A.2.5] is the only effective binding no-rebellion constraint. With $p'=p^*$ and $\mathcal{U}'_{\text{p}}=\mathcal{U}^*_{\text{p}}$, the binding constraint [A.2.5] implies

$$\mathcal{U}^*_{\text{p}}-\mathcal{U}_{\text{n}}=\delta\frac{p}{p^*}.$$

Combining this with the expressions for the non-elite expected payoff $\mathcal{U}_{\text{n}}$ from [5.10] and the worker payoff $\mathcal{U}_{\text{w}}$ from [5.6]:

$$\mathcal{U}_{\text{w}}=\mathcal{U}^*_{\text{p}}-\delta\frac{p}{p^*}-\alpha\mathcal{S}_{\text{i}}(\tilde{\theta})=\mathcal{U}^*_{\text{p}}-\delta\frac{p}{p^*}-\frac{\mu}{1-p}\mathcal{S}_{\text{i}}(\tilde{\theta}), \quad \text{[A.2.19]}$$

where the formula for the probability α of receiving an investment opportunity is taken from [5.9]. The tax $\tau_{\text{q}}=\text{q}-\mathcal{U}_{\text{w}}$ is then given by:

$$\tau_{\text{q}}=\text{q}+\delta\frac{p}{p^*}+\frac{\mu}{1-p}\mathcal{S}_{\text{i}}(\tilde{\theta})-\mathcal{U}^*_{\text{p}}.$$

Substituting this into the elite's objective function $\bar{\mathcal{U}}_{\mathrm{p}}$ in equation [A.2.6]:

$$\bar{\mathcal{U}}_{\mathrm{p}} = \frac{(1-p)\left(\mathrm{q} + \delta\frac{p}{p^*} - \mathcal{U}^*_{\mathrm{p}}\right) + \mu(\kappa - \tilde{\theta})s + \mu\mathcal{S}_{\mathrm{i}}(\tilde{\theta})}{p}. \qquad \text{[A.2.20]}$$

The first-order condition for maximizing $\bar{\mathcal{U}}_{\mathrm{p}}$ with respect to p (taking p^* and $\mathcal{U}^*_{\mathrm{p}}$ as given) is:

$$\frac{1}{p}\left((1-p)\left(\mathrm{q} + \delta\frac{p}{p^*} - \mathcal{U}^*_{\mathrm{p}}\right) + \mu(\kappa - \tilde{\theta})s + \mu\mathcal{S}_{\mathrm{i}}(\tilde{\theta})\right) = (1-p)\frac{\delta}{p^*} - \left(\mathrm{q} + \delta\frac{p}{p^*} - \mathcal{U}^*_{\mathrm{p}}\right),$$

and after imposing the Markovian equilibrium conditions $p = p^*$ and $\bar{\mathcal{U}}_{\mathrm{p}} = \mathcal{U}^*_{\mathrm{p}}$ and making use of [A.2.20]:

$$\mathcal{U}^*_{\mathrm{p}} = \delta\frac{(1-p^*)}{p^*} - \left(\mathrm{q} + \delta - \mathcal{U}^*_{\mathrm{p}}\right).$$

Solving this equation for p^* yields:

$$p^* = \frac{\delta}{\mathrm{q} + 2\delta} = p^\dagger,$$

given the expression for $p^\dagger$ in [5.5].

Using equation [A.2.19], the equilibrium payoff of workers is $\mathcal{U}^*_{\mathrm{w}} = \bar{\mathcal{U}}^*_{\mathrm{p}} - \delta - \alpha\mathcal{S}_{\mathrm{i}}(\tilde{\theta})$, which implies

$$\delta + \mathcal{U}^*_{\mathrm{w}} - \mathcal{U}^\dagger_{\mathrm{p}}(K) - \gamma = -(\mathcal{U}^\dagger_{\mathrm{p}}(K) - \bar{\mathcal{U}}^*_{\mathrm{p}}) - \alpha\mathcal{S}_{\mathrm{i}}(\tilde{\theta}) - \gamma. \qquad \text{[A.2.21]}$$

This is strictly negative because $\mathcal{U}^\dagger_{\mathrm{p}}(K) - \bar{\mathcal{U}}^*_{\mathrm{p}} > 0$ according to [A.2.18]. This finding, in combination with $p^* = p^\dagger$, implies that $\mathscr{N}(\lambda)$ is strictly negative for all λ using the formula in [A.2.11]. Note that $\mathscr{R}(\omega(\lambda)) \leq \mathscr{C}(\omega(\lambda)) = \mathscr{N}(\lambda) < 0$ given the definitions and [A.2.10a]. Therefore, the no-rebellion constraint for a rebel army at the post-investment stage including elite members, represented by the inequality $\mathscr{R}(\sigma) \geq 0$, is violated for some feasible value of σ. It follows that a Markovian equilibrium with this configuration of binding constraints does not exist.

The post-investment no-rebellion constraint including only workers cannot bind on its own

Now consider the possibility of a Markovian equilibrium featuring $s > 0$ with workers' post-investment constraint [A.2.1] as the only effective binding no-rebellion constraint. When the inequality in [A.2.1] binds:

$$\mathcal{U}_{\mathrm{w}} = \mathcal{U}^\dagger_{\mathrm{p}}(K) - \delta\frac{p}{p^\dagger}, \qquad \text{[A.2.22]}$$

from which can be found an expression for tax $\tau_{\mathrm{q}} = \mathrm{q} - \mathcal{U}_{\mathrm{w}}$ as a function of p:

$$\tau_{\mathrm{q}} = \mathrm{q} + \delta\frac{p}{p^\dagger} - \mathcal{U}^\dagger_{\mathrm{p}}(K).$$

This is substituted into the expression for the average elite payoff [A.2.6] to obtain:

$$\bar{\mathcal{U}}_{\mathrm{p}} = \frac{(1-p)\left(\mathrm{q} + \delta\frac{p}{p^\dagger} - \mathcal{U}^\dagger_{\mathrm{p}}(K)\right) + \mu(\kappa - \tilde{\theta})s}{p}. \qquad \text{[A.2.23]}$$

Taking the derivative of $\bar{\mathcal{U}}_{\mathrm{p}}$ with respect to p:

$$\frac{\partial\bar{\mathcal{U}}_{\mathrm{p}}}{\partial p} = \frac{1}{p}\left((1-p)\frac{\delta}{p^\dagger} - \left(\mathrm{q} + \delta\frac{p}{p^\dagger} - \mathcal{U}^\dagger_{\mathrm{p}}(K)\right)\right) - \frac{1}{p^2}\left((1-p)\left(\mathrm{q} + \delta\frac{p}{p^\dagger} - \mathcal{U}^\dagger_{\mathrm{p}}(K)\right) + \mu(\kappa - \tilde{\theta})s\right),$$

and by using the expressions for $\bar{\mathscr{U}}_{\mathrm{p}}$ and $p^{\dagger}$ from [A.2.23] and [5.5] respectively:

$$\frac{\partial \bar{\mathscr{U}}_{\mathrm{p}}}{\partial p} = \frac{1}{p}\left(\frac{\delta}{p^{\dagger}} - \mathrm{q} - 2\delta\frac{p}{p^{\dagger}} + (\mathscr{U}_{\mathrm{p}}^{\dagger}(K) - \bar{\mathscr{U}}_{\mathrm{p}})\right) = \frac{1}{p}\left(2\frac{\delta}{p^{\dagger}}(p^{\dagger} - p) + (\mathscr{U}_{\mathrm{p}}^{\dagger}(K) - \bar{\mathscr{U}}_{\mathrm{p}})\right). \quad \text{[A.2.24]}$$

Exploiting the fact that [A.2.22] holds in this case, equations [A.2.15] and [A.2.16] imply that:

$$\mathscr{U}_{\mathrm{p}}^{\dagger}(K) - \bar{\mathscr{U}}_{\mathrm{p}} = \frac{1}{p}\left(\frac{\delta}{p^{\dagger}}(p - p^{\dagger})^2 + \mu\tilde{\theta}s\right), \quad \text{[A.2.25]}$$

and by substituting this into [A.2.24]:

$$\frac{\partial \bar{\mathscr{U}}_{\mathrm{p}}}{\partial p} = \frac{1}{p^2}\left(2\frac{\delta}{p^{\dagger}}(p^{\dagger} - p)p + \frac{\delta}{p^{\dagger}}(p - p^{\dagger})^2 + \mu\tilde{\theta}s\right) = \frac{\delta}{p^2 p^{\dagger}}\left(p^{\dagger}\left(p^{\dagger} + \frac{\mu\tilde{\theta}s}{\delta}\right) - p^2\right). \quad \text{[A.2.26]}$$

Now make the following definition of the function $\pi(s)$:

$$\pi(s) \equiv p^{\dagger} + \frac{\mu\tilde{\theta}s}{\delta}. \quad \text{[A.2.27]}$$

The condition $\pi(s) < 1/2$ is equivalent to $\mu\tilde{\theta}s/\delta < 1/2 - p^{\dagger}$, and hence to:

$$\left(\frac{1}{2} - p^{\dagger} - \mu\right) + \left(1 - s\frac{\tilde{\theta}}{\delta}\right)\mu > 0.$$

After substituting the formula for $p^{\dagger}$ in [5.5] and using the bound on μ from [5.3], the first term is seen to be non-negative. Since $\theta \in [\psi, \kappa]$, it follows that $\tilde{\theta} \leq \kappa$. The restrictions in [5.3] then ensure $\tilde{\theta} < \delta$, which together with $0 \leq s \leq 1$ implies the second term above is strictly positive. It follows that $\pi(s) < 1/2$ for all $s \in [0, 1]$ with parameters consistent with the stated restrictions in [5.3].

Using the definition of $\pi(s)$ in [A.2.27] and equation [A.2.26], the derivative of $\bar{\mathscr{U}}_{\mathrm{p}}$ can be written as:

$$\frac{\partial \bar{\mathscr{U}}_{\mathrm{p}}}{\partial p} = \frac{\delta}{p^2 p^{\dagger}}\left(p^{\dagger}\pi(s) - p^2\right). \quad \text{[A.2.28]}$$

It follows that $\bar{\mathscr{U}}_{\mathrm{p}}$ is strictly increasing in p for $p < \sqrt{p^{\dagger}}\sqrt{\pi(s)}$, and strictly decreasing for $p > \sqrt{p^{\dagger}}\sqrt{\pi(s)}$. The first-order condition for maximizing the expression in [A.2.23] for $\bar{\mathscr{U}}_{\mathrm{p}}$ incorporating the binding constraint is therefore $p^* = \sqrt{p^{\dagger}}\sqrt{\pi(s)}$. It can be seen from [A.2.27] that $p^{\dagger} < \pi(s)$ for any $s > 0$, and it is known that $\pi(s) < 1/2$, so it follows that $p^{\dagger} < p^* < 1/2$.

Given that $p^* > p^{\dagger}$, it can be seen from the definitions in [A.2.4] and [A.2.12] that $\underline{\sigma} = 0$ and $\lambda^{\#} > 0$. Since [A.2.8] implies $\omega(0) = \underline{\sigma}$, it then follows from [A.2.11] that

$$\mathscr{N}(0) = \delta\left(\frac{p - p^{\dagger}}{p^{\dagger}}\right) - (\mathscr{U}_{\mathrm{p}}^{\dagger}(K) - \bar{\mathscr{U}}_{\mathrm{p}}).$$

Thus, the statement $\mathscr{N}(0) \geq 0$ is equivalent to

$$\mathscr{U}_{\mathrm{p}}^{\dagger}(K) - \bar{\mathscr{U}}_{\mathrm{p}} \leq \delta\left(\frac{p - p^{\dagger}}{p^{\dagger}}\right). \quad \text{[A.2.29]}$$

In the case under consideration where [A.2.1] is binding, an expression for the left-hand side of the above is given in [A.2.25], so $\mathscr{N}(0) \geq 0$ holds if and only if

$$\frac{\delta}{p^{\dagger}}(p - p^{\dagger})^2 + \mu s\tilde{\theta} \leq \frac{\delta}{p^{\dagger}}(p - p^{\dagger})p.$$

Simplification of the terms appearing in the inequality above show that it is equivalent to $p \geq \pi(s)$, with $\pi(s)$ as defined in [A.2.27]. But the first-order condition $p^* = \sqrt{p^\dagger}\sqrt{\pi(s)}$ implies $p^* < \pi(s)$ since $\pi(s) > p^\dagger$ for $s > 0$. Therefore, an equilibrium of this type requires $\mathscr{N}(0) < 0$.

Now note that the binding workers' no-rebellion constraint [A.2.22] implies

$$\delta + \mathcal{U}_\mathrm{w}^* - \mathcal{U}_\mathrm{p}^\dagger(K) - \gamma = -\delta\left(\frac{p^* - p^\dagger}{p^\dagger}\right) - \gamma. \qquad \text{[A.2.30]}$$

As $p^* > p^\dagger$, it is therefore established that $\delta + \mathcal{U}_\mathrm{w}^* - \mathcal{U}_\mathrm{p}^\dagger(K) - \gamma < 0$. From [A.2.11], in the range $\lambda \in [\lambda^\#, 1]$, this implies that $\mathscr{N}(\lambda)$ is strictly decreasing. In the range $\lambda \in [0, \lambda^\#)$, differentiating [A.2.11] reveals that $\mathscr{N}'(\lambda) = -(\mathcal{U}_\mathrm{p}^\dagger(K) - \bar{\mathcal{U}}_\mathrm{p} + \gamma)/(1-\lambda)^2$. Since [A.2.18] justifies $\mathcal{U}_\mathrm{p}^\dagger(K) - \bar{\mathcal{U}}_\mathrm{p}^* > 0$, this proves that the function $\mathscr{N}(\lambda)$ is strictly decreasing for all $\lambda \in [0,1]$. As $\mathscr{N}(0) < 0$, it follows that $\mathscr{N}(\lambda) < 0$ for all λ. Consequently, $\mathscr{R}(\omega(\lambda)) \leq \mathscr{C}(\omega(\lambda)) = \mathscr{N}(\lambda) < 0$, which violates $\mathscr{R}(\sigma) \geq 0$ (the post-investment no-rebellion constraint for elite members) for a feasible value of σ. This shows that there cannot be a Markovian equilibrium with the conjectured configuration of binding no-rebellion constraints.

Both pre- and post-investment no-rebellion constraints cannot be binding for non-elite members, nor can the maximum elite size constraint be binding

In analysing this case, consider first the possibility of a Markovian equilibrium with $s > 0$ in which the pre-investment constraint [A.2.5] is binding. This leads to the expression for $\bar{\mathcal{U}}_\mathrm{p}$ given in [A.2.20]. After imposing the Markovian equilibrium conditions $p = p^*$ and $\mathcal{U}_\mathrm{p}^* = \bar{\mathcal{U}}_\mathrm{p}$ in that equation:

$$\mathcal{U}_\mathrm{p}^* = (\mathrm{q} + \delta)(1 - p^*) + \mu(\kappa - \tilde{\theta})s + \mu\mathcal{S}_\mathrm{i}(\tilde{\theta}).$$

It has already been shown that when [A.2.5] binds, equation [A.2.19] for $\mathcal{U}_\mathrm{w}$ follows. By substituting the equation above into [A.2.19]:

$$\mathcal{U}_\mathrm{w}^* = \mathrm{q}(1 - p^*) - \delta p^* + \mu(\kappa - \tilde{\theta})s - \mu\frac{p^*}{1 - p^*}\mathcal{S}_\mathrm{i}(\tilde{\theta}).$$

This can be substituted into [A.2.1] to obtain a condition for p^* to satisfy the post-investment no-rebellion constraint for workers:

$$\mathcal{U}_\mathrm{p}^\dagger(K) - \mathrm{q}(1 - p^*) + \delta p^* - \mu(\kappa - \tilde{\theta})s + \mu\frac{p^*}{1 - p^*}\mathcal{S}_\mathrm{i}(\tilde{\theta}) \leq \delta\frac{p^*}{p^\dagger}.$$

Using the expressions for $\mathcal{U}_\mathrm{p}^\dagger(K)$ and $p^\dagger$ from [5.5], and $K = \mu\kappa s$ from [5.7], it can be seen that $\mathcal{U}_\mathrm{p}^\dagger(K) - \mathrm{q} = \delta p^\dagger + \mu\kappa s$. By substituting this into the inequality above it is seen to be equivalent to

$$\delta p^\dagger + (\mathrm{q} + \delta)p^* + \mu\tilde{\theta}s + \mu\frac{p^*}{1 - p^*}\mathcal{S}_\mathrm{i}(\tilde{\theta}) \leq \frac{\delta}{p^\dagger}p^*.$$

Noting that [5.5] implies $\delta/p^\dagger = \mathrm{q} + 2\delta$, the inequality above can be rearranged to yield

$$\left(1 - \frac{\mu}{1 - p^*}\frac{\mathcal{S}_\mathrm{i}(\tilde{\theta})}{\delta}\right)p^* \geq \pi(s), \qquad \text{[A.2.31a]}$$

where $\pi(s)$ is defined in [A.2.27].

Now start from the case where the post-investment no-rebellion constraint for workers [A.2.1] is binding at $p = p^*$. The condition needed for a Markovian equilibrium ($p' = p^*$ and $\mathcal{U}_\mathrm{p}' = \mathcal{U}_\mathrm{p}^*$) of this type to satisfy the pre-investment stage no-rebellion constraint [A.2.5] is

$$\mathcal{U}_\mathrm{p}^* - \mathcal{U}_\mathrm{n}^* \leq \delta.$$

The expected payoff $\mathcal{U}_\mathrm{n}$ for non-elite members is given in equation [5.10]. Using [5.6] and the formula for

α from [5.9], it follows that $\mathcal{U}_{\text{n}}^* = \mathcal{U}_{\text{w}}^* + (\mu/(1-p^*))\mathcal{S}_{\text{i}}(\tilde{\theta})$. Substituting this into the inequality above shows that [A.2.5] is satisfied if and only if

$$\mathcal{U}_{\text{p}}^* - \mathcal{U}_{\text{w}}^* - \frac{\mu}{1-p^*}\mathcal{S}_{\text{i}}(\tilde{\theta}) \leq \delta.$$

A binding post-investment constraint [A.2.1] for workers implies that $\mathcal{U}_{\text{w}}^* = \mathcal{U}_{\text{p}}^\dagger(K) - \delta p^*/p^\dagger$, which can be substituted into the condition above to obtain:

$$\delta\frac{p^*}{p^\dagger} - \delta - \frac{\mu}{1-p^*}\mathcal{S}_{\text{i}}(\tilde{\theta}) \leq \mathcal{U}_{\text{p}}^\dagger(K) - \mathcal{U}_{\text{p}}^*.$$

Equation [A.2.22] (the binding version of [A.2.1]) has already been shown to imply the expression for $\mathcal{U}_{\text{p}}^\dagger(K) - \bar{\mathcal{U}}_{\text{p}}$ in [A.2.25]. Since $\bar{\mathcal{U}}_{\text{p}}^* = \mathcal{U}_{\text{p}}^*$, this can be used to deduce that the inequality above is equivalent to:

$$\frac{\delta}{p^\dagger}(p^* - p^\dagger)p^* - \mu\frac{p^*}{1-p^*}\mathcal{S}_{\text{i}}(\tilde{\theta}) \leq \frac{\delta}{p^\dagger}(p^* - p^\dagger)^2 + \mu\tilde{\theta}s.$$

After some rearrangement, this condition can be written as

$$p^* - p^\dagger - \mu\frac{p^*}{1-p^*}\frac{\mathcal{S}_{\text{i}}(\tilde{\theta})}{\delta} \leq \frac{\mu\tilde{\theta}s}{\delta},$$

which can be stated in terms of the function $\pi(s)$ from [A.2.27]:

$$\left(1 - \frac{\mu}{1-p^*}\frac{\mathcal{S}_{\text{i}}(\tilde{\theta})}{\delta}\right)p^* \leq \pi(s). \qquad \text{[A.2.31b]}$$

Note that if both constraints [A.2.5] and [A.2.1] are binding in a Markovian equilibrium then both [A.2.31a] and [A.2.31b] must hold, hence:

$$\left(1 - \frac{\mu}{1-p^*}\frac{\mathcal{S}_{\text{i}}(\tilde{\theta})}{\delta}\right)p^* = \pi(s). \qquad \text{[A.2.32]}$$

The expected surplus to those receiving an investment opportunity is $\mathcal{S}_{\text{i}}(\tilde{\theta})$, which is defined in equation [5.10]. The probability distribution of the effort cost θ in [5.1] has density function $1/(\kappa - \psi)$ on support $[\psi, \kappa]$, so an explicit expression for the surplus is:

$$\mathcal{S}_{\text{i}}(\tilde{\theta}) = \int_{\theta=\psi}^{\tilde{\theta}} \frac{\tilde{\theta} - \theta}{\kappa - \psi}\text{d}\theta = \frac{1}{2}\frac{(\tilde{\theta} - \psi)^2}{\kappa - \psi}. \qquad \text{[A.2.33]}$$

The parameter restrictions in [5.3] require $\kappa < \delta$. Together with $\psi \leq \tilde{\theta} \leq \kappa$ and $0 < \psi < \kappa$, this implies

$$\mathcal{S}_{\text{i}}(\tilde{\theta}) \leq \frac{1}{2}(\kappa - \psi) < \frac{\kappa}{2} < \frac{\delta}{2}. \qquad \text{[A.2.34]}$$

Now define the function $\mathscr{M}(p)$ as follows and calculate its derivative:

$$\mathscr{M}(p) \equiv \left(1 - \frac{\mu}{1-p}\frac{\mathcal{S}_{\text{i}}(\tilde{\theta})}{\delta}\right)p, \quad \text{and} \quad \mathscr{M}'(p) = 1 - \frac{1}{1-p}\frac{\mu}{1-p}\frac{\mathcal{S}_{\text{i}}(\tilde{\theta})}{\delta}. \qquad \text{[A.2.35]}$$

Given the parameter restrictions in [5.3], the formula $\alpha = \mu/(1-p)$ from [5.9] always returns a well-defined probability for $p \leq 1/2$. Given that $p \leq 1/2$ implies $1/(1-p) \leq 2$ and [A.2.34] implies $\mathcal{S}_{\text{i}}(\tilde{\theta})/\delta < 1/2$, it follows from [A.2.35] that $\mathscr{M}'(p) > 0$ for all $p \in [0, 1/2]$. Noting that $\mathscr{M}(0) = 0$ and $\pi(s) > 0$ according to [A.2.27], the equation $\mathscr{M}(p^*) = \pi(s)$ has at most one solution p^* satisfying $0 < p^* < 1/2$. When such a solution exists, it follows immediately from [A.2.35] that $p^* > \pi(s)$ because $\mathcal{S}_{\text{i}}(\tilde{\theta})/\delta < 1/2$ and $\alpha = \mu/(1-p^*) \leq 1$. Observe that the equation $\mathscr{M}(p^*) = \pi(s)$ is equivalent to the condition in [A.2.32] for both constraints to be binding.

In the case where no solution $p^* \in (0, 1/2)$ exists, set $p^* = 1/2$, and since $\mathscr{M}(0) = 0$ and $\mathscr{M}'(p) > 0$ it must be the case that $\mathscr{M}(p^*) = \mathscr{M}(1/2) < \pi(s)$. Since it has been shown that $\pi(s) < 1/2$ for the function $\pi(s)$ defined in [A.2.27], it follows that $p^* > \pi(s)$. Note that this case is consistent with [A.2.31b], but not [A.2.31a], which means that it could only occur in conjunction with the post-investment no-rebellion constraint for workers being binding, while the pre-investment no-rebellion constraint is slack.

Now consider a deviation from either of the conjectured equilibria described above, an elite size satisfying $0 < p^* < 1$ and both no-rebellion constraints binding, or $p^* = 1/2$ with the pre-investment no-rebellion constraint slack. Workers' post-investment no-rebellion constraint binds in both cases. The deviation involves changing the elite size p, while taxes are adjusted so that the post-investment no-rebellion constraint for workers continues to bind, in which case the payoff of a worker is given in equation [A.2.22]. The condition for the new choice of p to be consistent with the pre-investment stage constraint (with $p' = p^*$ and $\mathcal{U}'_{\mathrm{p}} = \mathcal{U}^*_{\mathrm{p}}$ taken as given at their conjectured equilibrium values) is:

$$\mathcal{U}^*_{\mathrm{p}} - \mathcal{U}_{\mathrm{n}} \leq \delta \frac{p}{p^*}, \qquad \text{[A.2.36]}$$

where the expected non-elite payoff is obtained using equations [5.6], [5.10], and [A.2.22]:

$$\mathcal{U}_{\mathrm{n}} = \mathcal{U}_{\mathrm{w}} + \frac{\mu}{1-p}\mathcal{S}_{\mathrm{i}}(\tilde{\theta}) = \mathcal{U}^{\dagger}_{\mathrm{p}}(K) - \delta\frac{p}{p^{\dagger}} + \frac{\mu}{1-p}\mathcal{S}_{\mathrm{i}}(\tilde{\theta}).$$

Substituting this expression into [A.2.36] shows that the feasibility of the deviation requires:

$$\mathcal{U}^*_{\mathrm{p}} - \mathcal{U}^{\dagger}_{\mathrm{p}}(K) + \delta\frac{p}{p^{\dagger}} - \frac{\mu}{1-p}\mathcal{S}_{\mathrm{i}}(\tilde{\theta}) \leq \delta\frac{p}{p^*},$$

which can be rearranged as follows:

$$\left(\frac{1}{p^{\dagger}} - \frac{1}{p^*}\right)p - \frac{\mu}{1-p}\frac{\mathcal{S}_{\mathrm{i}}(\tilde{\theta})}{\delta} \leq \frac{\mathcal{U}^{\dagger}_{\mathrm{p}}(K) - \mathcal{U}^*_{\mathrm{p}}}{\delta}. \qquad \text{[A.2.37]}$$

This inequality is known to be satisfied at $p = p^*$ because p^* satisfies the pre-investment no-rebellion constraint.

Now define the function $\mathscr{K}(p)$, and note the following expression for its derivative:

$$\mathscr{K}(p) \equiv \left(\frac{1}{p^{\dagger}} - \frac{1}{p^*}\right)p - \frac{\mu}{1-p}\frac{\mathcal{S}_{\mathrm{i}}(\tilde{\theta})}{\delta}, \quad \text{and } \mathscr{K}'(p) = \frac{1}{p^{\dagger}} - \frac{1}{p^*} - \frac{\mu}{(1-p)^2}\frac{\mathcal{S}_{\mathrm{i}}(\tilde{\theta})}{\delta}. \qquad \text{[A.2.38]}$$

Evaluating the derivative at $p = p^*$:

$$\mathscr{K}'(p^*) = \frac{1}{p^{\dagger}p^*}\left(p^* - p^{\dagger} - \frac{\mu\mathcal{S}_{\mathrm{i}}(\tilde{\theta})}{\delta}\frac{p^{\dagger}}{1-p^*}\frac{p^*}{1-p^*}\right) > \frac{1}{p^{\dagger}p^*}\left(\pi(s) - p^{\dagger} - \frac{\mu\mathcal{S}_{\mathrm{i}}(\tilde{\theta})}{\delta}\frac{p^{\dagger}}{1-p^*}\frac{p^*}{1-p^*}\right),$$

where the inequality above follows from $p^* > \pi(s)$, which is true in all cases under consideration. Since [A.2.27] implies $\pi(s) - p^{\dagger} = \mu\tilde{\theta}s/\delta$, the inequality above implies:

$$\mathscr{K}'(p^*) > \frac{\mu}{\delta}\frac{1}{p^{\dagger}p^*}\left(\tilde{\theta}s - \frac{p^{\dagger}}{1-p^*}\frac{p^*}{1-p^*}\mathcal{S}_{\mathrm{i}}(\tilde{\theta})\right). \qquad \text{[A.2.39]}$$

Note that $s = \mathbb{P}_{\theta}[\theta \leq \tilde{\theta}]$ from [5.7], and use the definition of $\mathcal{S}_{\mathrm{i}}(\tilde{\theta})$ in [5.10] to deduce

$$\tilde{\theta}s - \mathcal{S}_{\mathrm{i}}(\tilde{\theta}) = \mathbb{E}_{\theta}\tilde{\theta}\mathbb{1}[\theta \leq \tilde{\theta}] - \mathbb{E}_{\theta}\max\{\tilde{\theta} - \theta, 0\} = \mathbb{E}_{\theta}\theta\mathbb{1}[\theta \leq \tilde{\theta}] \geq 0,$$

and hence $\mathcal{S}_{\mathrm{i}}(\tilde{\theta}) \leq \tilde{\theta}s$. Substituting this into [A.2.39] demonstrates that

$$\mathscr{K}'(p^*) > \frac{\mu}{\delta}\frac{\tilde{\theta}s}{p^{\dagger}p^*}\left(1 - \frac{p^{\dagger}}{1-p^*}\frac{p^*}{1-p^*}\right).$$

Since $p^{\dagger} < p^* \leq 1/2$, it follows that $p^{\dagger}/(1-p^*) < 1$ and $p^*/(1-p^*) \leq 1$, and thus $\mathscr{K}'(p^*) > 0$.

The pre-investment no-rebellion constraint [A.2.36] is equivalent to [A.2.37], which by comparison with the definition of $\mathscr{K}(p)$ in [A.2.38] is in turn equivalent to:

$$\mathscr{K}(p) \leq \frac{\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \mathcal{U}_{\mathrm{p}}^*}{\delta}.$$

It can be seen from [A.2.38] that $\mathscr{K}(p)$ is increasing in p. It is also known that the above inequality is satisfied at $p = p^*$. Observing that the right-hand side is unaffected by movements in p away from the conjectured equilibrium p^*, it then follows that it is possible to reduce p below p^* and still satisfy the pre-investment no-rebellion constraint.

Before considering any deviation in p, consider a change in the distribution of elite consumption that removes all inequality. This entails $\lambda = 0$ and $\check{\mathcal{U}}_{\mathrm{p}} = \bar{\mathcal{U}}_{\mathrm{p}}$ according to [A.2.7]. Since the utility function is linear, this redistribution of consumption among the elite has no direct effect on $\bar{\mathcal{U}}_{\mathrm{p}}$. Given that elite payoffs are now fully equalized, the functions $\mathscr{R}(\sigma)$ and $\mathscr{C}(\sigma)$ defined in [A.2.3] and [A.2.9] coincide. Since $\lambda = 0$, [A.2.4] and [A.2.8] imply that $\omega(\lambda) = \underline{\sigma}$. Furthermore, as long as it remains the case that $p \geq p^{\dagger}$ then $\underline{\sigma} = 0$ according to [A.2.4]. Using $\mathscr{R}(\sigma) = \mathscr{C}(\sigma)$ and equation [A.2.9]:

$$\mathscr{R}(0) = \delta\left(\frac{p - p^{\dagger}}{p^{\dagger}}\right) - (\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \bar{\mathcal{U}}_{\mathrm{p}}). \tag{A.2.40}$$

The constraint $\mathscr{R}(1) \geq 0$ is equivalent to the workers' post-investment no-rebellion constraint [A.2.1], as can be seen from the definition in [A.2.3]. This no-rebellion constraint is binding in the conjectured equilibrium and following the proposed deviation, hence $\mathscr{R}(1) = 0$. With payoff equalization, $\mathscr{R}(\sigma)$ inherits linearity in σ from $\mathscr{C}(\sigma)$, as can be seen by inspecting [A.2.9]. It follows that $\mathscr{R}(0) \geq 0$ implies $\mathscr{R}(\sigma) \geq 0$ for all σ. The constraint in [A.2.40] is the no-rebellion constraint [A.2.29] for elite members at the post-investment stage. It has been shown that whenever workers' post-investment constraint is binding (which is the case here), the post-investment no-rebellion constraint for the elite is equivalent to $p \geq \pi(s)$. Therefore, starting from the conjectured equilibrium at $p = p^* > \pi(s)$, it is feasible to reduce p by some positive amount and continue to satisfy all no-rebellion constraints. Moreover, it has been shown that when the workers' binding post-investment constraint is used to determine τ_{q}, the derivative of the elite's objective function $\bar{\mathcal{U}}_{\mathrm{p}}$ with respect to p is given by [A.2.28] and is thus strictly decreasing for $p > \sqrt{p^{\dagger}}\sqrt{\pi(s)}$. Since $p^* > \sqrt{p^{\dagger}}\sqrt{\pi(s)}$, the proposed deviation is both feasible and payoff-improving for the elite. Hence, there is no Markovian equilibrium with this configuration of binding constraints.

The pre-investment no-rebellion constraint and some post-investment no-rebellion constraint including elite members cannot both bind

The general set of no-rebellion constraints at the post-investment stage is given in [A.2.4]. Now conjecture there is an equilibrium with a binding no-rebellion constraint $\mathscr{R}(\sigma) = 0$ imposed for some $\sigma \in [\omega(\lambda), 1)$, in addition to the pre-investment no-rebellion constraint [A.2.5].

Suppose that the conjectured equilibrium features payoff inequality among the $1 - \lambda$ fraction of non-loyal elite members (those with payoff $\mathcal{U}_{\mathrm{p}}(\imath) \leq \mathcal{U}_{\mathrm{p}}^{\dagger}(K)$). Equalizing payoffs among this set of elite members implies that the function $\mathscr{R}(\sigma)$ becomes $\mathscr{C}(\sigma)$ from [A.2.9], and $\mathscr{C}(\sigma) > \mathscr{R}(\sigma)$ according to [A.2.10b] when there was payoff inequality among this group initially. This change leads to a slackening of the binding constraint and thus a strict increase in the elite's objective function (unless the constraint is redundant, in which case this problem would reduce to one with only the binding pre-investment no-rebellion constraint — that case has already been ruled out). Hence, there is no equilibrium of this type in which there is payoff inequality among the non-loyal elite members and where the constraint $\mathscr{R}(\sigma) = 0$ must be imposed for $\sigma \in (\omega(\lambda), 1)$. Thus, either the equilibrium in this case features payoff equalization among non-loyal elite members, or the post-investment no-rebellion constraint that needs to be imposed is with $\sigma = \omega(\lambda)$.

Consider an equilibrium of this type with payoff equalization among the $1-\lambda$ fraction of non-loyal elite members. In this case, the function $\mathscr{C}(\sigma)$ coincides with $\mathscr{R}(\sigma)$. As can be seen from [A.2.9], $\mathscr{C}(\sigma)$ is linear in σ, so it follows that the general post-investment no-rebellion constraint [A.2.4] cannot bind at some $\sigma(\omega(\lambda, 1)$ but not at $\sigma = \omega(\lambda)$ (otherwise the constraint would be violated for some other feasible σ value). Therefore, an equilibrium of this type can always be characterized by imposing the binding post-investment no-rebellion constraint $\mathscr{R}(\omega(\lambda)) = 0$.

The other binding constraint in this case is the pre-investment constraint [A.2.5]. It has already been seen that in a Markovian equilibrium with such a binding constraint, equation [A.2.21] must hold. Now consider the following deviation from the conjectured equilibrium. First, equalize consumption (and hence payoffs) among the $1-\lambda$ fraction of non-loyal elite members. This implies $\mathscr{R}(\sigma)$ becomes equal to the function $\mathscr{C}(\sigma)$, which prior to this change was such that $\mathscr{C}(\sigma) \geq \mathscr{R}(\sigma)$ according to [A.2.10a], so the general no-rebellion constraint $\mathscr{R}(\sigma) \geq 0$ for all $\sigma \in [\underline{\sigma}, 1]$ remains satisfied after the deviation. Following this, consider a reduction in λ. Given that [A.2.18] and [A.2.21] hold, it can be seen from [A.2.11] that the function $\mathscr{N}(\lambda)$ is strictly decreasing in λ. Since the definition of $\mathscr{C}(\sigma)$ from [A.2.9] is such that $\mathscr{N}(\lambda) = \mathscr{C}(\omega(\lambda))$, this means the binding no-rebellion constraint at the post-investment is slackened. As $\mathscr{C}(\sigma)$ is a linear combination of $\mathscr{C}(1)$ (unaffected by any of these changes) and $\mathscr{C}(\omega(\lambda))$ (now strictly larger), all other post-investment no-rebellion constraints remain satisfied. A higher value of the elite objective function $\bar{\mathcal{U}}_{\mathrm{p}}$ is therefore attainable, so there are no equilibria of this type with $\lambda > 0$.

Given that any equilibrium must feature $\lambda = 0$, it follows from [A.2.8] that $\omega(\lambda) = \underline{\sigma}$. If it were the case that $p^* \leq p^\dagger$ then [A.2.4] implies that $\underline{\sigma} = 1 - (p/p^\dagger)$, and hence by using [A.2.9] to evaluate $\mathscr{C}(\sigma)$ at $\sigma = \underline{\sigma}$:

$$\mathscr{R}(\underline{\sigma}) \leq \mathscr{C}(\underline{\sigma}) = -\frac{p}{p^\dagger}(\mathcal{U}_{\mathrm{p}}^\dagger(K) - \check{\mathcal{U}}_{\mathrm{p}}) - \left(1 - \frac{p}{p^\dagger}\right)(\mathcal{U}_{\mathrm{p}}^\dagger(K) - \mathcal{U}_{\mathrm{w}}) < 0,$$

because $\mathcal{U}_{\mathrm{p}}^\dagger(K) > \bar{\mathcal{U}}_{\mathrm{p}} = \check{\mathcal{U}}_{\mathrm{p}}$, as follows from [A.2.18] and [A.2.7] when $\lambda = 0$. This violates the general no-rebellion constraint [A.2.4]. Therefore, a Markovian equilibrium of this type requires $p^* > p^\dagger$.

When $p^* > p^\dagger$, any rebel army at the post-investment stage can only include a subset of the elite, so any equilibrium with a binding post-investment no-rebellion constraint with $\sigma < 1$ would be subject to a profitable deviation by equalizing elite members' payoffs. Attention can therefore be restricted to equilibria of this type in which elite payoffs are equalized. Since $p^* > p^\dagger$ and $\lambda = 0$, [A.2.4] and [A.2.8] imply that $\omega(\lambda) = \underline{\sigma} = 0$, so the binding post-investment no-rebellion constraint must be $\mathscr{R}(0) = 0$, using the expression from [A.2.40] evaluated at elite size $p = p^*$ and the payoff common to all elite members $\bar{\mathcal{U}}_{\mathrm{p}} = \mathcal{U}_{\mathrm{p}}^*$. That equation implies $\mathcal{U}_{\mathrm{p}}^*$ is given by:

$$\mathcal{U}_{\mathrm{p}}^* = \mathcal{U}_{\mathrm{p}}^\dagger(K) - \delta\frac{p^*}{p^\dagger} + \delta.$$

By substituting this expression into equation [A.2.19], which holds in a Markovian equilibrium where the pre-investment stage no-rebellion constraint is binding:

$$\mathcal{U}_{\mathrm{w}}^* = \left(\mathcal{U}_{\mathrm{p}}^\dagger(K) - \delta\frac{p^*}{p^\dagger} + \delta\right) - \delta - \alpha\mathcal{S}_{\mathrm{i}}(\tilde{\theta}) = \left(\mathcal{U}_{\mathrm{p}}^\dagger(K) - \delta\frac{p^*}{p^\dagger}\right) - \alpha\mathcal{S}_{\mathrm{i}}(\tilde{\theta}).$$

Since the investors' surplus $\mathcal{S}_{\mathrm{i}}(\tilde{\theta})$ is strictly positive in any equilibrium with $s > 0$, this equation implies workers' post-investment no-rebellion constraint [A.2.1] is violated. Therefore there is no Markovian equilibrium with this configuration of binding constraints.

The case where both post-investment no-rebellion constraints for workers and elite members are binding

Now consider the possibility of an equilibrium where the workers' post-investment no-rebellion constraint and one other post-investment no-rebellion constraint including elite members are binding. The constraint for workers [A.2.1] is equivalent to $\mathscr{R}(1) = 0$ when binding, using the definition of the function

$\mathscr{R}(\sigma)$ from [A.2.3]. Suppose that $\mathscr{R}(\sigma) = 0$ is also imposed for some $\sigma \in [\omega(\lambda), 1)$. Using the same arguments developed in the case of the post-investment no-rebellion constraint for elite members binding in combination with the pre-investment no-rebellion constraint, it is seen that the binding post-investment constraint including some elite members must be the one obtained from setting $\sigma = \omega(\lambda)$. Therefore, the two binding constraints in this case are $\mathscr{R}(\omega(\lambda)) = 0$ and $\mathscr{R}(1) = 0$.

The binding post-investment constraint for workers is [A.2.22], which has been shown to imply equation [A.2.30] when this constraint binds in a Markovian equilibrium. Now suppose the Markovian equilibrium features $\lambda > 0$. There are two cases to consider. First, if $p^* > p^\dagger$, equation [A.2.30] implies that $\delta + \mathcal{U}^*_{\mathrm{w}} - \mathcal{U}^\dagger_{\mathrm{p}}(K) - \gamma < 0$, which together with [A.2.18] demonstrates that the function $\mathscr{N}(\lambda)$ from [A.2.11] is strictly decreasing in λ. If payoffs among the $1-\lambda$ fraction of non-loyal elite members are equalized then the function $\mathscr{R}(\sigma)$ reduces to $\mathscr{C}(\sigma)$, which means all the no-rebellion constraints in [A.2.4] continue to hold given [A.2.10a]. Furthermore, since $\mathscr{R}(\omega(\lambda)) = \mathscr{C}(\omega(\lambda)) = \mathscr{N}(\lambda)$, decreasing λ slackens the binding constraint, allowing a higher average elite payoff to be obtained.

Now consider the case where $\lambda > 0$ and $p^* \leq p^\dagger$. At $\sigma = \omega(\lambda)$, all the $(1-\lambda)$ fraction of non-loyal elite members would be included in the rebel army, so the distribution of payoffs among them is irrelevant. Formally, this means that $\mathscr{R}(\omega(\lambda)) = \mathscr{C}(\omega(\lambda))$ even when payoffs among non-loyal elite members are not equalized. Next, substitute the expression for $\mathcal{U}^*_{\mathrm{w}}$ from the binding no-rebellion constraint [A.2.22] for workers into the function $\mathscr{C}(\sigma)$ from [A.2.9]:

$$\mathscr{C}(\sigma) = \delta\frac{p^*}{p^\dagger} - \sigma\left(\delta\frac{p^*}{p^\dagger}\right) - (1-\sigma)\left(\mathcal{U}^\dagger_{\mathrm{p}}(K) - \check{\mathcal{U}}^*_{\mathrm{p}} + \delta\right) = (1-\sigma)\left(\delta\frac{p^*}{p^\dagger} - \delta - (\mathcal{U}^\dagger_{\mathrm{p}}(K) - \check{\mathcal{U}}^*_{\mathrm{p}})\right).$$

Equation [A.2.8] shows that $\omega(\lambda) < 1$ in the case under consideration, and as $\mathscr{R}(\omega(\lambda)) = 0$, it must be the case that $\mathscr{C}(\omega(\lambda)) = 0$. Using the equation above, this implies:

$$\mathcal{U}^\dagger_{\mathrm{p}}(K) - \check{\mathcal{U}}_{\mathrm{p}} = \delta\frac{p^*}{p^\dagger} - \delta = -\delta\left(\frac{p^\dagger - p^*}{p^\dagger}\right).$$

Since $p^* \leq p^\dagger$, the right-hand side is less than or equal to zero. However, as $\mathcal{U}^\dagger_{\mathrm{p}}(K) > \bar{\mathcal{U}}_{\mathrm{p}}$ and $\bar{\mathcal{U}}_{\mathrm{p}} \geq \check{\mathcal{U}}_{\mathrm{p}}$ according to [A.2.18] and [A.2.7], the left-hand side is strictly positive. This is a contradiction, so there cannot be an equilibrium of this type with $\lambda > 0$.

With attention restricted to cases where $\lambda = 0$, equation [A.2.8] implies $\omega(\lambda) = \underline{\sigma}$. If it were the case that $p^* \leq p^\dagger$ then [A.2.4] reveals that $\underline{\sigma} = 1 - (p/p^\dagger)$, and hence by evaluating [A.2.9] at $\sigma = \underline{\sigma}$:

$$\mathscr{R}(\underline{\sigma}) \leq \mathscr{C}(\underline{\sigma}) = -\frac{p^*}{p^\dagger}(\mathcal{U}^\dagger_{\mathrm{p}}(K) - \check{\mathcal{U}}^*_{\mathrm{p}}) - \left(1 - \frac{p^*}{p^\dagger}\right)(\mathcal{U}^\dagger_{\mathrm{p}}(K) - \mathcal{U}^*_{\mathrm{w}}) < 0,$$

making use of [A.2.10a], $\mathcal{U}^\dagger_{\mathrm{p}}(K) > \bar{\mathcal{U}}_{\mathrm{p}} = \check{\mathcal{U}}_{\mathrm{p}}$ from [A.2.18] and [A.2.7], and $\mathcal{U}^\dagger_{\mathrm{p}}(K) > \mathcal{U}^*_{\mathrm{w}}$ from [A.2.22]. This constitutes a violation of one of the no-rebellion constraints [A.2.4], so any equilibrium of this type must feature $p^* > p^\dagger$.

If $p^* > p^\dagger$ then any rebel army at the post-investment stage can only include a subset of the elite. It then follows that any equilibrium with payoff inequality among elite members would be subject to a profitable deviation because equalizing payoffs would slacken a relevant binding no-rebellion constraint. In summary, if an equilibrium of this type exists, it must feature $\lambda = 0$, payoff equalization among elite members, and $p^* > p^\dagger$.

The post-investment no-rebellion constraint for members of the elite cannot bind on its own

Suppose the general post-investment no-rebellion constraint is binding for a rebel army that includes some positive measure of elite members, but the post-investment no-rebellion constraint including only workers and the pre-investment constraints are both slack (the cases where elite members' post-investment constraint is binding in combination with one or both of these two other no-rebellion constraints have already been analysed and ruled out). Thus, in terms of the general set of post-investment no-rebellion constraints [A.2.4], $\mathscr{R}(\sigma) \geq 0$ for all $\sigma \in [\omega(\lambda), 1]$ and $\mathscr{R}(\sigma) = 0$ for some $\sigma \in [\omega(\lambda), 1)$, while $\mathscr{R}(1) > 0$. The pre-investment stage constraint is also slack, so by evaluating [A.2.5] at a Markovian equilibrium and using [5.6] and [5.10]:

$$\mathcal{U}_{\mathrm{p}}^{*} - \mathcal{U}_{\mathrm{w}}^{*} - \alpha \mathcal{S}_{\mathrm{i}}(\tilde{\theta}) < \delta. \tag{A.2.41}$$

Starting from such a conjectured Markovian equilibrium, first consider a deviation whereby payoffs are equalized among the $1 - \lambda$ fraction of non-loyal elite members. Since $\mathscr{C}(\sigma) \geq \mathscr{R}(\sigma)$ holds prior to payoff equalization according to [A.2.10a], and as the functions $\mathscr{C}(\sigma)$ and $\mathscr{R}(\sigma)$ coincide following the payoff equalization, the general post-investment no-rebellion constraint [A.2.4] remains satisfied after this deviation. Therefore, $\mathscr{C}(\sigma) \geq 0$ for all $\sigma \in [\omega(\lambda), 1]$.

After the deviation described above, first consider the case where now $\mathscr{C}(\sigma) > 0$ for all $\sigma \in [\omega(\lambda), 1]$. It would now be possible to make a further deviation by raising the tax τ_{q} by some positive amount and distributing the proceeds to members of the elite. Given [A.2.41], if the change in τ_{q} is sufficiently small then the pre-investment no-rebellion constraint [A.2.5] continues to hold. Since the increase in elite members' payoffs cannot increase their fighting effort if they join a rebel army, the definition of $\mathscr{C}(\sigma)$ in [A.2.9] implies that $\mathscr{C}(\sigma)$ remains non-negative for all σ following this tax increase, as long as the change in τ_{q} is sufficiently small. Therefore, all the no-rebellion constraints remain satisfied following the deviation, and the elite objective function $\bar{\mathcal{U}}_{\mathrm{p}}$ is strictly increased. It must be the case that any equilibrium of this type features $\mathscr{C}(\sigma) = 0$ for some $\sigma \in [\omega(\lambda), 1]$ after the initial payoff-equalizing deviation.

The payoff equalization among non-loyal elite members does not affect the post-investment no-rebellion constraint for workers. Since this constraint remains slack, it follows that $\mathscr{C}(1) = \mathscr{R}(1) > 0$. Given that $\mathscr{C}(\sigma) \geq 0$ for all $\sigma \in [\omega(\lambda), 1]$, and $\mathscr{C}(\sigma)$ is linear in σ according to [A.2.9], if $\mathscr{C}(\sigma) = 0$ for some $\sigma \in [\omega(\lambda), 1)$, then it must be the case that $\mathscr{C}(\omega(\lambda)) = 0$.

In an equilibrium of this type where $\mathscr{C}(\omega(\lambda)) = 0$, consider an increase in the tax τ_{q} by a positive but sufficiently small amount such that the post-investment no-rebellion constraint for workers ($\mathscr{R}(1) = \mathscr{C}(1) \geq 0$) and the pre-investment no-rebellion constraint [A.2.5] remain satisfied. Since both constraints are initially slack, it is possible to find such a tax increase. Suppose the proceeds are distributed equally among only the $1 - \lambda$ fraction of non-loyal elite members. Since the tax is levied on all $1 - p^*$ non-elite members and the revenue is shared out among $(1 - \lambda)p^*$ non-loyal elite members, if $\mathcal{U}_{\mathrm{w}}$ falls by $\Delta\tau_{\mathrm{q}}$, $\check{\mathcal{U}}_{\mathrm{p}}$ rises by $((1 - p^*)/(1 - \lambda)p^*)\Delta\tau_{\mathrm{q}}$. Using equation [A.2.9], the net change in $\mathscr{R}(\omega(\lambda)) = \mathscr{C}(\omega(\lambda))$ is

$$\Delta\mathscr{C}(\omega(\lambda)) = (1 - \omega(\lambda)) \left(\frac{(1 - p)}{(1 - \lambda)p} \Delta\tau_{\mathrm{q}} \right) - \omega(\lambda)\Delta\tau_{\mathrm{q}}.$$

If $\omega(\lambda) = 0$, the expression above is clearly positive. If $\omega(\lambda) > 0$ then equation [A.2.8] implies $\omega(\lambda) = 1 - (1 - \lambda)(p/p^{\dagger})$, in which case this can be substituted into the equation above:

$$\Delta\mathscr{C}(\omega(\lambda)) = \left\{ \frac{(1 - \lambda)p}{p^{\dagger}} \frac{(1 - p)}{(1 - \lambda)p} - 1 + \frac{(1 - \lambda)p}{p^{\dagger}} \right\} \Delta\tau_{\mathrm{q}} = \left\{ \frac{1 - \lambda p}{p^{\dagger}} - 1 \right\} \Delta\tau_{\mathrm{q}}.$$

Note that $(1 - \lambda)p/p^{\dagger} = (1/p^{\dagger} - 2) + (1 - \lambda p)$, so since [5.5] implies $p^{\dagger} < 1/2$, it follows that the change in $\mathscr{C}(\omega(\lambda))$ is positive. Hence, after this deviation, $\mathscr{C}(\omega(\lambda)) \geq 0$ and $\mathscr{C}(1) \geq 0$. Given that $\mathscr{C}(\sigma)$ is linear in σ, it must then be the case that $\mathscr{C}(\sigma) \geq 0$ for all $\sigma \in [\omega(\lambda), 1]$. Given that $\mathscr{R}(\sigma) = \mathscr{C}(\sigma)$, all no-rebellion constraints continue to hold following the deviation, which also strictly increases $\check{\mathcal{U}}_{\mathrm{p}}$ and hence $\bar{\mathcal{U}}_{\mathrm{p}}$. Therefore there is no Markovian equilibrium with this configuration of binding no-rebellion constraints.

The equilibrium configuration of binding constraints

The analysis so far has ruled out all but one configuration of binding no-rebellion constraints: the case where the post-investment no-rebellion constraint for workers binds in combination with the no-rebellion constraint for elite members at the post-investment stage. Furthermore, it has been shown in this case that a Markovian equilibrium must feature no loyal elite members ($\lambda = 0$), full payoff equalization among all elite members, and a larger elite than would prevail in the equilibrium following a rebellion at the post-investment stage ($p^* > p^{\dagger}$). Therefore, the functions $\mathscr{R}(\sigma)$ and $\mathscr{C}(\sigma)$ from [A.2.3] and [A.2.9] are the same, and it can be seen from [A.2.4] and [A.2.8] that $\omega(\lambda) = 0$. The binding no-rebellion constraints are thus represented by the equations $\mathscr{R}(0) = \mathscr{C}(0) = 0$ and $\mathscr{R}(1) = \mathscr{C}(1) = 0$. Note also that when the post-investment no-rebellion constraint for workers is binding in a Markovian equilibrium, it has been shown that the constraint $\mathscr{C}(\omega(\lambda)) = \mathscr{C}(0) \geq 0$ is equivalent to $p^* \geq \pi(s)$, where $\pi(s)$ is the function defined in [A.2.27]. Since both constraints are binding, it follows that $p^* = \pi(s)$.

It remains to verify whether all other no-rebellion constraints are satisfied. Since $\mathscr{C}(\sigma)$ is linear in σ, the binding constraints at $\sigma = \underline{\sigma} = 0$ and $\sigma = 1$ imply that $\mathscr{C}(\sigma) = 0$ for all $\sigma \in [0, 1]$, hence [A.2.4] holds. It has already been shown that when the post-investment no-rebellion constraint for workers is binding, the pre-investment no-rebellion constraint [A.2.5] is equivalent to the inequality in [A.2.31b]. Hence, with $p^* = \pi(s)$ and $\mathcal{S}_{\mathrm{i}}(\tilde{\theta}) > 0$, this constraint is automatically satisfied.

The elite's payoff given the binding constraints

Using the definition in [A.2.9] and subtracting the equation $\mathscr{C}(1) = 0$ from $\mathscr{C}(0) = 0$, it follows that:

$$\mathcal{U}_{\mathrm{w}} = \mathcal{U}_{\mathrm{p}} - \delta.$$

Substituting this expression for $\mathcal{U}_{\mathrm{w}}$ into the elite's objective function $\bar{\mathcal{U}}_{\mathrm{p}}$ from equation [A.2.6], and noting that payoff equalization implies $\bar{\mathcal{U}}_{\mathrm{p}} = \mathcal{U}_{\mathrm{p}}$:

$$p\mathcal{U}_{\mathrm{p}} = (1 - p)(\mathrm{q} - (\mathcal{U}_{\mathrm{p}} - \delta)) + \mu(\kappa - \tilde{\theta})s,$$

and hence by rearranging the above to find an expression for $\mathcal{U}_{\mathrm{p}}$:

$$\mathcal{U}_{\mathrm{p}} = (\mathrm{q} + \delta)(1 - p) + \mu(\kappa - \tilde{\theta})s.$$

Now substituting for $p = \pi(s)$ using the formula for $\pi(s)$ from [A.2.27]:

$$\mathcal{U}_{\mathrm{p}} = (\mathrm{q} + \delta)(1 - p^\dagger) + \mu(\kappa - \tilde{\theta})s - \frac{\mu\tilde{\theta}s}{\delta}.$$

Simplifying this expression yields

$$\mathcal{U}_{\mathrm{p}} = (\mathrm{q} + \delta)(1 - p^\dagger) + \mu\left(\kappa - \frac{\tilde{\theta}}{p^\dagger}\right)s,$$

and by substituting the formula for $p^\dagger$ from [5.5], the expression for $\mathcal{U}_{\mathrm{p}}$ in [5.12b] is obtained. This completes the proof.

A.3 Proof of Proposition 3

First, note that the relationship between s and $\tilde{\theta}$ in [5.7] implies

$$\tilde{\theta} = \psi + (\kappa - \psi)s. \qquad [A.3.1]$$

Given that Proposition 2 shows that the general no-rebellion constraint [5.11] is binding for $\sigma_{\mathrm{w}}^\dagger = 1$ and for $\sigma_{\mathrm{p}}^\dagger = 1$, by subtracting these equations from one another, it follows that:

$$\mathcal{U}_{\mathrm{w}} = \mathcal{U}_{\mathrm{p}} - \delta. \qquad [A.3.2]$$

The level of investment in the Markovian equilibrium

By substituting the formula for $\tilde{\theta}$ from [A.3.1] into the expression for the elite's payoff in [5.12b]:

$$\mathcal{U}_{\mathrm{p}} = \frac{(\mathrm{q} + \delta)^2}{\mathrm{q} + 2\delta} + \mu\left(\kappa - \left(\frac{\mathrm{q} + 2\delta}{\delta}\right)\psi - \left(\frac{\mathrm{q} + 2\delta}{\delta}\right)(\kappa - \psi)s\right)s.$$

The derivative with respect to s is

$$\frac{\partial \mathcal{U}_{\mathrm{p}}}{\partial s} = \mu\left(\kappa - \left(\frac{q+2\delta}{\delta}\right)\psi - 2\left(\frac{q+2\delta}{\delta}\right)(\kappa-\psi)s\right).$$

Setting the derivative to zero and solving for s yields

$$s = \frac{\delta\kappa - (q+2\delta)\psi}{2(q+2\delta)(\kappa-\psi)} = \frac{1}{2}\frac{\delta\kappa-(q+2\delta)\psi}{(q+2\delta)\kappa-(q+2\delta)\psi}. \qquad [A.3.3]$$

Since $q+2\delta > \delta$, this expression can never be more than 1, but could be negative. Given that s is restricted to $s \in [0,1]$, the value of s that maximizes $\mathcal{U}_{\mathrm{p}}$ is

$$s^* = \max\left\{0, \frac{\delta\kappa-(q+2\delta)\psi}{2(q+2\delta)(\kappa-\psi)}\right\}. \qquad [A.3.4]$$

This is the solution given in [5.16]. It can be seen that s^* is positive whenever $\delta\kappa > (q+2\delta)\psi$, which is equivalent to $\kappa/\psi - 1 > 1 + q/\delta$.

To confirm that this is indeed a Markovian equilibrium, it is necessary to check whether several auxiliary constraints are satisfied. First, there is the condition from [3.6] that a member of the elite would find it rational to defend the existing institutions if there is a rebellion where the individual is not included in the rebel army. This requires $\mathcal{U}_{\mathrm{p}}^* > \mathcal{U}_{\mathrm{n}}' = \mathcal{U}_{\mathrm{n}}^*$ at the pre-investment stage, and $\mathcal{U}_{\mathrm{p}}^* > \mathcal{U}_{\mathrm{w}}^\dagger(K)$ at the post-investment stage.

Note that equations [5.6] and [5.9] imply the following expression for $\mathcal{U}_{\mathrm{n}}^*$:

$$\mathcal{U}_{\mathrm{n}}^* = \mathcal{U}_{\mathrm{w}}^* + \alpha\mathcal{S}_{\mathrm{i}}(\tilde{\theta}^*).$$

Given that $\mathcal{U}_{\mathrm{p}}^* = \mathcal{U}_{\mathrm{w}}^* + \delta$ according to [A.3.2], the condition $\mathcal{U}_{\mathrm{p}}^* > \mathcal{U}_{\mathrm{n}}^*$ is equivalent to

$$\alpha\mathcal{S}_{\mathrm{i}}(\tilde{\theta}^*) < \delta. \qquad [A.3.5]$$

Using the distribution of the effort cost θ from [5.1] and [5.7], it is known that $0 < \psi < \kappa$ and $\psi \leq \tilde{\theta}^* \leq \kappa$. Hence it must be the case that $\tilde{\theta}^* - \theta < \kappa$ for all $\theta \in [\psi, \kappa]$, and the definition of $\mathcal{S}_{\mathrm{i}}(\tilde{\theta})$ in [5.10] can then be used to deduce $\mathcal{S}_{\mathrm{i}}(\tilde{\theta}^*) < \kappa$. Since α is a probability and $\kappa < \delta$ according to the parameter restrictions in [5.3], the condition [A.3.5] is verified.

Using the expression for $\mathcal{U}_{\mathrm{p}}$ from [5.12b], the expression for $\mathcal{U}_{\mathrm{w}}^\dagger(K)$ from [5.5], and the formula for the capital stock K from [5.7], the condition $\mathcal{U}_{\mathrm{p}}^* > \mathcal{U}_{\mathrm{w}}^\dagger(K)$ is equivalent to:

$$\frac{(q+\delta)^2}{q+2\delta} + \mu\left(\kappa - \left(\frac{q+2\delta}{\delta}\right)\tilde{\theta}^*\right)s^* > \frac{(q+\delta)^2}{q+2\delta} - \delta + \mu\kappa s^*.$$

After cancelling terms and rearranging, this requirement reduces to:

$$\mu\left(\frac{q+2\delta}{\delta}\right)\tilde{\theta}^* s^* < \delta. \qquad [A.3.6]$$

Consider an equilibrium where $s^* > 0$, in which case equation [A.3.4] implies

$$(\kappa-\psi)s^* = \frac{\delta\kappa-(q+2\delta)\psi}{2(q+2\delta)},$$

and where it must be the case that $\delta\kappa > (q+2\delta)\psi$. Substituting this into the formula for $\tilde{\theta}$ in [A.3.1]:

$$\tilde{\theta}^* = \psi + \frac{\delta\kappa-(q+2\delta)\psi}{2(q+2\delta)} = \frac{1}{2}\left(\left(\frac{\delta}{q+2\delta}\right)\kappa + \psi\right).$$

Given that $\psi < \kappa$, it must be the case that $\psi < (\delta/(q+2\delta))\kappa$, and so the expression above for $\tilde{\theta}^*$ implies:

$$\tilde{\theta}^* < \left(\frac{\delta}{q+2\delta}\right)\kappa.$$

Since $\mu \leq 1$ and $s^* \leq 1$, and using the parameter restriction $\kappa < \delta$ from [5.3], the inequality above implies that [A.3.6] must hold, demonstrating that $\mathcal{U}_p^* > \mathcal{U}_w^\dagger(K)$.

As the utility function is linear, the link between worker and elite payoffs in [A.3.2] implies $C_p^* = C_w^* + \delta$. Therefore, all non-negativity constraints on consumption will hold if $C_w^* \geq 0$, which is equivalent to $U_w^* \geq 0$. Observe first that $s = 0$ is always a feasible choice for the elite in maximizing $\mathcal{U}_p$, so if $s^* > 0$, it follows from the expression in [5.12b] that:

$$\mu\left(\kappa - \left(\frac{q+2\delta}{\delta}\right)\tilde{\theta}^*\right)s^* \geq 0. \qquad \text{[A.3.7]}$$

Substituting the expression for $\mathcal{U}_p$ from [5.12b] into [A.3.2] yields:

$$\mathcal{U}_w^* = \left(\frac{(q+\delta)^2}{q+2\delta} - \delta\right) + \mu\left(\kappa - \left(\frac{q+2\delta}{\delta}\right)\tilde{\theta}^*\right)s^*, \qquad \text{[A.3.8]}$$

and since [A.3.7] shows the second term is non-negative, a sufficient condition for $\mathcal{U}_w^* \geq 0$ is

$$\frac{(q+\delta)^2}{q+2\delta} \geq \delta. \qquad \text{[A.3.9]}$$

The parameter restriction $\delta/q \leq \varphi$ from [5.3] implies that this inequality holds, so it is confirmed that $C_w^* \geq 0$. Therefore, the solution in [A.3.4] is shown to be the unique Markovian equilibrium.

The constrained efficient level of investment

Using the relationship between p and s from [5.12a], it follows that:

$$\delta(1-p) = \delta\left(1 - p^\dagger - \frac{\mu\tilde{\theta}s}{\delta}\right) = \delta(1-p^\dagger) - \mu\tilde{\theta}s = \frac{\delta(q+\delta)}{q+2\delta} - \mu\tilde{\theta}s,$$

where the formula for $p^\dagger$ from [5.5] is also substituted into the above expression. The definition of the average payoff $\bar{\mathcal{U}}$ from [5.14] is equivalent to [5.15], and the expression for $(1-\delta)p$ above can be used to obtain:

$$\bar{\mathcal{U}} = \frac{(q+\delta)^2}{q+2\delta} - \frac{\delta(q+\delta)}{q+2\delta} + \mu\left(\kappa - \left(\frac{q+\delta}{\delta}\right)\tilde{\theta}\right)s + \mu\tilde{\theta}s + \mu\mathcal{S}_i(\tilde{\theta}).$$

After simplification, this expression for $\bar{\mathcal{U}}$ reduces to:

$$\bar{\mathcal{U}} = \frac{q(q+\delta)}{q+2\delta} + \mu\left(\kappa - \left(\frac{q+\delta}{\delta}\right)\tilde{\theta}\right)s + \mu\mathcal{S}_i(\tilde{\theta}).$$

Equation [A.3.1] gives a relationship between $\tilde{\theta}$ and s, which can also be substituted into the above:

$$\bar{\mathcal{U}} = \frac{q(q+\delta)}{q+2\delta} + \mu\left(\kappa - \left(\frac{q+\delta}{\delta}\right) - \left(\frac{q+\delta}{\delta}\right)(\kappa-\psi)s\right)s + \mu\mathcal{S}_i(\tilde{\theta}). \qquad \text{[A.3.10]}$$

Using [5.1] and [5.10], an explicit expression for the expected surplus $\mathcal{S}_i(\tilde{\theta})$ from receiving an investment opportunity is given by:

$$\mathcal{S}_i(\tilde{\theta}) = \int_{\theta=\psi}^{\tilde{\theta}} \frac{\tilde{\theta}-\theta}{\kappa-\psi}\mathrm{d}\theta = \frac{1}{2}\frac{(\tilde{\theta}-\psi)^2}{\kappa-\psi} = \frac{1}{2}(\kappa-\psi)s^2,$$

where equation [5.7] has been used to write this solely in terms of s. This is then substituted into [A.3.10] to obtain an expression for $\bar{\mathcal{U}}$ in terms of s:

$$\bar{\mathcal{U}} = \frac{\mathrm{q}(\mathrm{q}+\delta)}{\mathrm{q}+2\delta} + \mu\left(\kappa - \left(\frac{\mathrm{q}+\delta}{\delta}\right) - \left(\frac{2\mathrm{q}+\delta}{2\delta}\right)(\kappa-\psi)s\right)s. \tag{A.3.11}$$

Using [A.3.11], the derivative of $\bar{\mathcal{U}}$ with respect to s is:

$$\frac{\partial\bar{\mathcal{U}}}{\partial s} = \mu\left(\kappa - \left(\frac{\mathrm{q}+\delta}{\delta}\right) - \left(\frac{2\mathrm{q}+\delta}{\delta}\right)(\kappa-\psi)s\right).$$

Setting the derivative to zero and solving for s yields:

$$s = \frac{\delta\kappa-(\mathrm{q}+\delta)\psi}{(2\mathrm{q}+\delta)(\kappa-\psi)} = \frac{\delta\kappa-(\mathrm{q}+\delta)\psi}{(\mathrm{q}+\delta)\kappa-(\mathrm{q}+\delta)\psi+\mathrm{q}(\kappa-\psi)}. \tag{A.3.12}$$

Since $\mathrm{q}+\delta > \delta$ and $\kappa > \psi$, this expression can never be greater than 1, but it could be negative. Therefore, if no auxiliary constraints are violated, the constrained efficient level of s is

$$s^{\diamond} = \max\left\{0, \frac{\delta\kappa-(\mathrm{q}+\delta)\psi}{(2\mathrm{q}+\delta)(\kappa-\psi)}\right\}. \tag{A.3.13}$$

This is the expression for $s^{\diamond}$ from [5.16]. It is positive whenever $\delta\kappa > (\mathrm{q}+\delta)\psi$, which is equivalent to $\kappa/\psi - 1 > \mathrm{q}/\delta$.

The auxiliary constraints to verify are $\mathcal{U}_{\mathrm{p}}^{\diamond} > \mathcal{U}_{\mathrm{n}}^{\diamond}$, $\mathcal{U}_{\mathrm{p}}^{\diamond} > \mathcal{U}_{\mathrm{w}}^{\dagger}(K)$, and $C_{\mathrm{w}}^{\diamond} \geq 0$. Given that there is the same configuration of binding no-rebellion constraints, the analysis leading to [A.3.5] also shows that $\mathcal{U}_{\mathrm{p}}^{\diamond} > \mathcal{U}_{\mathrm{n}}^{\diamond}$ is equivalent to $\alpha\mathcal{S}_{\mathrm{i}}(\tilde{\theta}^{\diamond}) < \delta$. Under the parameter restrictions from [5.3], this condition is necessarily satisfied.

Next, consider the constraint $\mathcal{U}_{\mathrm{p}}^{\diamond} > \mathcal{U}_{\mathrm{w}}^{\dagger}(K)$. Again, given that the configuration of binding no-rebellion constraints is the same, the analysis leading to [A.3.6] also applies in this case, so the requirement is equivalent to

$$\mu\left(\frac{\mathrm{q}+2\delta}{\delta}\right)\tilde{\theta}^{\diamond}s^{\diamond} < \delta. \tag{A.3.14}$$

Consider a case where $s^{\diamond} > 0$. Using [A.3.13], it must then be the case that

$$(\kappa-\psi)s^{\diamond} = \frac{\delta\kappa-(\mathrm{q}+\delta)\psi}{(2\mathrm{q}+\delta)},$$

noting $\delta\kappa > (\mathrm{q}+\delta)\psi$ is necessary. Substituting this into [A.3.1] yields:

$$\tilde{\theta}^{\diamond} = \psi + \frac{\delta\kappa-(\mathrm{q}+\delta)\psi}{(2\mathrm{q}+\delta)} = \frac{\mathrm{q}\psi+\delta\kappa}{2\mathrm{q}+\delta}. \tag{A.3.15}$$

Since $\psi < (\delta/(\mathrm{q}+\delta))\kappa$ in this case, it follows that

$$\tilde{\theta}^{\diamond} < \frac{\mathrm{q}\left(\frac{\delta}{\mathrm{q}+\delta}\right)\kappa+\delta\kappa}{2\mathrm{q}+\delta} = \frac{\delta(2\mathrm{q}+\delta)\kappa}{(\mathrm{q}+\delta)(2\mathrm{q}+\delta)} = \frac{\delta}{\mathrm{q}+\delta}\kappa,$$

which implies:

$$\left(\frac{\mathrm{q}+2\delta}{\delta}\right)\tilde{\theta}^{\diamond} < \frac{\mathrm{q}+2\delta}{\mathrm{q}+\delta}\kappa.$$

Using the parameter restrictions in [5.3] and the inequality above:

$$\mu\left(\frac{\mathrm{q}+2\delta}{\delta}\right)\tilde{\theta}^{\diamond} < \left(\frac{\mathrm{q}}{2(\mathrm{q}+2\delta)}\right)\left(\frac{\mathrm{q}+2\delta}{\mathrm{q}+\delta}\right)\kappa = \frac{1}{2}\frac{\mathrm{q}}{\mathrm{q}+\delta}\kappa < \kappa < \delta.$$

This demonstrates that [A.3.14] holds, which confirms that $\mathcal{U}_{\text{p}}^{\diamond} > \mathcal{U}_{\text{w}}^{\dagger}(K)$.

Finally, consider the non-negativity constraint $C_{\text{w}}^{\diamond} \geq 0$ for workers, which is equivalent to $\mathcal{U}_{\text{w}}^{\diamond} \geq 0$ given the utility function. Substituting the formula for $\mathcal{U}_{\text{p}}$ from [5.12b] into [A.3.2], and using [A.3.15] to obtain an expression for $\tilde{\theta}^{\diamond}$:

$$\mathcal{U}_{\text{w}}^{\diamond} = \left(\frac{(\mathrm{q}+\delta)^2}{\mathrm{q}+2\delta} - \delta\right) + \frac{\mu}{\delta(\mathrm{q}+2\delta)} \left(\delta(\mathrm{q}-\delta)\kappa - \mathrm{q}(\mathrm{q}+2\delta)\psi\right) s^{\diamond}. \qquad \text{[A.3.16]}$$

The sign of this expression is ambiguous for general parameters satisfying the restrictions in [5.3], so the non-negativity constraint for workers could be binding. When this expression is non-negative, the constrained efficient level of s is indeed given by the formula in [A.3.13] since all other auxiliary constraints are satisfied. More generally, since the non-negativity constraint is satisfied at $s = 0$, the possibility that it might be binding in equilibrium implies that $\kappa/\psi - 1 > \mathrm{q}/\delta$ is only a necessary condition for $s^{\diamond} > 0$.

Consider an equilibrium with $s^{\diamond} > 0$. In the case where the non-negativity constraint is not binding, the value of $s^{\diamond}$ is given by [A.3.13]. Comparison with the expression for s^* in [A.3.4] shows that $s^* < s^{\diamond}$. Now suppose the non-negativity constraint is binding. Since the non-negativity constraint is known to be satisfied at $s = 0$ and $s = s^*$, it follows that constrained efficient value of s must be strictly larger than s^*. Therefore, it is shown that $s^* < s^{\diamond}$ whenever $s^{\diamond} > 0$. This completes the proof.