A Model of Equilibrium Institutions*

Bernardo Guimaraes†
Sao Paulo School of Economics (FGV)

Kevin D. Sheedy
London School of Economics

First draft: 19th October 2009
This draft: 7th October 2011

Abstract

In order to understand inefficient institutions, one needs to understand what might cause the breakdown of a political version of the Coase Theorem. This paper considers an environment populated by ex-ante identical agents and develops a model of power and distribution where institutions (the "rules of the game") are set to maximize payoffs of those individuals in power. They are constrained by the threat of rebellion, where any rebels would be similarly constrained by further threats. Equilibrium institutions are the fixed point of this constrained maximization problem. This model can be applied to different economic environments. Private investment depends on credible limitations on expropriation, which can only be achieved if power is not as concentrated as those in power would like it to be, ex-post. Endogenously, this enables the group in power to act as government committed to protection of property rights, which would otherwise be time inconsistent. But the "political" Coase Theorem does not hold. Since sharing power implies sharing rents, capital taxation is inefficiently high.

JEL classifications: E02; O43; P48.

Keywords: government; political economy; public goods; capital taxation; time inconsistency.

---

*We thank Bruno Decreuse, Erik Eyster, Carlos Eduardo Goncalves, Ethan Ilzetzki, Per Krusell, Dirk Niepelt, Andrea Prat, Ronny Razin 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, ESSIM 2010, FGV–Rio, ISI–Delhi Economic Growth and Development conference 2010, Institute for International Economic Studies, 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, and U. St. Gallen for helpful comments.

†Corresponding author: Sao Paulo School of Economics (FGV), Email: bernardo.guimaraes@fgv.br.

> 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

Institutions are defined by North (1990) as the "rules of the game", or the "humanly devised constraints that shape human interaction". It has become commonplace to claim that institutions go a long way in explaining the huge disparities in income between countries — by affecting incentives to invest, produce, and exchange.[1] Failings are often ascribed to institutions serving the interests of an elite rather than the interests of society at large. This argument, however, raises a subtle question: while it is straightforward to see that elite control of institutions has distributional consequences, it is far less obvious why it should give rise to huge economic inefficiencies. Why fundamentally does the distribution of power prevent Pareto-improving policies being pursued? Or as phrased by Acemoglu (2003), why is there no "political" equivalent of the Coase theorem?

Moreover, even when such a "political" Coase theorem does not hold in general, it need not follow that it breaks down across the board. Empirical evidence suggests that where institutions are impeding economic growth, it is less because they are uniformly bad, but that they are inadequate along particular dimensions, with protections against expropriation of private property being especially important (Acemoglu, Johnson and Robinson, 2005). The question is then why institutions fail so badly to deliver efficient outcomes in some areas but not necessarily in others.

This paper builds a model to address these questions. Its starting point is to posit that institutions arise out of a power struggle — they exist in a world where any group of individuals can attempt to change them by force through rebellions and revolutions. Moreover, when new institutions are created, they are moulded not by a disinterested social planner, but by the self-interested elite currently triumphant in the power struggle. Equilibrium institutions are those that would be created and survive in such a world of *might*, not *right*.

To focus attention on the effects of the power struggle on institutions and the consequent economic outcomes, the analysis supposes an otherwise frictionless environment. Institutions can be created without any arbitrary restrictions, in general, laying down — for as long as they survive — a complete specification of the distribution of power and resources in a society. Furthermore, though institutions are set up to reflect the interests of the group in power, once they exist, those institutions represent more than just the unfettered will of the elite. To capture this idea, the analysis assumes that institutions do indeed define the "rules

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

of the game" that must be respected by everyone, subject only to no resumption of the power struggle through rebellion or revolution launched by any group.

In this environment, though the prospects for what might be considered a "fair" distribution of resources are bleak, the door is not necessarily closed on the weaker goal of economic efficiency. Achieving Pareto efficiency requires some means of supporting "deals" between the elite in power and those outside it. While the elite can shape new institutions as it wishes, it must consider the effects of its choices on others because of the threat of rebellions that can sweep it from power. The analysis places no arbitrary restrictions on which individuals can take part in a rebellion (including those inside the elite), so in principle, effects on all individuals' payoffs must be weighed — although generally not equally — when the elite sets up institutions. Making an analogy, the ability to fight in the power struggle plays the role of the well-defined legal property rights in the classic Coase theorem. The absence of restrictions on the form of the institutions the elite can create and the fact that these institutions do set the "rules of the game" are analogous to the absence of any *exogenous* transaction costs in writing and enforcing contracts.

This logic suggests that there is nothing in the nature of the environment here that necessarily precludes efficient institutions (a "political" Coase theorem) emerging in equilibrium. But what institutions prescribe is effective only if they are able to survive the power struggle. There is no extraneous force that can maintain a set of institutions if a group of individuals (possibly including those inside the elite) is willing and strong enough to overthrow them, in other words, there are no "meta-institutions" — nothing outside and above the institutions but the brute force of the power struggle. Thus, in spite of there being no exogenous restrictions on the deals that can be embodied in a set of institutions, the need to avoid any group of individuals having an incentive to rebel may *endogenously* place constraints on the set of possible deals, leading to a breakdown of the "political" Coase theorem.

The power struggle is modelled here by a simple linear "conflict technology" to which all individuals have access. This allows groups of individuals to exert fighting effort to overthrow the current institutions and create new ones, with those currently in power able to put up resistance. It is assumed that those defending the current institutions are at an advantage over attackers, but that fighting ability is otherwise the same across all individuals. If conflict occurs, there is an outright victory determined by which side has the greatest aggregate fighting strength. In this environment then, the goal of any elite is to design institutions in its own interests, subject to no group of individuals having an incentive to launch a viable rebellion. Given complete information about incentives to rebel, no conflict will occur in equilibrium. This means any inefficiencies that arise result not from destructive conflict per se, but rather the effect that the threat of conflict has on the choice of institutions.

In choosing other assumptions, the general aim is to approach as far as possible a *tabula* *rasa* by not assuming the prior existence of classes or social groups, or indeed any political ideologies. The model supposes instead a world populated by ex-ante identical and self-interested individuals. This means that the notion of equilibrium institutions is independent of the competence, benevolence, or factional affiliation of those comprising the elite — their only distinguishing feature is having the most recent victory in the power struggle. And such a victory does not put an end to that struggle: there are always opportunities to rebel against those who were hitherto the rebels. Thus, once a rebellion has succeeded in destroying the current institutions, those now in power will have exactly the same objectives and face exactly the same constraints as those formerly in power. As in George Orwell's *Animal Farm*, there is no essential difference between the "men" and the "pigs", but in equilibrium, some individuals will be "more equal" than others. Formally speaking, 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.

The model gives rise to a simple theory of distribution. The equilibrium 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 the more power poses a greater danger if he supports a rebellion because of his superior fighting strength. Since any rebellion will be made up of 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 disgruntlement with the institutions. Sharing power thus always entails sharing rents.

In an endowment economy, there is a basic trade-off 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. This allows them to levy higher taxes on those outside the elite even though that increases outsiders' incentive to rebel. On the other hand, the proceeds must then be divided more thinly among more individuals. The equilibrium elite size (which determines the distribution of power, and hence the distribution of resources) maximizes the payoff of an elite member by striking a balance between these two effects.

Do the equilibrium institutions lead to the elite acting as a "government" in any meaningful sense of the term? In other words, 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 agent. This outcome is consistent with a world characterized by a political Coase theorem. The ability of the elite to set down institutions prescribing public-good provision is analogous to the possibility of contracting in the classic Coasean analysis, while the notion of the elite constrained by threats of rebellion means that there is a "price" attached to policy actions affecting those outside the elite, analogous to property rights being defined.

A natural question is the extent to which this finding generalizes to other settings where institutions can influence economic outcomes. To address one important instance of this question, 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. With the effort cost of investment sunk, those who have already been expropriated will not find it rational to exert more or less fighting effort in a rebellion than those who never invested to begin with. Consequently, the equilibrium distribution of resources would reflect only the distribution of power, and payoffs would not depend on whether or not an individual had previously invested.

Thus, to provide appropriate incentives for individuals to invest, the institutions established prior to investment decisions must offer those who invest a higher payoff, and moreover, those institutions must survive rebellions so that what they prescribe is actually put into practice. The first requirement is straightforward because the elite is always free to choose institutions specifying any distribution of resources, or equivalently, any level of capital taxation. The challenge is the second requirement. In an endowment economy, the elite's principal concern is in avoiding a "popular uprising", a rebellion of outsiders. When offering incentives to investors, it becomes essential also to avoid a "coup d'état", a rebellion launched by insiders. The reason is that in rewarding the productive, rather than just the powerful, the elite deviates from the usual equilibrium distribution of resources that coincides with the distribution of power. Ex ante, it perceives such a deviation provides useful incentives. Ex post, once the effort costs of investing are sunk, this is not so.

The essential problem here is that institutions are binding on the elite and all other individuals only so long as not overthrown by force. Some way must therefore be found to stop those who hold power from using their positions to sweep aside institutions and build new ones that conform to their ex post interests. This might call to mind such notions as "independent judiciaries", "the rule of law", "political representation of investors", and the like. While it would be possible introduce such devices into the model by assumption, this would be as a *deus ex machina* that simply overrides the ability of elite members to initiate conflict and take part in the power struggle. As stated, the analysis does not place arbitrary restrictions on what institutions elites may choose to set up, but these institutions must be within the bounds of the "lower-level" technological primitives, namely, an unrestricted choice of the distributions of power and resources. Simply introducing by assumption a "higher-level" institutional technology that allows for the protection of property rights (perhaps at some cost) would not do justice to the problem 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.

If institutions are to provide credible incentives to investors in an environment where any groups can launch rebellions, it is necessary to reduce the attractiveness of rebellion to 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. If, however, the elite were large enough so that after any rebellion it would be optimal for the new group in power to be smaller, there would always be some individuals who lose power after a rebellion. Since these individuals would receive rents from their powerful position if the current institutions survived, they have both the means and the desire to defend them, thus raising the costs of rebellion for others.

Therefore, simply adding the possibility of investment to the model gives rise to an equilibrium with power sharing among a larger elite. Sharing power allows the elite to act as a government committed to a certain set of policies that would otherwise be time inconsistent. No ability to commit is assumed, but the model reveals how an *endogenous* commitment mechanism can be found instead. 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. It is important to note that power is not shared here among those individuals who are actually investing. The additional individuals in the elite in no sense represent or 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 provide for this efficient outcome, implying a breakdown of the political Coase theorem. 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 may remain unfulfilled. Members of the elite cannot share power without diluting their rents, so the distributional consequences of the necessary steps to protect investors against expropriation go against the interests of each individual in the elite.

While the model is quite abstract, it is congruent with a number of historical examples, some of which are discussed later in the paper: the disappearance of private corporations (the *societas publicanorum*) when power was concentrated under the Roman emperors; the need for a militarily strong leader (*podestà*) to guarantee stability in a society (medieval Genoa) where other strong groups could seize power; 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 developed (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 aiming at 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 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 people and to enforce contracts. The exogenous parameters are the extent of political turnover and institutional (or demographic) features that determine how much one group cares about the other. 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 ad-hoc restrictions on agents' choices. Hence there are no ex-ante groups, no ex-ante differences among individuals, and no relevant constraints on the choice of institutions besides the threat of conflict. In our view, this makes this paper more suited to studying which constraints on institutions emerge endogenously.

This paper shares important similarities with the literature on coalition formation, as analysed by 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 the paper here is the actual modelling of rebellions, since fighting is required if existing institutions are to be replaced by new ones.

The model assumes that the institutions established by the elite determine the allocation of resources once production has taken place. But how would these institutions manage to determine 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 institutions that survive the power struggle.

One possible interpretation of the view 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. For example, if the laws specify how much an individual must pay or receive from another, and if both expect to be harshly punished if they fail to comply (along with any "higher-order" deviators) then the laws will be self enforcing.

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 allocation of resources in society. This paper

[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).

takes a more cynical view of our fellow human beings. Here, the individuals in power choose the laws and institutions to maximize their own payoffs, and those institutions can only be destroyed by a rebellion — wiping out the old institutions, and making way for new ones. 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 they could agree on transfers to avoid conflict. This paper presupposes 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 over what has been produced. Moreover, they fight in groups, not as isolated individuals.

Acemoglu (2003) raises the question of the political Coase theorem and highlights the importance of commitment. Commitment is indeed the key issue in the main application of the model of this paper, but the question here is how commitment can endogenously arise and whether institutions that guarantee commitment will be consistent to a political Coase theorem. There are other theoretical models focused on political issues that lead to inefficiencies in protection of property rights. Examples include Glaeser, Scheinkman and Shleifer (2003), Acemoglu (2008), Guriev and Sonin (2009), and Myerson (2010). Here, the risk of capital expropriation 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 (see, for example, 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 equilibrium institutions in a simple endowment economy. Subsequent sections extend this analysis to richer environments where there is scope for institutions to affect economic outcomes.

3.1 Environment

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

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

and disutility if they exert fighting effort F:

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

where $u(\cdot)$ is a strictly increasing and weakly concave function.

Individuals who become *workers* ($\imath \in \mathcal{W}$) have access to a production technology that yields an exogenous quantity q of goods. Individuals who are currently *in power* ($\imath \in \mathcal{P}$) have a positive *fighting strength* (parameterized by δ) without needing to exert fighting effort F. These individuals are able to create new *institutions* (defined below) at no utility cost when they take power. Individuals who remain in power cannot simultaneously become workers.[4] It is assumed that a maximum of 50% of the population can hold positions of power.

Institutions (the "rules of the game") specify the identities of the individuals in power (the set $\mathcal{P}$), referred to as the *elite*, and the allocation of resources. Once institutions exist, the rules they specify laying down the allocation of resources are respected by all individuals unless a successful *rebellion* occurs. Rebellions are the only means of changing institutions. There is no fighting between isolated individuals over resources nor an ability for isolated individuals to resist the allocation of resources imposed by the prevailing institutions.

The institutions can specify any transfers between individuals subject only to leaving each individual $\imath$ with a non-negative quantity of consumption $C(\imath) \geq 0$ and satisfying an overall budget constraint. A worker $\imath \in \mathcal{W}$ facing an individual-specific tax $\tau(\imath)$ paid to the elite consumes

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

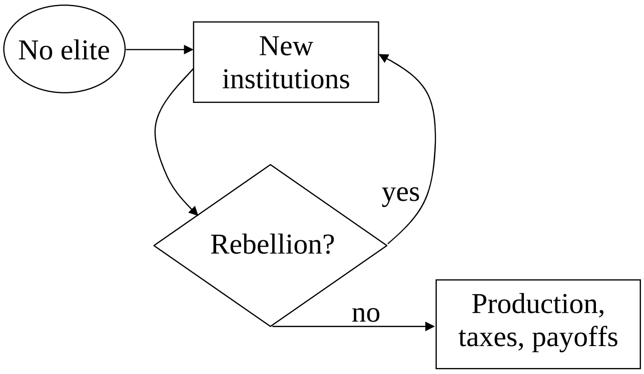
Tax revenue is used to finance the consumption of the elite. If $C_{\mathrm{p}}(\imath)$ is the individual-specific consumption of a member of the elite $\imath \in \mathcal{P}$ then the budget constraint is

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

The sequence of events is depicted in Figure 1. An elite takes power and establishes institutions. There are then opportunities for rebellion. If a successful rebellion occurs, new institutions are established, potentially changing the elite, with these institutions also being subject to subsequent threats of rebellion. When no rebellions occur, workers produce goods, the rules laid down by the prevailing institutions are implemented, and payoffs are received. Individuals have no ability to commit themselves to take actions at later stages of the game except where there is no utility loss in adhering to the commitment ex post.

[4]The assumption that those in power do not receive the same endowment as workers is not essential for the main results. However, it is not unreasonable to suppose there is some opportunity cost for individuals of being in power. Strictly speaking, the lost production for each individual in the elite means that this is not a pure endowment economy, however this "guns versus butter" inefficiency is *not* the focus of this paper.

Figure 1: *Sequence of events*



3.2 Institutions

The analysis begins by considering a stage of Figure 1 at which institutions have just been established. These institutions specify the distribution of power and the allocation of resources. Formally, institutions are a collection $\mathscr{I} = \{\mathcal{P}, \mathcal{W}, \tau(\imath), C_{\mathrm{p}}(\imath)\}$, where $\mathcal{P}$ is the set of individuals in power, $\mathcal{W}$ is the set of workers, $\tau(\imath)$ is a function specifying the distribution of taxes across workers, and $C_{\mathrm{p}}(\imath)$ is a function specifying the distribution of consumption within the elite.

The sets $\mathcal{P}$ and $\mathcal{W}$ partition the set of all individuals Ω. The taxes $\tau(\imath)$ and elite consumption levels $C_{\mathrm{p}}(\imath)$ must satisfy the budget constraint [3.3] and the non-negativity constraints on each individual's consumption.

Extensions of the basic model (some of which are considered in later sections of the paper) will have institutions specify other aspects of governance such as the degree of provision of public goods, the extent of protection of property rights, and potentially any other government policies.

3.3 Rebellions

A *rebellion* (if successful) destroys the current institutions, making way for the creation of new ones, which can rewrite the rules regarding the allocation of power and resources. A rebellion succeeds if the fighting strength of the rebels exceeds the fighting strength of those defending the current institutions. Rebellions can include any individuals, even those who belong the current elite.

The rebels cannot commit themselves in advance to take particular actions following the success of the rebellion (except ones for which there is no utility loss in honouring). The argument for this assumption is that there is no higher power to enforce contracts when institutions have been destroyed by force (no "meta-institutions"). Thus, although rebellions can destroy the current institutions, they do not include a binding manifesto for creating a particular set of new institutions in the future.

While a rebellion does not set down binding plans for the construction of subsequent institutions, beliefs about what institutions *would be* created following the destruction of the current ones influence incentives to take part in a rebellion. However, the fact that an individual gains from the success of a rebellion is *not* a sufficient reason for his fighting in support of it. Fighting is costly, while no single individual's fighting effort is pivotal in determining which side wins. Hence there are strong incentives for free-riding that armies must overcome.

Armies must therefore provide direct incentives for individuals to exert fighting effort, but there is a limit to what incentives they can credibly offer. Once the fighting is over, the effort cost of fighting is sunk, so unless the subsequent institutions can credibly treat two individuals differently, one of whom fought and one of whom shirked, who would otherwise have identical continuation utility, there can be no incentive for individuals to fight. As will be seen, such differences in payoffs would reduce the utility of the subsequent elite, so there is no incentive to honour past inducements to fight. This true for both financial rewards for fighting (the "carrot") and punishments for shirking (the "stick").

The only exception to this is that there will be incentives after a successful rebellion for power-sharing among a group that will comprise the new elite. Since all individuals are ex-ante identical and self-interested, any given individual in the elite does not care about the identities of the other individuals with whom he will be sharing power (though he does care about the total number of such people). As will be seen, those in the elite are able to obtain a higher payoff than those outside owing to their entrenched position if conflict occurs, so membership of the new elite can be offered as a credible incentive to fight. In the event of shirking, the place could be offered to someone else, which is a punishment that is costless for the elite to carry out on an isolated individual because of the existence of a pool of identical replacements who would like to join.

These arguments lead to the restriction that the rebel army can only include those who *expect* to have a place in the subsequent elite. The amount of fighting effort put in by each individual must also be individually rational. Let p'^{e} denote beliefs about post-rebellion elite size p', and $C_{\mathrm{p}}'^{e}(\cdot)$ beliefs about the distribution of consumption among members of the post-rebellion elite. The notation $'$ is used to denote an aspect of the institutions that would be created following a successful rebellion, with the superscript ^{e} denoting beliefs about these. It is not possible for particular individuals taking part in the rebellion to be credibly promised particular levels of consumption from the distribution $C_{\mathrm{p}}'^{e}(\cdot)$, so the expected payoff $\mathcal{U}_{\mathrm{p}}'^{e}$ from belonging to the post-rebellion elite is the following simple average:[5]

$$\mathcal{U}'^{e}_{\mathrm{p}} = \frac{1}{p'^{e}} \int_{\mathcal{P}'} u(C'^{e}_{\mathrm{p}}(\jmath))\mathrm{d}\jmath, \qquad [3.4]$$

assuming this elite will itself avoid losing power through a rebellion (as will be confirmed later).

Consider an individual $\imath$ who would receive utility $\mathcal{U}(\imath)$ under the prevailing institutions $\mathscr{I} = \{\mathcal{P}, \mathcal{W}, \tau(\imath), C_{\mathrm{p}}(\imath)\}$. The maximum amount of fighting effort this individual would find it individually rational to exert (assuming he expects a place in the subsequent elite) is denoted by $F(\imath)$, and the set of individuals willing to exert positive amounts of effort conditional on receiving a place in the post-rebellion elite is $\mathcal{F}$:

$$F(\imath) = \mathcal{U}'^{e}_{\mathrm{p}} - \mathcal{U}(\imath), \quad \text{and} \ \ \mathcal{F} = \{\imath \in \Omega \mid F(\imath) \geq 0\}, \qquad [3.5]$$

where $\mathcal{U}(\imath)$ is evaluated under the current institutions $\mathscr{I}$.

Formally, a rebellion is entirely characterized by an *elite selection function* $\mathscr{E}' : [0,1] \rightarrow \mathscr{P}$, which determines the identities of those who would have a place in the new elite if the rebellion succeeds in destroying the current institutions (it places no other restriction on the subsequent institutions). This is a mapping from the size p' of the subsequent elite to the set of (measurable) subsets of Ω, denoted by $\mathscr{P}$. The function $\mathscr{E}'(\cdot)$ has the property that $\mathscr{E}'(p')$ is a set of measure p' for all p', and if $p'_1 \leq p'_2$ then $\mathscr{E}'(p'_1) \subseteq \mathscr{E}'(p'_2)$, so all those included in a smaller elite would also belong a larger elite.

Given an elite selection function $\mathscr{E}'(\cdot)$, the prevailing institutions $\mathscr{I} \equiv \{\mathcal{P}, \mathcal{W}, \tau(\imath), C_{\mathrm{p}}(\imath)\}$, and beliefs p'^{e} and $\mathcal{U}'^{e}_{\mathrm{p}}$ about the post-rebellion institutions, the *rebel army* $\mathcal{R}$ and the *incumbent army* $\mathcal{A}$ are the sets

$$\mathcal{R} = \mathscr{E}'(p'^{e}) \cap \mathcal{F}, \quad \text{and} \ \ \mathcal{A} = \mathcal{P} \backslash \mathscr{E}'(p'^{e}). \qquad [3.6]$$

The rebel army can include any individuals willing to exert a positive amount of fighting effort conditional on being members of the subsequent elite. This can include those who belong to the current elite, since individuals cannot commit themselves to defend the current institutions if they receive a credible opportunity to join a rebellion and obtain a higher payoff than the current institutions grant them. The incumbent army includes those individuals who are in power according to the current institutions and do not join the rebel army, and so would lose their position in the elite were the rebellion to succeed. Owing to the free-riding problems discussed earlier, individuals who are neither part of the current elite nor expect to have a

[5] If there is a non-degenerate distribution of payoffs, it is assumed payoffs are allocated to particular individuals in the elite by lottery. This assumption avoids the need to study hierarchies within the elite.

place in a subsequent elite play no part in the fighting.

Individual $\imath \in \mathcal{R}$ in the rebel army exerts the individually rational fighting effort $F(\imath)$ from [3.5], which translates into fighting strength $F(\imath)$. Those in the incumbent army have fighting strength δ each. The rebellion succeeds if and only if

$$\int_{\mathcal{R}} F(\imath)\mathrm{d}\imath > \int_{\mathcal{A}} \delta \mathrm{d}\imath, \qquad [3.7]$$

that is, if the fighting strength of the rebel army exceeds that of the incumbent army. For simplicity, the fighting strength of each army is linear in the fighting strength of its members, and the fighting strength of the rebels is equal to their fighting effort (which decreases utility linearly). There is no uncertainty about the amount of fighting effort exerted given beliefs about the post-rebellion institutions, nor about the outcome given the fighting strength of the armies. Thus the outcome of any conflict is non-stochastic.

This approach to modelling the threat of conflict allows for a simple representation of the constraints on institutions if they are to avoid rebellions, without accounting explicitly for the punches and sword thrusts. The parameter δ measures the fighting strength of an individual in power who defends the current institutions. Unlike those in rebel army, members of the incumbent army have this defensive fighting strength without needing to exert fighting effort.

There are two differences in the treatment of the incumbent army and the rebel army in [3.7], one essential, and one an inessential simplification. The simplification is that each individual currently in power who does not defect the rebels has fighting strength δ, and that this is inelastic with respect to fighting effort F. As will be seen, the current elite has a several margins along which it can design institutions to ensure it remains in power, such as varying the number of people in power, or offering transfers to those who might join a rebel army. Adding an extra margin of being able to increase the fighting strength of its members (the intensive margin) at some cost does not fundamentally change the nature of the problem. On the other hand, it is *essential* that there is an asymmetry between the mapping from fighting effort to fighting strength for the incumbents and for the rebels. If both were identical then the notion of being "in power" would be meaningless and thus all individuals would be treated symmetrically in equilibrium.

One interpretation of the parameter δ is that the individuals in power under the current institutions possess some defensive fortifications, such as a castle, which place them at an immediate fighting advantage over any rebels, assuming those doing the defending do not defect to the rebels (sabotaging the fortifications as they go). A broader interpretation is that any existing institutions feature a customary chain of authority, that is, the implementation of any rules in a society depends on individuals knowing from whom they are to take orders, in the expectation of punishment if they disobey. A rebellion must supplant one system of authority with another, and the rebels face a more severe coordination problem than those already in power. This is because they must convince enough people that they are the new source of authority, reaching a tipping point where people come to expect others to start obeying the rebels rather than the existing elite, who start from a privileged position in this respect. Defections from the elite, where those in power join the rebels, are thus helpful not just for the extra fighting effort, but also for the failure of these individuals to play their expected role in supporting the current institutions.[6]

Although the term *rebellion* has been used to describe the process of destroying the current institutions, the formal definition encompasses revolutions, coups d'état, suspensions of constitutions, as well as rebellions in the conventional sense of the term. This is because rebellion is rebellion against the current institutions, which comprise the rules allocating resources as well as the identities of the elite, and the participants in rebellions are not restricted to those outside the elite. The only difference between these different types of rebellion is in who the participants are in the rebel and incumbent "armies". The model is set up with one general notion of rebellion that nests all these cases.

For example, when the rebel army comprises only workers, the rebellion can be interpreted as a "popular uprising". When the rebel army is drawn solely from the elite, the rebellion can be interpreted as a "coup d'état". When the rebel army includes all those in the current elite, the interpretation is similar to a "suspension of the constitution". In cases where the rebel army includes a mixture of workers and members of the current elite, the interpretation could be a "revolution" that receives the backing of some insiders from the current regime.

3.4 Establishing institutions

New institutions can be created once a rebellion has swept away any existing institutions, or starting from a point where no institutions previously existed. The group of individuals who will be in power has the ability to create these new institutions, so it is assumed the institutions will be set up to maximize the payoff of the elite associated with them.

The new institutions will maximize the average of the payoffs $\mathcal{U}_{\mathrm{p}}(\imath)$ of those who will be in elite ($\imath \in \mathcal{P}$), through choosing taxes $\tau(\imath)$ on workers ($\imath \in \mathcal{W}$), consumption $C_{\mathrm{p}}(\imath)$ of members of the elite, and determining the status of each individual, that is, assigning each $\imath \in \Omega$ to one of the sets $\mathcal{P}$ or $\mathcal{W}$, the set of those in power and the set of workers, respectively.[7]

The only constraint that must be respected by the choice of new institutions is the elite

[6] It is also possible to interpret condition [3.7] in a world where no actual fighting takes place. Under this interpretation, the rebels must incur a sunk effort cost to demonstrate they have the strength and are sufficiently well-organized to overcome the physical defences of the incumbent and the coordination problems inherent in launching a rebellion. Once the remaining elite members see this tipping point is reached, they surrender without a fight.

[7] Moving away from the assumption that the elite maximizes the average payoff of its members would require modelling its hierarchy, which is beyond the scope of this paper. See Myerson (2008) for a model addressing that question.

selection function $\mathscr{E}(\cdot)$. This function determines the identities of those who are to be in the elite, *conditional on* the elite having size p. Formally, if p is the measure of the set $\mathcal{P}$ of those in power then it is required that $\mathcal{P} = \mathscr{E}(p)$. The elite selection function is predetermined and was set down at the time of the previous rebellion, or in the case of starting from a blank slate, was randomly drawn by nature. To reiterate, an earlier rebellion does *not* determine the size of the elite, nor the allocation of resources laid down by the institutions, but does determine the *identities* of those who will hold power, conditional on the elite being of a given size.

Taking account of the elite selection function reduces the choice of institutions $\mathscr{I} = \{\mathcal{P}, \mathcal{W}.\tau(\imath), C_{\mathrm{p}}(\imath)\}$ to a choice of elite size p, taxes $\tau(\imath)$, and elite consumption levels $C_{\mathrm{p}}(\imath)$, with the sets $\mathcal{P}$ and $\mathcal{W}$ determined by

$$\mathcal{P} = \mathscr{E}(p), \quad \text{and} \;\; \mathcal{W} = \Omega \backslash \mathscr{E}(p). \tag{3.8}$$

If the institutions survive then members of elite receive continuation payoffs $\mathcal{U}_{\mathrm{p}}(\imath) = u(C_{\mathrm{p}}(\imath))$ since any past fighting effort F is sunk. However, once new institutions are created, there are opportunities for rebellion as described earlier. Thus the elite maximizes the average payoff of its members subject to no group having an incentive to launch a successful rebellion. As described in section 3.3, given the current institutions and beliefs about institutions after a rebellion, a particular rebellion is entirely characterized by a new elite selection function $\mathscr{E}'(\cdot)$. This means that the elite maximizes subject to the successful rebellion condition [3.7] not holding for *any* new elite selection function $\mathscr{E}'(\cdot)$, where the corresponding incumbent and rebel armies would be determined according to [3.5] and [3.6].

The incentives to rebel and the size of the rebel army depend on beliefs about the post-rebellion elite, its size p'^e and average payoff $\mathcal{U}'^e_{\mathrm{p}}$. These beliefs are taken as given for now, both by the rebels themselves and the current elite. The total fighting effort exerted by those in the rebel army is $\mathcal{U}'^e_{\mathrm{p}} - \mathcal{U}(\imath)$ summed (where positive) over the set of those ordered in the first p'^e places in the post-rebellion elite, determined by the new elite selection function $\mathscr{E}'(\cdot)$. The total fighting effort exerted by the incumbent army is δ multiplied by the measure p of individuals in the elite minus those who would also be in the post-rebellion elite and for whom there is an incentive to defect: $\mathcal{U}'_{\mathrm{p}} > \mathcal{U}(\imath)$. Therefore, the new institutions solve the following constrained maximization problem:

$$\max_{p,\tau(\imath),C_{\mathrm{p}}(\imath)} \frac{1}{p}\int_{\mathscr{E}(p)} \mathcal{U}_{\mathrm{p}}(\imath)\mathrm{d}\imath \;\text{ s.t. } \int_{\mathscr{E}'(p'^e)} \max\{\mathcal{U}'^e_{\mathrm{p}} - \mathcal{U}(\imath), 0\}\mathrm{d}\imath \leq \delta p - \int_{\mathscr{E}(p)\cap\mathscr{E}'(p'^e)} \delta \mathbb{1}[\mathcal{U}'_{\mathrm{p}} \geq \mathcal{U}(\imath)]\mathrm{d}\imath, \tag{3.9}$$

for all $\mathscr{E}'(\cdot)$, and subject to the budget constraint [3.3].

The new institutions maximize the average payoff of the elite without honouring any past commitments other than the elite selection function $\mathscr{E}(\cdot)$. This is consistent with the assumption that there are no means of enforcing commitments when there is an ex post gain from deviation. Under the assumption that all individuals are ex ante identical, the maximized elite payoff in [3.9] is the same for all elite selection functions $\mathscr{E}(\cdot)$, thus it is not possible to increase the average elite payoff by deviating from the elite selection function after a successful rebellion. The intuition is that an individual member of the elite does not care about the identities of those individuals with whom he shares power, only the total number of such people and the policies laid down by the institutions.

A consequence of this feature of the model is that there is no non-arbitrary way of choosing the identities of those inside the elite as the solution to a maximization problem over some function of the distribution of elite payoffs. The elite selection function thus resolves an essential indeterminacy regarding the identities of those who will fill certain positions in society, which provides an argument for its credibility as an incentive for individuals to participate in a rebellion. This credibility derives not from an ability of individuals to enter into binding commitments, but from the absence of any incentive to deviate from it ex post. Hence the elite selection function works as a coordination device, not as a commitment device.

While the elite selection function is an important state variable so far as individuals are concerned, it is an irrelevant state variable in relation to the problem of determining new institutions. Furthermore, while the size of the rebel army that triumphed over the previous incumbent will be equal to the size of the new elite in equilibrium, the elite's size is not constrained by the nature of the rebellion that brought it to power.[8] The maximization problem [3.9] is thus purely *forward looking* in depending only on beliefs p'^{e} and $\mathcal{U}_{\mathrm{p}}'^{e}$ about the future elite size and average elite payoff. These beliefs do influence the size of the rebel army and its incentive to fight. Determining equilibrium institutions then requires endogenizing these beliefs.

3.5 Equilibrium

In the absence of any aggregate uncertainty, beliefs about subsequent institutions must coincide with outcomes, hence $p'^{e} = p'$ and $\mathcal{U}_{\mathrm{p}}'^{e} = \mathcal{U}_{\mathrm{p}}'$. Determining beliefs then simply requires solving the maximization problem of the post-rebellion elite. However, this elite would also be subject to threats of rebellion, so its constrained maximization problem is of an identical form to that in [3.9], with p' and $\mathcal{U}_{\mathrm{p}}'$ now being determined as functions of p'' and $\mathcal{U}_{\mathrm{p}}''$, the beliefs about the institutions that would be set up following a successful rebellion against it. This shifts the original problem to one of determining the post-post-rebellion beliefs, and so on recursively, *ad infinitum*.

Nonetheless, at all stages of this sequence of (hypothetical) events, all elites are solving

[8]This implicitly assumes that 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. This would add a significant complication to the model without obviously delivering any new insights.

a maximization problem of exactly the same form, the only potential difference being beliefs about the actions of subsequent elites were they to come to power through rebellions. Given that individuals are ex ante identical, there is no fundamental reason for elites to make different choices regarding institutions (though of course the identities of those inside and outside the elite may change). Therefore it is natural to focus upon equilibria where outcomes (and hence beliefs) are functions only of the fundamentals, in other words, Markovian equilibria.

Formally, a Markovian equilibrium is a solution $\{p^*, \tau^*(\imath), C^*_\mathrm{p}(\imath)\}$ of the constrained maximization problem [3.9] where beliefs about the post-rebellion institutions $\{p', \tau'(\imath), C'_\mathrm{p}(\imath)\}$ coincide with the current institutions (up to a permutation of identities): $p^* = p'$; an identical distribution of taxes on workers: $\mathbb{P}[\tau^*(\imath) \leq \uptau] = \mathbb{P}[\tau'(\imath) \leq \uptau]$ for all $\uptau$; and an identical distribution of consumption received by members of the elite: $\mathbb{P}[C^*_\mathrm{p}(\imath) \leq \mathrm{C}] = \mathbb{P}[C'_\mathrm{p}(\imath) \leq \mathrm{C}]$ for all C. The following result demonstrates some features of any Markovian 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* $\imath$ *(with measure one)*

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

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

(iv) *The equilibrium institutions can be characterized by restricting the maximization problem in* [3.9] *subject only to a single "no-rebellion" constraint:*

$$\mathcal{U}'_\mathrm{p} - \mathcal{U}_\mathrm{w} \leq \updelta \frac{p}{p'}. \qquad [3.10]$$

(v) *A Markovian equilibrium always exists and is unique. The equilibrium elite size* p^* *is positive and bounded above by* $2 - \upvarphi$*, where* $\upvarphi \equiv (1+\sqrt{5})/2$ *is the Golden ratio.*[9] *The condition* $\updelta/\mathrm{q}u'(0) \leq 1$ *is sufficient for an interior solution where the non-negativity constraints of workers are slack, while the condition* $\updelta/\mathrm{q}u'(0) \geq \upvarphi$ *is sufficient for a corner solution where the non-negativity constraints bind. The necessary and sufficient condition for an interior solution is* $\updelta/\mathrm{q}u'(0) < 1 + \mathrm{q}/u^{-1}(u(0)+\updelta)$.

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, equalizing payoffs within

[9] The *Golden ratio*, or the *mean of Phidias*, is a well-known mathematical constant that appears in a number of contexts in pure mathematics and the natural sciences, and which some also claim has aesthetic properties that explain its occurrence in the arts and architecture. Algebraically, the Golden ratio is the dominant eigenvalue of the Fibonacci sequence. Geometrically, it is the ratio of the longest side to the shortest side of a rectangle where the sum of the two is in the same proportion to the longest side.

the set of workers and the set of members of the elite. These results do not depend on the utility function being strictly concave in consumption: the elite retains a strict preference for within-group equality even when utility is linear.

The intuition for the payoff-equalization results is that the composition of the rebel army most dangerous to the elite is the one which includes the individuals who have the greatest incentive to fight. Since a rebel army will always be a subset of the whole population, the rebel army that is willing to exert the greatest combined fighting effort will not include those workers who receive a relatively high payoff when there is inequality among workers. If this were the case, the elite could then reduce the maximum fighting effort it would face from rebels by redistributing from relatively well-off workers to those worse off, allowing it to achieve a higher payoff. Basically, the elite's tax policy should maximize the utility of the subset of workers with the minimum utility, which requires equalizing the payoffs of all workers.[10]

An analogous argument implies that heterogeneity in elite payoffs is undesirable from the standpoint of maximizing the average elite payoff. If some members of the elite received a payoff below the average then these individuals would be willing to defect from the incumbent army and fight alongside the rebels. Redistribution from the better-off members of the elite to the less well-off members does not directly lead to a lower average payoff (and would increase it in the case utility is concave), while does reduce the relative strength of the most dangerous rebel army, thus allowing for a higher average elite payoff. Notice that because defections from the elite weaken the incumbent army in proportion to the power parameter δ, there is no version of this argument that calls for equalization of payoffs *between* workers and the elite.

The results in Proposition 1 provide the underpinning of the earlier claim that promises to reward fighting in the rebel army by payments made once a new elite was in power, or punishment (whether pecuniary or otherwise) for shirking, lack credibility. The new elite would have a strict preference to renege on commitments that led to inequality between individuals who had otherwise identical continuation utility (owing to fighting effort being a sunk cost). This argument does not apply where rebels receive membership of the new elite as a reward for fighting because being in power allows a higher payoff to be extracted, not as a result of a past promise, but from a credible threat of withdrawing fighting strength δ from the incumbent army.

The payoff equalization results in Proposition 1 also allow for a considerable simplification of the elite's maximization problem over institutions. 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

[10]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.

be found using the budget constraint [3.3]. Thus

$$\mathcal{U}_{\mathrm{w}} = u(C_{\mathrm{w}}) = u(\mathrm{q} - \tau), \quad \text{and } \mathcal{U}_{\mathrm{p}} = u(C_{\mathrm{p}}) = u\left(\frac{(1-p)\tau}{p}\right). \tag{3.11}$$

A more important simplification is to the set of constraints that must be satisfied to disincentivize all possible groups of rebels. If under the current institutions, all workers receive the same payoff, and all elite members receive the same payoff, all that matters for the composition of a rebel army is the fraction of its total numbers drawn from workers and the fraction drawn from the current elite. Denote the former by σ_{w} and the latter by σ_{p}. The equilibrium institutions that are the solution to [3.9] in a Markovian equilibrium are then the solution of the simpler constrained maximization problem

$$\max_{p,\tau} \mathcal{U}_{\mathrm{p}} \text{ s.t. } \sigma_{\mathrm{w}} \max\{\mathcal{U}'_{\mathrm{p}} - \mathcal{U}_{\mathrm{w}}, 0\} + \sigma_{\mathrm{p}}(\mathcal{U}'_{\mathrm{p}} - \mathcal{U}_{\mathrm{p}} + \delta)\mathbb{1}[\mathcal{U}'_{\mathrm{p}} > \mathcal{U}_{\mathrm{p}}] \leq \delta \frac{p}{p'}, \tag{3.12}$$

for all 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, with $\mathcal{U}_{\mathrm{w}}$ and $\mathcal{U}_{\mathrm{p}}$ as given in [3.11], and beliefs p' and $\mathcal{U}'_{\mathrm{p}}$ determined in accordance with Markovian equilibrium.

Proposition 1 allows a further simplification of [3.12]. The fourth claim states that the equilibrium institutions can be characterized by a single "no-rebellion" constraint [3.10]. This constraint is equivalent to setting $\sigma_{\mathrm{w}} = 1$ and $\sigma_{\mathrm{p}} = 0$ in the general constraint of [3.12] (and noting that $\mathcal{U}'_{\mathrm{p}}$ will exceed $\mathcal{U}_{\mathrm{w}}$ in a Markovian equilibrium). Thus, satisfaction of [3.10] is clearly necessary. The substantive content of Proposition 1 here is that this single constraint is also *sufficient* for the general constraint to hold in a Markovian equilibrium.

The idea is that the elite simply needs to avoid a "peasant revolt" since this will be the only binding "no-rebellion" constraint in a Markovian equilibrium. This finding is specific to the simple endowment economy model of this section; subsequent extensions of the model will find that other more interesting no-rebellion constraints become binding. Which composition or compositions of the rebel army are associated with binding no-rebellion constraints in equilibrium is of course endogenous and will depend on the environment being analysed.

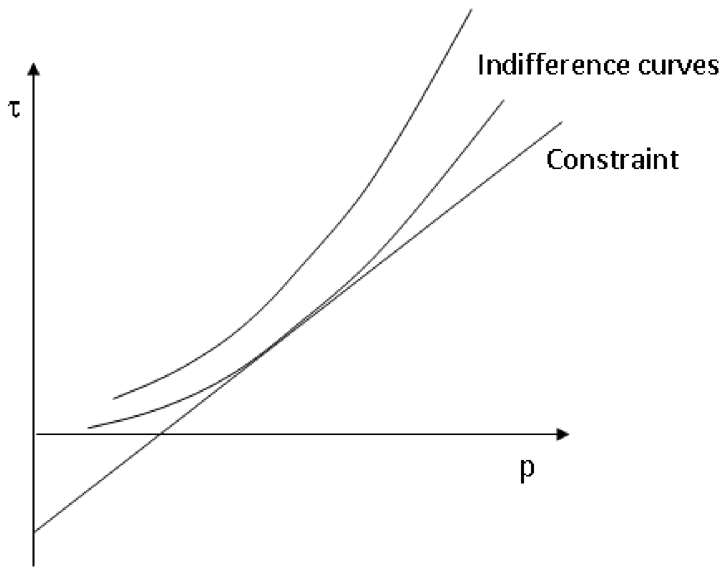
After all the simplifications justified by Proposition 1, the maximization problem characterizing the equilibrium institutions is

$$\max_{p,\tau} \mathcal{U}_{\mathrm{p}} \text{ s.t. } \mathcal{U}'_{\mathrm{p}} - \mathcal{U}_{\mathrm{w}} \leq \delta \frac{p}{p'}, \tag{3.13}$$

where $\mathcal{U}_{\mathrm{w}}$ and $\mathcal{U}_{\mathrm{p}}$ are as given in [3.11], and beliefs p' and $\mathcal{U}'_{\mathrm{p}}$, though taken as given in the maximization problem, are equal to the corresponding values p^* and $\mathcal{U}^*_{\mathrm{p}}$ that solve the maximization problem.

As seen from [3.11], the payoff of those in the elite is increasing in the tax τ and decreasing in the size of the elite p because fewer workers are available to tax and the total tax revenue must be distributed more widely. The resulting indifference curves of the elite over τ and p are plotted in Figure 2 (the elite payoff is increasing in the direction of the top-left corner of the diagram).

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



An increase in τ reduces the payoff of workers, making them more willing to fight in a rebellion, while an increase in the size of the elite increases the fighting strength of the incumbent army in the event of rebellion, making rebellion less attractive for 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 an upward-sloping "no-rebellion" constraint as depicted in Figure 2 (with points avoiding rebellion lying below the constraint). If the elite maximizes its payoff subject to remaining in power then the no-rebellion constraint binds and the maximum is at the tangency point. With utility linear in consumption, the no-rebellion constraint is a straight line, as shown in the diagram. With utility strictly concave in consumption, the constraint is a concave function of p.

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 basic trade-off in determining its optimal size. On the one hand, a larger size will strengthen the elite and allow higher taxes to be extracted from workers while still avoiding rebellion. On the other hand, a large size will spread the proceeds of these taxes more thinly among a larger number of individuals (and also reduce the tax base). Proposition 1 shows it is not possible in equilibrium to add extra individuals to the elite to augment its power while at the same time not 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, so the allocation of power 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 depend only on the single exogenous power parameter δ.

3.6 Example: linear utility

There are three exogenous parameters in the model: the power parameter δ, the endowment q of a worker, and the utility function $u(C)$ over consumption. This section illustrates the workings of the model for a linear utility function. With $u(C) = C$, the maximization problem [3.13] of the elite becomes

$$\max_{p,\tau} \frac{(1-p)\tau}{p} \quad \text{s.t.} \quad C'_{\mathrm{p}} - (\mathrm{q} - \tau) \leq \delta \frac{p}{p'}, \tag{3.14}$$

after substituting the expressions for $\mathcal{U}_{\mathrm{p}}$ and $\mathcal{U}_{\mathrm{w}}$ from [3.11]. The single no-rebellion constraint is binding, and can be used to solve explicitly for tax $\tau = \mathrm{q} - C'_{\mathrm{p}} + \delta p/p'$. Substituting this tax level into the objective function yields:

$$C_{\mathrm{p}} = \frac{1-p}{p}\left(\mathrm{q} - C'_{\mathrm{p}} + \delta \frac{p}{p'}\right). \tag{3.15a}$$

This is now an unconstrained maximization problem in p with beliefs p' and C'_{p} taken as given. The first-order condition is

$$\frac{C^*_{\mathrm{p}}}{1-p^*} = (1-p^*)\frac{\delta}{p'}. \tag{3.15b}$$

Now the Markovian equilibrium conditions are imposed ($p^* = p'$, $C^*_{\mathrm{p}} = C'_{\mathrm{p}}$) in [3.15a]:

$$C^*_{\mathrm{p}} = (\mathrm{q} + \delta)(1-p^*). \tag{3.15c}$$

Combining equations [3.15b] and [3.15c] (imposing $p^* = p'$ again) yields the Markovian equilibrium:

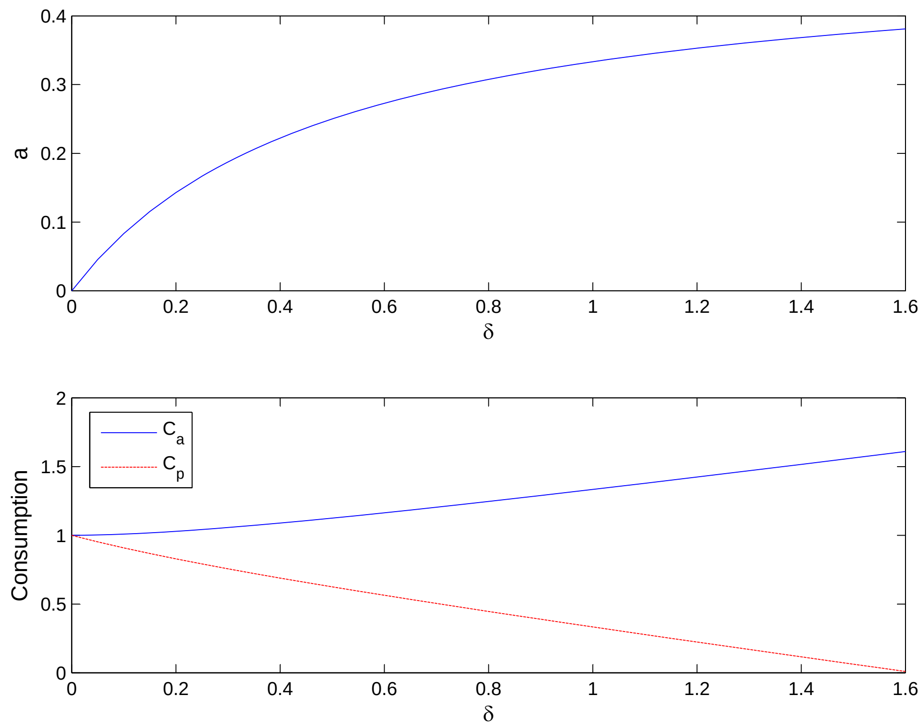
$$p^* = \frac{\delta}{\mathrm{q} + 2\delta}, \quad C^*_{\mathrm{p}} = \frac{(\mathrm{q}+\delta)^2}{\mathrm{q} + 2\delta}, \quad \text{and} \quad C^*_{\mathrm{w}} = \frac{(\mathrm{q}+\delta)^2}{\mathrm{q} + 2\delta} - \delta. \tag{3.16}$$

Notice in this case that the size of the elite is a function of the ratio δ/q.[11] The relationship between the power parameter δ and the endogenous variables of the model is shown in Figure 3 for $\mathrm{q} = 1$.

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 it. This

[11]For the linear utility function, the parameter restriction $\delta/\mathrm{q} \leq \varphi$ (where φ is the Golden ratio) is necessary and sufficient to obtain an equilibrium in which the non-negativity constraint for workers is not binding.

Figure 3: *The case of linear utility*



"income effect" leads to an increase in τ and a decrease in p. Second, the payoff that the rebels will receive once in power increases as their position would also be stronger once they have supplanted the current elite, making rebellion more attractive. This effect makes the position of the elite weaker, leading to an offsetting "income effect" that 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 workers' consumption remains positive, the third effect dominates and the size of the elite is increasing in δ.

4 Public goods

In the previous section there was no scope for the elite to do what governments are customarily thought to do, such as the provision of public goods. This section introduces a technology that allows for production of public goods. It is then natural to ask whether such public goods would be provided, since unlikely atomistic individuals, the elite can set up institutions that determine spending on public goods together with the taxes to finance such spending. The question is then whether a political Coase theorem will arise in this setting.

The new technology converts units of output into public goods. If g units of goods per capita are converted using the technology then everyone receives an extra $\Gamma(g)$ units of the consumption good. The model is otherwise identical to that of section 3. The utility of an individual is now

$$\mathcal{U} = u(C + \Gamma(g)) - F. \tag{4.1}$$

Given the aggregate resource constraint, per capita consumption is

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

A benevolent social planner would choose g such that

$$\Gamma_g(\hat{g}) = 1, \tag{4.3}$$

to maximize the total amount of goods available for consumption. Note that the choice of $\hat{g}$ is independent of p.

The definition of institutions $\mathscr{I}$ from section 3.2 is now modified to specify the provision g of public goods: $\mathscr{I} = \{\mathcal{P}, \mathcal{W}, \tau(\imath), C_{\mathrm{p}}(\imath), g\}$. All individuals observe the choice of g and take it into account when deciding whether to participate in a rebellion and how much fighting effort to exert (just as with all other dimensions of the institutions).

The insights of Proposition 1 continue to apply in this new environment when attention is restricted to Markovian equilibria, so it is possible without loss of generality to focus on institutions specifying the size of the elite p, the tax τ levied on workers, and the necessarily common public-good provision g. Under a particular set of institutions, the consumption of a worker is

$$C_{\mathrm{w}} = \mathrm{q} - \tau + \Gamma(g), \tag{4.4}$$

and using the budget constraint [4.2], the consumption of a member of the elite is

$$C_{\mathrm{p}} = \frac{(1-p)\tau - g}{p} + \Gamma(g). \tag{4.5}$$

The argument of Proposition 1 that there is a single binding "no-rebellion" constraint sufficient to characterize the equilibrium institutions also carries over to this new environment. As in section 3, this no-rebellion constraint is for the case where the rebel army comprises only workers. With these results in hand, the equilibrium institutions are the solution of the following constrained maximization problem

$$\max_{p,\tau,g} \; u\left(\frac{(1-p)\tau - g}{p} + \Gamma(g)\right) \;\; \text{s.t.} \;\; \mathcal{U}'_{\mathrm{p}} - u\left(q - \tau + \Gamma(g)\right) \le \delta \frac{p}{p'}, \tag{4.6}$$

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

$$\frac{u_C(C_\mathrm{p}^*)}{\lambda^* u_C(C_\mathrm{w}^*)}\left(\frac{1-p^*}{p^*}\right) = 1, \quad \text{and} \quad \frac{u_C(C_\mathrm{p}^*)}{\lambda^* u_C(C_\mathrm{w}^*)}\left(\frac{1}{p^*} - \Gamma_g(g^*)\right) = \Gamma_g(g^*). \qquad [4.7]$$

By eliminating the term $u_C(C_\mathrm{p}^*)/(\lambda^* u_C(C_\mathrm{w}^*))$ from the equations above, public-good provision g^* under the equilibrium institutions is determined by

$$\Gamma_g(g^*) = 1. \qquad [4.8]$$

This is identical to the equation [4.3] determining the public-good provision $\hat{g}$ of a benevolent social planner, so $g^* = \hat{g}$. The equilibrium institutions deliver efficient public-good provision.

The distribution of total output between workers and the elite depends on the other parameters of the model, including the utility function $u(C)$. In general, the possibility of providing public goods will benefit everyone. The elite is extracting rents from workers, but this does not preclude it from acting as if it were benevolent in other contexts. An implication is that 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 through the threat of conflict. This reflects the ambivalent effects of having a ruling elite on ordinary people.[12]

The no-rebellion constraint implies that the elite cannot disregard the interests of the workers. Provision of public goods slackens the "no-rebellion" constraint, while the taxes raised to finance them tighten the constraint. By optimally trading off the benefits of the public good against the cost of provision, the elite effectively maximizes the size of the pie, making use of transfers to ensure everyone is indifferent between rebelling or not. By not rebelling, those outside the elite essentially acquiesce to the "offer" made by the elite, analogous to the contracting that underlies the regular Coase theorem.

The result is far from surprising and can be obtained in several other settings. This is discussed by Persson and Tabellini (2000) in the context of voting and elections. Here the result provides a benchmark where a political Coase theorem holds.

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 realized. During

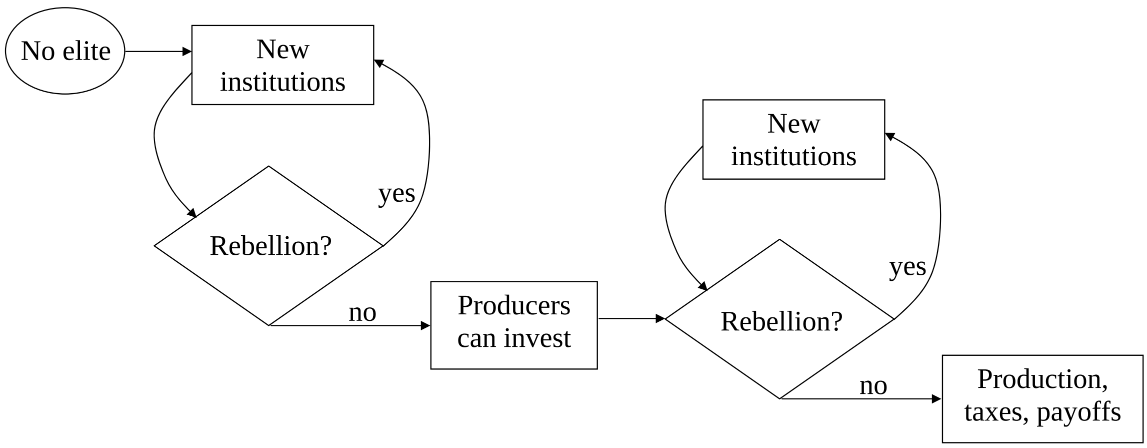
[12] This trade-off is present in the Bible, in the book of Samuel 8:10–20. People want a king to provide them public goods, despite being warned by prophet Samuel that the king would use his power on his own interest. Many centuries later, in far too many cases, the warnings of Samuel remain as relevant as ever.

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 the elite's design of institutions. The following analysis considers to what extent the equilibrium institutions will provide incentives for individuals to invest, and whether these institutions are efficient, that is, consistent with a political Coase theorem.

5.1 Environment

The sequence of events is depicted in Figure 4. Before any investment decisions are made, institutions are first established through a process identical to that described in section 3 (compare Figure 1). Institutions now specify the tax that will be levied on holdings of capital in addition to determining who holds power and other taxes and transfers. Once institutions have been established, 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 trigger any further rebellion, endowments are received, transfers are made, and goods are consumed.

Figure 4: *Sequence of events in model with investment*



Individuals who are in power (members of the elite, denoted by $\imath \in \mathcal{P}$) have fighting strength δ in the event of conflict, 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* that are randomly distributed among the economically active individuals at the investment stage. 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 sunk by the time the capital is produced. Capital here simply means more units of the consumption good. For simplicity, there is no intensive margin to investment: an individual receives at most one opportunity of a fixed size. An individual's effort cost θ is a random draw from the distribution

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

where $0 < \psi < \kappa$.[13] The receipt of an investment opportunity and whether it is taken are private information, as is the individual's required effort cost θ, while possession of capital is common knowledge. Moreover, whether an opportunity will be received, and the corresponding θ value if so, are not known to the individual himself before the investment stage.[14]

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

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

where $I \in \{0, 1\}$ indicates whether an investment opportunity is received and taken, and F denotes any fighting effort, as in the model of section 3.[15]

For analytical tractability, agents' preferences are assumed to be linear in consumption:

$$u(C) = C.$$

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. As in the model of section 3, the number of individuals holding positions of power is limited to less than 50%. The following parameter restrictions are also imposed:

$$\frac{\delta}{q} \leq \varphi \equiv \frac{1+\sqrt{5}}{2} \approx 1.618, \quad \mu \leq \frac{q}{2(q+2\delta)}, \quad \text{and} \;\; \kappa < \delta. \qquad [5.3]$$

The first restriction is the usual Golden-ratio condition from section 3 necessary to ensure non-negativity constraints do not bind in equilibrium. The second restriction states that the measure μ of individuals who receive an investment opportunity is not too large, which ensures

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

[14] 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 constraint is slack for those individuals outside the elite at the pre-investment stage, so this assumption need not significantly affect the results.

[15] 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.

that capitalists are not the predominant group by numbers.[16] The third restriction places a physical limit on the economy's maximum capital stock.

5.2 Equilibrium institutions

Characterizing the equilibrium institutions requires working backwards, starting from the post-investment stage and determining the equilibrium institutions if a rebellion were to occur at that point, and then analysing what institutions will be chosen by the elite 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 in the first place. Since holdings of capital are common knowledge, an argument similar to Proposition 1 shows that both of these individuals must necessarily secure the same payoff through their potential participation in a rebellion. The institutions chosen by the elite in the unique Markovian equilibrium would therefore equalize continuation payoffs for economically active individuals outside the elite ($\imath \in \mathcal{N}$), and also equalize payoffs within the elite ($\imath \in \mathcal{P}$). 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.

The total amount of capital held by all individuals is

$$K = \int_{\Omega} K(\imath) \mathrm{d}\imath.$$

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

$$pC_{\mathrm{p}} = (1-p)\tau_{\mathrm{q}} + \int_{\Omega} \tau_{\kappa}(\imath) \mathrm{d}\imath = (1-p)\tau_{\mathrm{q}} + K,$$

where C_{p} is the equal consumption of each member of the elite ($\imath \in \mathcal{P}$). The elite has size p,

[16] If the parameter restrictions in [5.3] do not hold, the nature of the binding constraints might change and the problem becomes 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.

with the $1 - p$ economically active agents outside the elite ($\imath \in \mathcal{N}$) receiving utility

$$\mathcal{U}_{\mathrm{n}} = C_{\mathrm{n}} = (\mathrm{q} - \tau_{\mathrm{q}}) + (K(\imath) - \tau_{\kappa}(\imath)) = \mathrm{q} - \tau_{\mathrm{q}}.$$

Combining this with the budget constraint yields an expression for the utility of the elite $\mathcal{U}_{\mathrm{p}} = C_{\mathrm{p}}$ in terms of the utility $\mathcal{U}_{\mathrm{n}}$ of those outside the elite:

$$\mathcal{U}_{\mathrm{p}} = \frac{(1-p)(\mathrm{q} - \mathcal{U}_{\mathrm{n}}) + K}{p}. \qquad [5.4]$$

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. 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 included no relevant state variables). 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:

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

The equilibrium elite size $p^{\dagger}$ is independent of K and is the same as that found in the endowment model with linear utility from section 3.6.[17] 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 presence of capital increases incentives for rebellions, which leads the elite to distribute the expropriated capital among all individuals.

5.2.2 Pre-investment stage institutions

Institutions chosen at this stage specify the set of individuals in power, denoted as usual by $\mathcal{P}$, and the set of economically active individuals outside the elite, denoted by $\mathcal{N}$. Depending on the institutions, some of the individuals in $\mathcal{N}$ will become *investors*, denoted by $\mathcal{I}$. Those economically active individuals who do not become investors (and so only receive the endowment q) are referred to as *workers*, denoted by $\mathcal{W}$. Since all aspects of investment opportunities are

[17] This analytically convenient finding is owing to the linearity of utility in consumption. It does not substantively affect the results that follow.

private information to those who receive them, institutions specifying a command economy where individuals must undertake investments by decree are infeasible. Instead, individuals choose voluntarily whether to invest based on beliefs about how much of any investment proceeds they will be able to keep, and the effort cost of investing.

It is shown formally in Proposition 2 below that the elite's payoff is strictly lower when there 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.[18] 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.

The identities of the individuals in the elite are determined by the usual elite selection function given the elite size p, and the identities of those who become investors depend on the arrival of investment opportunities and the realization of the effort costs. Once the pre-investment stage institutions survive rebellion, those economically active individuals who receive an investment opportunity must decide whether to take it. The elite will want to design institutions that also survive rebellion at the post-investment stage, so the incentive to take an investment opportunity is assessed under the assumption that the capital tax τ_κ specified by the current institutions will be implemented.

If an investment opportunity is not received or not taken then the individual becomes a worker ($\imath \in \mathcal{W}$) and obtains utility

$$\mathcal{U}_\mathrm{w} = C_\mathrm{w} = \mathrm{q} - \tau_\mathrm{q}, \qquad [5.6]$$

assuming (correctly in equilibrium) that this individual will not want to participate in a rebellion at the post-investment stage. Any economically active individual $\imath$ holding capital $K(\imath) = \kappa$ also receives the endowment q, but now pays total tax $\tau_\mathrm{q} + \tau_\kappa$. Thus, all investors have consumption $C_\mathrm{i} = \mathrm{q} + \kappa - \tau_\mathrm{q} - \tau_\kappa$. If an individual receives an investment opportunity

[18] The arguments relating to the distribution of consumption among investors work differently from those for workers or elite members. As will be seen, providing incentives to investors requires giving them greater consumption than workers. This means that investors' continuation payoffs at the post-investment stage will be higher than those of workers. Since these individuals have the same fighting strength (investors are outside the elite), investors will not be included in the rebel army with the greatest fighting strength. As a result, there is no equivalent result stating that ex-post inequality among investors' payoffs reduces the payoff of the elite because the utility function $u(\cdot)$ is linear. 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, or indeed any effects on 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 small, but would not change any of the results regarding the aggregate amount of investment in equilibrium.

with effort cost θ then the utility received from taking it ($\imath \in \mathcal{I}$) is

$$\mathcal{U}_{\mathrm{i}}(\theta) = C_{\mathrm{i}} - \theta = (\mathrm{q} - \tau_{\mathrm{q}}) + (\kappa - \tau_{\kappa}) - \theta. \quad [5.7]$$

An individual invests if it is individually rational to do so given beliefs about what capital tax will prevail after investment. The condition $\mathcal{U}_{\mathrm{i}}(\theta) \geq \mathcal{U}_{\mathrm{w}}$ is equivalent to the effort cost θ being not more than a threshold $\tilde{\theta}$, where $\mathcal{U}_{\mathrm{i}}(\tilde{\theta}) = \mathcal{U}_{\mathrm{w}}$:

$$\tilde{\theta} = \kappa - \tau_{\kappa}. \quad [5.8]$$

The proportion of those receiving an investment opportunity who take it is denoted by s. 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. It follows that the total measure i of investors and the total capital stock K are

$$i = \mu s, \quad \text{where} \;\; s = \mathbb{P}_{\theta}[\theta \leq \tilde{\theta}], \quad \text{and} \;\; K = i\kappa. \quad [5.9]$$

Given the uniform distribution of the effort cost θ between ψ and κ, the relationship between the fraction s taking the opportunity and the threshold $\tilde{\theta}$ is

$$s = \frac{\tilde{\theta} - \psi}{\kappa - \psi}. \quad [5.10]$$

For institutions chosen at the pre-investment stage to prevail, they must 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 aggregate environment that occurs as a result of investment (capital is a state variable), and also because individuals' incentives to rebel are altered by the information revealed to them at the investment stage and the choices they make at that point.

At the post-investment stage, there are three groups of individuals: investors ($\mathcal{I}$) who have already incurred the sunk effort cost, workers ($\mathcal{W}$), and those in power ($\mathcal{P}$). To distinguish the payoffs of investors before and after the effort cost is incurred, the later are referred to as *capitalists*, having continuation payoff $\mathcal{U}_{\mathrm{k}}$:

$$\mathcal{U}_{\mathrm{k}} = C_{\mathrm{i}} = (\mathrm{q} - \tau_{\mathrm{q}}) + (\kappa - \tau_{\kappa}). \quad [5.11]$$

Continuation utility is the same for all capitalists because there is no heterogeneity in capital held or in capital taxes that need to be paid. Workers receive the payoff $\mathcal{U}_{\mathrm{w}}$ given in equation

[5.6]. The payoff of a member of the elite $\imath \in \mathcal{P}$ is derived from the budget constraint:

$$\mathcal{U}_{\mathrm{p}} = C_{\mathrm{p}} = \frac{(1-p)\tau_{\mathrm{q}} + i\tau_{\kappa}}{p}. \qquad [5.12]$$

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 individuals: those in power ($\mathcal{P}$), and those economically active individuals ($\mathcal{N}$) outside the elite (who do not yet know whether they will become workers or investors). Let $\mathcal{U}_{\mathrm{n}}$ 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:

$$\mathcal{U}_{\mathrm{n}} = (1-\alpha)\mathcal{U}_{\mathrm{w}} + \alpha\mathbb{E}_{\theta}\max\{\mathcal{U}_{\mathrm{i}}(\theta), \mathcal{U}_{\mathrm{w}}\}, \quad \text{where } \alpha = \frac{\mu}{1-p}. \qquad [5.13]$$

The formula for α is valid because the measure of economically active individuals $1-p$ always exceeds the measure of investment opportunities μ. The payoff can be written in terms of the expected surplus $\mathcal{S}_{\mathrm{i}}(\tilde{\theta})$ of those receiving an investment opportunity:

$$\mathcal{U}_{\mathrm{n}} = (\mathrm{q} - \tau_{\mathrm{q}}) + \alpha\mathcal{S}_{\mathrm{i}}(\tilde{\theta}), \quad \text{where } \mathcal{S}_{\mathrm{i}}(\tilde{\theta}) \equiv \mathbb{E}_{\theta}\max\{\tilde{\theta} - \theta, 0\}. \qquad [5.14]$$

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 have already been characterized as a function of the aggregate capital stock K, which is the only relevant state variable. The subsequent elite would be of size $p^{\dagger}$ and receive utility $\mathcal{U}_{\mathrm{p}}^{\dagger}(K)$ as given in equation [5.5], with the capital stock K predetermined according to [5.9]. At the pre-investment stage, there will be beliefs p' and $\mathcal{U}_{\mathrm{p}}'$ 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 Markovian equilibrium restriction (noting that there are no relevant state variables at the pre-investment stage).

The general problem of the elite's problem of choosing institutions is thus

$$\max_{p,\tau_{\mathrm{q}},\tau_{\kappa}} \mathcal{U}_{\mathrm{p}} \;\text{ s.t. }\; \sigma_{\mathrm{n}}\max\{\mathcal{U}_{\mathrm{p}}' - \mathcal{U}_{\mathrm{n}}, 0\} + \sigma_{\mathrm{p}}\max\{\mathcal{U}_{\mathrm{p}}' - \mathcal{U}_{\mathrm{p}}, 0\} \leq \delta\left(\frac{p}{p'} - \sigma_p\right) \;\text{ and } \qquad [5.15a]$$

$$\sigma_{\mathrm{w}}^{\dagger}\max\{\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \mathcal{U}_{\mathrm{n}}, 0\} + \sigma_{\mathrm{p}}^{\dagger}\max\{\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \mathcal{U}_{\mathrm{p}}, 0\} + \sigma_{\mathrm{i}}^{\dagger}\max\{\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \mathcal{U}_{\mathrm{k}}, 0\} \leq \delta\left(\frac{p}{p^{\dagger}} - \sigma_{\mathrm{p}}^{\dagger}\right). \qquad [5.15b]$$

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

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

Proposition 2 *Consider an arbitrary choice of the capital tax τ_{κ}, which determines $\tilde{\theta}$ and s according to [5.8] and [5.10]. 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 for $(\sigma^{\dagger}_{\mathrm{w}}, \sigma^{\dagger}_{\mathrm{p}}, \sigma^{\dagger}_{\mathrm{i}}) = (1, 0, 0)$ and $(\sigma^{\dagger}_{\mathrm{w}}, \sigma^{\dagger}_{\mathrm{p}}, \sigma^{\dagger}_{\mathrm{i}}) = (0, 1, 0)$.

(iii) The binding no-rebellion constraints imply that

$$p = p^{\dagger} + \frac{\mu\tilde{\theta}s}{\delta}, \qquad [5.16a]$$

so sustaining investment requires a larger elite.

(iv) After expanding the elite size in accordance with [5.16a], the payoff of a member of the elite is

$$\mathcal{U}_{\mathrm{p}} = \frac{(q+\delta)^2}{q+2\delta} + \mu\left(\kappa - \left(\frac{q+2\delta}{\delta}\right)\tilde{\theta}\right)s. \qquad [5.16b]$$

Proof See appendix A.2. ■

The proposition shows that no investor belongs to a rebel army 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 investment 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], [5.8] and [5.11]:

$$\mathcal{U}_{\mathrm{k}} = \mathcal{U}_{\mathrm{w}} + \tilde{\theta},$$

where $\tilde{\theta} > 0$ when $s > 0$. The analysis of section 5.2.2 shows that the institutions following a rebellion at the post-investment stage will not respect individual holdings of capital prior to the rebellion, again because the effort cost of investing is sunk. Thus, what investors stand to receive following rebellion (net of fighting costs) is no different from that of workers (their power is identical), while what they lose is superior. Accordingly, they are less willing to fight to replace the current institutions. This means that 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 a rebellion.

The fundamental problem here is that the distribution of income needed to support investment incentives diverges from the distribution of income consistent with the distribution of power, and so there are incentives for groups to rebel against institutions incentivizing investment, thereby bringing the distribution of income into line with the distribution of power. 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 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. On the one hand, the elite would like to design institutions encouraging investment by not taxing capital too heavily, but on the other hand, there is 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.[19]

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 the elite, a rebellion backed by all members of the elite succeeds without requiring any fighting effort. This means that the elite can destroy the current institutions through a "suspension of the constitution", allowing new institutions to be created while leaving the current elite members in power, essentially granting the elite full discretion to rewrite completely the "rules of the game" ex post. If the elite size p were equal to $p^\dagger$, where $p^\dagger$ is the equilibrium elite size characterized in section 5.2.1 following a rebellion after investments are made, this would certainly be true. However, matters are different in the case $p > p^\dagger$ where 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, therefore the willingness of individuals to exercise their discretion to rebel might be curtailed.

The formal analysis in Proposition 2 confirms that satisfaction of the no-rebellion constraints for workers and elite members is equivalent to ensuring the elite size p is large in relation to $p^\dagger$. As the number of investors s rises, the required elite size p increases. The choice of capital tax τ_κ (which determines s via equations [5.8] and [5.10]) can be interpreted

[19] The no-rebellion constraint for the elite places a lower bound on $\mathcal{U}_\mathrm{p}$ 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 prevailing institutions through rebellion ex post.

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 [5.16a] is then that credible limitations on expropriation require an elite size that is sufficiently larger 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 the elite size *sufficient* for credible protection of property rights; it is also *necessary*. That is, there is no other design of institutions which 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.[20]

Sharing power among a wider group thus allows the elite to act as a government committed to policies that would otherwise be time inconsistent. Even though all individuals act 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. This way, changes to the status quo face powerful opponents.[21]

This analysis rests on two key assumptions made earlier about institutions and the power struggle. First, institutions lay down the rules of the game followed by everyone, the sole exception being when they are defied by a resumption of the power struggle. For institutions to function in supporting cooperation among individuals who have chosen to stop fighting, it is essential that particular groups cannot arbitrarily modify aspects of those institutions while managing at the same time to avoid calling into question the whole structure.[22] This assumption is necessary for the sheer existence of an environment where rules are followed.

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

[21]The extra members of the elite are in no way intrinsically different from the existing members and have no access to any technology directly protecting property rights. Note also that by assumption, power is not shared with those who are actually investing. This is a simplifying assumption, but it does ensure that this is not responsible for the results obtained in Proposition 2. Furthermore, it far from clear that including investors in the elite, while keeping the overall elite size constant, would actually provide credible incentives. Increasing the power of capitalists increases their ability to extract rents (including from other capitalists), rather than just defend their own individual property.

[22]While there are no aggregate shocks in the model here, what institutions prescribe could be made state contingent in more general settings. So modification of institutions means something other than those institutions simply reacting to the realizations of exogenous shocks.

It captures the idea that elites can in principle create institutions that specify, for example, known tax rates and known levels of public-goods provision (to which individuals acquiesce when they decide not to rebel). Second, were a rebellion to occur, there is no means of enforcing commitments made prior to the rebellion. The underlying idea is that institutions can support deals between individuals, but there are no “meta-institutions” to enforce deals concerning the choice of the 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.[23] In particular, following a rebellion after investment decisions have been made, the degree of power sharing is reoptimized as well as the tax system.[24]

As mentioned above, economic development ultimately requires rewarding the productive rather than just the strong. For this to happen, institutions must credibly protect the property rights of investors. It is an endogenous feature of the model that institutions with more power sharing can achieve that goal. Since the elite has the freedom to set up institutions as it wishes, does it have an incentive to build institutions conducive to investment?

5.3 The equilibrium and efficient choices of capital taxes

The elite has a choice of three institutional variables p, τ_{q} and τ_{κ}. The two binding no-rebellion constraints at the post-investment stage can be used to eliminate two of these variables. Proposition 2 also shows that no pre-investment stage constraints are binding: groups would always have an incentive to defer rebellion until after investments have already been made. This leaves one degree of freedom to maximize the elite payoff $\mathcal{U}_{\mathrm{p}}$. The third part of Proposition 2 yields an expression for $\mathcal{U}_{\mathrm{p}}^*$ as a function of s and $\tilde{\theta}$. Using the relationship between $\tilde{\theta}$ and s implied by [5.10] and differentiating the expression in [5.16b] yields the equilibrium value of s^*:

$$s^* = \max\left\{0, \frac{\delta\kappa - (q+2\delta)\psi}{2(q+2\delta)(\kappa-\psi)}\right\}. \qquad [5.17]$$

As confirmed below in Proposition 3, this solution satisfies all individual rationality and non-negativity constraints, in addition to all the no-rebellion constraints, so this is the unique Markovian equilibrium of the model. But is this an efficient level of investment?

Since θ lies in the range $[\psi, \kappa]$, it is easy to see that investors’ surplus $\mathcal{S}_{\mathrm{i}}(\tilde{\theta})$ from [5.14] is maximized when $\hat{\tilde{\theta}} = \kappa$. This is the first-best level of investment. From [5.8] and [5.10], it follows that $\hat{\tau}_{\kappa} = 0$ and $\hat{s} = 1$, so capital taxes should be zero and all individuals receiving

[23] This does not of course mean that all features of the old institutions *necessarily* change, but that the elite cannot choose *not* to reconsider all aspects of the institutions.

[24] If it were possible for the rebels to commit to restrict optimization following a rebellion to certain areas this might paradoxically make 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 limitations on expropriation of private property on the second page. Being able to rebel against page two, but at the same time being able somehow to commit not to touch page one, would annihilate the credibility of the second page.

investment opportunities should take them. From a public finance perspective in a world where the government needs to raise a particular amount of revenue, since lump-sum taxes and transfers are available, this is also the capital tax that would be chosen optimally. It is clear from [5.17] that $s^* < 1$, so equilibrium investment always falls short of the first-best level.

However, the first best may not be the most interesting welfare benchmark. As discussed earlier, the model predicts that power sharing is required for protection of property rights to survive the power struggle. An increase in power sharing requires a larger elite, which 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 opportunity cost of adding more individuals to the elite, compared to the hypothetical first-best world where there is no need to provide credible incentives, and hence no cost of investment other than the direct effort cost?

To address this question, consider the following notion of *constrained efficiency*. Suppose that it were possible exogenously to impose some level of the capital tax τ_κ on all possible institutions could be chosen by elites. However, all other dimensions of institutions would be chosen by elites to maximize their own payoffs subject to surviving the power struggle as before. The constrained efficient level of the capital tax is what would be chosen by a benevolent agent, taking into account the constraints imposed by the power struggle. The benevolent agent would then appreciate that more investment requires greater protection of property rights, and thus a larger elite if this is to survive the power struggle. The concept of constrained efficiency then requires setting the benefit of more investment against the resource cost of the larger elite.[25] In section 4, a benevolent agent could not create a better economic outcome with a different choice of public-good provision from what prevails in equilibrium.[26] Here, the question is whether the equilibrium protection of property rights coincides with the constrained efficient level.

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

$$\bar{\mathcal{U}} \equiv \int_{\Omega} \mathcal{U}(\imath)\mathrm{d}\imath. \tag{5.18}$$

All aspects of institutions other than τ_κ are determined in equilibrium as before, and Proposition 2 continues to apply with the value of s resulting from the benevolent agent's choice of τ_κ, using equations [5.8] and [5.10] as usual. The elite will choose its size p, which together with s determines the number of investors $i = \mu s$ (all of whom have $\theta \leq \tilde{\theta}$) and workers

[25] If there were no resource cost in increasing the size of the elite then the notion of constrained efficiency would coincide with the first best.

[26] In the model of section 4, the notions of first-best and constrained efficient are equivalent because it is assumed there that the elite can use the public-good technology without affecting any other aspect of the environment.

$w = 1 - p - i$. Writing average utility in terms of the payoffs of each of these groups:

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

For the marginal investor with effort cost $\tilde{\theta}$, $\mathcal{U}_{\mathrm{i}}(\tilde{\theta}) = \mathcal{U}_{\mathrm{w}}$. Average utility can then be rewritten in terms of the investors' surplus $\mathcal{S}_{\mathrm{i}}(\tilde{\theta})$ from [5.14]:

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

Proposition 2 implies that the only two binding constraints are for workers and members of the elite at the post-investment stage. A consequence of these being the two binding constraints is that worker and elite payoffs are tied together by $\mathcal{U}_{\mathrm{w}} = \mathcal{U}_{\mathrm{p}} - \delta$. Since the benevolent agent takes these constraints into account, this relationship is substituted into the expression for $\bar{\mathcal{U}}$:

$$\bar{\mathcal{U}} = \mathcal{U}_{\mathrm{p}} - \delta(1 - p) + \mu\mathcal{S}_{\mathrm{i}}(\tilde{\theta}). \qquad [5.19]$$

There are two differences between the expressions for $\bar{\mathcal{U}}$ and $\mathcal{U}_{\mathrm{p}}$. The second term in [5.19] is related to the distribution of resources in the economy, and the third term reflects the investors' surplus.

Proposition 3 *(i) The unique Markovian equilibrium s^* is given by the expression in* [5.17].

(ii) The Markovian equilibrium capital tax τ_{κ}^* exceeds the tax that maximizes total capital tax revenue, and so lies to the right of the peak of the Laffer curve.

(iii) The value of s that maximizes $\bar{\mathcal{U}}$ from [5.19] *in the constrained efficient problem, denoted by $s^{\diamond}$, satisfies:*

$$s^{\diamond} \leq \max\left\{0, \frac{\delta\kappa - (\mathrm{q} + \delta)\psi}{(2\mathrm{q} + \delta)(\kappa - \psi)}\right\}. \qquad [5.20]$$

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

Proof See appendix A.3. ■

The constrained efficient choice of s leads to more investment than the equilibrium choice for two reasons. The first (and more interesting) distortion follows from the distributional effects of protecting against expropriation (the second term in [5.19]). Protection of property rights requires sharing power. Sharing power requires sharing rents, because a rebellions including members of the current elite have an advantage in any fighting. Thus while a larger elite might allow for profitable investment and higher output, it also reduces the fraction of output that can be appropriated by a member of the elite.

The association between power and rents creates an endogenous limit on the set of possible transfers between individuals, leading to a breakdown of the political Coase theorem. The elite could implement the constrained-optimal allocation, which requires a larger p, but the constraints imposed by the power struggle mean that they would have to offer the extra elite members rents. Therefore, the cost to elite members of expanding their number is not simply the lost output from diverting individuals from directly productive activities.

Second, the equilibrium choice of capital taxes does not take into account investors' surpluses (the third term in [5.19]). As the effort cost θ is not public information, it is impossible for the elite to extract rents from those who invested, at the margin. Consequently, the no-rebellion constraint for those who invested is slack, so the elite obtains no benefit from marginal increases in investors' payoffs. A political Coase theorem does not hold here because of the non-observability of effort, θ, which makes it impossible to engineer transfers contingent on effort.

The inefficiently high capital taxes can be interpreted as insufficient protection against expropriation of property. Recent empirical work has highlighted the importance of institutions that prevent the government from expropriating individuals' resources. But why would expropriation of property be so susceptible to political failures? Why would a political Coase theorem fail to apply in this particular case? The model here sheds light on this question.

Lastly, a note on the effect of the power parameter δ. In section 4, the welfare of workers in an economy with institutions and a self-interested elite might be larger than the welfare of workers in an economy where no transfers could be enforced, and hence with no public-good provision. The possibility of establishing institutions allowed for public-good provision in equilibrium. However, a larger δ could only be harmful to workers since it allows for 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 cannot be any investment in equilibrium. If δ is low enough, no protection of property rights is even constrained efficient because it would require such a large increase in the size of the elite, which has a large opportunity cost. A larger δ parameter makes it easier for the elite to remain in power, which directly benefits them, but might also allow them partly to offer protection of property rights.

5.4 Analogies with historical examples

The results show the importance of sharing power to stimulate investment and prevent changes in the rules of the game that lead to expropriation. They also show that rulers will not share power as much as would be efficient. Although the model is too abstract to match any given historical case precisely, the results resonate with a number of episodes.

Broadly speaking, the extra individuals in power required to protect 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 from outsiders, and also especially insiders. Parliaments are usually thought of as representing those who elected their members, but so are democratically elected presidents, and the presence of large numbers of members of parliaments is not only useful in 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 power is distributed and on the limits to taxation. Once power is 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 accepted that power would be shared. North and Weingast (1989) argue that the Glorious Revolution led to secure property rights and elimination of confiscatory government. This allowed the English government to borrow much more, and at substantially lower rates. This was certainly in the interest of the king, yet the 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 pie becomes larger, with a lesser share, the amount received by the king might end up being smaller.

Malmendier (2009) studies the Roman *societas publicanorum*, perhaps the earliest precursors of the modern business corporation. 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 property rights of the *societas publicanorum*, 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.

Greif (2006) analyses the importance of the *podesteria* system for interclan cooperation in medieval Italy. The *podestà*, an individual coming from another city who would be in power for a year, was generously paid, and played an important role in the development of cities such as Genoa by allowing for cooperation and investment. Interestingly, Greif argues that the *podestà* had to be sufficiently strong because otherwise, if one clan had defeated its rival, it could easily defeat the *podestà* as well. In the language of the model, the strength of the *podestà* (the difference between p^* and $p^\dagger$) had to be large enough to ensure he could not be defeated by a clan, which would then be able to change the institutions (Greif, 2006, pp.

239–240).

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 different 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 under analysis. A unified framework that can be used to analyse 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, under the same rules.

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 mechanism. 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 large elite is endogenously formed. But the same conflict mechanism that explains how power sharing can overcome the commitment problem also implies that sharing power entails sharing rents. This imposes endogenous limits on the set of possible transfers and leads to a breakdown of the "political" version of the Coase theorem. In equilibrium, there is too little power sharing, and thus not enough institutional stability to offer investors the protection from expropriation that would support the efficient level of investment.

References

Acemoglu, D. (2003), "Why not a political Coase theorem? Social conflict, commitment, and politics", *Journal of Comparative Economics*, **31(4)**:620–652. 1, 8

——— (2008), "Oligarchic versus democratic societies", *Journal of the European Economic Association*, **6(1)**:1–44 (March). 8

Acemoglu, D., Egorov, G. and Sonin, K. (2008), "Coalition formation in non-democracies", *Review of Economic Studies*, **75(4)**:987–1009 (October). 7

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

Acemoglu, D. and Robinson, J. A. (2006), *Economic Origins of Dictatorship and Democracy*, Cambridge University Press. 6

——— (2008), "Persistence of power, elites, and institutions", *American Economic Review*, **98(1)**:267–293 (March). 6

Baron, D. P. and Ferejohn, J. A. (1989), "Bargaining in legislatures", *American Political Science Review*, **83(4)**:1181–1206 (December). 7

Basu, K. (2000), *Prelude to Political Economy*, Oxford University Press. 7

Besley, T. and Persson, T. (2009a), "The origins of state capacity: Property rights, taxation, and politics", *American Economic Review*, **99(4)**:1218–1244 (September). 7

——— (2009b), "Repression or civil war?", *American Economic Review*, **99(2)**:292–297 (May). 7

——— (2010), "State capacity, conflict and development", *Econometrica*, forthcoming. 7

Downs, A. (1957), "An economic theory of political action in a democracy", *Journal of Political Economy*, **65(2)**:135–150 (April). 6

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. 8

Glaeser, E., Scheinkman, J. and Shleifer, A. (2003), "The injustice of inequality", *Journal of Monetary Economics*, **50(1)**:199–222 (January). 8

Greif, A. (2006), *Institutions and the Path to the Modern Economy*, Cambridge University Press. 6, 39

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). 8

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). 8

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). 8

Levy, G. (2004), "A model of political parties", *Journal of Economic Theory*, **115(2)**:250–277 (April). 7

Mailath, G., Morris, S. and Postlewaite, A. (2001), "Laws and authority", working paper, Princeton University. 7

Malmendier, U. (2009), "Law and finance at the origin", *Journal of Economic Literature*, **47(4)**:1076–1108. 39

Myerson, R. B. (1993), "Incentives to cultivate favored minorities under alternative electoral rules", *American Political Science Review*, **87(4)**:856–869 (December). 18

——— (2004), "Justice, institutions and multiple equilibria", *Chicago Journal of International Law*, **5(1)**:91–107. 7

——— (2008), "The autocrat's credibility problem and foundations of the constitutional state", *American Political Science Review*, **102(1)**:125–139. 14

——— (2009), "Learning from Schelling's *Strategy of Conflict*", *Journal of Economic Literature*, **47(4)**:1109–1125. 7

——— (2010), "Capitalist investment and political liberalization", *Theoretical Economics*, **5(1)**:73–91. 8

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, 39

Persson, T. and Tabellini, G. (2000), *Political Economics: Explaining Economic Policy*, MIT Press. 6, 8, 24

Piccione, M. and Razin, R. (2009), "Coalition formation under power relations", *Theoretical Economics*, **4(1)**:1–15 (March). 7

Ray, D. (2007), *A Game-Theoretic Perspective on Coalition Formation*, Oxford University Press. 7

Schelling, T. C. (1960), *The Strategy of Conflict*, Harvard University Press. 7

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. This means that such institutions maximize $\bar{\mathcal{U}}_{\mathrm{p}}$ subject to the general no-rebellion constraint (with $\mathbb{1}[\cdot]$ being the indicator function):

$$\begin{aligned}&\int_{\mathcal{W}\cap\mathscr{E}'(p')} \max\{\mathcal{U}'_{\mathrm{p}} - \mathcal{U}_{\mathrm{w}}(\imath), 0\}\mathrm{d}\imath + \int_{\mathcal{P}\cap\mathscr{E}'(p')} \max\{\mathcal{U}'_{\mathrm{p}} - \mathcal{U}_{\mathrm{p}}(\imath), 0\}\mathrm{d}\imath \\ \leq &\int_{\mathcal{P}\cap\mathscr{E}'(p')} \delta\mathbb{1}[\mathcal{U}_{\mathrm{p}}(\imath) > \mathcal{U}'_{\mathrm{p}}]\mathrm{d}\imath + \int_{\mathcal{P}\setminus\mathscr{E}'(p')} \delta\mathbb{1}[\mathcal{U}_{\mathrm{p}}(\imath) \geq \mathcal{U}'_{\mathrm{w}}]\mathrm{d}\imath \ \text{ for all } \mathscr{E}'(\cdot),\end{aligned} \tag{A.1.1}$$

and the conditions $p = p' = p^*$, $\bar{\mathcal{U}}_{\mathrm{p}} = \mathcal{U}'_{\mathrm{p}} = \bar{\mathcal{U}}^*_{\mathrm{p}}$, and $\bar{\mathcal{U}}_{\mathrm{w}} = \mathcal{U}'_{\mathrm{w}} = \bar{\mathcal{U}}^*_{\mathrm{w}}$. The terms $\bar{\mathcal{U}}_{\mathrm{w}}$ and $\bar{\mathcal{U}}_{\mathrm{p}}$ are the average utilities of workers and current elite members:

$$\bar{\mathcal{U}}_{\mathrm{w}} \equiv \frac{1}{|\mathcal{W}|}\int_{\mathcal{W}} \mathcal{U}_{\mathrm{w}}(\imath)\mathrm{d}\imath, \quad \text{and } \bar{\mathcal{U}}_{\mathrm{p}} \equiv \frac{1}{|\mathcal{P}|}\int_{\mathcal{P}} \mathcal{U}_{\mathrm{p}}(\imath)\mathrm{d}\imath, \tag{A.1.2}$$

with $|\cdot|$ denoting the measure of a set.

Observe that if $\mathcal{U}_{\mathrm{w}}(\imath) > \mathcal{U}'_{\mathrm{p}}$ for a positive mass 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.1.1] held for all $\mathscr{E}'(\cdot)$ then it continues to do so after this deviation, so the deviation increasing $\bar{\mathcal{U}}_{\mathrm{p}}$ is feasible, so the original institutions are not an equilibrium. All equilibrium institutions must therefore feature $\mathcal{U}_{\mathrm{w}}(\imath) \leq \mathcal{U}'_{\mathrm{p}}$ for all $\imath \in \mathcal{W}$. Consequently, $\bar{\mathcal{U}}_{\mathrm{w}} \leq \mathcal{U}'_{\mathrm{p}}$, so Markovian equilibria must feature $\bar{\mathcal{U}}^*_{\mathrm{w}} \leq \bar{\mathcal{U}}^*_{\mathrm{p}}$.

Next, consider the case where $\mathcal{U}_{\mathrm{p}}(\imath) < \mathcal{U}'_{\mathrm{w}}$ for a positive mass 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 not replacing them with others) would strictly increase $\bar{\mathcal{U}}_{\mathrm{p}}$. Suppose that the alternative institutions grant the expelled individuals the same consumption as when they were elite members. Since those outside the elite can produce, there are also extra resources to distribute. Supposing these are distributed among elite members, $\mathcal{U}_{\mathrm{p}}(\imath)$ is no lower for anyone who remains within the elite. It follows that the left-hand side of [A.1.1] is no higher for any $\mathscr{E}'(\cdot)$. 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.1.1] is unaffected for all $\mathscr{E}'(\cdot)$. The deviation raising $\bar{\mathcal{U}}_{\mathrm{p}}$ is feasible, so 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 $\mathcal{U}_{\mathrm{p}}(\imath) \geq \mathcal{U}'_{\mathrm{w}}$, the general no-rebellion constraint [A.1.1] is equivalent to

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

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 $\mathcal{E}_{\mathrm{p}} = \mathcal{P} \cap \mathscr{E}'(p')$. 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

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

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, that is, $\sigma = |\mathcal{E}_{\mathrm{w}}|/p'$, and also $1-\sigma = |\mathcal{E}_{\mathrm{p}}|/p'$. 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 $\overline{\sigma}$, defined as follows:

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

Now define the following functions of $\sigma \in [\underline{\sigma}, \overline{\sigma}]$:

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

The general no-rebellion constraint [A.1.4] imposes an upper bound on the sum of the effective rebellion strength of workers and current elite members for all compositions of the rebel army. This is therefore equivalent to asking whether the upper bound is satisfied by the maximum over the sets of possible rebel army compositions. Constraint [A.1.4] holds if and only if

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

with bounds $\underline{\sigma}$ and $\overline{\sigma}$ from [A.1.5]. This is in turn equivalent to the single consolidated constraint

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

The equilibrium institutions are those maximizing $\bar{\mathcal{U}}_{\mathrm{p}}$ subject to the above constraint.

Note that the constraint [A.1.8] must be binding. 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 Markovian equilibrium. Since $p' = p = p^*$, and $p^* < 1/2$, it follows from the expressions in [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 values of these functions evaluated at 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 mean $\bar{\mathcal{U}}^*_{\mathrm{w}}$. Since $\mathscr{F}^{*\ddagger} \geq \mathscr{F}^*(0)$, there are two possible cases to consider.

Suppose first that $\mathscr{F}^*(0) < \mathscr{F}^{*\ddagger}$. Let $\bar{C}_{\mathrm{w}}$ denote the level of consumption required to give a worker the average utility of all workers under the Markovian equilibrium, that is, $u(\bar{C}_{\mathrm{w}}) = \bar{\mathcal{U}}^*_{\mathrm{w}}$. Each worker's utility under the Markovian equilibrium institutions is $\mathcal{U}^*_{\mathrm{w}}(\imath) = u(C^*_{\mathrm{w}}(\imath))$. Given the weak concavity of the utility function $u(\cdot)$, Jensen's inequality implies

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

Given that the utility function $u(\cdot)$ is strictly increasing, this yields

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

Hence consider a deviation from the Markovian equilibrium institutions where every worker now receives utility $\mathcal{U}_{\mathrm{w}} = u(\bar{C}_{\mathrm{w}}) = \bar{\mathcal{U}}^*_{\mathrm{w}}$. The analysis above shows that this is feasible, and there may even be resources left over that can be distributed among the elite. With $C_{\mathrm{p}}(\imath) \geq C^*_{\mathrm{p}}(\imath)$, it must be the case that $\mathcal{U}_{\mathrm{p}}(\imath) \geq \mathcal{U}^*_{\mathrm{p}}(\imath)$ for all $\imath \in \mathcal{P}$. It can then be seen from the definition of $\mathscr{F}_{\mathrm{p}}(\sigma)$ in [A.1.6] that $\mathscr{F}_{\mathrm{p}}(\sigma) \leq \mathscr{F}^*_{\mathrm{p}}(\sigma)$ for all $\sigma \in [0,1]$.

Under the new institutions, $\mathcal{U}_{\mathrm{w}}(\imath) = \bar{\mathcal{U}}^*_{\mathrm{w}}$ for all $\imath \in \mathcal{W}$. It follows directly from the definition of $\mathscr{F}_{\mathrm{w}}(\sigma)$ in [A.1.6] that $\mathscr{F}_{\mathrm{w}}(\sigma) = \sigma p^*(\bar{\mathcal{U}}^*_{\mathrm{p}} - \bar{\mathcal{U}}^*_{\mathrm{w}})$ for all $\sigma \in [0,1]$. Now note that since $p^* < 1/2$, $\sigma p^* < 1 - p^* = |\mathcal{W}|$ for all $\sigma \in [0,1]$. In the case of payoff inequality among workers, it follows that

$$\min_{\substack{\mathcal{E}_{\mathrm{w}} \subseteq \mathcal{W} \\ |\mathcal{E}_{\mathrm{w}}| = \sigma p^*}} \int_{\mathcal{E}_{\mathrm{w}}} \mathcal{U}^*_{\mathrm{w}}(\imath)\mathrm{d}\imath < \sigma p^* \bar{\mathcal{U}}^*_{\mathrm{w}},$$

for all $\sigma \in (0,1]$, and hence

$$\max_{\substack{\mathcal{E}_{\mathrm{w}} \subseteq \mathcal{W} \\ |\mathcal{E}_{\mathrm{w}}| = \sigma p^*}} \int_{\mathcal{E}_{\mathrm{w}}} (\bar{\mathcal{U}}_{\mathrm{p}}^* - \mathcal{U}_{\mathrm{w}}^*(\imath)) \mathrm{d}\imath > \sigma p^* (\bar{\mathcal{U}}_{\mathrm{p}}^* - \bar{\mathcal{U}}_{\mathrm{w}}^*),$$

again for all $\sigma \in (0,1]$. In terms of the function $\mathscr{F}_{\mathrm{w}}(\sigma)$ from [A.1.6], $\mathscr{F}_{\mathrm{w}}(\sigma) < \mathscr{F}_{\mathrm{w}}^*(\sigma)$ for all $\sigma \in (0,1]$. Together with the earlier result $\mathscr{F}_{\mathrm{p}}(\sigma) \leq \mathscr{F}_{\mathrm{p}}^*(\sigma)$ for all $\sigma \in [0,1]$, it is established that for all $\sigma \in (0,1]$, $\mathscr{F}(\sigma) < \mathscr{F}^*(\sigma)$. 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, $\sigma^\ddagger > 0$. The analysis above has shown $\mathscr{F}(\sigma^\ddagger) < \mathscr{F}^*(\sigma^\ddagger)$ for any such $\sigma^\ddagger$, and therefore $\mathscr{F}^\ddagger < \mathscr{F}^{*\ddagger} = \delta p^*$. The no-rebellion constraint [A.1.9] is now slack, allowing taxes to be raised, strictly increasing $\bar{\mathcal{U}}_{\mathrm{p}}$.

The second case to consider is $\mathscr{F}(0) = \mathscr{F}^{*\ddagger}$. Observe from the definition of $\mathscr{F}_{\mathrm{w}}(\sigma)$ in [A.1.9] that $\mathscr{F}_{\mathrm{w}}(\sigma)/\sigma$ is weakly decreasing in σ, and $\mathscr{F}_{\mathrm{w}}(0) = 0$. This is equivalent to $\mathscr{F}_{\mathrm{w}}(\sigma)$ being concave in σ. Similarly, $\mathscr{F}_{\mathrm{p}}(\sigma)/(1-\sigma)$ is weakly decreasing in σ, and $\mathscr{F}_{\mathrm{p}}(1) = 0$, which implies $\mathscr{F}_{\mathrm{p}}(\sigma)$ is concave. Therefore, the sum $\mathscr{F}(\sigma) = \mathscr{F}_{\mathrm{w}}(\sigma) + \mathscr{F}_{\mathrm{p}}(\sigma)$ is also concave. The combination of $\mathscr{F}(\sigma) \leq \mathscr{F}(0)$ for all σ and $\mathscr{F}(\sigma)$ being concave implies $\mathscr{F}(\sigma)$ is weakly decreasing for all $\sigma \in [0,1]$.

If $\mathscr{F}_{\mathrm{w}}(\sigma) + \mathscr{F}_{\mathrm{p}}(\sigma)$ is weakly decreasing for all σ then the definitions of these functions in [A.1.6] imply

$$\begin{aligned} &\min_{\imath \in \mathcal{W}} \left\{ \bar{\mathcal{U}}_{\mathrm{p}}^* - \mathcal{U}_{\mathrm{w}}^*(\imath) \,\middle|\, |\{\jmath \in \mathcal{W} \,|\, \mathcal{U}_{\mathrm{w}}^*(\imath) > \mathcal{U}_{\mathrm{w}}^*(\jmath)\}| = 0 \right\} \\ \leq &\max_{\imath \in \mathcal{P}} \left\{ \mathbb{1}[\mathcal{U}_{\mathrm{p}}^*(\imath) \leq \bar{\mathcal{U}}_{\mathrm{p}}^*](\bar{\mathcal{U}}_{\mathrm{p}}^* - \mathcal{U}_{\mathrm{p}}^*(\imath) + \delta) \,\middle|\, |\{\jmath \in \mathcal{P} \,|\, \mathcal{U}_{\mathrm{p}}^*(\imath) < \mathcal{U}_{\mathrm{p}}^*(\jmath)\}| = 0 \right\}, \end{aligned} \qquad \text{[A.1.10]}$$

using the fact that with $\sigma = 0$, $\mathcal{E}_{\mathrm{p}} = \mathcal{P}$ satisfies the conditions $\mathcal{E}_{\mathrm{p}} \subseteq \mathcal{P}$ and $|\mathcal{E}_{\mathrm{p}}| = (1-\sigma)p^*$ since $p = p' = p^*$. The existence of payoff inequality among workers implies there must be a positive measure of workers $\imath \in \mathcal{W}$ such that

$$(\bar{\mathcal{U}}_{\mathrm{p}}^* - \mathcal{U}_{\mathrm{w}}^*(\imath)) < \min_{\imath \in \mathcal{W}} \left\{ \bar{\mathcal{U}}_{\mathrm{p}}^* - \mathcal{U}_{\mathrm{w}}^*(\imath) \,\middle|\, |\{\jmath \in \mathcal{W} \,|\, \mathcal{U}_{\mathrm{w}}^*(\imath) > \mathcal{U}_{\mathrm{w}}^*(\jmath)\}| = 0 \right\}.$$

Consider individual-specific tax increases on the workers satisfying the above inequality up to the point where all workers $\imath \in \mathcal{W}$ have a payoff $\mathcal{U}_{\mathrm{w}}(\imath)$ such that

$$(\bar{\mathcal{U}}_{\mathrm{p}}^* - \mathcal{U}_{\mathrm{w}}(\imath)) = \min_{\imath \in \mathcal{W}} \left\{ \bar{\mathcal{U}}_{\mathrm{p}}^* - \mathcal{U}_{\mathrm{w}}^*(\imath) \,\middle|\, |\{\jmath \in \mathcal{W} \,|\, \mathcal{U}_{\mathrm{w}}^*(\imath) > \mathcal{U}_{\mathrm{w}}^*(\jmath)\}| = 0 \right\}.$$

Given these new payoffs for workers, it follows from [A.1.10] that

$$(\bar{\mathcal{U}}_{\mathrm{p}}^* - \mathcal{U}_{\mathrm{w}}(\imath)) \leq \max_{\imath \in \mathcal{P}} \left\{ \mathbb{1}[\mathcal{U}_{\mathrm{p}}^*(\imath) \leq \bar{\mathcal{U}}_{\mathrm{p}}^*](\bar{\mathcal{U}}_{\mathrm{p}}^* - \mathcal{U}_{\mathrm{p}}^*(\imath) + \delta) \,\middle|\, |\{\jmath \in \mathcal{P} \,|\, \mathcal{U}_{\mathrm{p}}^*(\imath) < \mathcal{U}_{\mathrm{p}}^*(\jmath)\}| = 0 \right\},$$

for all $\imath \in \mathcal{W}$. This shows 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 is distributed among the elite, ensuring $\mathcal{U}_{\mathrm{p}}(\imath) \geq \mathcal{U}_{\mathrm{p}}^*(\imath)$ and thus $\mathscr{F}_{\mathrm{p}}(\sigma) \leq \mathscr{F}_{\mathrm{p}}^*(\sigma)$ for all $\sigma \in [0,1]$, as well as strictly increasing $\bar{\mathcal{U}}_{\mathrm{p}}$ given that taxes are increased by strictly positive amounts for a positive measure of workers.

Putting the above 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, demonstrating that it is feasible.

Therefore, payoff inequality for workers is not consistent with Markovian equilibrium. Any equilibrium must feature $\mathcal{U}_{\mathrm{w}}^*(\imath) = \bar{\mathcal{U}}_{\mathrm{w}}^* = \mathcal{U}_{\mathrm{w}}^*$.

Payoff equalization among members of the elite

Given the payoff equalization for workers established above, the function $\mathscr{F}_{\mathrm{w}}(\sigma)$ from [A.1.6] reduces to

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

Since $\mathscr{F}_{\mathrm{p}}^*(1) = 0$, it follows from feasibility of the institutions that $\mathscr{F}_{\mathrm{w}}^*(1) = \mathscr{F}^*(1) \leq \mathscr{F}^{*\ddagger} = \delta p^*$, and hence

$$\bar{\mathcal{U}}_{\mathrm{p}}^* - \mathcal{U}_{\mathrm{w}}^* \leq \delta. \qquad \text{[A.1.11]}$$

Now consider the possibility of an equilibrium with $\bar{\mathcal{U}}_{\mathrm{p}}^* - \mathcal{U}_{\mathrm{w}}^* < \delta$. Starting from such institutions, consider the following changes. First, redistribute consumption among elite members equally so that all elite members receive the average $\bar{C}_{\mathrm{p}}^*$ specified by the initial institutions. With $C_{\mathrm{p}}(\imath) = \bar{C}_{\mathrm{p}}^*$ and the concavity of the utility function $u(\cdot)$, Jensen's inequality implies

$$\mathcal{U}_{\mathrm{p}}(\imath) = u(\bar{C}_{\mathrm{p}}^*) = u\left(\frac{1}{|\mathcal{P}|}\int_{\mathcal{P}} C_{\mathrm{p}}^*(\imath)\mathrm{d}\imath\right) \geq \frac{1}{|\mathcal{P}|}\int_{\mathcal{P}} u(C_{\mathrm{p}}^*(\imath))\mathrm{d}\imath = \bar{\mathcal{U}}_{\mathrm{p}}^*. \qquad \text{[A.1.12]}$$

for all $\imath \in \mathcal{P}$. The second change is a tax increase imposed on all workers, one that is strictly positive, but which nonetheless ensures that $\bar{\mathcal{U}}_{\mathrm{p}}^* - \mathcal{U}_{\mathrm{w}} \leq \delta$ holds. The proceeds are distributed among members of the elite. In combination with the earlier redistribution of elite consumption that implied [A.1.12], it must be the case that $\mathcal{U}_{\mathrm{p}}(\imath) \geq \bar{\mathcal{U}}_{\mathrm{p}}^*$ for all $\imath \in \mathcal{P}$. Furthermore, the increase in total tax revenue implies that $\bar{\mathcal{U}}_{\mathrm{p}}$ is strictly larger.

After these changes, since $\mathcal{U}_{\mathrm{p}}(\imath) \geq \bar{\mathcal{U}}_{\mathrm{p}}^*$, it follows that

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

for all $\imath \in \mathcal{P}$, since in a Markovian equilibrium, $\mathcal{U}_{\mathrm{p}}' = \bar{\mathcal{U}}_{\mathrm{p}}^*$. As a result, the function $\mathscr{F}_{\mathrm{p}}(\sigma)$ from [A.1.6] can be bounded above:

$$\mathscr{F}_{\mathrm{p}}(\sigma) \leq (1-\sigma)p^*\delta,$$

for all $\sigma \in [0,1]$. Following the tax increase, worker payoffs remain equalized, so $\mathscr{F}_{\mathrm{w}}(\sigma) = \sigma p^*(\bar{\mathcal{U}}_{\mathrm{p}}^* - \mathcal{U}_{\mathrm{w}})$ for all σ, with $\mathcal{U}_{\mathrm{p}}' = \bar{\mathcal{U}}_{\mathrm{p}}^*$ unchanged. Given that $\bar{\mathcal{U}}_{\mathrm{p}}^* - \mathcal{U}_{\mathrm{w}} \leq \delta$, therefore $\mathscr{F}_{\mathrm{w}}(\sigma) \leq \sigma p^*\delta$. Putting the bounds for $\mathscr{F}_{\mathrm{w}}(\sigma)$ and $\mathscr{F}_{\mathrm{p}}(\sigma)$ together:

$$\mathscr{F}(\sigma) = \mathscr{F}_{\mathrm{w}}(\sigma) + \mathscr{F}_{\mathrm{p}}(\sigma) \leq \sigma p^*\delta + (1-\sigma)p^*\delta = \delta p^*$$

The changed institutions thus satisfy the general no-rebellion constraint [A.1.9], but strictly increase $\bar{\mathcal{U}}_{\mathrm{p}}$. Therefore, it must be the case that any Markovian equilibrium features

$$\bar{\mathcal{U}}_{\mathrm{p}}^* - \mathcal{U}_{\mathrm{w}}^* = \delta. \qquad \text{[A.1.13]}$$

It has been shown that any Markovian equilibrium must feature payoff equalization for workers and satisfy equation [A.1.13]. Consider the possibility that payoffs are not equalized for members of the elite. This implies that there is a positive measure of elite members $\imath \in \mathcal{P}$ with payoffs $\mathcal{U}_{\mathrm{p}}^*(\imath)$ strictly below the average $\bar{\mathcal{U}}_{\mathrm{p}}^*$ (which must equal $\mathcal{U}_{\mathrm{p}}'$ in a Markovian equilibrium). For these members of the elite:

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

Given the definition of $\mathscr{F}_{\mathrm{p}}(\sigma)$ from [A.1.6], it must be the case for all $\sigma \geq \bar{\sigma}$ with $\bar{\sigma} < 1$ sufficiently close to 1:

$$\mathscr{F}_{\mathrm{p}}^*(\sigma) > (1-\sigma)p^*\delta. \qquad \text{[A.1.14]}$$

With payoff equalization for workers, $\mathcal{U}'_{\mathrm{p}} = \bar{\mathcal{U}}^*_{\mathrm{p}}$, and equation [A.1.13]:

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

Putting this together with the inequality from [A.1.14], it follows that for some $\sigma \in [0, 1]$:

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

This shows that $\mathscr{F}^{*\ddagger} > \delta p^*$, violating the constraint in [A.1.9]. Therefore, given the features of any Markovian equilibrium already established, payoff inequality among elite members is not feasible. Thus, elite members' payoffs must also be equalized in a Markovian equilibrium.

Power determines rents

Payoff equalization for elite members means that $\mathcal{U}^*_{\mathrm{p}}(\imath) = \bar{\mathcal{U}}^*_{\mathrm{p}} = \mathcal{U}^*_{\mathrm{p}}$ for all $\imath \in \mathcal{P}$. It follows from equation [A.1.13] that $\mathcal{U}^*_{\mathrm{p}} = \mathcal{U}^*_{\mathrm{w}} + \delta$.

Reduction to a single no-rebellion constraint

Given that any Markovian equilibrium must feature payoff equalization among workers and among members of the elite, the general no-rebellion constraint is

$$\sigma(\mathcal{U}'_{\mathrm{p}} - \mathcal{U}_{\mathrm{w}}) + (1 - \sigma)\mathbb{1}[\mathcal{U}_{\mathrm{p}} \leq \mathcal{U}'_{\mathrm{p}}](\mathcal{U}'_{\mathrm{p}} - \mathcal{U}_{\mathrm{p}} + \delta) \leq \delta \frac{p}{p'} \text{ for all } \sigma \in [\underline{\sigma}, 1], \qquad \text{[A.1.15]}$$

where the bound $\underline{\sigma}$ is defined in [A.1.5], and $\overline{\sigma} = 1$ given that $p < 1/2$. The choice of institutions reduces to a choice of p and τ. A Markovian equilibrium is a solution (p^*, τ^*) to the problem of maximizing $\mathcal{U}_{\mathrm{p}}$ subject to [A.1.15] (taking p' and $\mathcal{U}'_{\mathrm{p}}$ as given) and any other constraints (e.g. non-negativity constraints), but with $p' = p^*$ and $\mathcal{U}'_{\mathrm{p}} = \mathcal{U}^*_{\mathrm{p}}$ in equilibrium.

Now consider the problem of maximizing $\mathcal{U}_{\mathrm{p}}$ subject to [A.1.15] holding for only $\sigma = 1$, that is:

$$\mathcal{U}'_{\mathrm{p}} - \mathcal{U}_{\mathrm{p}} \leq \delta \frac{p}{p'}. \qquad \text{[A.1.16]}$$

A Markovian equilibrium of this alternative problem is defined similarly as a solution (p^*, τ^*) to the problem of maximizing $\mathcal{U}_{\mathrm{p}}$ subject to [A.1.16] (also taking p' and $\mathcal{U}'_{\mathrm{p}}$ as given) and any other constraints, but with $p' = p^*$ and $\mathcal{U}'_{\mathrm{p}} = \mathcal{U}^*_{\mathrm{p}}$.

Start by considering a Markovian equilibrium (p^*, τ^*) of problem [A.1.16]. Since this point must be feasible and $p' = p^*$ and $\mathcal{U}'_{\mathrm{p}} = \mathcal{U}^*_{\mathrm{p}}$, it must be the case that $\mathcal{U}^*_{\mathrm{p}} - \mathcal{U}^*_{\mathrm{w}} \leq \delta$. Hence for any $\sigma \in [0, 1]$:

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

This shows that (p^*, τ^*) is a feasible subject to the constraints in [A.1.15]. Now take any other feasible point (p, τ) satisfying [A.1.15] with $p' = p^*$ and $\mathcal{U}'_{\mathrm{p}} = \mathcal{U}^*_{\mathrm{p}}$. This alternative point must satisfy [A.1.15] for $\sigma = 1$ in particular, so [A.1.16] must hold. Given that (p^*, τ^*) maximizes $\mathcal{U}_{\mathrm{p}}$ over all feasible points according to [A.1.16], it must be the case that $\mathcal{U}_{\mathrm{p}} \leq \mathcal{U}^*_{\mathrm{p}}$ for this alternative choice. Therefore, (p^*, τ^*) maximizes over the set defined by [A.1.15], so it is also a Markovian equilibrium of this problem as well.

Now consider the converse. Take a Markovian equilibrium (p^*, τ^*) of the original problem [A.1.15]. This point is obviously feasible subject to [A.1.15], of which [A.1.16] is the special case $\sigma = 1$, so (p^*, τ^*) is feasible subject to [A.1.16]. Suppose for contradiction that (p^*, τ^*) is not a Markovian equilibrium of the problem [A.1.16]. Since it is feasible, it must therefore be the case that there exists another feasible point (p, τ) satisfying [A.1.16] such that $\mathcal{U}_\text{p} > \mathcal{U}_\text{p}^*$. Therefore, $\mathbb{1}[\mathcal{U}_\text{p} \leq \mathcal{U}_\text{p}^*](\mathcal{U}_\text{p}^* - \mathcal{U}_\text{p} + \delta) = 0$. Now take any $\sigma \in [\underline{\sigma}, 1]$ and multiply both sides of the inequality [A.1.16] by this number:

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

Hence, (p, τ) satisfies [A.1.15] for all $\sigma \in [\underline{\sigma}, 1]$, and is thus a feasible point in the original maximization problem. Therefore, $\mathcal{U}_\text{p} \leq \mathcal{U}_\text{p}^*$, which contradicts the inequality $\mathcal{U}_\text{p} > \mathcal{U}_\text{p}^*$ already obtained. This establishes that (p^*, τ^*) must then be a Markovian equilibrium of the problem with the sole constraint [A.1.16]

In summary, these arguments show that the set of Markovian equilibria subject to the full set of no-rebellion constraints in [A.1.15] is identical to the set of Markovian equilibrium subject only to the constraint [A.1.16].

Existence and uniqueness of the Markovian equilibrium

Given the earlier results, any Markovian equilibrium (p^*, τ^*) must be the solution to the maximization problem:

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

with p^* and $\mathcal{U}_\text{p}^* = u((1-p^*)\tau^*/p^*)$ taken as given. There are non-negativity constraints to satisfy for workers' and elite members' consumption, which imposes the constraint $0 \leq \tau \leq \text{q}$. The defensive strength of the elite is effective only up to $p < 1/2$. The utility function $u(\cdot)$ is strictly increasing, weakly concave, and differentiable.

In this maximization problem, the single no-rebellion constraint must bind, which provides an expression for τ as a function of p:

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

The problem is equivalent to maximizing $C_\text{p} = (1-p)\tau/p$ substituting for τ as above:

$$\max_{p} \frac{(1-p)}{p}\left(\text{q} - u^{-1}\left(\mathcal{U}_\text{p}^* - \delta\frac{p}{p^*}\right)\right).$$

Taking the derivative of C_p with respect to p:

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

The second derivative is

$$\frac{\partial^2 C_\text{p}}{\partial p^2} = -\frac{2}{u'\left(u^{-1}\left(\mathcal{U}_\text{p}^* - \delta\frac{p}{p^*}\right)\right)}\frac{\delta p}{p^*} + \frac{u''\left(u^{-1}\left(\mathcal{U}_\text{p}^* - \delta\frac{p}{p^*}\right)\right)}{\left\{u'\left(u^{-1}\left(\mathcal{U}_\text{p}^* - \delta\frac{p}{p^*}\right)\right)\right\}^3} - \frac{2}{p}\frac{\partial C_\text{p}}{\partial p},$$

which given the concavity of $u(\cdot)$ is strictly negative at all points where the first derivative is zero, ensuring the appropriate second-order condition is always satisfied.

In a Markovian equilibrium, $p = p^*$, so the binding no-rebellion constraint becomes:

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

or equivalently

$$u\left(\frac{(1-p^*)(\mathrm{q}-C_{\mathrm{w}}^*)}{p^*}\right)=\mathscr{U}_{\mathrm{p}}^*=u(C_{\mathrm{w}}^*)+\delta,$$

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

$$p^*u^{-1}(u(C_{\mathrm{w}}^*)+\delta)=(1-p^*)(\mathrm{q}-C_{\mathrm{w}}^*).$$

Evaluating the derivative of C_{p} at the Markovian equilibrium $p=p^*$:

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

Since $u^{-1}(\mathscr{U}_{\mathrm{p}}^*-\delta)=C_{\mathrm{w}}^*$, this can be simplified as follows:

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

In addition to the no-rebellion constraint, there is also the constraint $p<1/2$ and the non-negativity constraints, which require $0\leq C_{\mathrm{w}}\leq \mathrm{q}$ in terms of worker consumption C_{w}. Notice that the no-rebellion constraint implies that τ is increasing in p (holding p^* and $\mathscr{U}_{\mathrm{p}}^*$ constant). Therefore, the full set of cases for the first-order condition are:

$$\left.\frac{\partial C_{\mathrm{p}}}{\partial p}\right|_{p=p^*}\begin{cases}\leq 0 & \text{if } p^*=0 \text{ or } C_{\mathrm{w}}^*=\mathrm{q}\\ =0 & \text{if } 0<p^*<1/2 \text{ or } 0<C_{\mathrm{w}}^*<\mathrm{q}\\ \geq 0 & \text{if } p^*=1/2 \text{ or } C_{\mathrm{w}}^*=0\end{cases}.$$

Now define the functions:

$$\mathscr{G}_{\mathrm{c}}(p,C_{\mathrm{w}})\equiv pu^{-1}(u(C_{\mathrm{w}})+\delta)-(1-p)(\mathrm{q}-C_{\mathrm{w}}),\quad \mathscr{G}_{\mathrm{d}}(p,C_{\mathrm{w}})\equiv\delta(1-p)-(\mathrm{q}-C_{\mathrm{w}})u'(C_{\mathrm{w}}).$$

The no-rebellion constraint is equivalent to $\mathscr{G}_{\mathrm{c}}(p^*,C_{\mathrm{w}}^*)=0$. The other function is related to the derivative of C_{p}:

$$\left.\frac{\partial C_{\mathrm{p}}}{\partial p}\right|_{p=p^*}=\frac{1}{p^{*2}}\frac{1}{u'(C_{\mathrm{w}}^*)}\mathscr{G}_{\mathrm{d}}(p^*,C_{\mathrm{w}}^*).$$

Taking partial derivatives of the function $\mathscr{G}_{\mathrm{c}}(p,C_{\mathrm{w}})$:

$$\frac{\partial\mathscr{G}_{\mathrm{c}}}{\partial p}=u^{-1}(u(C_{\mathrm{w}})+\delta)+(\mathrm{q}-C_{\mathrm{w}}),\quad \frac{\partial\mathscr{G}_{\mathrm{c}}}{\partial C_{\mathrm{w}}}=p\frac{u'(C_{\mathrm{w}})}{u'\left(u^{-1}(u(C_{\mathrm{w}})+\delta)\right)}+(1-p).$$

Since $u(\cdot)$ is strictly increasing, $u'(C_{\mathrm{w}})>0$ and $u'(u^{-1}(u(C_{\mathrm{w}})+\delta))>0$, and also $u^{-1}(u(C_{\mathrm{w}})+\delta)>C_{\mathrm{w}}$. It follows that both of the above partial derivatives are strictly positive for all $0\leq p\leq 1$ and $0\leq C_{\mathrm{w}}\leq\mathrm{q}$.

Similarly, the partial derivatives of the function $\mathscr{G}_{\mathrm{d}}(p,C_{\mathrm{w}})$ are:

$$\frac{\partial\mathscr{G}_{\mathrm{d}}}{\partial p}=-\delta,\quad \frac{\partial\mathscr{G}_{\mathrm{d}}}{\partial C_{\mathrm{w}}}=u'(C_{\mathrm{w}})-(\mathrm{q}-C_{\mathrm{w}})u''(C_{\mathrm{w}}).$$

The properties of $u(\cdot)$ ensure that $u'(C_{\mathrm{w}})>0$ and $u''(C_{\mathrm{w}})\leq 0$, so $\mathscr{G}_{\mathrm{d}}(p,C_{\mathrm{w}})$ is strictly decreasing in p and strictly increasing in C_{w}.

Now consider two functions $\mathscr{H}_{\mathrm{c}}(p)$ and $\mathscr{H}_{\mathrm{d}}(p)$ defined implicitly by the equations $\mathscr{G}_{\mathrm{c}}(p,\mathscr{H}_{\mathrm{c}}(p))=0$

and $\mathscr{G}_{\mathrm{d}}(p, \mathscr{H}_{\mathrm{d}}(p)) = 0$. Where these functions are defined, their derivatives are:

$$\mathscr{H}_{\mathrm{c}}'(p) = -\frac{\partial \mathscr{G}_{\mathrm{c}}}{\partial p} \Big/ \frac{\partial \mathscr{G}_{\mathrm{c}}}{\partial C_{\mathrm{w}}} < 0, \quad \text{and} \ \mathscr{H}_{\mathrm{d}}'(p) = -\frac{\partial \mathscr{G}_{\mathrm{d}}}{\partial p} \Big/ \frac{\partial \mathscr{G}_{\mathrm{d}}}{\partial C_{\mathrm{w}}} > 0.$$

Observe that $\mathscr{G}_{\mathrm{c}}(0, C_{\mathrm{w}}) = -(\mathrm{q} - C_{\mathrm{w}})$, from which $\mathscr{H}_{\mathrm{c}}(0) = \mathrm{q}$ follows. Similarly, $\mathscr{G}_{\mathrm{d}}(1, C_{\mathrm{w}}) = -(\mathrm{q} - C_{\mathrm{w}})u'(C_{\mathrm{w}})$ implies $\mathscr{H}_{\mathrm{d}}(1) = \mathrm{q}$ since $u'(\cdot) > 0$. As $\mathscr{H}_{\mathrm{d}}(p)$ is strictly increasing in p, it follows by continuity that there exists a $p_{\mathrm{d}} \in [0, 1)$ such that $\mathscr{H}_{\mathrm{d}}(p_{\mathrm{d}}) = 0$ if $p_{\mathrm{d}} > 0$, or $\mathscr{H}_{\mathrm{d}}(0) \geq 0$ if $p_{\mathrm{d}} = 0$. Let $\underline{C}_{\mathrm{w}} \equiv \mathscr{H}_{\mathrm{d}}(p_{\mathrm{d}})$ (with $0 \leq \underline{C}_{\mathrm{w}} < \mathrm{q}$). The function $\mathscr{H}_{\mathrm{d}}(p)$ is then well-defined on the interval $[p_{\mathrm{d}}, 1]$ in the sense of returning a C_{w} in the interval $[\underline{C}_{\mathrm{w}}, \mathrm{q}]$.

Since $u(\cdot)$ is strictly increasing, so is its inverse $u^{-1}(\cdot)$. It follows that $u^{-1}(u(C_{\mathrm{w}}) + \delta) > C_{\mathrm{w}}$ and thus

$$\mathscr{G}_{\mathrm{c}}(p, C_{\mathrm{w}}) > pC_{\mathrm{w}} - (1 - p)(\mathrm{q} - C_{\mathrm{w}}) = C_{\mathrm{w}} - \mathrm{q}(1 - p).$$

Hence, $0 = \mathscr{G}_{\mathrm{c}}(p, \mathscr{H}_{\mathrm{c}}(p)) > \mathscr{H}_{\mathrm{c}}(p) - \mathrm{q}(1 - p)$, which implies

$$\mathscr{H}_{\mathrm{c}}(p) < \mathrm{q}(1 - p),$$

for all p for which $\mathscr{H}_{\mathrm{c}}(p)$ is well defined. Since $\mathscr{H}_{\mathrm{c}}(p)$ is strictly decreasing in p, given $\mathscr{H}_{\mathrm{c}}(0) = \mathrm{q}$ and the bound above, it follows there exists a $p_{\mathrm{c}} \in (0, 1)$ such that $\mathscr{H}_{\mathrm{c}}(p_{\mathrm{c}}) = 0$. The function $\mathscr{H}_{\mathrm{c}}(p)$ is then well-defined on the interval $[0, p_{\mathrm{c}}]$ in the sense of returning a C_{w} in the interval $[0, \mathrm{q}]$.

Let $\mathscr{H}_{\mathrm{c}}^{-1}(C_{\mathrm{w}})$ denote the inverse function of $\mathscr{H}_{\mathrm{c}}(p)$, defined on $[0, \mathrm{q}]$. Similarly, $\mathscr{H}_{\mathrm{d}}^{-1}(C_{\mathrm{w}})$ is the inverse function of $\mathscr{H}_{\mathrm{d}}(p)$, defined on $[\underline{C}_{\mathrm{w}}, \mathrm{q}]$, where $\underline{C}_{\mathrm{w}} \equiv \mathscr{H}_{\mathrm{d}}(p_{\mathrm{d}}) \geq 0$. Since $\mathscr{H}_{\mathrm{c}}(p)$ is strictly decreasing, and $\mathscr{H}_{\mathrm{d}}(p)$ is strictly increasing, the inverse functions inherit these properties. Now define the following function $\mathscr{A}(C_{\mathrm{w}})$ on $[\underline{C}_{\mathrm{w}}, \mathrm{q}]$:

$$\mathscr{A}(C_{\mathrm{w}}) \equiv \mathscr{H}_{\mathrm{d}}^{-1}(C_{\mathrm{w}}) - \mathscr{H}_{\mathrm{c}}^{-1}(C_{\mathrm{w}}).$$

This function is strictly increasing.

Consider first the case where $p_{\mathrm{d}} < p_{\mathrm{c}}$.

$$\mathscr{A}(\underline{C}_{\mathrm{w}}) = p_{\mathrm{d}} - \mathscr{H}_{\mathrm{c}}^{-1}(\underline{C}_{\mathrm{w}}).$$

If $\underline{C}_{\mathrm{w}} > 0$ then $p_{\mathrm{d}} = 0$, so $\mathscr{A}(\underline{C}_{\mathrm{w}}) = -\mathscr{H}_{\mathrm{c}}^{-1}(\underline{C}_{\mathrm{w}})$. Since $\underline{C}_{\mathrm{w}} < \mathrm{q}$, $\mathscr{H}_{\mathrm{c}}^{-1}(\underline{C}_{\mathrm{w}}) > 0$, so $\mathscr{A}(\underline{C}_{\mathrm{w}}) < 0$. If $\underline{C}_{\mathrm{w}} = 0$ then $\mathscr{A}(\underline{C}_{\mathrm{w}}) = p_{\mathrm{d}} - p_{\mathrm{c}} < 0$. Now note that

$$\mathscr{A}(\mathrm{q}) = 1 - 0 > 0.$$

As $\mathscr{A}(C_{\mathrm{w}})$ is strictly increasing and changes sign over the interval $[\underline{C}_{\mathrm{w}}, \mathrm{q}]$, there exists a unique solution C_{w}^* (with $0 < C_{\mathrm{w}}^* < \mathrm{q}$) to the equation $\mathscr{A}(C_{\mathrm{w}}) = 0$.

Now consider the case where $p_{\mathrm{d}} \geq p_{\mathrm{c}}$. As before, $\mathscr{A}(\mathrm{q}) = 1$, and $\mathscr{A}(\underline{C}_{\mathrm{w}}) = \mathscr{A}(0) = p_{\mathrm{d}} - p_{\mathrm{c}}$. It follows that either $C_{\mathrm{w}}^* = 0$ is the only possible solution of $\mathscr{A}(C_{\mathrm{w}}) = 0$, or there is no solution. Whether or not a solution of the equation exists, set $C_{\mathrm{w}}^* = 0$, and note that $\mathscr{A}(C_{\mathrm{w}}^*) \geq 0$.

Taking the C_{w}^* constructed in the appropriate case above, let $p^* = \mathscr{H}_{\mathrm{c}}^{-1}(C_{\mathrm{w}}^*)$. Since $C_{\mathrm{w}}^* < \mathrm{q}$, it must be the case that $p^* > 0$. By definition, it also implies $\mathscr{G}_{\mathrm{c}}(p^*, C_{\mathrm{w}}^*) = 0$, and hence:

$$\mathrm{q} - C_{\mathrm{w}}^* = \frac{p^*}{1 - p^*} u^{-1}(u(C_{\mathrm{w}}^*) + \delta).$$

Since $\mathscr{A}(C_{\mathrm{w}}^*) \geq 0$ in all cases, $\mathscr{H}_{\mathrm{d}}^{-1}(C_{\mathrm{w}}^*) \geq \mathscr{H}_{\mathrm{c}}^{-1}(C_{\mathrm{w}}^*) = p^*$. Since $\mathscr{G}_{\mathrm{d}}(p, C_{\mathrm{w}})$ is decreasing in p and

$\mathscr{G}_{\mathrm{d}}(\mathscr{H}_{\mathrm{d}}^{-1}(C_{\mathrm{w}}), C_{\mathrm{w}}) = 0$, it follows from this that $\mathscr{G}_{\mathrm{d}}(p^*, C_{\mathrm{w}}^*) \geq 0$. This implies

$$\mathrm{q} - C_{\mathrm{w}}^* \leq \frac{\delta(1-p^*)}{u'(C_{\mathrm{w}}^*)}.$$

Combining this with the earlier equation:

$$\frac{p^*}{1-p^*} u^{-1}(u(C_{\mathrm{w}}^*) + \delta) \leq \frac{\delta(1-p^*)}{u'(C_{\mathrm{w}}^*)},$$

and hence:

$$u^{-1}(u(C_{\mathrm{w}}^*) + \delta) < \frac{(1-p^*)^2}{p^*} \frac{\delta}{u'(C_{\mathrm{w}}^*)}.$$

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 $\mathscr{U}_{\mathrm{w}}^* = u(C_{\mathrm{w}}^*)$. This implies

$$u^{-1}(u(C_{\mathrm{w}}^*) + \delta) \geq u^{-1}(u(C_{\mathrm{w}}^*)) + \frac{1}{u'(u^{-1}(u(C_{\mathrm{w}}^*)))}\delta = C_{\mathrm{w}}^* + \frac{\delta}{u'(C_{\mathrm{w}}^*)} \geq \frac{\delta}{u'(C_{\mathrm{w}}^*)},$$

since $C_{\mathrm{w}}^* \geq 0$. Combining this with the earlier inequality yields:

$$\frac{\delta}{u'(C_{\mathrm{w}}^*)} \leq \frac{(1-p^*)^2}{p^*} \frac{\delta}{u'(C_{\mathrm{w}}^*)},$$

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 the unit interval (the smallest root), and it must be case that $p^* \leq \bar{p}$. This root is

$$\bar{p} = \frac{3-\sqrt{5}}{2} = 2 - \left(\frac{1+\sqrt{5}}{2}\right) = 2 - \varphi,$$

where φ is the *Golden ratio*:

$$\varphi \equiv \frac{1+\sqrt{5}}{2} \approx 1.62.$$

Given the bound on p^*:

$$\frac{1-p^*}{p^*} = \frac{1}{p^*} - 1 \geq \frac{1}{\bar{p}} - 1 = \frac{1}{2-\varphi} - 1 = \frac{\varphi - 1}{2-\varphi}.$$

Since the Golden ratio is a root of the quadratic $\varphi^2 - \varphi - 1 = 0$, it follows that $(\varphi-1)/(2-\varphi) = \varphi$, and hence for all possible p^*:

$$\frac{1-p^*}{p^*} \geq \varphi.$$

Now turn to question of whether a Markovian equilibrium exists, and if so, what type it is. Consider first the possibility that $p^* = 0$. This requires $C_{\mathrm{w}}^* = \mathscr{H}_{\mathrm{c}}(0) = \mathrm{q}$. Notice that $\mathscr{G}_{\mathrm{d}}(p, \mathrm{q}) = \delta(1-p)$, so $\mathscr{G}_{\mathrm{d}}(p^*, C_{\mathrm{w}}^*) = \mathscr{G}_{\mathrm{d}}(0, \mathrm{q}) = \delta > 0$. It then follows from the earlier expression for the derivative that $\partial C_{\mathrm{p}}/\partial p \to \infty$ at $p^* = 0$. At this corner, the first-order conditions require $\partial C_{\mathrm{p}}/\partial p \leq 0$, so $p^* = 0$ cannot be a Markovian equilibrium. Next, consider the possibility of another corner solution, $C^*_{\mathrm{w}} = \mathrm{q}$. Recall that $\mathrm{q} = \mathscr{H}_{\mathrm{c}}(0)$ and $\mathscr{H}_{\mathrm{c}}(p)$ is strictly decreasing in p, so $p^* = 0$ is the only value of p consistent with $C^*_{\mathrm{w}} = \mathrm{q}$, but that case has just been ruled out.

Now consider an equilibrium with $0 < p^* < 1/2$ and $0 < C^*_{\mathrm{w}} < \mathrm{q}$. Given the first-order condition in this case, this requires $\mathscr{G}_{\mathrm{c}}(p^*, C^*_{\mathrm{w}}) = 0$ and $\mathscr{G}_{\mathrm{d}}(p^*, C^*_{\mathrm{w}}) = 0$, or that $p^* = \mathscr{H}_{\mathrm{c}}^{-1}(C^*_{\mathrm{w}})$ and $p^* = \mathscr{H}_{\mathrm{d}}^{-1}(C^*_{\mathrm{w}})$. This is equivalent to the equation $\mathscr{A}(C^*_{\mathrm{w}}) = 0$ studied earlier. In the case $p_{\mathrm{d}} < p_{\mathrm{c}}$, such an equilibrium has been shown to exist, and to be unique. If this condition does not hold, an equilibrium in this range does not exist.

Now consider equilibria with either $C^*_{\mathrm{w}} = 0$ or $p^* = 1/2$. Such an equilibrium must satisfy the binding no-rebellion constraint equivalent to $\mathscr{G}_{\mathrm{c}}(p^*, C^*_{\mathrm{w}}) = 0$, but the first-order condition now requires $\mathscr{G}_{\mathrm{d}}(p^*, C^*_{\mathrm{w}}) \geq 0$. Take the case $C^*_{\mathrm{w}} = 0$ first. Since p^* must satisfy $p^* = \mathscr{H}_{\mathrm{c}}^{-1}(C^*_{\mathrm{w}}) = 0$ and $\mathscr{G}_{\mathrm{d}}(p, C_{\mathrm{w}})$ is strictly decreasing in p, $\mathscr{G}_{\mathrm{d}}(p^*, C^*_{\mathrm{w}}) = \mathscr{G}_{\mathrm{d}}(\mathscr{H}_{\mathrm{c}}^{-1}(C^*_{\mathrm{w}}), C^*_{\mathrm{w}}) \geq 0$ if and only if $\mathscr{H}_{\mathrm{d}}^{-1}(C^*_{\mathrm{w}}) \geq \mathscr{H}_{\mathrm{c}}^{-1}(C^*_{\mathrm{w}})$ because $\mathscr{G}_{\mathrm{d}}(\mathscr{H}_{\mathrm{d}}^{-1}(C^*_{\mathrm{w}}), C^*_{\mathrm{w}}) = 0$. This condition is equivalent to $\mathscr{A}(C^*_{\mathrm{w}}) \geq 0$. The earlier analysis has shown that $C^*_{\mathrm{w}} = 0$ and p^* as defined satisfy this condition if and only if $p_{\mathrm{d}} \geq p_{\mathrm{c}}$.

Finally, consider the case $p^* = 1/2$. Such an equilibrium would need to satisfy $\mathscr{G}_{\mathrm{c}}(p^*, C^*_{\mathrm{w}}) = 0$ and $\mathscr{G}_{\mathrm{d}}(p^*, C^*_{\mathrm{w}}) \geq 0$. But it has been shown for all points p^* satisfying these conditions that $p^* \leq 2 - \varphi = (3 - \sqrt{5})/2$. As this bound is less than 0.5, there are no equilibrium at this boundary.

In summary, it has been shown that a Markovian equilibrium always exists and is always unique. The elite size is always positive, but less than two minus the Golden ratio: $0 < p^* \leq 2 - \varphi$. In the case $p_{\mathrm{d}} < p_{\mathrm{c}}$ the equilibrium features slack non-negativity constraints: $0 < C^*_{\mathrm{w}} < \mathrm{q}$. In the remaining case $p_{\mathrm{d}} \geq p_{\mathrm{c}}$, the equilibrium features a binding non-negativity constraint: $C^*_{\mathrm{w}} = 0$.

Note that:

$$\mathscr{G}_{\mathrm{d}}(p, 0) = \delta(1 - p) - u'(0)\mathrm{q},$$

and that $\mathscr{G}_{\mathrm{d}}(0, 0) \leq 0$ if $\delta/(u'(0)\mathrm{q}) \leq 1$. This proves that $p_{\mathrm{d}} = 0$ in this case, and hence that $p_{\mathrm{d}} < p_{\mathrm{c}}$, in which case it has been shown there is an interior equilibrium. Now consider cases where $\delta/(u'(0)\mathrm{q}) > 1$, and thus $\mathscr{G}_{\mathrm{d}}(0, 0) > 0$. It follows that $p_{\mathrm{d}} > 0$, and the definition in this case implies $\mathscr{G}_{\mathrm{d}}(p_{\mathrm{d}}, 0) = 0$, which yields:

$$\frac{1}{1 - p_{\mathrm{d}}} = \beta, \quad \text{where } \beta \equiv \frac{\delta}{u'(0)\mathrm{q}}.$$

Rearranging this leads to an explicit expression for p_{d} in terms of β:

$$p_{\mathrm{d}} = \frac{\beta - 1}{\beta}.$$

The definition of p_{c} is such that $\mathscr{G}_{\mathrm{c}}(p_{\mathrm{c}}, 0) = 0$, and hence:

$$p_{\mathrm{c}} u^{-1}(u(0) + \delta) = (1 - p_{\mathrm{c}})\mathrm{q},$$

which can be stated as:

$$\frac{p_{\mathrm{c}}}{1 - p_{\mathrm{c}}} = \frac{\mathrm{q}}{u^{-1}(u(0) + \delta)}.$$

The condition $p_{\mathrm{d}} < p_{\mathrm{c}}$ is equivalent to $p_{\mathrm{d}}/(1 - p_{\mathrm{d}}) < p_{\mathrm{c}}/(1 - p_{\mathrm{c}})$, and thus:

$$\left(\frac{\beta - 1}{\beta}\right)\beta < \frac{\mathrm{q}}{u^{-1}(u(0) + \delta)}.$$

This can be written as:

$$\frac{\delta}{u'(0)\mathrm{q}} < 1 + \frac{\mathrm{q}}{u^{-1}(u(0) + \delta)},$$

using the definition of β. Notice that since $u(\cdot)$ is strictly increasing, $u^{-1}(u(0)+\delta) > 0$, so $\beta \leq 1$ is sufficient for this to hold. The condition is thus necessary and sufficient for an interior equilibrium.

Now make use of the convexity of the inverse function $u^{-1}(\cdot)$ to deduce:

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

Therefore:

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

This leads to the following necessary condition for an interior equilibrium:

$$\beta < 1 + \beta^{-1},$$

which is equivalent to $\beta^2 - \beta - 1 < 0$. Analysis of this quadratic shows that this condition reduces to $\beta < \varphi$, where φ is the Golden ratio defined earlier. Since an equilibrium is known to exist in all cases, it follows that $\beta \geq \varphi$ is sufficient for a corner equilibrium. This completes the proof.

A.2 Proof of Proposition 2

Consider a Markovian equilibrium with $s^* > 0$.

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

At the pre-investment stage, individuals outside the elite do not yet know whether they will be investors or workers. Thus, individuals at this stage are only distinguished by being inside or outside the elite. The number of non-elite members is $n = 1 - p$, so given that $p < 1/2$, n will always exceed p', the size of the rebel army at the pre-investment stage.

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

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

and since the formula for $p^\dagger$ implies

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

it follows that

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

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

Payoff equalization for workers

Suppose the Markovian equilibrium features different taxes $\tau_q^*(\imath)$ levied on different workers, and hence dispersion in worker payoffs $\mathcal{U}_w^*(\imath)$. 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 (the increase in elite payoffs cannot increase fighting effort from elite members who join a rebel army). This cannot be an equilibrium. The second case is where some workers belong to rebel army associated with a binding no-rebellion constraint. Since workers (or non-elite members) are larger in size than the rebel army, no 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 slackening of the binding no-rebellion constraints allows the elite to raise taxes.

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 workers:

$$\mathcal{U}_{\mathrm{w}} = \mathrm{q} - \tau_{\mathrm{q}}. \tag{A.2.1}$$

Taxes can always be raised to ensure $\mathcal{U}_{\mathrm{w}} \leq \mathcal{U}_{\mathrm{p}}^{\dagger}(K)$ without any danger of rebellion. Given that there are always sufficient workers at the post-investment stage, the following no-rebellion constraint for a rebel army including only workers must hold:

$$\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \mathcal{U}_{\mathrm{w}} \leq \delta \frac{p}{p^{\dagger}}. \tag{A.2.2}$$

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

Ex post, investors receive a common payoff

$$\mathcal{U}_{\mathrm{k}} = (\mathrm{q} - \tau_{\mathrm{q}}) + (\kappa - \tau_{\kappa}), \tag{A.2.3}$$

which given the incentive compatibility condition [5.8], is related to the worker payoff [A.2.1] as follows:

$$\mathcal{U}_{\mathrm{k}} = \mathcal{U}_{\mathrm{w}} + \tilde{\theta}. \tag{A.2.4}$$

Since $\psi \leq \tilde{\theta} \leq \kappa$ and $\psi > 0$, any Markovian equilibrium must feature $\tilde{\theta}^{*} > 0$. Hence, $\mathcal{U}_{\mathrm{k}}^{*} = \mathcal{U}_{\mathrm{w}}^{*} + \tilde{\theta}^{*} > \mathcal{U}_{\mathrm{w}}^{*}$, so investors would exert less fighting effort. Since there is no shortage of workers (all receiving the same payoff), 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 $\sigma_{\mathrm{i}}^{\dagger} = 0$.

Now consider the set of no-rebellion constraints at the post-investment stage. Let σ denote the fraction of places in the rebel army (of size $p^{\dagger}$) filled by workers, with the remaining places filled by those from the current elite. Define the following function of σ:

Write the no-rebellion constraint as:

$$\mathscr{R}(\sigma) \equiv \delta \frac{p}{p^{\dagger}} - \frac{1}{p^{\dagger}} \max_{\substack{\mathcal{E}_{\mathrm{p}} \subseteq \mathcal{P} \\ |\mathcal{E}_{\mathrm{p}}| = (1-\sigma) p^{\dagger}}} \int_{\mathcal{E}_{\mathrm{p}}} \mathbb{1}[\mathcal{U}_{\mathrm{p}}(\imath) \leq \mathcal{U}_{\mathrm{p}}^{\dagger}(K)]\{\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \mathcal{U}_{\mathrm{p}}(\imath) + \delta\} \mathrm{d}\imath - \sigma(\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \mathcal{U}_{\mathrm{w}}), \tag{A.2.5}$$

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. This is defined for a general distribution of elite payoffs $\mathcal{U}_{\mathrm{p}}(\imath)$. The post-investment no-rebellion constraints are equivalent to:

$$\mathscr{R}(\sigma) \geq 0 \ \text{ for all } \ \sigma \in [\underline{\sigma}, 1], \quad \text{where } \ \underline{\sigma} \equiv \max\left\{0, \frac{p^{\dagger} - p}{p^{\dagger}}\right\}, \tag{A.2.6}$$

where $\overline{\sigma} = 1$ because there are always sufficient workers to fill the rebel army. The term $\underline{\sigma}$ is the physical lower bound for the worker share given the relative sizes of p and $p^{\dagger}$.

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 stage 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 that for a rebel army drawn solely from non-elite members (of whom there is always a sufficient number).

Thus, the pre-investment stage no-rebellion constraint is

$$\mathcal{U}'_{\mathrm{p}} - \mathcal{U}_{\mathrm{n}} \leq \delta \frac{p}{p'}, \tag{A.2.7}$$

where the expected utility of a non-elite member is

$$\mathcal{U}_{\mathrm{n}} = (1-\alpha)\mathcal{U}_{\mathrm{w}} + \alpha \mathbb{E}_{\theta} \max\{\mathcal{U}_{\mathrm{i}}(\theta), \mathcal{U}_{\mathrm{w}}\} = \mathcal{U}_{\mathrm{w}} + \alpha \mathcal{S}_{\mathrm{i}}(\tilde{\theta}), \tag{A.2.8}$$

with the investors' surplus $\mathcal{S}_{\mathrm{i}}(\tilde{\theta})$ given by

$$\mathcal{S}_{\mathrm{i}}(\tilde{\theta}) = \mathbb{E}_{\theta} \max\{\mathcal{U}_{\mathrm{i}}(\theta) - \mathcal{U}_{\mathrm{i}}(\tilde{\theta}), 0\} = \mathbb{E}_{\theta} \max\{\tilde{\theta} - \theta, 0\}, \tag{A.2.9}$$

and where for the marginal investor, $\mathcal{U}_{\mathrm{i}}(\tilde{\theta}) = \mathcal{U}_{\mathrm{w}}$.

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

The resource constraint implies that the average consumption $\bar{C}_{\mathrm{p}}$ of elite members is:

$$\bar{C}_{\mathrm{p}} = \frac{(1-p)\tau_{\mathrm{q}} + i\tau_{\kappa}}{p}.$$

The incentive compatibility condition [5.8] for investors implies $\tau_{\kappa} = \kappa - \tilde{\theta}$, so this is determined as a function of $\tilde{\theta}$ (and implicitly, of s). The number of investors is $i = \mu s$. The utility function $u(C) = C$ is linear, so average elite utility (the elite's objective function) is equal to average elite consumption: $\bar{\mathcal{U}}_{\mathrm{p}} = \bar{C}_{\mathrm{p}}$. Hence:

$$\bar{\mathcal{U}}_{\mathrm{p}} = \frac{(1-p)\tau_{\mathrm{q}} + \mu(\kappa - \tilde{\theta})s}{p}, \tag{A.2.10}$$

and this function is not directly affected by the distribution of elite consumption.

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

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

This equation can be solved for $\tilde{\mathcal{U}}_{\mathrm{p}}$:

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

and therefore:

$$\mathcal{U}^{\dagger}_{\mathrm{p}}(K) - \tilde{\mathcal{U}}_{\mathrm{p}} = \frac{\mathcal{U}^{\dagger}_{\mathrm{p}}(K) - \bar{\mathcal{U}}_{\mathrm{p}} + \lambda\gamma}{1-\lambda}. \tag{A.2.11}$$

Now define $\sigma^{\ddagger}$ as the minimum number of workers that must be included in the rebel army if the

λ fraction of loyal elite members are not to be used at all:

$$\sigma^{\ddagger}(\lambda) \equiv \max\left\{0, \frac{p^{\dagger} - (1-\lambda)p}{p^{\dagger}}\right\}, \qquad \text{[A.2.12]}$$

where $\sigma^{\ddagger}(\lambda) \geq \underline{\sigma}$. For any $\sigma \in [\underline{\sigma}, \sigma^{\ddagger}(\lambda))$, $\mathscr{R}(\sigma) \geq \mathscr{R}(\sigma^{\ddagger})$, so the no-rebellion constraints are redundant in this range. It is necessary and sufficient only to verify $\mathscr{R}(\sigma) \geq 0$ for $\sigma \in [\sigma^{\ddagger}(\lambda), 1]$.

Now define a hypothetical equivalent of $\mathscr{R}(\sigma)$ assuming that all elite members with payoff less than or equal to $\mathcal{U}^{\dagger}_{\mathrm{p}}(K)$ receive the average for those below this threshold, that is, $\tilde{\mathcal{U}}_{\mathrm{p}}$. In other words, payoffs are equalized among the fraction $1-\lambda$ of elite members who are not loyal (but payoffs need not be completely equalized among all elite members unless $\lambda = 0$). Denote the equivalent function by $\bar{\mathscr{R}}(\sigma)$:

$$\bar{\mathscr{R}}(\sigma) = \delta\frac{p}{p^{\dagger}} - \sigma(\mathcal{U}^{\dagger}_{\mathrm{p}}(K) - \mathcal{U}_{\mathrm{w}}) - \begin{cases}(1-\sigma)(\mathcal{U}^{\dagger}_{\mathrm{p}}(K) - \tilde{\mathcal{U}}_{\mathrm{p}} + \delta) & \text{if } \sigma \in [\sigma^{\ddagger}(\lambda), 1] \\ (1-\sigma^{\ddagger})(\mathcal{U}^{\dagger}_{\mathrm{p}}(K) - \tilde{\mathcal{U}}_{\mathrm{p}} + \delta) & \text{if } \sigma \in [\underline{\sigma}, \sigma^{\ddagger}(\lambda))\end{cases} \qquad \text{[A.2.13]}$$

Since payoff equalization among the $1-\lambda$ non-loyal members of the elite cannot increase overall fighting effort, it follows that

$$\mathscr{R}(\sigma) \leq \bar{\mathscr{R}}(\sigma) \ \text{ for all } \ \sigma \in [\underline{\sigma}, 1]. \qquad \text{[A.2.14]}$$

If there is a positive measure of elite members with $\mathcal{U}_{\mathrm{p}}(\imath) \leq \mathcal{U}^{\dagger}_{\mathrm{p}}(K)$ and $\mathcal{U}_{\mathrm{p}}(\imath) \neq \tilde{\mathcal{U}}_{\mathrm{p}}$ then:

$$\mathscr{R}(\sigma) < \bar{\mathscr{R}}(\sigma) \ \text{ for all } \ \sigma \in (\sigma^{\ddagger}(\lambda), 1),$$

since $\sigma < 1$ implies there are some elite members included in the rebel army, while $\sigma > \sigma^{\ddagger}(\lambda)$ implies $(1-\sigma)p^{\dagger} < (1-\lambda)p$, so the subset of most dissatisfied elite members can be included.

Now define $\mathscr{N}(\lambda) \equiv \bar{\mathscr{R}}(\sigma^{\ddagger}(\lambda))$ and note that this implies:

$$\mathscr{N}(\lambda) = \delta\left(\frac{p - (1-\underline{\sigma})p^{\dagger}}{p^{\dagger}}\right) - (1-\underline{\sigma})(\mathcal{U}^{\dagger}_{\mathrm{p}}(K) - \tilde{\mathcal{U}}_{\mathrm{p}}) - \underline{\sigma}(\mathcal{U}^{\dagger}_{\mathrm{p}}(K) - \mathcal{U}_{\mathrm{w}}), \qquad \text{[A.2.15]}$$

for all $\lambda \in [0,1]$. Let the threshold $\lambda^{\ddagger}$ be defined as follows:

$$\lambda^{\ddagger} \equiv \max\left\{0, 1 - \frac{p^{\dagger}}{p}\right\},$$

observing that $0 \leq \lambda^{\ddagger} < 1$. If $\lambda \in [0, \lambda^{\ddagger})$ then necessarily $\lambda^{\ddagger} > 0$, so $1 - \lambda^{\ddagger} = p^{\dagger}/p$. It then follows that $\lambda < \lambda^{\ddagger}$ implies $(1-\lambda)p > p^{\dagger}$, so $\sigma^{\ddagger}(\lambda) = 0$. In this case, $\mathscr{N}(\lambda)$ reduces to:

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

where the second equality uses the relationship between $\mathcal{U}^{\dagger}_{\mathrm{p}}(K) - \tilde{\mathcal{U}}_{\mathrm{p}}$ and $\mathcal{U}^{\dagger}_{\mathrm{p}}(K) - \bar{\mathcal{U}}_{\mathrm{p}}$ from [A.2.11]. The other case is $\lambda \in [\lambda^{\ddagger}, 1]$. In this range, it is known that $1 - \lambda \leq p^{\dagger}/p$, so $(1-\lambda)p \leq p^{\dagger}$ and hence:

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

Substituting this into function $\mathscr{N}(\lambda)$:

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

Making some simplifications and using [A.2.11] to substitute for $\tilde{\mathcal{U}}_\mathrm{p}$ in terms of $\bar{\mathcal{U}}_\mathrm{p}$:

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

then collecting terms in λ:

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

Therefore, in summary, the function $\mathscr{N}(\lambda)$ in all cases:

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

In the case where $\lambda^\ddagger > 0$, the function is continuous at $\lambda = \lambda^\ddagger$.

Time inconsistency problem for members of the elite

The equation for the payoff of a worker implies $\tau_\mathrm{q} = \mathrm{q} - \mathcal{U}_\mathrm{w}$, so the average elite payoff [A.2.10] 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 restated as follows:

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

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

$$\mathcal{U}_\mathrm{p}^\dagger(K) - \bar{\mathcal{U}}_\mathrm{p} = \frac{1}{p}\left(\mathcal{U}_\mathrm{p}^\dagger(K) - \left((1-p)\left(\mathrm{q} + \delta\frac{p}{p^\dagger}\right) + \mu(\kappa - \tilde{\theta})s\right)\right) + \left(\frac{1-p}{p}\right)\left(\mathcal{U}_\mathrm{p}^\dagger(K) - \mathcal{U}_\mathrm{w} + \delta\frac{p}{p^\dagger}\right). \tag{A.2.16}$$

Examining the first term in brackets in more detail:

$$\begin{aligned}
\mathcal{U}_{\mathrm{p}}^{\dagger}(K)-\left((1-p)\left(\mathrm{q}+\delta\frac{p}{p^{\dagger}}\right)+\mu(\kappa-\tilde{\theta})s\right) &= \frac{(\mathrm{q}+\delta)^2}{\mathrm{q}+2\delta}+K-(1-p)\left(\mathrm{q}+\delta\frac{p}{p^{\dagger}}\right)-\mu\kappa s+\mu\tilde{\theta}s \\
&= (\mathrm{q}+\delta)(1-p^{\dagger})+\mu\kappa s-\mathrm{q}(1-p)-\frac{\delta}{p^{\dagger}}p+\frac{\delta}{p^{\dagger}}p^2-\mu\kappa s+\mu\tilde{\theta}s \\
&= \frac{\delta}{p^{\dagger}}\left(p^2-p+\frac{\mathrm{q}p^{\dagger}}{\delta}p+\frac{p^{\dagger}}{\delta}\left((\mathrm{q}+\delta)(1-p^{\dagger})-\mathrm{q}\right)\right)+\mu\tilde{\theta}s \\
&= \frac{\delta}{p^{\dagger}}\left(p^2-\left(1-\frac{\mathrm{q}}{\mathrm{q}+2\delta}\right)p+\frac{p^{\dagger}}{\delta}\left(\frac{(\mathrm{q}+\delta)^2-\mathrm{q}(\mathrm{q}+2\delta)}{\mathrm{q}+2\delta}\right)\right)+\mu\tilde{\theta}s \\
&= \frac{\delta}{p^{\dagger}}\left(p^2-2\left(\frac{\delta}{\mathrm{q}+2\delta}\right)p+p^{\dagger}\left(\frac{\delta}{\mathrm{q}+2\delta}\right)\right)+\mu\tilde{\theta}s \\
&= \frac{\delta}{p^{\dagger}}\left(p^2-2p^{\dagger}p+p^{\dagger 2}\right)+\mu\tilde{\theta}s=\frac{\delta}{p^{\dagger}}\left(p-p^{\dagger}\right)^2+\mu\tilde{\theta}s.
\end{aligned} \quad [A.2.17]$$

Whenever $s>0$, since $\tilde{\theta}>0$ as well, the final term above is strictly positive. The other is non-negative, hence:

$$\mathcal{U}_{\mathrm{p}}^{\dagger}(K)-\left((1-p)\left(\mathrm{q}+\delta\frac{p}{p^{\dagger}}\right)+\mu(\kappa-\tilde{\theta})s\right)>0. \quad [A.2.18]$$

The no-rebellion constraint for workers at the post-investment stage [A.2.2] implies

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

therefore together with [A.2.16] and [A.2.18] it implies for $s>0$:

$$\mathcal{U}_{\mathrm{p}}^{\dagger}(K)>\bar{\mathcal{U}}_{\mathrm{p}}. \quad [A.2.19]$$

One consequence of this is that it is impossible to have $\lambda=1$. Since $\tilde{\mathcal{U}}_{\mathrm{p}}$ must be bounded below, in particular by $\mathcal{U}_{\mathrm{w}}^{\dagger}(K)=\mathcal{U}_{\mathrm{p}}^{\dagger}(K)-\delta$, it must be the case that $\lambda\leq\overline{\lambda}$, where

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

with $0<\overline{\lambda}<1$.

The pre-investment constraint cannot bind on its own

Consider first the possibility of a Markovian equilibrium with $s>0$ in which the only effective binding constraint that needs to be imposed is the pre-investment stage constraint [A.2.7]. With $p'=p^*$ and $\mathcal{U}_{\mathrm{p}}'=\mathcal{U}_{\mathrm{p}}^*$:

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

Using the expression for the non-elite expected payoff $\mathcal{U}_{\mathrm{n}}$ from [A.2.8]:

$$\mathcal{U}_{\mathrm{w}}=\mathcal{U}_{\mathrm{p}}^*-\delta\frac{p}{p^*}-\alpha\mathcal{S}_{\mathrm{i}}(\tilde{\theta})=\mathcal{U}_{\mathrm{p}}^*-\delta\frac{p}{p^*}-\frac{\mu}{1-p}\mathcal{S}_{\mathrm{i}}(\tilde{\theta}), \quad [A.2.20]$$

using the formula for the probability α of receiving an investment opportunity. If this is the binding constraint then the tax $\tau_{\mathrm{q}} = \mathrm{q} - \mathcal{U}_{\mathrm{w}}$ is given by:

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

Substituting this into the elite's objective function $\bar{\mathcal{U}}_{\mathrm{p}}$ from [A.2.10] yields:

$$\bar{\mathcal{U}}_{\mathrm{p}} = \frac{(1-p)\left(\mathrm{q} + \delta\frac{p}{p^*} - \mathcal{U}^*_{\mathrm{p}}\right) + \mu(\kappa - \tilde{\theta})s + \mu\mathcal{S}_{\mathrm{i}}(\tilde{\theta})}{p}. \qquad \text{[A.2.21]}$$

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

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

and after imposing Markovian equilibrium $p = p^*$ and $\bar{\mathcal{U}}_{\mathrm{p}} = \mathcal{U}^*_{\mathrm{p}}$:

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

Solving this equation for p^* yields:

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

Using equation [A.2.20], 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

$$\delta + \mathcal{U}^*_{\mathrm{w}} - \mathcal{U}^\dagger_{\mathrm{p}}(K) - \gamma = -(\mathcal{U}^\dagger_{\mathrm{p}}(K) - \bar{\mathcal{U}}^*_{\mathrm{p}}) - \alpha\mathcal{S}_{\mathrm{i}}(\tilde{\theta}) - \gamma. \qquad \text{[A.2.22]}$$

This is strictly negative because $\mathcal{U}^\dagger_{\mathrm{p}}(K) - \bar{\mathcal{U}}^*_{\mathrm{p}} > 0$, which in combination with $p^* = p^\dagger$ implies $\mathscr{N}(\lambda)$ is strictly negative for all λ. Since $\mathscr{R}(\sigma^\ddagger(\lambda)) \leq \bar{\mathscr{R}}(\sigma^\ddagger(\lambda) = \mathscr{N}(\lambda) < 0$, the no-rebellion constraint including elite members at the post-investment stage $\mathscr{R}(\sigma) \geq 0$ is violated for some feasible σ. A Markovian equilibrium of this type does not exist.

The post-investment constraint including only workers cannot bind on its own

Now consider the possibility of a Markovian equilibrium featuring $s > 0$ with the workers' post-investment no-rebellion constraint [A.2.2] as the only effective constraint. When this binds:

$$\mathcal{U}_{\mathrm{w}} = \mathcal{U}^\dagger_{\mathrm{p}}(K) - \delta\frac{p}{p^\dagger}, \qquad \text{[A.2.23]}$$

from which an expression for the tax τ_{q} can be obtained 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.2.10]

$$\bar{\mathcal{U}}_{\mathrm{p}} = \frac{(1-p)\left(\mathrm{q} + \delta\frac{p}{p^\dagger} - \mathcal{U}^\dagger_{\mathrm{p}}(K)\right) + \mu(\kappa - \tilde{\theta})s}{p}. \qquad \text{[A.2.24]}$$

Taking the derivative with respect to p:

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

which can be stated as (using the expression for $\bar{\mathcal{U}}_{\mathrm{p}}$ from [A.2.24]):

$$\frac{\partial \bar{\mathcal{U}}_{\mathrm{p}}}{\partial p} = \frac{1}{p}\left(\frac{\delta}{p^{\dagger}} - \mathrm{q} - 2\delta\frac{p}{p^{\dagger}} + (\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \bar{\mathcal{U}}_{\mathrm{p}})\right) = \frac{1}{p}\left(2\frac{\delta}{p^{\dagger}}(p^{\dagger} - p) + (\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \bar{\mathcal{U}}_{\mathrm{p}})\right). \qquad \text{[A.2.25]}$$

Since [A.2.23] holds, by combining equations [A.2.16] and [A.2.17]:

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

Substituting this into [A.2.25]:

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

Define the function $\pi(s)$ according to:

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

The condition $\pi(s) < 1/2$ is equivalent to

$$\frac{\mu\tilde{\theta}s}{\delta} < \frac{1}{2} - p^{\dagger},$$

and hence

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

Using the expression for $1/2 - p^{\dagger}$ and the bound on μ from [5.3], the first term is 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 is strictly positive. So $\pi(s) < 1/2$ for all values of $s \in [0, 1]$ and parameters consistent with the stated restrictions in [5.3].

The derivative of $\bar{\mathcal{U}}_{\mathrm{p}}$ can be written in terms of $\pi(s)$ as follows:

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

Thus, $\bar{\mathcal{U}}_{\mathrm{p}}$ 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 an equilibrium of this type is therefore:

$$p^* = \sqrt{p^{\dagger}}\sqrt{\pi(s)}.$$

For any $s > 0$, equation [A.2.27] implies $p^{\dagger} < \pi(s)$, so

$$p^{\dagger} < p^* < \pi(s) < \frac{1}{2}.$$

Since $p^* > p^\dagger$, it follows that $\underline{\sigma} = 0$ and $\lambda^\ddagger > 0$. It then follows from [A.2.15] that

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

and hence that $\mathscr{N}(0) \geq 0$ is equivalent to

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

In the case under consideration where [A.2.2] is binding, an expression for the left-hand side is given in [A.2.26], so the required condition can be written as:

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

Simplification establishes this entails

$$p \geq p^\dagger + \frac{\mu s\tilde{\theta}}{\delta} \equiv \pi(s),$$

and is therefore equivalent to $p \geq \pi(s)$. But since $p^* < \pi(s)$, it is shown that $\mathscr{N}(0) < 0$. Given the binding workers' no-rebellion constraint [A.2.23]:

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

As $p^* > p^\dagger$, it is therefore shown that $\delta + \mathcal{U}^*_{\mathrm{w}} - \mathcal{U}^\dagger_{\mathrm{p}}(K) - \gamma < 0$. Along with $\mathcal{U}^\dagger_{\mathrm{p}}(K) - \bar{\mathcal{U}}^*_{\mathrm{p}} > 0$, this proves that the function $\mathscr{N}(\lambda)$ is strictly decreasing in λ. Since $\mathscr{N}(0) < 0$, it follows that $\mathscr{N}(\lambda) < 0$ for all λ. Consequently, $\mathscr{R}(\sigma^\ddagger(\lambda)) \leq \bar{\mathscr{R}}(\sigma^\ddagger(\lambda) = \mathscr{N}(\lambda) < 0$, which violates the post-investment no-rebellion constraint for elite members $\mathscr{R}(\sigma) \geq 0$ for some feasible σ value. There are no Markovian equilibria in this case either.

No equilibrium with both pre- and post-investment constraints binding for non-elite members

In analysing this case, consider first a Markovian equilibrium in which the pre-investment constraint [A.2.7] is binding. This leads to the expression for $\bar{\mathcal{U}}_{\mathrm{p}}$ in [A.2.21]. Imposing Markovian equilibrium ($p = p^*$ and $\mathcal{U}^*_{\mathrm{p}} = \bar{\mathcal{U}}_{\mathrm{p}}$) in that equation:

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

Substituting into [A.2.20] yields:

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

The condition for p^* to satisfy the post-investment no-rebellion constraint for workers is

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

which, using $\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \mathrm{q} = \delta p^{\dagger} + \mu\kappa s$, is equivalent to:

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

Since $\delta/p^{\dagger} = \mathrm{q}+2\delta$, this can be rearranged as follows:

$$\left(1 - \frac{\mu}{1-p^{*}}\frac{\mathcal{S}_{\mathrm{i}}(\tilde{\theta})}{\delta}\right)p^{*} \geq p^{\dagger} + \frac{\mu\tilde{\theta}s}{\delta} \equiv \pi(s). \qquad \text{[A.2.30]}$$

Now start from the case where the post-investment no-rebellion constraint for workers [A.2.2] is binding at $p = p^{*}$. The condition needed for a Markovian equilibrium of this type to satisfy the pre-investment stage no-rebellion constraint [A.2.7] is

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

The expected payoff for non-elite members from [A.2.8] is

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

Hence, for the constraint to be satisfied:

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

Given that the post-investment constraint for workers is assumed to be binding in this case:

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

and by substituting this into the other constraint:

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

Since $\mathcal{U}_{\mathrm{p}}^{*} = \bar{\mathcal{U}}_{\mathrm{p}}^{*}$, the expression for $\mathcal{U}_{\mathrm{p}}^{\dagger}(K) - \bar{\mathcal{U}}_{\mathrm{p}}$ from [A.2.26] can be used because [A.2.2] is binding. This means the inequality above is equivalent to:

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

After some simplification:

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

and hence:

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

In summary, satisfying both constraints simultaneously requires [A.2.30] and [A.2.31], hence the equation:

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

The investors' surplus is

$$\mathcal{S}_{\mathrm{i}}(\tilde{\theta}) \equiv \mathbb{E}_{\theta} \max\{\tilde{\theta} - \theta, 0\} = \int_{\theta=\psi}^{\tilde{\theta}} \frac{\tilde{\theta} - \theta}{\kappa - \psi} \mathrm{d}\theta = \frac{1}{2} \frac{(\tilde{\theta} - \psi)^2}{\kappa - \psi}. \tag{A.2.33}$$

Since $\psi \leq \tilde{\theta} \leq \kappa$, $0 < \psi < \kappa$ and $\kappa < \delta$:

$$\mathcal{S}_{\mathrm{i}}(\tilde{\theta}) \leq \frac{1}{2}(\kappa - \psi) < \frac{\kappa}{2} < \frac{\delta}{2}.$$

Define the following function:

$$\mathscr{M}(p) \equiv \left(1 - \frac{\mu}{1-p} \frac{\mathcal{S}_{\mathrm{i}}(\tilde{\theta})}{\delta}\right) p,$$

which has derivative:

$$\mathscr{M}'(p) = 1 - \frac{1}{1-p} \frac{\mu}{1-p} \frac{\mathcal{S}_{\mathrm{i}}(\tilde{\theta})}{\delta}.$$

Since $\mathscr{S}_{\mathrm{i}}(\tilde{\theta})/\delta < 1/2$, $\alpha = \mu/(1-p) < 1$ for $p < 1/2$, and $1-p > 1/2$, it follows that $\mathscr{M}'(p) > 0$ for all $p \in [0, 1/2]$. As $\mathscr{M}(0) = 0$ and since $\pi(s) > 0$, the equation $\mathscr{M}(p^*) = \pi(s)$, equivalent to [A.2.32], has at most one solution p^* satisfying $0 < p^* < 1/2$. Again, since $\mathcal{S}_{\mathrm{i}}(\tilde{\theta})/\delta < 1/2$ and $\alpha = \mu/(1-p^*) < 1$ when $p^* < 1/2$, it follows that $0 < \mathscr{M}(p^*)/p^* < 1$ where a solution exists, and hence that $p^* > \pi(s)$.

In the case where no solution $p^* \in (0, 1/2)$ exists, then setting $p^* = 1/2$ implies

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

This is consistent with the post-investment constraint for workers binding and the pre-investment constraint being satisfied, but not the case where the pre-investment constraint binds and the post-investment constraint for workers is satisfied. Therefore, there is either an interior solution $p^* \in (0, 1/2)$ or a corner solution $p^* = 1/2$, with the post-investment constraint for workers binding in both cases, and the pre-investment constraint binding in the former case, but slack in the latter.

Now consider a deviation from either the interior or corner equilibrium described above. The elite size p will be changed, with taxes adjusted so that the post-investment constraint for workers continues to bind. This means that the worker payoff is

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

The condition for the new choice of p to be consistent with the pre-investment stage constraint (with expectations remaining consistent with the conjectured equilibrium, $p' = p^*$, $\mathcal{U}_{\mathrm{p}}' = \mathcal{U}_{\mathrm{p}}^*$) is:

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

where the expected non-elite payoff is (using the binding constraint for workers):

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

Therefore, the required condition is

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

which can be rearranged as follows:

$$\left(\frac{1}{p^\dagger}-\frac{1}{p^*}\right)p-\frac{\mu}{1-p}\frac{\mathcal{S}_\mathrm{i}(\tilde{\theta})}{\delta}\leq\frac{\mathcal{U}^\dagger_\mathrm{p}(K)-\mathcal{U}^*_\mathrm{p}}{\delta}. \qquad [A.2.34]$$

This is known to be satisfied at $p=p^*$. Now define:

$$\mathscr{C}(p)\equiv\left(\frac{1}{p^\dagger}-\frac{1}{p^*}\right)p-\frac{\mu}{1-p}\frac{\mathcal{S}_\mathrm{i}(\tilde{\theta})}{\delta},$$

which has derivative:

$$\mathscr{C}'(p)=\frac{1}{p^\dagger}-\frac{1}{p^*}-\frac{\mu}{(1-p)^2}\frac{\mathcal{S}_\mathrm{i}(\tilde{\theta})}{\delta}.$$

Evaulating this derivative at $p=p^*$:

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

since $p^*>\pi(s)$ in all cases under consideration. Since [A.2.27] implies $\pi(s)-p^\dagger=\mu\tilde{\theta}s/\delta$, the derivative can be written as:

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

Note that

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

and hence $\mathcal{S}_\mathrm{i}(\tilde{\theta})\leq\tilde{\theta}s$, which demonstrates that

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

Since $p^\dagger<p^*\leq 1/2$, it follows that $p^\dagger/(1-p^*)<1$ and $p^*/(1-p^*)\leq 1$, so $\mathscr{C}'(p^*)>0$. The no-rebellion constraint [A.2.34] is equivalent to

$$\mathscr{C}(p)\leq\frac{\mathcal{U}^\dagger_\mathrm{p}(K)-\mathcal{U}^*_\mathrm{p}}{\delta},$$

and since the right-hand side is unaffected by p, $\mathscr{C}'(p^*)>0$, and the constraint is satisfied at $p=p^*$, it follows that it is possible to reduce p below p^* and still satisfy the constraint.

Before considering this deviation, suppose that all elite member payoffs are equalized, which entails $\lambda=0$ and $\tilde{\mathcal{U}}_\mathrm{p}=\bar{\mathcal{U}}_\mathrm{p}$. Since the utility function is linear, this redistribution of consumption among the elite has no direct effect on $\bar{\mathcal{U}}_\mathrm{p}$. Given payoff equalization, $\mathscr{R}(\sigma)=\bar{\mathscr{R}}(\sigma)$. Since $\lambda=0$, $\sigma^\ddagger(\lambda)=\underline{\sigma}$, and if $p\geq p^\dagger$ then $\underline{\sigma}=0$, thus:

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

The constraint $\mathscr{R}(1)\geq 0$ is equivalent to the workers' post-investment constraint, which is binding by construction here, even when p deviates from p^*. As $\mathscr{R}(\sigma)$ is linear in σ, it follows that $\mathscr{R}(\sigma)\geq 0$ for all σ if $\mathscr{R}(0)\geq 0$. When the workers' post-investment constraint is binding, it has been shown that $\mathscr{R}(0)\geq 0$ (the condition [A.2.29]) is satisfied if and only if $p\geq\pi(s)$. Hence, starting from a conjectured equilibrium at p^*, because $p^* > \pi(s)$, it is feasible to reduce p and still 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 elite objective function $\bar{\mathcal{U}}_{\text{p}}$ is strictly decreasing in p for $p > \sqrt{p^\dagger}\sqrt{\pi(s)}$. As $p^* > \sqrt{p^\dagger}\sqrt{\pi(s)}$, the deviation is both feasible and payoff-improving for the elite. Hence, there are no Markovian equilibria with this configuration of binding constraints.

The pre-investment constraint and some post-investment constraint including elite members cannot both bind

Suppose the binding constraint $\mathscr{R}(\sigma) = 0$ is imposed for some $\sigma \in [\sigma^\ddagger(\lambda), 1)$, in addition to the pre-investment constraint. If $\sigma \in (\sigma^\ddagger(\lambda), 1)$ then equalizing payoffs among the $1 - \lambda$ non-loyal elite leads to a slackening of the binding constraint and a strict increase in the elite's objective function (unless the constraint is redundant, which would make this case equivalent to that where only the pre-investment constraint binding, as has already been analysed and ruled out). Hence, there is no equilibrium in which there is payoff inequality among this group and where the constraint $\mathscr{R}(\sigma) = 0$ must be imposed for $\sigma \in (\sigma^\ddagger(\lambda), 1)$. Hence, either the equilibrium in this case features payoff equalization, or the constraint that needs to be imposed is for $\sigma = \sigma^\ddagger(\lambda)$. If elite payoffs are equalized then $\mathscr{R}(\sigma) = \bar{\mathscr{R}}(\sigma)$, and the latter is linear in σ. It follows it cannot bind at an interior point, but not bind at $\sigma = \sigma^\ddagger(\lambda)$ (otherwise the constraint would be violated for some other feasible σ value). Therefore, in all cases, an equilibrium of this type can be characterized by imposing the binding constraint $\mathscr{R}(\sigma^\ddagger(\lambda)) = 0$.

The other constraint supposed to be binding in this case is the pre-investment constraint [A.2.7]. It has already been seen that in a Markovian equilibrium, such a binding constraint implies [A.2.22]. Now consider the following payoff-improving deviation from the conjectured equilibrium. First, equalize payoffs among the $1-\lambda$ non-loyal elite. This now implies $\mathscr{R}(\sigma)$ becomes equal to $\bar{\mathscr{R}}(\sigma)$, which is feasible since prior to the payoff equalization $\mathscr{R}(\sigma) \geq 0$ and $\mathscr{R}(\sigma) \leq \bar{\mathscr{R}}(\sigma)$ for all $\sigma \in [\sigma^\ddagger(\lambda), 1]$. Then consider a reduction in λ. Given that condition [A.2.22] holds, the function $\mathscr{N}(\lambda)$ is strictly decreasing in λ. Since $\mathscr{N}(\lambda) = \bar{\mathscr{R}}(\sigma^\ddagger(\lambda))$, this means the binding no-rebellion constraint at the post-investment is slackened. As $\bar{\mathscr{R}}(\sigma)$ is a linear combination of $\bar{\mathscr{R}}(1)$ (unaffected by any of these changes) and $\bar{\mathscr{R}}(\sigma^\ddagger(\lambda))$ (now strictly larger), all other post-investment constraints remain satisfied. A higher value of the elite objective function $\bar{\mathcal{U}}_{\text{p}}$ is now attainable.

Therefore, an equilibrium of this type must feature $\lambda = 0$, and thus $\sigma^\ddagger(\lambda) = \underline{\sigma}$. If $p^* \leq p^\dagger$ then $\underline{\sigma} = 1 - (p/p^\dagger)$ and hence:

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

as a result of $\mathcal{U}_{\text{p}}^\dagger(K) > \bar{\mathcal{U}}_{\text{p}} = \tilde{\mathcal{U}}_{\text{p}}$. Thus, a Markovian equilibrium of this type is only feasible if $p^* > p^\dagger$, in which case, any rebel army can now only include a subset of the elite, so any equilibrium without payoff equalization among the elite would feature a profitable deviation. Attention can therefore be restricted to equilibria in which elite payoffs are equalized. Since $p^* > p^\dagger$ and $\lambda = 0$, $\sigma^\ddagger(\lambda) = \underline{\sigma} = 0$, and so the binding post-investment constraint must be:

$$\mathcal{U}_{\text{p}}^\dagger(K) - \mathcal{U}_{\text{p}}^* = \delta\left(\frac{p^* - p^\dagger}{p^\dagger}\right),$$

where $\mathcal{U}_{\text{p}}^* = \bar{\mathcal{U}}_{\text{p}}$ is the equalized elite payoff. Hence:

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

Given the binding pre-investment constraint:

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

Since the investors' surplus $\mathcal{S}_{\mathrm{i}}(\tilde{\theta})$ is strictly positive in any equilibrium with $s > 0$, this means

$$\mathscr{U}_{\mathrm{w}} < \mathscr{U}_{\mathrm{p}}^{\dagger}(K) - \delta\frac{p^{*}}{p^{\dagger}},$$

which violates the workers' post-investment no-rebellion constraint. Therefore there is no Markovian equilibrium with this configuration of binding constraints.

The case where both post-investment constraints for workers and the elite are binding

Now consider the possibility of an equilibrium where the workers' post-investment constraint and one other post-investment constraint including elite members are binding. The constraint for workers [A.2.2] is equivalent to $\mathscr{R}(1) = 0$ when binding, using the definition of the function $\mathscr{R}(\sigma)$ from [A.2.5]. Suppose that $\mathscr{R}(\sigma) = 0$ is also imposed for some $\sigma \in [\sigma^{\ddagger}(\lambda), 1)$.

First, consider the case where $\sigma \in (\sigma^{\ddagger}(\lambda, 1)$. If there is payoff inequality among the fraction $1 - \lambda$ of non-loyal elite members then $\mathscr{R}(\sigma) < \bar{\mathscr{R}}(\sigma)$, so $\bar{\mathscr{R}}(\sigma) > 0$. Equalizing payoffs among the $1 - \lambda$ fraction of elite members changes $\mathscr{R}(\sigma)$ to $\bar{\mathscr{R}}(\sigma)$. Since $\mathscr{R}(\sigma) \leq \bar{\mathscr{R}}(\sigma)$ for all $\sigma \in [\sigma^{\ddagger}(\lambda), 1]$, all no-rebellion constraints continue to be satisfied after this change, while the one that was binding for $\sigma^{\ddagger}(\lambda) < \sigma < 1$ is now slackened. This allows the elite's objective function to be increased (otherwise the constraint would be redundant, in which case the equilibrium can be characterized by a single binding constraint, as already studied). Therefore, any equilibrium must feature payoff equalization among the $1 - \lambda$ fraction of elite members if the binding constraint is for $\sigma \in (\sigma^{\ddagger}(\lambda), 1)$, otherwise $\sigma = \sigma^{\ddagger}(\lambda)$. If payoffs are equalized then $\mathscr{R}(\sigma) = \bar{\mathscr{R}}(\sigma)$, which is linear in σ. The constraint $\mathscr{R}(\sigma) \geq 0$ cannot then bind at an interior point in the interval $[\sigma^{\ddagger}(\lambda), 1)$ but not bind at the endpoint $\sigma^{\ddagger}(\lambda)$. Therefore, in all cases, $\mathscr{R}(\sigma^{\ddagger}(\lambda)) = 0$ can be taken as the binding constraint in addition to $\mathscr{R}(1) = 0$.

The binding constraint for workers implies

$$\mathscr{U}_{\mathrm{w}}^{*} = \mathscr{U}_{\mathrm{p}}^{\dagger}(K) - \delta\frac{p^{*}}{p^{\dagger}},$$

from which it follows that:

$$\delta + \mathscr{U}_{\mathrm{w}}^{*} - \mathscr{U}_{\mathrm{p}}^{\dagger}(K) - \gamma = -\delta\left(\frac{p^{*} - p^{\dagger}}{p^{\dagger}}\right) - \gamma.$$

Now suppose the Markovian equilibrium features $\lambda > 0$. There are two cases to consider. First, if $p^{*} > p^{\dagger}$, the above equation shows that $\delta + \mathscr{U}_{\mathrm{w}}^{*} - \mathscr{U}_{\mathrm{p}}^{\dagger}(K) - \gamma < 0$, and hence that the function $\mathscr{N}(\lambda)$ from [A.2.15] is strictly decreasing in λ. If payoffs among the $1 - \lambda$ fraction of elite members are equalized then $\mathscr{R}(\sigma$ becomes $\bar{\mathscr{R}}(\sigma)$, which means all the no-rebellion constraints continue to hold. Furthermore, since $\mathscr{R}(\sigma^{\ddagger}(\lambda)) = \bar{\mathscr{R}}(\sigma^{\ddagger}(\lambda)) = \mathscr{N}(\lambda)$, decreasing λ slackens the binding constraint, allowing a higher average elite payoff to be obtained.

The only case in which the above deviation would not be available when $\lambda > 0$ is $p^{*} \leq p^{\dagger}$. In this case, at $\sigma = \sigma^{\ddagger}(\lambda)$, all the $(1 - \lambda)p$ non-loyal elite members would be included in the rebel army, so the distribution of payoffs among them is irrelevant. Formally, $\mathscr{R}(\sigma^{\ddagger}(\lambda)) = \bar{\mathscr{R}}(\sigma^{\ddagger}(\lambda))$ even when payoffs among this group are not equalized. Substituting the expression for $\mathscr{U}_{\mathrm{w}}^{*}$ when the workers' constraint is binding into the function $\bar{\mathscr{R}}(\sigma)$ from [A.2.13]:

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

Therefore, since $\bar{\mathscr{R}}(\sigma^\ddagger(\lambda)) = 0$ and $\sigma^\ddagger(\lambda) < 1$:

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

Since $p^* \leq p^\dagger$, the right-hand side is less than or equal to zero. However, since $\mathcal{U}_\mathrm{p}^\dagger(K) > \bar{\mathcal{U}}_\mathrm{p}$ and $\bar{\mathcal{U}}_\mathrm{p} \geq \tilde{\mathcal{U}}_\mathrm{p}$, the left-hand side is strictly positive. This is a contradiction. Therefore, there cannot be an equilibrium in this case with $\lambda > 0$. With attention restricted to $\lambda = 0$, $\sigma^\ddagger(\lambda) = \underline{\sigma}$. If it were the case that $p^* \leq p^\dagger$ then $\underline{\sigma} = 1 - (p/p^\dagger)$ and hence:

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

as a result of $\mathcal{U}_\mathrm{p}^\dagger(K) > \bar{\mathcal{U}}_\mathrm{p} = \tilde{\mathcal{U}}_\mathrm{p}$. This constitutes a violation of one of the no-rebellion constraints, so any equilibrium of this type must feature $p^* > p^\dagger$. As then a rebel army can only include a subset of the elite, any equilibrium without payoff equalization would be subject to a profitable deviation. In summary, an equilibrium of this type is feasible only if $\lambda = 0$ and there is full payoff equalization among elite members.

The post-investment constraint for the elite cannot bind on its own

Suppose the post-investment constraint is binding for a rebel army that includes some positive measure of elite members, but the worker-only post-investment constraint and the pre-investment constraints are slack (the cases where the constraint is binding in combination with one or both of these two have already been analysed). Thus, $\mathscr{R}(\sigma) \geq 0$ for all $\sigma \in [\sigma^\ddagger(\lambda), 1]$ and $\mathscr{R}(\sigma) = 0$ for some $\sigma \in [\sigma^\ddagger(\lambda), 1)$, while $\mathscr{R}(1) > 0$. The pre-investment stage constraint is also slack:

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

Starting from such a conjectured Markovian equilibrium, first consider an equalization of payoffs among the $1-\lambda$ fraction of elite members. Since $\bar{\mathscr{R}}(\sigma) \geq \mathscr{R}(\sigma)$ before the equalization, this change is feasible once $\mathscr{R}(\sigma)$ coincides with $\bar{\mathscr{R}}(\sigma)$ after the equalization.

If $\bar{\mathscr{R}}(\sigma) > 0$ for all $\sigma \in [\sigma^\ddagger(\lambda), 1]$ then it would be feasible to raise the tax τ_q by some positive amount, which when distributed among elite members increases $\bar{\mathcal{U}}_\mathrm{p}$ and can not increase elite-members' fighting effort in a rebellion. The tax change can also be sufficiently small so that the pre-investment constraint remains satisfied.

Now suppose $\bar{\mathscr{R}}(\sigma) = 0$ for some σ. As $\bar{\mathscr{R}}(\sigma)$ is linear in σ, since $\bar{\mathscr{R}}(1) = \mathscr{R}(1) > 0$ and as the allocation with payoff equalization among the $1-\lambda$ satisfies all no-rebellion constraints, the only possibility is $\bar{\mathscr{R}}(\sigma^\ddagger(\lambda)) = 0$. Then consider an increase in the tax τ_q by a positive but sufficiently small amount such that $\mathscr{R}(1)$ remains non-negative and the pre-investment constraint remains satisfied. The proceeds of this tax are distributed equally among only the $1-\lambda$ fraction of non-loyal elite members. Since the tax is levied on $1-p$ non-elite members (workers and investors) and the proceeds are shared out among $(1-\lambda)p$ individuals, if $\mathcal{U}_\mathrm{w}$ falls by $\Delta\tau_\mathrm{q}$, $\tilde{\mathcal{U}}_\mathrm{p}$ rises by $((1-p)/(1-\lambda)p)\Delta\tau_\mathrm{q}$. The net effect on $\mathscr{R}(\sigma^{\ddagger}(\lambda))$ is

$$\Delta\bar{\mathscr{R}}(\sigma^{\ddagger}(\lambda)) = (1-\sigma^{\ddagger})\left(\frac{(1-p)}{p(1-\lambda)}\Delta\tau_{\mathrm{q}}\right) - \sigma^{\ddagger}(\lambda)\Delta\tau_{\mathrm{q}}.$$

In the case $\sigma^{\ddagger}(\lambda) = 0$, the expression above is clearly positive. If $\sigma^{\ddagger}(\lambda) > 0$ then $\sigma^{\ddagger}(\lambda) = 1 - (1 - \lambda)(p/p^{\dagger})$, in which case:

$$\Delta\bar{\mathscr{R}}(\sigma^{\ddagger}(\lambda)) = \left\{\frac{(1-\lambda)p}{p^{\dagger}}\frac{(1-p)}{p(1-\lambda)} - 1 + \frac{(1-\lambda)p}{p^{\dagger}}\right\}\Delta\tau_{\mathrm{q}} = \left\{\frac{1-\lambda p}{p^{\dagger}} - 1\right\}.$$

Since $0 < p^{\dagger} < 1/2$ and $0 < p < 1$, it follows that the effect is positive. Hence, after these changes, $\bar{\mathscr{R}}(\sigma^{\ddagger}(\lambda)) \geq 0$ and $\bar{\mathscr{R}}(1) \geq 0$, and since $\bar{\mathscr{R}}(\sigma)$ is linear in σ, it must be the case that $\bar{\mathscr{R}}(\sigma) \geq 0$ for all $\sigma \in [\sigma^{\ddagger}(\lambda), 1]$. Therefore, it is always feasible to increase the average elite payoff in these cases, so no Markovian equilibrium with only this one binding constraint exists.

The equilibrium configuration of binding constraints

The analysis so far has ruled out all but one configuration of binding constraints: the case where the post-investment constraint binds for workers in combination with a binding constraint at the post-investment stage for a rebel army including members of the elite. Furthermore, it has been shown in this case that such a conjectured 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, $\mathscr{R}(\sigma) = \bar{\mathscr{R}}(\sigma)$ and $\mathscr{R}(0) = 0$ (since $p^* > p^{\dagger}$ and $\lambda = 0$ imply $\sigma^{\ddagger}(\lambda) = 0$). Finally, the elite size in this case must be $p^* = \pi(s)$, where $\pi(s)$ is the function defined in [A.2.27].

It remains only to check whether all other constraints are satisfied. Since $\bar{\mathscr{R}}(\sigma) = \mathscr{R}(\sigma)$ is linear in σ, the binding constraints at the end-points $\sigma = 0$ and $\sigma = 1$ imply $\mathscr{R}(\sigma) = 0$ for all $\sigma \in [0, 1]$. Earlier analysis has shown that when the post-investment constraint for workers is binding, the pre-investment constraint is satisfied if and only if

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

Hence, with $p^* = \pi(s)$ and $\mathcal{S}_{\mathrm{i}}(\tilde{\theta}) > 0$, this constraint is automatically satisfied. Therefore, the Markovian equilibrium must exist, and must feature this configuration of binding constraints.

The elite's payoff given the binding constraints

By combining the equations $\bar{\mathscr{R}}(0) = 0$ and $\mathscr{R}(1) = 0$, it follows that

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

Equation [A.2.10] for the elite payoff $\bar{\mathcal{U}}_{\mathrm{p}} = \mathcal{U}_{\mathrm{p}}$ implies:

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

and hence:

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

Substituting for $p = \pi(s)$ using [A.2.27]:

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

Therefore:

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

which by substituting the formula for $p^\dagger$ from [5.5] yields the expression for $\mathcal{U}_\mathrm{p}$ in [5.16b]. This completes the proof.

A.3 Proof of Proposition 3

First, note that the relationship between s and $\tilde{\theta}$ in [5.10] implies

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

Using the specification of the general post-investment no-rebellion constraint in [5.15b], the two binding constraints identified by Proposition 2 imply that

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

The level of investment in the Markovian equilibrium

Begin by substituting the equation [A.3.1] into the expression for the elite's payoff in [5.16b]:

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

Deriving the first-order condition for the value of s maximizing $\mathcal{U}_\mathrm{p}$:

$$\mu\left(\kappa - \left(\frac{\mathrm{q}+2\delta}{\delta}\right)\psi - 2\left(\frac{\mathrm{q}+2\delta}{\delta}\right)(\kappa-\psi)s\right) = 0.$$

Solving for s yields:

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

Since $\mathrm{q}+2\delta > \delta$, this expression can never be more than 1, but could be negative. Since s is restricted to the unit interval:

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

Confirmation that this is indeed the solution then requires checking that all remaining constraints are satisfied. These are each elite member's participation constraint $\mathcal{U}^*_\mathrm{p} \geq \mathcal{U}^*_\mathrm{n}$; the requirement $\mathcal{U}^*_\mathrm{p} \geq \mathcal{U}'_\mathrm{n}$ that each elite member is willing to defend the institutions at the pre-investment stage if a rebellion occurs in which he does not participate (which in a Markovian equilibrium is equivalent to the participation constraint); the condition $\mathcal{U}^*_\mathrm{p} \geq \mathcal{U}^\dagger_\mathrm{w}(K)$ for those in the elite being willing to defend the institutions at the post-investment stage (following a rebellion in which an elite member does not participate); and finally the set of non-negativity constraints on all individuals' consumption, which, given that workers always receive the lowest quantity of consumption, is equivalent to checking $\mathcal{U}^*_\mathrm{w} = C^*_\mathrm{w} \geq 0$.

Considering the elite's participation constraint first, equations [5.6] and [5.13] imply:

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

Since $\mathcal{U}_{\mathrm{p}}^* = \mathcal{U}_{\mathrm{w}}^* + \delta$ according to [A.3.2], it follows that $\mathcal{U}_{\mathrm{p}}^* \geq \mathcal{U}_{\mathrm{n}}^*$ is equivalent to

$$\alpha \mathcal{S}_{\mathrm{i}}(\tilde{\theta}^*) \leq \delta. \tag{A.3.5}$$

As $0 < \psi < \kappa$ and $\psi \leq \tilde{\theta}^* \leq \kappa$, it must be the case that $\tilde{\theta}^* - \theta < \kappa$ for all $\theta \in [\psi, \kappa]$. Therefore, the definition of $\mathcal{S}_{\mathrm{i}}(\tilde{\theta})$ in [5.14] implies $\mathcal{S}_{\mathrm{i}}(\tilde{\theta}^*) < \kappa$. Since α is a probability, this means the condition above is necessarily satisfied. In a Markovian equilibrium, $\mathcal{U}_{\mathrm{n}}' = \mathcal{U}_{\mathrm{n}}^*$, so this demonstrates $\mathcal{U}_{\mathrm{p}}^* \geq \mathcal{U}_{\mathrm{n}}'$ as well.

Next, turning to the condition $\mathcal{U}_{\mathrm{p}}^* \geq \mathcal{U}_{\mathrm{w}}^\dagger(K)$ that states that elite members have an incentive to defend the institutions at the post-investment stage following a rebellion in which they do not participate. Using the expression for $\mathcal{U}_{\mathrm{p}}$ from [5.16b], the expression for $\mathcal{U}_{\mathrm{w}}^\dagger(K)$ from [5.5], and the capital stock equation from [5.9], this condition is equivalent to:

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

After cancelling terms, this reduces to:

$$\mu\left(\frac{q+2\delta}{\delta}\right)\tilde{\theta}^* s^* \leq \delta. \tag{A.3.6}$$

Consider an equilibrium with $s^* > 0$ where the solution in [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 $\delta\kappa > (q+2\delta)\psi$. Using the formula for $\tilde{\theta}$ in [A.3.1]:

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

Since $\psi < (\delta/(q+2\delta))\kappa$, this expression for $\tilde{\theta}^*$ implies:

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

Given that $\mu \leq 1$ and $s^* \leq 1$ and $\kappa \leq \delta$, this condition implies that [A.3.6] must hold, demonstrating that $\mathcal{U}_{\mathrm{p}}^* \geq \mathcal{U}_{\mathrm{w}}^\dagger(K)$.

Finally, turning to the non-negativity constraint $\mathcal{U}_{\mathrm{w}}^* \geq 0$ for workers. Observe first that $s = 0$ is always a feasible choice for the elite in maximizing $\mathcal{U}_{\mathrm{p}}$, so from the expression in [5.16b] it follows that:

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

Then substituting [5.16b] into [A.3.2] yields:

$$\mathcal{U}_{\mathrm{w}}^* = \left(\frac{(q+\delta)^2}{q+2\delta} - \delta\right) + \mu\left(\kappa - \left(\frac{q+2\delta}{\delta}\right)\tilde{\theta}^*\right)s^*, \tag{A.3.8}$$

and since the second term has already been seen to be non-negative, a sufficient condition for $\mathcal{U}^*_{\mathrm{w}} \geq 0$ is

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

However, it has already been shown that $\delta/\mathrm{q} \leq \varphi$ (where φ is the Golden ratio) implies that this inequality holds. As this restriction is imposed in [5.3], all the non-negativity constraints must be satisfied, confirming that [A.3.4] is the Markovian equilibrium.

Let $T_\kappa = \tau_\kappa i$ denote the total amount of revenue derived from taxes on capital. Equation [5.8] implies that $\tau_\kappa = \kappa - \tilde{\theta}$. Together with [5.9]:

$$T_\kappa = \mu s(\kappa - \tilde{\theta}) = \mu(\kappa - \psi)s(1-s),$$

where [A.3.1] has also been used to substitute for $\tilde{\theta}$ in terms of s. The function $s(1-s)$ is maximized at $s = 1/2$, so the value of s associated with the peak of the Laffer curve is $s^\ell = 1/2$. In the case where $s^* > 0$, the expression for s^* coincides with that given in [A.3.3]. Since $\mathrm{q}+2\delta > \delta$, it follows from [A.3.3] that $s^* < 1/2$, and therefore $s^* < s^\ell$. From [A.3.1], this means that $\tilde{\theta}^* < \tilde{\theta}^\ell$, and from [5.8] that $\tau^*_\kappa > \tau^\ell_\kappa$.

The constrained efficient level of investment

Proposition 2 characterizes which no-rebellion constraints are binding even when s is not chosen to maximize the elite's payoff (but the other institutional variables are). This means that the relationship between p and s in [5.16a] still holds, hence:

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

using the expression for $p^\dagger$ from [5.5]. Therefore, substituting [5.16b] and the above equation into [5.19] yields:

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

which simplifies to

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

Substituting for $\tilde{\theta}$ using [A.3.1]:

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

Using [5.14], the expected surplus $\mathcal{S}_{\mathrm{i}}(\tilde{\theta})$ from receiving an investment opportunity is given by

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

where equation [5.10] 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:

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

The first-order condition for maximizing [A.3.11] is:

$$\mu\left(\kappa - \left(\frac{q+\delta}{\delta}\right) - \left(\frac{2q+\delta}{\delta}\right)(\kappa-\psi)s\right) = 0.$$

Solving for s yields:

$$s = \frac{\delta\kappa - (q+\delta)\psi}{(2q+\delta)(\kappa-\psi)} = \frac{\delta\kappa - (q+\delta)\psi}{(q+\delta)\kappa - (q+\delta)\psi + q(\kappa-\psi)}. \qquad \text{[A.3.12]}$$

Since $q+\delta > \delta$ and $\kappa > \psi$, this expression can never be greater than 1, but it could be negative. Therefore:

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

To show this is the constrained efficient level of investment, it is necessary to verify that the auxiliary constraints hold.

As before, the elite's participation constraint (and individual rationality condition for defence of the institutions at the pre-investment stage) $\mathcal{U}_p^\diamond \geq \mathcal{U}_n^\diamond$ is equivalent to [A.3.5]. Since $\mathcal{S}_i(\tilde{\theta}) < \kappa < \delta$ for any $\tilde{\theta} \in [\psi, \kappa]$, it follows that $\alpha\mathcal{S}_i(\tilde{\theta}^\diamond) \leq \delta$, so this condition holds.

Now consider the individual rationality condition for defence of the institutions at the post-investment stage by those elite members not included in a rebel army. This requires $\mathcal{U}_p^\diamond \geq \mathcal{U}_w^\dagger(K)$, which reduces to [A.3.6] as before. Consider a case where $s^\diamond > 0$, which requires $\delta\kappa > (q+\delta)\psi$. Using equation [A.3.13]:

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

and substituting this into [A.3.1]:

$$\tilde{\theta}^\diamond = \psi + \frac{\delta\kappa - (q+\delta)\psi}{(2q+\delta)} = \frac{q\psi + \delta\kappa}{2q+\delta}. \qquad \text{[A.3.14]}$$

When $s^\diamond > 0$, it must be the case that $\psi < (\delta/(q+\delta))\kappa$, and hence:

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

which implies:

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

Since the parameter restrictions in [5.3] imply $\mu \leq q/(2(q+2\delta))$:

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

This demonstrates that $\mathcal{U}_p^\diamond \geq \mathcal{U}_w^\dagger(K)$.

Finally, consider the non-negativity constraint $\mathcal{U}_w^\diamond \geq 0$ for workers. Combining equation [5.16b] with [A.3.2] yields:

$$\mathcal{U}_w^\diamond = \left(\frac{(q+\delta)^2}{q+2\delta} - \delta\right) + \mu\left(\kappa - \left(\frac{q+2\delta}{\delta}\right)\tilde{\theta}^\diamond\right)s^\diamond,$$

and then substituting for $\tilde{\theta}^{\diamond}$ using [A.3.14]:

$$\mathcal{U}_{\mathrm{w}}^{\diamond} = \left(\frac{(\mathrm{q}+\delta)^2}{\mathrm{q}+2\delta} - \delta\right) + \frac{\mu}{\delta(\mathrm{q}+2\delta)}\left(\delta(\mathrm{q}-\delta)\kappa - \mathrm{q}(\mathrm{q}+2\delta)\psi\right) s^{\diamond}. \qquad \text{[A.3.15]}$$

The sign of this expression is ambiguous without further parameter restrictions.

Now consider the relationship between s^* and $s^{\diamond}$ and $\hat{s}$. When $s^{\diamond} > 0$, the expression for $s^{\diamond}$ coincides with that from [A.3.12], from which it can be seen that $s^{\diamond} < 1$. Since $\hat{s} = 1$, it follows that $s^{\diamond} < \hat{s}$.

In the case $\kappa/\psi > (\mathrm{q}+2\delta)/\delta$, equation [A.3.4] confirms that $s^* > 0$. This condition implies $\kappa/\psi > (\mathrm{q}+\delta)/\delta$, so a positive $s^{\diamond}$ value is obtained according to [A.3.13]. Thus from [A.3.4] and [A.3.13] in this case:

$$s^* = \frac{\delta\kappa - (\mathrm{q}+2\delta)\psi}{(2\mathrm{q}+4\delta)(\kappa-\psi)}, \quad \text{and} \quad s^{\diamond} = \frac{\delta\kappa - (\mathrm{q}+\delta)\psi}{(2\mathrm{q}+\delta)(\kappa-\psi)},$$

and comparison of these expressions reveals that $s^* < s^{\diamond}$. Therefore, as it follows from [A.3.7]–[A.3.9] that $\mathcal{U}_{\mathrm{w}}^* > 0$ when $s^* > 0$, given that the sign of [A.3.15] is ambiguous, it must be the case that:

$$s^* < s^{\diamond} \leq \frac{\delta\kappa - (\mathrm{q}+\delta)\psi}{(2\mathrm{q}+\delta)(\kappa-\psi)}.$$

The $s^{\diamond}$ value here takes account of the fact that the non-negativity constraint for workers may become binding somewhere strictly to the right of s^* but before the point in [A.3.13]. Therefore, in all cases except where $s^* = 0$, it has been shown that $s^* < s^{\diamond}$. Using [A.3.1] and [5.8], it follows that $\tilde{\theta}^* < \tilde{\theta}^{\diamond}$ and $\tau_{\kappa}^* > \tau_{\kappa}^{\diamond}$. This completes the proof.