

Emergence and Persistence of Inefficient States*

Daron Acemoglu
MIT

Davide Ticchi
University of Urbino

Andrea Vindigni
Princeton

February 2007.

Abstract

Inefficiencies in the bureaucratic organization of the state are often viewed as important factors in retarding economic development. Why certain societies choose or end up with such inefficient organizations has received very little attention, however. In this paper, we present a simple theory of the emergence and persistence of inefficient states based on patronage politics. The society consists of rich and poor individuals. The rich are initially in power, but expect to transition to democracy, which will choose redistributive policies. Taxation requires the employment of bureaucrats. We show that, under certain circumstances, by choosing an inefficient state structure, the rich may be able to use patronage and capture democratic politics. This enables them to reduce the amount of redistribution and public good provision in democracy. Moreover, the inefficient state creates its own constituency and tends to persist over time. Intuitively, an inefficient state structure creates more rents for bureaucrats than would an efficient state structure. When the poor come to power in democracy, they will reform the structure of the state to make it more efficient so that higher taxes can be collected at lower cost and with lower rents for bureaucrats. Anticipating this, when the society starts out with an inefficient organization of the state, bureaucrats support the rich, who set lower taxes but also provide rents to bureaucrats. We show that in order to generate enough political support, the coalition of the rich and the bureaucrats may not only choose an inefficient organization of the state, but they may expand the size of bureaucracy “excessively” so as to gain additional votes. The model shows that an equilibrium with an inefficient state is more likely to arise when there is greater inequality between the rich and the poor, when bureaucratic rents take intermediate values and when individuals are sufficiently forward-looking.

Keywords: bureaucracy, corruption, democracy, patronage politics, political economy, public goods, redistributive politics.

JEL Classification: P16, H11, H26, H41.

*We thank Timothy Besley, Bruno Contini, Miriam Golden, Gilles Saint-Paul, Guido Tabellini and seminar participants at IGER-Bocconi, the Latin American Meetings of the Econometric Society, Toulouse University, and Yale University for useful comments and suggestions. Acemoglu gratefully acknowledges financial support from the National Science Foundation and the hospitality of the Yale Economics Department. Vindigni gratefully acknowledges the hospitality of Collegio Carlo Alberto and of the Yale Leitner Program in International Political Economy.

1 Introduction

There are large cross-country differences in the extent of bureaucratic corruption and the efficiency of the state organization (e.g., World Bank, 2004). An influential argument, dating back at least to Tilly (1990), maintains that differences in “state capacity” are an important determinant of economic development.¹ The evidence that many less-developed economies in sub-Saharan Africa, Asia and Latin America only have a small fraction of their GDP raised in tax revenue and invested by the government (e.g., Acemoglu, 2005) and the correlation between measures of state capacity and economic growth (e.g., Rauch and Evans, 2000) are also consistent with this view. Societies with limited state capacity are also those that invest relatively little in public goods and do not adopt policies that redistribute resources to the poor.² Brazil provides a typical example of a society, where the state sector has been relatively inefficient and democratic politics has generated only limited public goods and benefits for the poor (e.g., Gay, 1990, Evans, 1992, Weyland 1996, Roett, 1999).

In this paper, we construct a political economy model, which links the emergence and persistence of inefficient states to the strategic use of *patronage politics* by the elite as a means of capturing democratic politics. Democratic capture enables the elite to limit the provision of public goods and redistribution, but at the cost of aggregate inefficiencies. Our approach therefore provides a unified answer to the questions of: (i) why inefficient states emerge in some societies; (ii) what explains the prevalence of patronage politics and its association with greater inefficiencies in bureaucracy; and (iii) why many democracies pursue relatively pro-elite policies. Our model also suggests a potential reason why certain democracies may exhibit relatively poor economic performance and adopt various inefficient policies.

Our model economy consists of two groups, the rich elite and poor citizens. Linear taxes can be imposed on both groups, with the proceeds used to finance public good investments. The rich are generally opposed to high levels of taxes and public good investments. Tax collection requires that the state employs bureaucrats to prevent individuals from evading taxes, but bureaucrats themselves also need to be given incentives so that they exert effort (or do not accept bribes). The efficiency with which a central authority can monitor the bureaucrats is our measure of the “organization of the state”. Political competition is modeled either by assuming the existence of two parties, respectively aligned with the rich and the poor, or by allowing free entry into the political arena by citizen candidates (Osborne and Slivinski, 1996, Besley and Coate, 1997). In both cases, there is no commitment to policies before elections and the party

¹See, for example, Evans (1989, 1995), Levi (1989), Migdal (1988), Epstein (2000), Herbst (2000), Centeno (2002) and Kohli (2004).

²See, for example, Etzioni-Halevy (1983) on the importance of state capacity and bureaucratization for the development of the welfare state in the West, and Rothstein and Uslaner (2005) on the importance of state capacity for income redistribution.

that comes to power chooses the policy vector, including taxes, public good provision, and bureaucratic wages, and whether to reform the efficiency of the state institutions. Democratic political competition is made interesting by the fact that bureaucrats may support either the rich or poor parties (candidates) and their support may be pivotal in the outcome of elections.³

We consider two possible organizations of the state: the first is an “efficient” organization, in which bureaucrats will be detected easily if they fail to exert effort, while the second is an “inefficient” one in which monitoring bureaucrats is difficult. In equilibrium, when the state is inefficient, bureaucrats need to be paid rents in order to induce them to perform their roles of tax collection and inspection. The presence of rents creates the possibility of patronage politics, whereby bureaucrats may support the party that will maintain the inefficient structure.⁴

In a society that is always dominated by the rich elite or that is permanently in democracy (with a poor citizen as the median voter), the political process produces an efficient organization of bureaucracy, since an inefficient state creates additional costs and no benefits for those holding power. Our main result is that when the society starts out as nondemocratic (under the control of the rich elite) and is expected to transition to democracy, the rich may find it beneficial to choose an inefficient organization of the state so as to exploit patronage politics to limit redistribution. In particular, bureaucrats realize that once the poor median voter comes to power in democracy, there will be bureaucratic reform, reducing their rents from then on. Therefore, if the rich elite, when in power, choose an inefficient organization of the state, the current bureaucrats—who are receiving rents—prefer to support the rich rather than vote with the poor. Consequently, an inefficient state organization emerges as a political instrument for the rich elite to capture the democratic decision-making process by fostering a coalition between themselves and the bureaucrats. It is also noteworthy that the inefficient state not only emerges in equilibrium, but also persists; when the state is inefficient, the bureaucrats vote for the party of the rich, which chooses not to reform the bureaucracy and continues to maintain the support of the existing bureaucrats and thus its political power.

Our analysis shows that patronage politics typically leads not only to the emergence and persistence of an inefficient state apparatus, but also to the *overemployment* of bureaucrats. This is because the rich may prefer to hire additional (unnecessary) bureaucrats so as to boost their party’s votes. Consequently, a captured democracy will typically feature an inefficient state (bureaucracy), provide relatively few public goods and employ an excessive number of

³Throughout the paper, we focus on Markov perfect equilibria (where strategies only depend on payoff-relevant state variables). This equilibrium concept implies that there is no commitment to policies before elections and thus enables us to illustrate the logic of patronage politics in the clearest fashion. Equilibria with patronage politics are easier to support as non-Markovian subgame perfect equilibria.

⁴The assumption that the main role of bureaucrats is tax inspection is not essential in our basic model. The important feature is that an inefficient state organization must pay bureaucrats rents in order to provide them with the right incentives. Bureaucrats’ role as tax inspectors becomes important in the extension in subsection 5.3, where they can be bribed by producers for tax evasion. We simplify the presentation by assuming that bureaucrats’ main role is tax inspection throughout.

bureaucrats. This pattern of bureaucratic inefficiency is consistent with the stylized view of corrupt and low-capacity bureaucracies in many less developed countries (e.g., Geddes, 1991, and Rauch and Evans, 2000).⁵

We also show that the equilibrium with an inefficient state is more likely when there is greater inequality. This is because greater inequality raises the equilibrium tax rate in democracy and makes it more appealing for the rich to create an inefficient state apparatus to prevent democratic outcomes. An inefficient state also requires intermediate levels of rents/“efficiency wages” for bureaucrats; when rents are limited, bureaucrats would not support the rich, while too high rents would make the inefficient state equilibrium prohibitively costly for the rich elite. Finally, an inefficient state is more likely to arise when agents are more forward-looking, because bureaucrats support the inefficient state in order to obtain future rents.

It is worth emphasizing that in our model patronage politics and the creation of an inefficient state are the only way for the rich to increase their votes and influence in politics. This is not meant to imply that there are no other, complementary methods of democratic capture. Instead, our purpose is to isolate a particular mechanism that appears to be important in practice and investigate its implications.⁶

The rest of the paper is organized as follows. Section 2 provides a brief discussion of a number of case studies that illustrate how patronage politics has been used to limit redistribution towards the poor and also discusses the related literature. Section 3 outlines the basic economic and political environment. Section 4 characterizes the equilibria of the baseline model both under permanent nondemocratic and democratic regimes as benchmarks, and more importantly, under a regime that starts out as nondemocratic and becomes democratic thereafter. We show that in this last political environment the rich elite may choose an inefficient state organization and a sufficiently large bureaucracy in order to create a majority coalition. Section 5 generalizes this framework in a number of directions; in particular, it allows for more general contracts with bureaucrats, considers a citizen-candidate setup for political competition, and allows producers to bribe bureaucrats to evade taxes. Section 6 briefly investigates a distinctive implication of our approach about the relationship between relative

⁵Even with the overemployment of bureaucrats, bureaucrats and the rich elite are unlikely to have an absolute majority in the electorate. In practice, the elite may be able to gain additional influence using other, complementary methods such as lobbying, vote buying or use of paramilitaries (see footnote 6). In addition, the lower turnout rates of relatively poor individuals may further limit the political power of parties favoring less pro-elite policies.

⁶Other methods via which a rich elite may have disproportionate effect in a democracy include lobbying, vote buying and use of force and paramilitaries. Lobbying is discussed extensively in the literature, but typically not in this context, see, for example, Austen-Smith (1987), Baron (1994), and Grossman and Helpman (1996). See also Acemoglu and Robinson (2006b) on democratic capture and use of force and paramilitaries in democracies.

Dixit and Londregan (1998) and Lizzeri and Persico (2001) construct models in which parties can make credible promises to “vote buying” transfers or to group-specific public goods. Our assumption that there is no commitment to future actions combined with our focus on Markov perfect equilibria imply that such strategies are not feasible in our model.

wages of public sector employees and the amount of public good provision in democracies. We report cross-country correlations consistent with this implication. Section 7 concludes, while the Appendix contains some of the proofs omitted from the text.

2 Motivation and Related Literature

In this section, we briefly discuss a number of case studies that motivate our analysis and also relate our paper to the existing literature in political economy.

2.1 Patronage Politics, Inefficient States and Elite Control

The experiences of many societies in Latin America, Asia and Africa illustrate the link between patronage politics, inefficient states and elite control. Here we briefly mention three cases.

Perhaps the most transparent example of inefficient and oversized bureaucracy comes from Brazil. Several authors (e.g. Gay, 1990, Weyland 1996, Roett, 1999) have argued that the distribution of large numbers of public jobs, both in the public administration and in parastatal organizations, has created a pattern of patronage politics in Brazil.⁷ The control over these jobs has enabled traditional elites to preserve their political power and limit the amount of public good provision and redistribution. In fact, despite the high level of inequality in Brazil, elites have been able to control politics for much of the 20th century with only limited amount of repression and relatively short periods of military rule. Interestingly, the amount of redistribution and public good provision does not show marked differences between military and democratic periods.

Patronage politics has often ensured that even those in poorest neighborhoods of Rio have supported the traditional parties rather than socialist or social democratic parties running on platforms of greater public good provision and redistribution (Gay, 1990). Students of Brazilian politics have noted the role of public sector employees in this process. For example, Roett (1999 p. 91) writes “state company employees emerged as being among the strongest supporter of the patrimonial order”. In return, successive governments have withstood external pressures from the IMF and have not reformed the public sector, despite the “public perception that public-sector workers were overpaid and underworked” (Roett, p. 97). The process of reforming the public sector has started only recently and progressed slowly.

Another example of effective patronage politics is provided by the policies of Parti Socialiste (PS) of President Leopold Sedar Senghor in Senegal. After independence, PS faced increasing challenges from various different opposition groups, including urban workers and

⁷In the early 1980s, about 4 million people had a job in some branch of the Brazilian public sector. Evans (1992) observes that the Brazilian state is commonly recognized as a huge *cabide de emprego* (source of jobs) and remarks that in contrast to the Weberian conception, recruitment of public employees in Brazil is not related to merit but to political connections.

farmers. Nevertheless, it managed to preserve its power, with relatively limited amount of repression, largely owing to its use of patronage politics. In fact, Senghor promoted some amount of political liberalization and allowed the creation of a multi-party system. However, PS also exploited its incumbency advantage to manipulate the democratic process by creating an extensive patronage network centered on the state apparatus and parastatal sector. Interestingly, it was precisely during this period of democratization that the size of the public sector grew substantially (Boone 1990, Beck, 1997). Boone (1990 p. 347) describes this process as: “The strategic allocation of government jobs coopted restive intellectuals and professionals and incorporated them into political factions anchored in the state bureaucracy.” Thanks to this successful implementation of patronage politics, PS retained much of its power following the transition to democracy.

A final example of the rise of patronage politics in the face of political competition comes from Italy. The evolution of the Italian bureaucracy in the post-WWII decades demonstrates that the mechanism that our model identifies may operate even in relatively developed countries. A significant extension in the Italian bureaucracy was initiated by the Italian Christian Democratic Party (DC) in the 1950s after the electoral challenge from the Communist Party increased sharply, especially following the 1953 political elections. Until the defeat in World War II, Italian politics was dominated by Mussolini’s dictatorship. After the war, DC emerged as the dominant party. In the 1950s, faced by electoral challenge from the left, DC created a highly disorganized and oversized bureaucracy, which subsequently became a natural source of political support and patronage for the party. Golden (2003 p. 199) describes the motive for the expansion of the bureaucracy as:

“The massive system of political patronage that the leaders of the DC constructed after 1953 was their aggregate answer ... to enlarging the party’s aggregate vote share while protecting the incumbency advantage of individual legislators.”

In part with the support of the bureaucracy that it created, DC’s dominance of Italian politics continued until the 1980s and prevented the formation of a left-wing government.

2.2 Related Literature

Our paper is related to a number of different strands in the literature. However, to the best of our knowledge, there are no other papers in the literature that derive an equilibrium explanation for the various inefficiencies in the organization of bureaucracies and provide a range of comparative statics showing when such inefficiencies are more likely.⁸ We are also

⁸The political science and sociology literature on the organization of the state and the bureaucracy mentioned above discusses related issues, but does not provide formal models or emphasize the mechanisms we focus on in this paper. The small economics literature on the internal organization of the state and bureaucracy, for

not aware of any approaches that derive a micro-founded model of patronage politics, despite the ubiquity of this type of political interactions in practice. In addition, while some papers emphasize the impact of lobbying on political decisions,⁹ the issue of how elites can capture decision-making in democracies has only been discussed in a very different context in Acemoglu and Robinson (2006b).

Perhaps most closely related to our work is the small literature on how politicians may distort policies for strategic reasons. Papers in this literature include models where inefficient policies (such as excessive state employment) are chosen in order to gain votes (e.g., Fiorina and Noll, 1978, Geddes, 1991, Shleifer and Vishny, 1994, Lizzeri and Persico, 2001, Robinson and Torvik, 2005). Other papers focus on how certain inefficient choices (including wasteful investments, large budget deficits, and inefficient fiscal systems) could be made in order to constrain future politicians (e.g., Glazer, 1989, Persson and Svensson, 1989, Tabellini and Alesina, 1990, Aghion and Bolton, 1990, Cukierman, Edwards, and Tabellini, 1992). None of these papers feature the mechanism of an elite creating an inefficient state structure to maintain their political power in the face of an emerging democracy or derive a rationale for patronage politics.¹⁰

The literature on the inefficiency of the form of redistribution is also related to our work. Wilson (1990) and Becker and Mulligan (2003) argue that inefficient methods of redistribution are chosen as a way of limiting the amount of redistribution (see also Coate and Morris, 1995, Rodrik, 1995). There is a close connection between this idea and the main mechanism in our paper, whereby an inefficient state is chosen by the rich in order to limit the amount of future redistribution. Nevertheless, there is also an important distinction; in the basic Becker-Mulligan-Wilson story, it is not clear why the society can commit to the form of redistribution and *not* to the amount of redistribution. In contrast, in our model the choice of an inefficient bureaucracy is a way of affecting the future political equilibrium so as to bring the party aligned with the interests of the rich to political power, and via this channel, to limit the provision of public goods and taxation. As such, our mechanism is more closely related to the models of inefficient redistribution in Saint-Paul (1996) and Acemoglu and Robinson (2001), where a politically powerful group may push for inefficient forms of transfers in order to maintain its

example, Acemoglu and Verdier (1998), Dixit (2002), Egorov and Sonin (2005) and Debs (2006), does not investigate the relationship between patronage politics and the emergence of the inefficient state as a method of limiting redistribution.

⁹See, among others, Austen-Smith (1987), Baron (1994), Dixit, Grossman and Helpman (1997), and Grossman and Helpman (1996).

¹⁰The literature on comparative politics and public finance (e.g., Persson and Tabellini, 2003, Ticchi and Vindigni, 2005) also investigates sources of differences in fiscal policies among democracies, but does not focus on democratic capture, patronage politics or the possibility of inefficient bureaucracies. Therefore, relative to this literature, our approach suggests an alternative, but complementary, source of variation, related to the desire and the ability of the economic elite to dominate democratic politics, which can generate both differences in the level of public goods provision and in the efficiency of the state.

future political power. However, these papers do not model or explain how an existing elite can capture democratic politics by creating an inefficient state structure or provide a mechanism for patronage politics.

Finally, our approach is related to sociological analyses of “cooptation” in democracy by existing elites in some Marxist sociology and political science literatures. In particular, Therborn (1980 pp. 228-234) argues that the control of the state apparatus is a crucial objective of the economic elites in democracy, which they achieve using strategies including cooptation (see also the discussion of hegemony in Gramsci, 1971). However, this literature neither articulates a mechanism through which the elite may accomplish these objectives nor models the costs of such a strategy relative to other options.¹¹

3 Basic Model

3.1 Description of the Economic Environment

Consider the following discrete time infinite-horizon economy populated by a continuum 1 of agents, each of which has the following risk-neutral preferences

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(c_t^j + G_t - h e_t^j \right),$$

at time $t = 0$, where \mathbb{E}_0 is the expectation at time $t = 0$, $\beta \in (0, 1)$ is the discount factor, $c_t^j \geq 0$ denotes the consumption of the agent in question (agent j), $G_t \geq 0$ is the level of public good enjoyed by all agents, $e_t^j \in \{0, 1\}$ is the effort decision of the agent (which will be necessary in some occupations), and $h > 0$ is the cost of effort.

There are two types of agents: $n > 1/2$ are poor (low-skill), while $1 - n$ are rich (high-skill). We denote poor agents by the symbol L (corresponding to low-productivity), and rich agents by H , and also use \mathcal{L} and \mathcal{H} to denote the set of poor and rich agents.

There are two occupations: producer and bureaucrat. In each period, as long as some amount of investment in infrastructure, $K > 0$, is undertaken, each producer generates an income depending on his skill; A^L for poor agents and $A^H > A^L$ for rich agents. If the investment in infrastructure K is not undertaken at time t , then no agent can produce within that period. Producers receive and consume their income net of taxes.

A set of agents denoted by \mathcal{X}_t are bureaucrats at time t . These agents do not produce, but receive a *net* wage of $w_t \geq 0$ from the government (i.e., they do not pay taxes on their

¹¹Another major difference between the Marxist approaches and ours is that in our model bureaucrats can side either with rich or poor agents, whereas in most Marxist approaches, the state apparatus is, ultimately, controlled by the economic elite (e.g., Miliband, 1969, Poulantzas, 1978, Therborn, 1980). In this respect, the notion of bureaucracy and state apparatus in our model is also different from that of Max Weber, which views bureaucracy as an “apolitical” organization, with no goals or interests. See also Alford and Friedland (1985) for a critical discussions of Marxist and non-Marxist theories of the state.

wage income). The role of bureaucrats is tax collection. In particular, we will allow for a linear tax rate $\tau_t \in [0, 1]$ on earned incomes in order to finance the infrastructure investment K , additional spending on the public good G_t and the wages of bureaucrats. This tax rate is the same irrespective of whether the individual is rich or poor. To simplify the discussion, we assume that only poor agents can become bureaucrats. This assumption is not necessary for the results, since it will be evident below that low-productivity poor agents always prefer bureaucracy more than the high-productivity rich agents (see Remark 2 below).

Both rich and poor agents can try to evade taxes. We assume that if an individual tries to evade taxes, he gets caught with probability $p(x_t)$, where $p : [0, 1] \rightarrow [0, 1]$ is an increasing, twice continuously differentiable, and strictly concave function with $p(0) = 0$, and x_t denotes the number of bureaucrats exerting positive effort at time t . More formally, this is defined as $x_t = \int_{j \in \mathcal{X}_t} e_t^j dj$. This expression incorporates the fact that bureaucrats who do not exert effort are not useful.¹²

If an individual is caught evading taxes, all of his income during that period is lost. For simplicity, we assume that this income does not accrue to the government either (though this is not an important assumption). We also assume that there is *full anonymity* in the market, so that the past history of individual producers is not observed. This implies that future punishments on tax evaders are not possible. Moreover, because of limited liability, i.e., $c_t^j \geq 0$, more serious punishments are not possible.

Since effort is costly, bureaucrats will exert effort only if their compensation depends on their effort decision. We assume that if they do not exert effort, bureaucrats are caught with probability q_t at time t . If they are not caught, they receive the wage w_t , and if they are caught shirking, they lose their wage, but are not fired from the bureaucracy. This assumption simplifies the algebra and the exposition considerably and is relaxed in subsection 5.1 below.

The probability of detection q_t depends on the “organization of the state”. We allow two types of organizations, represented by $I_t \in \{0, 1\}$, which correspond to different types/degrees of monitoring within the bureaucracy. In particular, we assume $q(I = 1) = 1$, so that one of the possible organizations of the state allows for perfect monitoring of bureaucrats. In contrast, $q(I = 0) = q_0 < 1$, so that the alternative organization involves an imperfect monitoring technology and shirking bureaucrats are not necessarily detected. Clearly, imperfect monitoring will lead to equilibrium distortions, and for this reason we refer to $I = 0$ as an “inefficient organization of the state” and to $I = 1$ as “efficient state”. To simplify the analysis we assume

¹²Alternatively, instead of inducing bureaucrats to exert effort, it may be important to ensure that they do not accept bribes from the individuals supposed to pay taxes (e.g., Acemoglu and Verdier, 1998, 2000). We investigate a variant of our model with corruption in subsection 5.3.

that $I = 1$ has no cost relative to $I = 0$.¹³

At each date, the political system chooses the following policies:

- A tax rate on all earned income $\tau_t \in [0, 1]$.
- The wage rate for bureaucrats $w_t \in \mathbb{R}_+$.
- A level of public good $G_t \in \mathbb{R}_+$.
- The number of bureaucrats hired, $X_t \in [0, 1]$.
- A decision on the organization of the state for the next date, $I_t \in \{0, 1\}$ —the efficiency of the state at the current date, I_{t-1} , is part of the state variable, determined by choices in the previous period.

The additional restrictions on these policies are as follows:

1. The government budget constraint (specified below) has to be satisfied at every date.
2. If $X_t \geq X_{t-1}$, then existing bureaucrats cannot be fired (although each bureaucrat can decide to quit if he finds this beneficial). Moreover, if $X_t < X_{t-1}$, then no new bureaucrats are hired and a fraction $(X_t - X_{t-1})/X_t$ of the bureaucrats is fired (those fired being randomly chosen irrespective of their past history).

We denote a vector of policies satisfying these restrictions by $\rho_t \equiv (\tau_t, w_t, G_t, X_t, I_t) \in \mathcal{R}$.

3.2 Description of the Political System

We will consider three different political environments:

1. Permanent nondemocracy: the rich elite are in power at all dates, meaning that only the rich can vote, and since all rich agents have the same policy preferences over the available set of policies, the policy vector most preferred by a representative rich elite will be implemented.
2. Permanent Democracy: the citizens, who form the majority, are in power at all dates starting at $t = 0$ (or at all dates there are elections as described below).
3. Emerging Democracy: the rich elite are in power at $t = 0$, and in all future dates, the regime will be democratic with majoritarian elections.

¹³In general, one can imagine that setting up a more efficient state apparatus may involve additional expenditures. We ignore those both to simplify the algebra and also to highlight that inefficient states can arise *even when* an efficient organization is costlessly available.

The first two environments are for comparison. The third one is our main focus in this paper. It is a simple way of capturing the idea that some decisions are originally taken by elites, anticipating that democracy will arrive at some point—in this case right at date $t = 1$.¹⁴

To start with, we model the democratic system in a very simple way, by assuming that there are two parties, one run by an elite agent and one run by a poor agent, and that bureaucrats cannot run for office. We use the symbols P and R to denote these parties and $d_t = P$ denotes that party P is elected to office at date t . Parties are unable to make commitments to the policies they will implement once they come to power. Thus whichever party receives the majority of the votes comes to power and the agent in control of the party chooses the policy vector that maximizes his own utility. This last assumption departs from the standard Downsian models of political competition where parties commit to their policy platform before the election. Instead, it is closer to the literature on citizen-candidate models, which will be discussed further in subsection 5.2 (see also Alesina, 1988). Specifically, in subsection 5.2, we will consider a richer model of democratic politics, where each agent can run as a citizen-candidate, and we will show that the same results apply with this richer setup. Nevertheless, it is useful to start with the simpler environment with only two parties to highlight the main economic forces.

3.3 Timing of Events

To recap, the timing of events within each date is:

- The society starts with some political regime, nondemocracy or democracy, i.e., $s_t \in \{N, D\}$, a set $\mathcal{X}_{t-1} \subset \mathcal{L}$ of agents who are already bureaucrats (since, by assumption, the set of bureaucrats \mathcal{X}_{t-1} must be a subset of the set of poor agents), and a level of efficiency of the state, $I_{t-1} \in \{0, 1\}$. Then:
 1. In democracy, all individuals $j \in [0, 1]$ vote for either party P or party R , i.e., individual j decides $v_t^j \in \{P, R\}$.
 2. In democracy, the elected party or in nondemocracy the representative elite agent decides the policy vector $\rho_t \equiv (\tau_t, w_t, G_t, X_t, I_t) \in \mathcal{R}$.
 3. Observing this vector, each individual $j \notin \mathcal{X}_{t-1}$ decides whether to apply to become a bureaucrat, $\chi_t^j \in \{0, 1\}$, and each individual $j \in \mathcal{X}_{t-1}$ decides whether to quit bureaucracy, $\chi_t^j \in \{0, 1\}$ (which is denoted by the same symbol without any risk of confusion).

¹⁴In this case, the society is nondemocratic at date $t = 0$, and we assume that it will become democratic for exogenous reasons at date $t = 1$. It is possible to model democratization as equilibrium institutional change along the lines of the models of endogenous democratization in the literature (see Acemoglu and Robinson, 2006a, for a discussion and references), but doing so would complicate the analysis without generating additional economic insights in the current context.

Naturally, by assumption, $\chi_t^j = 0$ for all the rich agents. The number of bureaucrats at time t is then $\min \left\{ X_t, \int_0^1 \chi_t^j dj \right\}$, i.e., the minimum of the number of bureaucrats chosen by the polity in power and the number of people applying to or remaining in bureaucracy. This also determines the current set of bureaucrats, \mathcal{X}_t .

4. Each bureaucrat decides whether to exert effort, $e_t^j \in \{0, 1\}$, which determines $x_t = \int_{j \in \mathcal{X}_t} e_t^j dj$, and thus the probability of detection of individuals evading taxes.
5. Production takes place and each producer decides whether to evade taxes or not, denoted by $z_t^j \in \{0, 1\}$.
6. A fraction $p(x_t)$ of producers evading taxes are caught.
7. A fraction $q_t = q(I_{t-1})$ of shirking bureaucrats are caught and punished.
8. Taxes are collected, remaining bureaucrats are paid their wage, w_t , and the public good G_t is supplied.

Naturally, the society starts with $\mathcal{X}_{-1} = \emptyset$, i.e., in the initial date there are no incumbent bureaucrats. We also suppose that $I_{-1} = 0$ (though this has no bearing on any of our results except the actions at time $t = 0$, since the choice of $I_t \in \{0, 1\}$ is without any costs).

4 Characterization of Equilibria

We now characterize the equilibrium of the environments described above.

4.1 Definition of Equilibrium

Throughout, we focus on pure strategy Markov Perfect Equilibria (MPE).¹⁵

Recall that Markovian strategies condition only on the payoff-relevant state variables (and on the prior actions within the same stage game). An MPE is defined as a set of Markovian strategies that are best responses to each other given every history. In the current game, the aggregate state vector can be represented as $S_t \equiv (s_t, I_{t-1}, X_{t-1}) \in \mathcal{S}$, where $s_t \in \{N, D\}$ is the political regime at time t , $I_{t-1} \in \{0, 1\}$ is the efficiency of the bureaucracy inherited from the previous period, and X_{t-1} is the size of the bureaucracy inherited from the previous period.¹⁶ Individual actions will be a function of the aggregate state vector S_t and the individual's identity, in particular, $a_t \in \{L, H, B\}$ representing whether the individual is a poor producer,

¹⁵We focus on MPE both because in the current context the MPE is unique and is relatively straightforward to characterize and also because the focus on MPE makes the emergence of a coalition between the rich and the bureaucrats more difficult (since there cannot be "commitment" to future rents for bureaucrats).

¹⁶In addition, for each individual we could specify whether the individual is currently a bureaucrat, i.e., whether $j \in \mathcal{X}_{t-1}$ and whether he is a party leader as part of the individual-specific state vector. Nevertheless, Markovian strategies can be defined without doing this, which simplifies the notation.

rich producer or a bureaucrat. Thus as a function of S_t and a_t , each individual will decide which party to vote for, i.e., $v_t^j \in \{P, R\}$, whether to apply (or to remain) in bureaucracy, $\chi_t^j \in \{0, 1\}$, whether to evade taxes, $z_t^j \in \{0, 1\}$, if the individual is a producer, and whether to exert effort, $e_t^j \in \{0, 1\}$, if the individual is a bureaucrat. Finally, strategies also include the choice of $I_t \in \{0, 1\}$, $\tau_t \in [0, 1]$, $X_t \in [0, n]$, and $G_t \in \mathbb{R}_+$ when the individual is the party leader. Thus Markovian strategies can be represented by the following mapping

$$\sigma : \mathcal{S} \times \{L, H, B\} \rightarrow \{P, R\} \times \{0, 1\}^3 \times [0, 1] \times [0, n] \times \mathbb{R}_+.$$

An MPE is a mapping σ^* that is best response to itself at every possible history.

We will often refer to subcomponents of σ rather than the entire strategy profile, and with a slight abuse of notation, we will use $v(I | a)$ to denote the voting strategy of an individual of group $a \in \{L, H, B\}$ as a function of the efficiency of the state institutions. Moreover, when there is no risk of confusion, we will use the index j to denote individuals or groups interchangeably.

4.2 Preliminary Results

We now state a number of results that will be useful throughout the analysis. Even though many of these results are straightforward, listing and proving them is useful for the exposition.

Lemma 1 *If $p(x_t) < \tau_t$, then $z_t^j = 0$ for all $j \notin \mathcal{X}_t$, i.e., all producers evade taxes at time t .*

Proof. The proof of this lemma is straightforward and is provided for completeness in Appendix A. ■

Since with tax evasion there is no government revenue, Lemma 1 implies that in equilibrium we need to have the following *incentive compatibility constraint* for producers

$$p(x_t) \geq \tau_t$$

to be satisfied. Alternatively, defining

$$\pi(\tau) \equiv p^{-1}(\tau), \tag{1}$$

producers' incentive compatibility constraint can be expressed as:¹⁷

$$x_t \geq \pi(\tau_t). \tag{2}$$

This condition requires the number of bureaucrats exerting effort to be greater than $\pi(\tau_t)$. This constraint is sufficient to ensure that all individuals choose not to evade taxes.

Since $p(\cdot)$ is strictly increasing, continuously differentiable and strictly concave, $\pi(\cdot)$ defined in (1) is strictly increasing, continuously differentiable and strictly convex.

¹⁷This condition can also be interpreted as a “state capacity constraint” since, given the effective size of the bureaucracy, it determines the maximum tax rate.

Lemma 2 *If*

$$w_t < \frac{h}{q_t}, \quad (3)$$

then $e_t^j = 0$ for all $j \in \mathcal{X}_t$, i.e., all bureaucrats will shirk at time t .

Proof. See Appendix A. ■

Remark 1 In subsection 5.1 below, we will allow bureaucrats caught shirking to be fired from bureaucracy. In this case, it is clear that the optimal contract involves firing a bureaucrat if he is caught shirking. Given this, the condition in Lemma 2 will have to be forward-looking, taking into account the future rents that the bureaucrat will lose if caught shirking. In particular, imagine a stationary equilibrium, where today and in all future periods the tax rate is equal to $\hat{\tau}$, the wage rate for bureaucrats is \hat{w} , and the probability of getting caught is \hat{q} , then the necessary condition (3) would become

$$\frac{\hat{w} - h}{1 - \beta} < \hat{q}\beta \frac{(1 - \hat{\tau})A^L}{1 - \beta} + (1 - \hat{q}) \left(\hat{w} + \beta \frac{\hat{w} - h}{1 - \beta} \right),$$

since the left-hand side is what the individual would receive by exerting effort at every date, whereas the right-hand side is the payoff to deviating for one period, and then switching to exerting effort from then on (implicitly using the one-step ahead deviation principle, see Fudenberg and Tirole, 1991, Chapter 4). In particular, the right-hand side has the individual getting caught with probability \hat{q} , receiving nothing today and the wage of a low-skill producer from then on, and not getting caught with probability $1 - \hat{q}$, in which case he receives \hat{w} today and then receives the discounted version of the left-hand side (as he switches back to exerting effort). A bureaucrat who loses his job always receives the wage of a low-skill producer from then on, since along the equilibrium path, there will be no further hiring into bureaucracy. Rearranging terms, the above inequality can be expressed as:

$$\hat{w} < \beta(1 - \hat{\tau})A^L + \frac{(1 - \beta(1 - \hat{q}))h}{\hat{q}}. \quad (4)$$

In a stationary equilibrium where bureaucrats are fired when caught shirking, condition (4) will replace (3), and when it is satisfied, all bureaucrats will shirk. Correspondingly, the incentive compatibility constraint, (5), below will change to the converse of this condition. We return to a further analysis of this case in subsection 5.1.

If bureaucrats are expected to shirk, all individuals will evade taxes and there will be no tax revenues. Consequently, the infrastructure investment K could not be financed and there would be no production. Thus, the society also needs to satisfy the *incentive compatibility constraint* of the bureaucrats given by

$$w_t \geq \frac{h}{q_t}, \quad (5)$$

where $q_t = q(I_{t-1})$. This constraint is necessary and sufficient to ensure that all bureaucrats choose to exert effort. In addition, (poor) individuals must prefer to become bureaucrats. That is, the participation constraint

$$w_t \geq (1 - \tau_t) A^L + h, \quad (6)$$

needs to be satisfied so that bureaucrats receive at least as much as they would obtain in private production.

Remark 2 If rich agents could become bureaucrats, the equivalent participation constraint, corresponding to (6), for rich agents would be

$$w_t \geq (1 - \tau_t) A^H + h.$$

Comparison of this inequality with condition (6) makes it clear that poor agents are always more willing to enter bureaucracy than rich agents. Our assumption that rich agents cannot become bureaucrats therefore enables us to avoid imposing explicit conditions to ensure that this inequality is not satisfied and (6) is.

The above discussion, in particular Lemmas 1 and 2, immediately establishes the following lemma (proof omitted):

Lemma 3 *In any MPE, conditions (2), (5) and (6) must hold and $e_t^j = 1$ for all $j \in \mathcal{X}_t$ and all t , and $z_t^j = 1$ for all $j \notin \mathcal{X}_t$ and all t .*

In other words, in any equilibrium the incentive compatibility constraints of producers and bureaucrats and the participation constraint of bureaucrats are satisfied, and no producer evades taxes and all bureaucrats exert effort.

From Lemma 3 (and the fact that only poor agents become bureaucrats), it immediately follows that, as long as the constraints (2) and (5) are satisfied, the government budget constraint can be written as:

$$K + G_t + w_t X_t \leq (1 - n) \tau_t A^H + (n - X_t) \tau_t A^L, \quad (7)$$

where the left-hand side is government expenditures, consisting of the investment in infrastructure, spending on public goods and bureaucrats' wages, while the right-hand side is total government tax receipts. This expression takes into account that all bureaucrats exert effort and no producer evades taxes. Moreover, (7) highlights that in our model, taxation reduces output through a particular general equilibrium mechanism; the government can raise taxes only by hiring bureaucrats and bureaucrats themselves do not produce any output.

Finally, the following lemma is immediate and is stated without proof:

Lemma 4 *Rich agents always vote for party R, i.e., for all $j \in \mathcal{H}$, $v_t^j = R$, and poor producers always vote for party P, i.e., for all $j \in \mathcal{L}$ and $j \notin \mathcal{X}_{t-1}$, $v_t^j = P$.*

4.3 Equilibria under Permanent Democracy and Nondemocracy

Equilibria under permanent democracy and permanent nondemocracy are of interest as a comparison to our main political environment, which involves the society starting as nondemocratic and then transitioning to democracy. The following results are straightforward:

Proposition 1 *Under permanent democracy, there exists a unique MPE. In this equilibrium, at each $t \geq 0$ $d_t = P$ and the following policy vector is implemented at each $t > 0$:*

$$I_t = 1, \quad w_t = (1 - \tau^D) A^L + h, \quad X_t = \pi(\tau^D),$$

$$G_t = G^D \equiv (1 - n) \tau^D A^H + [n - \pi(\tau^D)] \tau^D A^L - [(1 - \tau^D) A^L + h] \pi(\tau^D) - K, \quad (8)$$

and τ^D is the unique solution to the maximization problem:

$$\max_{\tau, G} (1 - \tau) A^L + G \quad (9)$$

subject to

$$G = (1 - n) \tau A^H + [n - \pi(\tau)] \tau A^L - [(1 - \tau) A^L + h] \pi(\tau) - K.$$

Proof. By Lemma 4, for all $j \in \mathcal{L}$, $v_t^j = P$. Under permanent democracy, the poor can vote and form the majority starting at $t = 0$, thus $d_t = P$ for all t . Then the payoff to the decisive voter $j' \in \mathcal{L}$ can be written as

$$V_t^{j'} = (1 - \tau) A_t + G_t + \beta V_{t+1}^{j'}(\sigma^*),$$

where again σ^* is the optimal policy and $\beta V_{t+1}^{j'}(\sigma^*)$ is the discounted optimal continuation value for this individual. The continuation value $\beta V_{t+1}^{j'}(\sigma^*)$ is unaffected by current policies, thus the optimal policy can be determined as a solution to the following program:

$$\max_{\tau, w, X, I, G} (1 - \tau) A^L + G \quad (10)$$

subject to

$$\pi(\tau) \leq X$$

$$\max \left\{ \frac{h}{q(I_t)}, (1 - \tau) A^L + h \right\} \leq w$$

$$G \leq (1 - n) \tau A^H + [n - X] \tau A^L - wX - K$$

$$0 \leq G.$$

It is evident that $I_t = 1$ relaxes the second constraint relative to $I_t = 0$, so will always be chosen in all periods $t \geq 0$. Moreover, there cannot be a solution in which any one of the first three constraints is slack (since this would allow an increase in G , raising the value of the objective function), so we have $X = \pi(\tau)$ and $w = \max \{h, (1 - \tau) A^L + h\} = (1 - \tau) A^L + h$. Substituting these equalities yields (9) for all periods where $I_t = 1$, i.e., for all $t > 0$. Strict convexity of $\pi(\cdot)$ then ensures that τ^D is uniquely defined. ■

Proposition 2 *Under permanent nondemocracy, there exists a unique MPE. In this equilibrium, the following policy vector is implemented at each $t > 0$:*

$$I_t = 1, \quad w_t = (1 - \tau^N) A^L + h, \quad X_t = \pi(\tau^N), \quad G_t = G^N \equiv 0,$$

and τ^N is the unique solution to the equation

$$[(1 - \tau) A^L + h] \pi(\tau) - (1 - n) \tau A^H - [n - \pi(\tau)] \tau A^L + K = 0. \quad (11)$$

Proof. Under permanent nondemocracy, the rich retain political power forever. Then the payoff to the representative rich individual $j' \in \mathcal{H}$ can be written as

$$V_t^{j'} = (1 - \tau) A^H + G_t + \beta V_{t+1}^{j'}(\sigma^*),$$

where σ^* is the optimal policy and $\beta V_{t+1}^{j'}(\sigma^*)$ is the discounted optimal continuation value for this individual. Because the continuation value $\beta V_{t+1}^{j'}(\sigma^*)$ is unaffected by current policies, the optimal policy can be determined as a solution to the following program:

$$\begin{aligned} \max_{\tau, w, X, I, G} \quad & (1 - \tau) A^H + G & (12) \\ \text{subject to} \quad & \\ & \pi(\tau) \leq X \\ \max \left\{ \frac{h}{q(I_t)}, (1 - \tau) A^L + h \right\} & \leq w \\ G & \leq (1 - n) \tau A^H + [n - X] \tau A^L - wX - K \\ 0 & \leq G. \end{aligned}$$

It is again evident that $I_t = 1$ relaxes the second constraint relative to $I_t = 0$, so will always be chosen. Moreover, the first three constraints must again hold as equalities, so we have $X = \pi(\tau)$ and $w = \max\{h, (1 - \tau) A^L + h\} = (1 - \tau) A^L + h$. Substituting for these equalities in program (12), it follows immediately that $G = 0$, and the strict convexity of $\pi(\cdot)$ again ensures the uniqueness of the solution to (11). ■

The main conclusion from both of these benchmark political environments is that the politically decisive agents choose a policy vector consistent with their own interests, and this always involves an *efficient* organization of the state, i.e., $I_t = 1$ for all $t \geq 0$. There is no reason to make the state inefficient. Consequently, both consolidated democratic and nondemocratic regimes involve $I = 1$. Moreover, in both regimes the capacity of the state is fully utilized in the sense that constraint (2) holds as equality and the minimum number of bureaucrats necessary to prevent tax evasion are employed.

It is straightforward to see that the unique solution (τ^D, G^D) in (9) involves $\tau^D > 0$, since infrastructure spending, $K > 0$, has to be financed (and for the same reason, $\tau^N > 0$

in Proposition 2). However, because raising further revenues involves the employment of bureaucrats which is costly, it is possible that the solution to (9) involves $G^D = 0$. If this were the case, there would be no difference between the political bliss points of poor and rich agents given in Propositions 1 and 2 and thus no interesting political conflict. Therefore, throughout we are more interested in the case where the following condition is satisfied:

Condition 1 *The solution to (9) involves $G^D > 0$.*

It can be verified that if the gap between A^H and A^L is small and $\pi'(\tau)$ is large, this condition will be violated. Therefore, this condition imposes that there is a certain degree of inequality in society and raising taxes is not excessively costly, so that the poor would like a higher level of public good provision than the rich. When Condition 1 is satisfied, it also follows that $\tau^D > \tau^N$, and since $\pi(\cdot)$ is strictly increasing, $\pi(\tau^D) > \pi(\tau^N)$ and the size of the bureaucracy is larger in permanent democracy than in permanent nondemocracy.

4.4 Political Equilibrium with Regime Change

We now look at the more interesting case with regime change—i.e., where at date $t = 0$, the rich are in power and from then on there will be elections. We start with a series of lemmas. Our first result shows that with efficient state institutions, the rich will choose their political bliss point as in Proposition 2:

Lemma 5 *In an MPE, if $d_t = R$ and $I_t = 1$, then $w_t = (1 - \tau^N)A^L + h$, $X_t = \pi(\tau^N)$, $G_t = G^N \equiv 0$, and τ^N is given by (11).*

Proof. See Appendix A. ■

The next lemma establishes that the party representing the poor, party P , being elected to office is an “absorbing state,” meaning that once the party of the poor is elected, the results of Proposition 1 apply subsequently:

Lemma 6 *If $d_t = P$, then $d_{t'} = P$ for all $t' \geq t$, and we have the following equilibrium policy vector at all dates $t' > t$:*

$$\begin{aligned} I_t &= 1, \quad w_t = (1 - \tau^D)A^L + h, \quad X_t = \pi(\tau^D), \\ G_t &= G^D \equiv (1 - n)\tau^D A^H + [n - \pi(\tau^D)]\tau^D A^L - [(1 - \tau^D)A^L + h]\pi(\tau^D) - K, \end{aligned} \tag{13}$$

and τ^D is given by (9).

Proof. See Appendix A. ■

The intuition for this result is as follows. Once the party of the poor wins an election, they will choose their preferred policy vector, which includes $I_t = 1$, and given an efficient state,

bureaucrats will have no reason to support the rich party and the poor will continue to win elections in all future periods and the organization of the state will continue to be efficient. An efficient organization of the state ensures that bureaucrats receive no rents and receive the same payoff as poor producers. Thus they will also support party P , and the political bliss point of the poor will be implemented in all future periods. This lemma also implies that when $I_{t-1} = 1$ —i.e., when the state is efficient—the rich will not be able to win a majority. This is related to the basic idea of our approach: the rich can only convince bureaucrats to vote for their party by *committing* to giving them rents and this can only be achieved when the organization of the state is inefficient, i.e., $I_{t-1} = 0$.

We next investigate whether or not the rich may be able to convince the bureaucrats to vote for their party starting with $I_{t-1} = 0$. Since there is no commitment to policies, the party of the rich, when in power, will choose policies in line with its (the rich agents') preferences. The next lemma characterizes these policies starting with $I_{t-1} = 0$.

Lemma 7 *Suppose that $I_{t-1} = 0$, then $w_t = h/q_0$. Moreover, if $d_t = R$, then $G_t = G^E \equiv 0$, and if $d_t = P$, then $G_t = \hat{G}^D$ given by the solution to the following maximization program:*

$$\begin{aligned} & \max_{\tau, G} (1 - \tau) A^L + G & (14) \\ & \text{subject to} \\ G & = (1 - n) \tau A^H + [n - \pi(\tau)] \tau A^L - \frac{h}{q_0} \pi(\tau) - K. \end{aligned}$$

Proof. See Appendix A. ■

Remark 3 As with the solution to the maximization problem (9), the solution to (14) may involve $\hat{G}^D = 0$. With the same reasoning as there, when the level of inequality between the rich and the poor is sufficiently high, the solution to the program (14) will involve $\hat{G}^D > 0$.

The next lemma provides necessary conditions for the party of the rich to win an election starting with $I_{t-1} = 0$:

Lemma 8 *In an MPE, $d_t = R$, i.e., the rich will win the election at time t , if $I_{t-1} = 0$,*

$$(1 - q_0) \frac{h}{q_0} > (1 - \tau^D) A^L + G^D + \frac{1 - \beta}{\beta} \hat{G}^D, \quad (15)$$

and

$$X_t \geq n - \frac{1}{2}, \quad (16)$$

where G^D is given by (8), \hat{G}^D is given by (14), and τ^D is given by (9).

Proof. See Appendix A. ■

Lemma 8 determines the conditions under which the bureaucrats will support party R (a rich agent running for office) and will be numerous enough to give them the majority. Condition (16) requires the size of the bureaucracy to be sufficient to give the majority to party R when all bureaucrats vote with the rich. Nevertheless, $n - 1/2$ may not be the actual size of bureaucracy. In particular, at $X = n - 1/2$, the government budget may not balance. To ensure that it does, we need to consider two cases separately.

Let us first define τ^E as the tax rate that party R would choose as its unconstrained optimal policy to finance the investment in infrastructure, K , given that bureaucratic wages are equal to $w = h/q_0$. Clearly, τ^E is given by the unique solution to the equation

$$\pi(\tau^E) \frac{h}{q_0} - (1-n)\tau^E A^H - [n - \pi(\tau^E)]\tau^E A^L + K = 0. \quad (17)$$

In other words, τ^E balances the government budget when the minimum number of bureaucrats necessary to avoid tax evasion, $X = \pi(\tau^E)$, are employed.

The first case corresponds to the one where $\pi(\tau^E) \geq n - 1/2$, so that the unconstrained optimal size of bureaucracy for party R is also sufficient to make sure that condition (16) is satisfied and the rich have a majority.

The second case applies when this inequality does not hold, i.e., when $\pi(\tau^E) < n - 1/2$. In this case, the unconstrained optimal policy for the rich would *not* satisfy (16), and party R cannot win the election with the minimum number of bureaucrats. Instead, party R can win an election only if $X \geq n - 1/2$, and with this larger size of bureaucracy, budget balance requires the greater tax rate $\hat{\tau}^E$ given by the solution to

$$\left(n - \frac{1}{2}\right) \frac{h}{q_0} - (1-n)\hat{\tau}^E A^H - \frac{1}{2}\hat{\tau}^E A^L + K = 0. \quad (18)$$

It can be verified that whenever $n - 1/2 > \pi(\tau^E)$, we also have $\hat{\tau}^E > \tau^E$, and whenever $n - 1/2 \leq \pi(\tau^E)$, $\hat{\tau}^E \leq \tau^E$. This implies that the size of the bureaucracy necessary for the rich to form a winning coalition is the maximum of $\pi(\tau^E)$ and $n - 1/2$, and correspondingly, the tax rate that party R needs to set is $\max\{\tau^E, \hat{\tau}^E\}$.

The results so far have provided the necessary conditions for the rich to be able to generate sufficient votes from the bureaucrats to remain in power. It remains to check whether the rich prefer to pursue this strategy and commit to an inefficient state in order to maintain political power in democracy. The following lemma answers this question:

Lemma 9 *Suppose that condition (15) holds. Then the rich prefer to set $I_t = 0$ for all t if*

the following condition is satisfied:

$$\begin{aligned} \text{either } \tau^E \geq \hat{\tau}^E, \quad \text{and} \quad & (\tau^D - \tau^E) A^H > G^D, \\ \text{or } \tau^E < \hat{\tau}^E, \quad \text{and} \quad & (1 - \beta) (1 - \tau^E) A^H > \\ & (1 - \beta) (1 - \tau^E) A^H + \beta [(1 - \tau^D) A^H + G^D], \end{aligned} \tag{19}$$

where G^D is given by (8), τ^D is given by (9), τ^E is given by (17), and $\hat{\tau}^E$ is given by (18).

Proof. See Appendix A. ■

Remark 4 If Condition 1 were not satisfied, the conditions in Lemma 9 could never be satisfied. In particular, when Condition 1 does not hold, we have $G^D = 0$ and $\tau^D = \tau^E$, so that neither part of condition (19) could hold. This is a direct consequence of the fact that a significant conflict in policies between the rich and the poor is necessary for the rich to set up an inefficient system of patronage politics.

Now putting all these lemmas together we obtain the main result of this section:

Proposition 3 *Consider the political environment with emerging democracy. If conditions (15) and (19) hold, then there exists a unique MPE. In this equilibrium, the rich elite choose $I_t = 0$ for all $t \geq 0$, the rich party R always remains in power and the following policies are implemented:*

$$\begin{aligned} w_t &= \frac{h}{q_0}, \quad X_t = \max \{ \pi(\tau^E), n - 1/2 \}, \\ G_t &= G^E \equiv 0, \quad \text{and } \tau_t = \max \{ \tau^E, \hat{\tau}^E \}, \end{aligned}$$

where τ^E is given by (17) and $\hat{\tau}^E$ is given by (18).

If, on the other hand, one or both of conditions (15) and (19) hold with the reverse inequality, the unique MPE involves $I_t = 1$ in the initial period, and for all $t \geq 1$, $d_t = P$ and the unique policy vector is

$$\begin{aligned} w_t &= (1 - \tau^D) A^L + h, \quad X_t = \pi(\tau^D), \\ G_t &= G^D \equiv (1 - n) \tau^D A^H + [n - \pi(\tau^D)] \tau^D A^L - [(1 - \tau^D) A^L + h] \pi(\tau^D) - K, \end{aligned}$$

and τ^D is given by (9).

Proof. The first part of the proposition follows immediately from combining Lemma 8 and Lemma 9, which provide the conditions, summarized by (15) and (19), under which the party of the rich, R , can convince the bureaucrats to vote for them, and this is desirable for the rich relative to living under the rule of party P . When (15) or (19) does not hold, then party P is in power and the second part of the proposition follows immediately from Lemma 6 and Proposition 1. ■

Remark 5 Proposition 3 does not cover the case in which one of conditions (15) and (19) holds as equality; in this case the MPE is no longer unique. It is straightforward to see that in such a case, either the rich or the poor party could receive the majority of the votes, or the rich could be indifferent between maintaining an inefficient and an efficient state. We do not describe the equilibrium in these cases to avoid repetition and to save space.

Remark 6 It can also be verified that the set of parameter values where $I_t = 0$ emerges as an equilibrium in Proposition 3 is nonempty. A straightforward way of doing this is to consider high values of β as in the proof of Proposition 5 in the Appendix.

Proposition 3 is our first major result. It establishes the possibility that the rich elite, who are in power *temporarily* at time $t = 0$, may choose an inefficient state organization and a large (inefficient) bureaucracy as a way of credibly committing to providing rents to bureaucrats. This enables them to create a majority coalition consisting of themselves and the bureaucrats, and thus capture democratic politics. This coalition implements policies that support low redistribution and low provision of public goods, but creates high rents for bureaucrats. Perhaps more interestingly, after $t = 1$, even when the society is democratic, the inefficient state institutions *persist* and the rule of the rich continues. This is in spite of the fact that at any date these inefficient institutions can be reformed at no cost and made more efficient. The reasoning is related to the formation of the coalition between the rich and the bureaucrats in the first place. The rich realize that they will be able to maintain power only by keeping an inefficient state structure and creating sufficient rents for bureaucrats. If these rents disappear, bureaucrats will ally themselves with the poor, since their net income will be the same as the net income of poor producers (recall Lemmas 5 and 6). It is precisely the presence of inefficient state institutions creating rents for the bureaucrats that induces them to support the policies of the rich. Recognizing this, when in power the rich choose to maintain the inefficient state structure. At the next date, the party representing the rich receives the support of the bureaucrats and the rich; consequently, the rich remain in power and the cycle continues. The model therefore generates a political economy theory for both the emergence and the persistence of inefficient state institutions.¹⁸

It is also noteworthy that even though taxes are lower in the equilibrium with inefficient state than they would have been under permanent democracy (recall Proposition 1 and Lemma 9), the size of the bureaucracy can be greater than under permanent democracy. This could be the case when the rich elite hire more bureaucrats than necessary for preventing tax evasion in order to create a majority in favor of the persistence of the inefficient state—i.e., in the case

¹⁸The nature of persistence here is different from the persistence of policies arising in Coate and Morris (1999), Hassler et al. (2003), or Gomes and Jehiel (2005), because the focus is not on persistence of a certain set of collective decisions within a given institutional framework, but on the persistence of the inefficiency of state institutions.

where $X > \pi(\tau^E)$. In particular, note that bureaucracy will be more numerous under the control of the elite than in democracy whenever

$$\pi(\tau^D) < n - 1/2.$$

Since in this case equation (19) implies that $\tau^E < \tau^D$, we must also have $\pi(\tau^E) < \pi(\tau^D) < n - 1/2$ and thus

$$X > \pi(\tau^E).$$

Consequently, the rich not only choose an inefficient state organization, but they also choose *overemployment* of bureaucrats, in the sense that bureaucracy is now unnecessarily large and the number of bureaucrats is strictly greater than that necessary for tax inspection. The capture of democratic politics by the rich elite therefore creates an inefficient state, with poorly monitored and overpaid bureaucrats, and also leads to a situation in which the capacity of the state is not fully utilized. These inefficiencies imply that the allocation of resources in a captured democracy is worse than in a nondemocracy (or than in a perfectly functioning democracy). Naturally, these inefficiencies have a political rationale, which is to increase the number of bureaucrats that will vote for the party aligned with the rich, so that the rich can maintain political power in the future.

Interestingly, because creating an inefficient bureaucracy is more costly than creating an efficient one (which is smaller and gives bureaucrats no rents), the citizens are *worse off* in a nonconsolidated (emerging) democracy, where they are taxed at rate $\max\{\tau^E, \hat{\tau}^E\}$, than they would be under a consolidated nondemocracy, where they are only taxed at rate $\tau^N < \max\{\tau^E, \hat{\tau}^E\}$. Moreover, the rich are also worse off in this equilibrium than they would be in a permanent nondemocracy, since they are paying higher wages to bureaucrats and possibly employing an excessive number of them.

4.5 Comparative Statics

We next investigate the conditions under which the equilibrium involves the emergence and persistence of inefficient state institutions. The following proposition establishes that a certain degree of inequality between the poor and the rich (i.e., a high level of A^H/A^L), a sufficiently high discount factor, β , and intermediate bureaucratic rents, $(1 - q_0)h/q_0$, are necessary for the emergence of inefficient state institutions.

Proposition 4 *Consider an economy characterized by the parameters $(\beta, n, A^L, A^H, K, h, q_0)$ and the function $p(\cdot)$. Holding all other parameters constant, we have:*

1. *there exists $a > 1$ such that if $A^H/A^L \leq a$, then the state is always efficient, i.e., $I_t = 1$;*
2. *there exist $a' > 1$ and $\bar{\beta} \in (0, 1)$ such that as long as $A^H/A^L \geq a'$, $\beta \leq \bar{\beta}$ implies $I_t = 1$;*

3. there exists $\underline{\theta} > 0$ and $\bar{\theta}$ such that if $(1 - q_0)h/q_0 \notin (\underline{\theta}, \bar{\theta})$, then $I_t = 1$.

Proof. For the first part simply recall Remark 4; inspection of the maximization problem (9) immediately shows that as $A^L \rightarrow A^H$, Condition 1 will be violated and the conditions in (19) cannot hold. Then the result follows from Lemma 9 and Proposition 3.

For the second part, recall from Remark 3 that some minimal level of inequality, say $A^H/A^L \geq a'$, is necessary for $\hat{G}^D > 0$. Suppose this is the case. From Proposition 3, condition (15) is necessary for $I_t = 0$. Since $\hat{G}^D > 0$, there exists $\beta_0 \in (0, 1)$ such that $(1 - q_0)h/q_0 = \beta_0 \hat{G}^D / (1 - \beta_0)$. Since the sum of the other terms on the right hand side of (15) is positive, this implies that there exists $\bar{\beta} < \beta_0$ such that for all $\beta \leq \bar{\beta}$ (15) will be violated and thus $I_t = 1$.

For the third part, note that bureaucratic rents are equal to $h/q_0 - h = (1 - q_0)h/q_0$, which needs to be greater than or equal to the right hand side of (15). Let this right hand side be denoted by $\underline{\theta}$ (and note that $\underline{\theta} > 0$). If $(1 - q_0)h/q_0 < \underline{\theta}$, then (15) will be violated and $I_t = 1$. This implies that we need $(1 - q_0)h/q_0 \geq \underline{\theta} > 0$. Next observe from (17) that there exists a value of $(1 - q_0)h/q_0$, say $\bar{\theta}_0$, such that $\hat{\tau}_E = 1$. It is evident that when $\hat{\tau}_E = 1$, condition (19) cannot be satisfied, thus $I_t = 1$. This implies that for $I_t = 0$, we need $h/q_0 \leq \bar{\theta}_0$ and thus $(1 - q_0)h/q_0 \leq \bar{\theta}$. ■

The first part of the proposition implies that a certain level of inequality is necessary for the emergence of an inefficient state. This is intuitive; with limited inequality, democracy will not be redistributive and it will not be worthwhile for the rich to set up an inefficient bureaucracy in order to keep the poor away from power. The second part implies that the high discount factor is also necessary for the emergence of the inefficient state. This follows because bureaucrats vote for party R as an “investment”, that is, to obtain higher returns in the future. Instead, if they deviate and vote for party P , in the current period they receive both the same high wages (since $I_t = 0$) and the positive level of public good provided by party P , $\hat{G}^D > 0$. If their discount factor were very small, it would be impossible for rich agents to convince bureaucrats to support their party.¹⁹ Finally, the third part of the proposition implies that bureaucratic rents need to take intermediate values. If bureaucratic rents are very small, bureaucrats would not support the party of the rich. If they are very large, it becomes prohibitively costly for the rich to control democratic politics.

While Proposition 4 shows that a certain degree of inequality is necessary for $I_t = 0$, it does not establish that inequality has a monotonic effect on the likelihood of an inefficient state. The next proposition establishes this result under somewhat more restrictive assumptions.

¹⁹Robinson (2001) and Acemoglu and Robinson (2006b) also obtain the result that higher discount factors may lead to greater inefficiencies. However, in these models the source of inefficiency is very different. In particular, inefficient political equilibria arise when pivotal agents—elites or rulers—are sufficiently patient and thus take inefficient actions in order to secure their future political survival.

In this proposition, by greater inequality we mean a mean-preserving spread of the income distribution in the economy, i.e., a simultaneous increase in A^H and decrease in A^L such that mean income, $Y = (1 - n)A^H + nA^L$ remains constant.

Proposition 5 *Suppose that $\pi(\tau)$ is log-concave in τ and τ^D given by (9) satisfies $\tau^D < 1 - \pi(\tau^D) < 1$. Then there exists $\tilde{\beta} \in (0, 1)$ such that for all $\beta \geq \tilde{\beta}$, greater inequality makes the inefficient state equilibrium, i.e., $I_t = 0$, more likely.*

Proof. See Appendix A. ■

Remark 7 The condition that $\pi(\tau)$ is log-concave is not very restrictive. For example, any $p(x)$ that takes the power function form, i.e., $p(x) = P_0x^\alpha$ for $P_0 > 0$ and $\alpha \in (0, 1)$, satisfies this condition. The condition that $\tau^D < 1 - \pi(\tau^D) < 1$ is also natural; if this condition were violated, we would have that the utility of the poor in democracy $(1 - \tau^D)A^L + G^D$ would be non-increasing in A^L (see the Appendix).

In addition to generalizing the first part of Proposition 4, this result implies that taxes (and public spending) can be higher in more equal societies, because unequal societies are more likely to create inefficient bureaucracies to limit taxation and public spending. This result therefore presents an alternative explanation to the often-discussed negative cross-sectional correlation between inequality and redistribution (e.g. Perotti, 1996, Bénabou, 2000).

5 Extensions

In this section, we discuss a number of extensions of our benchmark model. First, we allow bureaucrats to be fired when they are caught shirking, so that the incentive compatibility constraint of bureaucrats is forward-looking and takes into account the rents that a bureaucrat will lose when he gets caught not exerting effort. Second, we allow a richer political environment where each individual can run for office (form a party) as a citizen-candidate, so that bureaucrats can also form their own party and compete against the party of the poor and the rich. Third, we consider the case where the moral hazard problem of bureaucrats arises from their temptation to accept bribes that might be offered by taxpayers.

5.1 Equilibrium When Bureaucrats Can Be Fired

The main result of the previous section, Proposition 3, was derived under the assumption that bureaucrats cannot be fired when they are caught shirking. This simplified the analysis by enabling us to write the incentive compatibility constraint of bureaucrats in the form of condition (5). As discussed in Remark 1 this was mainly for expositional reasons. We now allow bureaucrats to be fired when they are caught shirking. It is clear that from the viewpoint

of discouraging shirking, a contract which commits to firing bureaucrats when they are caught shirking is optimal. The discussion in Remark 1 establishes that, in a stationary equilibrium, the incentive compatibility constraint of bureaucrats, the equivalent of (5), in this case, would be:

$$\hat{w} \geq \beta (1 - \hat{\tau}) A^L + \frac{(1 - \beta (1 - \hat{q})) h}{\hat{q}}. \quad (20)$$

Given this condition, all of the results from the previous section apply with appropriate modifications. In particular we have (proof omitted):

Lemma 10 *Consider the environment where bureaucrats can be fired for shirking. Then in any MPE, if $d_t = R$ and $I_{t-1} = 0$, we have $w_t = \beta (1 - \tilde{\tau}^E) A^L + (1 - \beta (1 - q_0)) h/q_0$ and $G_t = G^E \equiv 0$, where $\tilde{\tau}^E$ is the solution to*

$$\lambda_m \left[\beta (1 - \tilde{\tau}^E) A^L + \frac{(1 - \beta (1 - q_0)) h}{q_0} \right] - (1 - n) \tilde{\tau}^E A^H - [n - \lambda_m] \tilde{\tau}^E A^L + K = 0 \quad (21)$$

where $\lambda_m \equiv \max \{ \pi (\tilde{\tau}^E), n - 1/2 \}$.

Moreover, we have the following generalization of Lemma 9 (proof omitted):

Lemma 11 *Consider the environment where bureaucrats can be fired for shirking. Then in any MPE, the rich will win the election at time t only if there is an inefficient state, i.e., $I_{t-1} = 0$; if bureaucrats prefer to support the party of the rich, i.e., if*

$$\beta (1 - \tilde{\tau}^E) A^L + \frac{(1 - \beta) (1 - q_0) h}{q_0} > (1 - \tau^D) A^L + G^D + \frac{1 - \beta}{\beta} \hat{G}^D, \quad (22)$$

and if the rich-bureaucrat coalition has the majority, i.e., if

$$X_t \geq n - \frac{1}{2}, \quad (23)$$

where $\tilde{\tau}^E$ is given by (21), G^D is given by (8), τ^D is given by (9), and \hat{G}^D is given by the solution to

$$\begin{aligned} & \max_{\tau, G} (1 - \tau) A^L + G \\ & \text{subject to} \\ G &= (1 - n) \tau A^H + [n - \pi(\tau)] \tau A^L - \left[\beta (1 - \tilde{\tau}^E) A^L + (1 - \beta (1 - q_0)) \frac{h}{q_0} \right] \pi(\tau) - K. \end{aligned}$$

Furthermore, the rich prefer this equilibrium and choose $I_t = 0$ at time t only if

$$(\tau^D - \tilde{\tau}^E) A^H > G^D. \quad (24)$$

These two lemmas give the following analogue to Proposition 3:

Proposition 6 *Consider the political environment with emerging democracy and suppose that bureaucrats can be fired if caught shirking. Then, if conditions (22) and (24) hold, the unique MPE is one in which the rich elite choose $I_t = 0$ in the initial period and for all t thereafter, the rich party always remains in power and the following policies are implemented at all dates: $w_t = \beta (1 - \tilde{\tau}^E) A^L + (1 - \beta (1 - q_0)) h/q_0$, $X_t = \max \{ \pi (\tilde{\tau}^E), n - 1/2 \}$, $G_t = G^E \equiv 0$, and $\tau_t = \tilde{\tau}^E$, where $\tilde{\tau}^E$ is given by (21).*

If one or both of conditions (22) and (24) hold with the reverse inequality, the unique MPE involves $I_t = 1$ in the initial period, and for all $t \geq 1$, $d_t = P$ and the unique policy vector $w_t = (1 - \tau^D) A^L + h$, $X_t = \pi (\tau^D)$, $G_t = G^D \equiv (1 - n) \tau^D A^H + [n - \pi (\tau^D)] \tau^D A^L - K - [(1 - \tau^D) A^L + h] \pi (\tau^D)$, and τ^D as given by (9).

Proof. Combining Lemma 10 and Lemma 11 provides the conditions, (22) and (24), under which the party of the rich, R , can convince the bureaucrats to vote for them, and this is desirable for the rich relative to living under the rule of party P . When instead (22) or (24), or both, do not hold, then party P is in power and the second part of the proposition follows immediately from Lemma 6 and Proposition 1. ■

Proposition 6 demonstrates that the results in Proposition 3 generalize to the environment where bureaucrats can be fired if caught shirking. One important difference is worth noting, however. In our main analysis, Proposition 4 showed that a higher discount factor, β , makes the emergence of an inefficient state more likely. Instead, when bureaucrats can be fired, the relationship between the discount factor and the emergence of inefficient states is more complex. Higher β again increases the importance that bureaucrats attach to future rents, but it also reduces the level of rents, because being fired from bureaucracy becomes more costly.

5.2 Political Equilibrium Citizen-Candidates

The previous analysis limited the political system under democracy to a two-party competition between P and R , the two parties representing the interests of the poor and the rich. We justified this by assuming that bureaucrats are not allowed to run for office. Even if bureaucrats are not allowed to run for office, it is possible that a party representing their interest might form. If such a party forms, bureaucrats may vote for that party, and the coalition between the rich and the bureaucrats, choosing low public good provision and low taxes, may not materialize. We now investigate whether in general we expect this to be the case or not when multiple parties can enter the political system.

We follow Osborne and Slivinski's (1996) and Besley and Coate's (1997) citizen-candidate model, where each individual agent can run as a candidate and upon election chooses his most-preferred policy vector. This setup is quite similar to the one we used above, since parties could not make credible policy promises and the policy vector was chosen after a politician (party)

was elected office. The problem with the citizen-candidate models in general is that when more than two parties compete, coordination among the citizens regarding which party has a chance to win the election is important for the outcomes and typically lead to multiple equilibria in the voting stage. To avoid these problems, we consider the following modification of the standard citizen-candidate model:

1. Each individual can decide to form a party and run for office, and this has cost ε , which is taken to be small (in particular, we will consider the case where $\varepsilon \downarrow 0$). Individuals derive no utility from coming to power, but simply benefit from being in power by implementing policies that are in line with their interests.

2. Given all parties that are running for office, individuals vote using ballots with transferable votes, meaning that each individual ranks all parties in strict order of preference. In particular, the vote of individual j can be represented as $\mathbf{v}_t^j = i_1 i_2 i_3$, where i_1, i_2 and i_3 are distinct elements of $\{R, P, B\}$, e.g., $\mathbf{v}_t^j = RBP$. In the first stage, parties are allocated votes according to the first preferences of the voters. Then as is standard with this type of voting rule, the party that gets the lowest fraction of votes is eliminated, and its votes are allocated to the second-ranked choice of the voters who had originally voted for this party. This process continues until one of the parties has a majority.

To simplify the discussion, in this section we assume that bureaucrats cannot be fired if caught shirking, so the incentive compatibility constraint for bureaucrats is given by (5)—though this has no effect on any of the results in the section.

Given this setup, the notion of Markov Perfect Equilibrium is modified accordingly. The analysis in this case is still tractable thanks to the following series of lemmas:

Lemma 12 *Truthful ranking is a weakly dominant strategy for each individual.*

Proof. The transferable votes imply that at any stage of the elimination process, either an individual is pivotal, has a choice between two options, and thus is better off ranking his more preferred outcome above a less preferred outcome. Alternatively, the individual is not pivotal, any choice is a best response. This establishes that truthful ranking is weakly dominant. ■

Lemma 13 *In any MPE, there will never be more than one party operated by an individual of the same group. Thus the maximum number of parties is three.*

Proof. The result follows since the policies chosen by two parties run by two poor agents (or two rich agents or two bureaucrats) will be identical. Moreover, from Lemma 12, each agent ranks parties truthfully, thus the addition of a new party will not change the equilibrium probability that a party run by a poor individual, a rich individual or a bureaucrat wins the election. Thus conditional on a party run by a poor agent existing, there is no point for any other poor agent to incur the cost $\varepsilon > 0$ and form a party. ■

Lemma 13 then enables us to simply look at the (truthful) preference ranking of each individual over at three parties $\{P, R, B\}$, corresponding to parties run by a poor individual, a rich individual and a bureaucrat (there is no source of confusion in this notation, since there can at most be one party run by a poor agent, one run by a rich agent, and one run by a bureaucrat). To do this, we need to know the policies that will be chosen by the three types of parties. Our previous analysis already establishes the policies that will be chosen by parties P and R (provided that party R is trying to come to office by attracting the votes of bureaucrats). We therefore only need to look at the policy choice of a party run by a bureaucrat. The following lemma characterizes this choice:

Lemma 14 *Taking future election results as given, the party $d_t = B$ would choose the following policy vector: $G_t = 0$, $X_t = \min \{X_{t-1}, \pi(\tau^B)\}$, and (τ^B, w^B) such that*

$$\begin{aligned} w_B &= \arg \max_{\tau, w} w \\ &\text{subject to} \\ \min \{X_{t-1}, \pi(\tau)\} w + K &\leq (1-n)\tau A^H + [n - \min \{X_{t-1}, \pi(\tau)\}] \tau A^L. \end{aligned}$$

Proof. This immediately follows by writing the program to maximize the return to a bureaucrat (without allowing firing of existing bureaucrats):

$$\max_{\tau, w, X_t, I, G} w + G$$

subject to

$$\begin{aligned} \min \{X_{t-1}, \pi(\tau)\} &\leq X \\ \max \left\{ \frac{h}{q(I_t)}, (1-\tau)A^L + h \right\} &\leq w \\ G &\leq (1-n)\tau A^H + [n - X]\tau A^L - wX - K \\ 0 &\leq G. \end{aligned}$$

■

Intuitively, bureaucrats would maximize their wages subject to the government budget constraint. Notice that Lemma 14 applies taking the results of future elections as given. If the current bureaucratic government could influence the outcome of future elections, this could be beneficial for it only by increasing X_t above $\min \{X_{t-1}, \pi(\tau^B)\}$, which would (from the government budget constraint) make this policy vector even less attractive to poor and rich agents.

The key to the results in this section is the following observation: because a bureaucratic government will maximize wages paid to bureaucrats (and provide no public goods), it yields

a lower utility to poor agents than a rich government would do. As a result, we will see that a bureaucratic government will never get elected. To show this more formally, let us denote the vote of individual j at time t by \mathbf{v}_t^j , which is a ranking over $\{P, R, B\}$. For example, $\mathbf{v}_t^j = PRB$ means that the individual ranks the poor party first, the rich party second in the bureaucratic party last.

We now have the following rankings for individuals:

Lemma 15 *If $j \in \mathcal{H}$, then $\mathbf{v}_t^j = RPB$.*

If $j \in \mathcal{L}$ and $j \notin \mathcal{X}$, then $\mathbf{v}_t^j = PRB$.

If $j \in \mathcal{X}$, then $\mathbf{v}_t^j = BRP$.

Proof. We have already established that voters rank parties truthfully, so that all voters rank their own party first. Assuming that party R implements the policy characterized in Proposition (3) to attract the bureaucrats, we have that bureaucrats indeed prefer the rich to the poor as second choice; hence, if $j \in \mathcal{X}$, then $\mathbf{v}_t^j = BRP$. Moreover, the poor prefer the rich to the bureaucrats since neither of them offers any public good, but the rich tax less than the bureaucrats. This follows since both the rich and the bureaucrats choose to finance K , and party B chooses a wage w^B for bureaucrats higher than the wage h/q_0 that the bureaucrats get if the rich are in power. Hence, if $j \in \mathcal{L}$ and $j \notin \mathcal{X}$, then $\mathbf{v}_t^j = PRB$. Finally, the second choice of the rich is for the poor, both because the poor would provide a positive amount of the public good rather than zero as the bureaucrats would, and because the poor would tax less than the bureaucrats given that the marginal cost of taxation for the poor is positive, and zero for the bureaucrats (who are not taxed by assumption). It follows that if $j \in \mathcal{H}$, then $\mathbf{v}_t^j = RPB$. ■

Lemma 15 implies that the poor, when they cannot have a majority by themselves, will support the rich party, thus as long as the bureaucrats are not in majority by themselves, i.e., $X_t < 1/2$ and the rich pursue the policy in Proposition 3, we will have $d_t = R$. This implies that the rich can continue to use same political strategies as in the previous section to control political decision-making in democracy.

Now combining the previous lemmas, we have the following proposition, which mirrors Proposition 3.

Proposition 7 *Consider the political environment with emerging democracy and free political entry by citizen candidates. Suppose $\varepsilon \downarrow 0$ that and that conditions (15) and (19) and $\pi(\tau^E) < 1/2$, where τ^E is defined by (17) above. Then, in any MPE of the citizen-candidate political game, only a party run by a rich agent is active. The unique equilibrium policy vector is given by $I_t = 0$, $w_t = h/q_0$, $X_t = \max\{\pi(\tau^E), n - 1/2\}$, $G_t = G^E \equiv 0$, and $\tau_t = \max\{\tau^E, \hat{\tau}^E\}$, for all t , where $\hat{\tau}^E$ is given by (18).*

If one or both of conditions (15) and (19) holds with the reverse inequality and $\pi(\tau^D) < 1/2$, where τ^D is defined by (9), then the unique MPE involves only a party run by a poor agent is active, and the unique equilibrium policy vector involves $I_t = 1$ for all t , and for all $t \geq 1$, $w_t = (1 - \tau^D) A^L + h$, $X_t = \pi(\tau^D)$, $G_t = G^D \equiv (1 - n) \tau^D A^H + [n - \pi(\tau^D)] \tau^D A^L - [(1 - \tau^D) A^L + h] \pi(\tau^D) - K$.

Proof. This proposition can be proved by backward induction. First, suppose that conditions (15) and (19) hold. Then, from Lemma 15, when $X < 1/2$, the bureaucratic party will never win an election. The assumption that $\mathcal{X}_{-1} = \emptyset$ implies that in the initial period $X_{-1} < 1/2$, and the assumption that $\pi(\tau^E) < 1/2$ ensure that $X < 1/2$ continues to be the case when the rich party is in power. Therefore, when the rich party is in power, no bureaucrat incurs the cost $\varepsilon \downarrow 0$ to form a party, and thus bureaucrats support party R by the same argument as in the proof of Proposition 3. Next, knowing that bureaucrats support party R , no poor agent incurs the cost $\varepsilon \downarrow 0$ to compete against party R as long as party R is choosing the policy in Proposition 3 (if they did deviate from this policy, then a poor party can win an election, and thus a poor agent will find it beneficial to enter and form a party since $\varepsilon \downarrow 0$; thus despite the fact that party P would not be running, party R has to adopt the same policy vector as in Proposition 3). Finally, since $\varepsilon \downarrow 0$, it cannot be an equilibrium for no rich agents to form a party, since such a party would create strictly positive gains for each rich agent, and the cost of creating a party is $\varepsilon \downarrow 0$.

The proof of the cases where one or both of conditions (15) and (19) hold with the reverse inequality is similar. ■

This proposition therefore shows that our main results regarding the use of an inefficient state as a way by the rich elite to control the democratic political process continue to apply even when the political structure is enriched to allow free entry by citizen-candidates of any occupation. The additional insights that is interesting in this case is that when the poor producers prefer to support the party of the rich, R , rather than the party of the bureaucrats, B , since the latter would impose high taxes and provide no public goods (spending all the proceeds on bureaucratic wages).

5.3 Bureaucratic Corruption

We now briefly discuss an extension of our basic model in which the moral hazard problem on the side of bureaucrats is not related to their effort, but to whether or not they accept bribes from producers evading taxes. This source of moral hazard problem is arguably as important as the effort choice of bureaucrats. Moreover, we will see below that it leads to an interesting pattern of *de facto regressive taxation* as a result of successful patronage politics by the rich elite.

The economic and political environment is similar to the baseline version of the model with a two-party system. The only difference is that the bureaucrats no longer have an effort choice. Instead, producers that have evaded taxes can pay a bribe $b \geq 0$ to the bureaucrat inspecting them in order to avoid paying taxes.

Similar to the baseline model, we allow for two levels of monitoring efficiency, described by the state variable $I_t \in \{0, 1\}$. When $I = 1$, there is an efficient organization of the state and corruption is detected with probability $q(I = 1) = 1$. When $I = 0$, the state organization is inefficient and corruption is detected with probability $q(I = 0) = q_0 < 1$. We make a number of assumptions to simplify the exposition. First, we assume that a bureaucrat caught accepting bribes loses his wage and the bribe, but the punishment is limited to only one period; the producer paying the bribes loses the bribe but receives no other punishment. Second, all bribe payments and other income confiscated are lost and thus do not enter the government budget constraint. Third, we assume that after matching with a bureaucrat, the producer has all the bargaining power and makes a take-it-or-leave-it bribe offer to the bureaucrat. All of these assumptions can be relaxed without changing our main results.

Finally, we assume that each bureaucrat can be matched with at most one producer and that, for the relevant part of the parameter values, $p(x) < x/(1-x)$. Note that the function $p(x)$ is concave while $x/(1-x)$ is convex and both are equal to zero for $x = 0$. Therefore, there is a range for $x \in [0, x_m]$ such that $p(x) \geq x/(1-x)$. We assume that x_m is lower than the minimum size of the bureaucracy necessary to finance the infrastructure K , which ensures that the region where $p(x) \geq x/(1-x)$ is irrelevant for the equilibrium.

Let us start with the case where the state is inefficient so that $q = q_0$ and characterize the most preferred policies of the rich. The participation constraint of the bureaucrat is slightly different from (6), since there is no cost of effort. It requires that

$$w_t \geq (1 - \tau_t) A^L. \quad (25)$$

The incentive compatibility constraint for bureaucrats (5) is now replaced by the following “no bribe constraint”:

$$w_t \geq (1 - q_0) (w_t + b_t), \quad (26)$$

where b_t is the bribe offered to the bureaucrat by a producer. Intuitively, the right hand side of (26) represents the expected return of a bureaucrat that accepts a bribe b_t , given by the sum of the wage and the bribe, weighted by the probability of not being detected. If condition (26) does not hold, it is not possible to prevent the corruption of bureaucrats by producers.²⁰ Condition (26) implies that, given the public sector wage w_t , only bribes higher

²⁰Van Rijckeghem and Weder (2001) provide evidence that higher public sector wages relative to manufacturing wages reduce the scope for the corruption of the public administration.

than a threshold $\underline{b}(w_t)$ will be accepted, where

$$\underline{b}(w) \equiv \frac{q_0}{1 - q_0} w. \quad (27)$$

In what follows, we drop time subscripts to simplify notation. When in power, the rich maximize their per-period utility with respect to τ , w , X , G , and the decision variable $z \in \{0, 1\}$, which, as before, designates their decision of whether to pay taxes. The expected utility of the rich when they do not pay taxes is

$$u^H(z = 0) = p(x) \max \left\{ A^H - \frac{q_0}{1 - q_0} w, 0 \right\} + [1 - p(x)] A^H + G. \quad (28)$$

Expression (28) incorporates the following facts: (i) producers are inspected by a bureaucrat with probability $p(x)$; (ii) bribing is detected with probability q_0 ; (iii) the bribe offered by the rich to bureaucrats is equal to the lowest acceptable bribe $\underline{b}(w) \equiv q_0 w / (1 - q_0)$ defined in (27); and (iv) when inspected, the income of a rich producer is $\max \{ A^H - q_0 w / (1 - q_0), 0 \}$. Expression (28) is maximized subject to the following constraints

$$p(x) \max \left\{ A^L - \frac{q_0}{1 - q_0} w, 0 \right\} + [1 - p(x)] A^L \leq (1 - \tau) A^L, \quad (29)$$

$$xw + G + K \leq (n - x) \tau A^L, \quad (30)$$

and subject to the participation constraint (25) of the bureaucrats. Constraint (29) requires that the poor prefer to pay taxes to tax evasion. This constraint has to be satisfied since at least one class must pay taxes, otherwise it would not be possible to finance the infrastructure investment, K (this is because, if the poor prefer to evade taxes, the rich will do so a fortiori). Constraint (30) implies that the government budget constraint is satisfied, taking account of the fact that public revenues come from the taxation of the poor only.

Lemma 16 *Suppose that the rich prefer not to pay taxes. Then their optimal policies involve*

$$\tilde{w}^E \equiv (1 - q_0) A^L / q_0, \quad (31)$$

$$p(x) = \tau, \quad (32)$$

and $\tilde{G}^E = 0$ for some $\tau \in [0, 1]$.

Proof. See Appendix A. ■

We will next show that given (31) and (32), the equilibrium involves tax evasion by the rich. Substituting for these expressions, we obtain the utility of the rich when they evade taxes as

$$u^H(z = 0) = (1 - \tau) A^H + \tau (A^H - A^L). \quad (33)$$

In contrast, when a rich agent pays the tax rate (while all others evade taxes) his utility would be

$$\begin{aligned} u^H(z=1) &= (1-\tau)A^H, \\ &< (1-\tau)A^H + \tau(A^H - A^L), \\ &= u^H(z=0). \end{aligned} \tag{34}$$

Next let $\check{\tau}^E$ denote the unique value of τ satisfying the government budget constraint (30), at the candidate equilibrium with the rich agents evading taxes

$$\pi(\check{\tau}^E) \frac{1-q_0}{q_0} A^L + K = [n - \pi(\check{\tau}^E)] \check{\tau}^E A^L, \tag{35}$$

where $\pi(\cdot)$ is again defined in (1).

As in the main analysis, there are two cases to consider depending on whether $n - \pi(\check{\tau}^E)$ is greater than or less than $1/2$. Here we simplify the analysis by focusing on the case where there are sufficiently many bureaucrats so that, together with the rich, they are the absolute majority, i.e., $n - \pi(\check{\tau}^E) \leq 1/2$. The converse case with $n - \pi(\check{\tau}^E) > 1/2$ necessitates that the rich create an inefficiently large bureaucracy in order to win the election. Since the results in this case are again similar, we do not discuss them in this extension.

Lemmas 5 and 6 continue apply in this modified environment. In particular, if bureaucrats ever vote for the poor, there is a permanent transition to an equilibrium with an efficient state with the poor in power within one period from the election. The following lemma characterizes the policy vector that the poor would implement in the period they win the election when the existing organization of the state is inefficient and also the policy vector that they will choose when the state is efficient.

Lemma 17 *Suppose that $d_t = P$ and consider the following maximization program:*

$$\begin{aligned} &\max_{\tau, G, w} (1-\tau)A^L + G \\ &\text{subject to} \\ G &= z(1-n)\tau A^H + [n - \pi(\tau)]\tau A^L - w\pi(\tau) - K \\ p(\pi(\tau)) \max \left\{ A^L - \frac{q(I)}{1-q(I)}w, 0 \right\} &+ [1 - p(\pi(\tau))]A^L \leq (1-\tau)A^L, \\ p(\pi(\tau)) \max \left\{ A^H - \frac{q(I)}{1-q(I)}w, 0 \right\} &+ [1 - p(\pi(\tau))]A^H \leq (1-\tau)A^H \quad \text{and } z = 1, \text{ or } z = 0, \end{aligned}$$

and $(1-\tau)A^L \leq w$, where $z \in \{0, 1\}$ denotes the decision of the rich whether to pay taxes. Then the policy vector that the poor would choose when $I_{t-1} = 0$, $\tau_t = \hat{\tau}^D$, $G_t = \hat{G}^D$, $w_t = \hat{w}^D$, is given by the solution to this program when $q(I) = q_0$. The policy vector that the poor would choose when $I_{t-1} = 1$, is given by $\tau_t = \check{\tau}^D$, $G_t = \check{G}^D$, $w = \check{w}^D$, when $q(I) = 1$ and the first term in the last two inequalities is equal to zero.

The penultimate inequality in the maximization program in this lemma represents the “no tax evasion” constraint for the poor, while the last constraint allows the program to choose whether or not to satisfy the no tax evasion constraint of the rich. Notice that if this last constraint is satisfied, the penultimate one will also be satisfied automatically (since $A^H > A^L$). When $I_{t-1} = 1$ and the state is efficient, bribery is not possible and the max term in the last two inequalities becomes zero.

The next two lemmas are the analogues of Lemmas 8 and 9 and determine the conditions under which the bureaucrats are willing to vote for the rich, and the rich prefer the allocation in which they are in power to the one in which the poor are in power. Since their proofs are similar to those of Lemmas 8 and 9, they are omitted.

Lemma 18 *In an MPE, the rich will win the election at time t (i.e., $d_t = R$) if only if $I_{t-1} = 0$,*

$$\frac{1 - q_0}{q_0} A^L > (1 - \beta) (\hat{w}^D + \hat{G}^D) + \beta \left[(1 - \check{\tau}^D) A^L + \check{G}^D \right], \quad (36)$$

where \hat{w}^D , \hat{G}^D , $\check{\tau}^D$ and \check{G}^D are defined in Lemma 17.

Condition (36) implies that the bureaucrats prefer to be in an inefficient state under the rule of the rich, given the “wage policy” that is optimal for the rich, rather than voting for the poor. In fact, if they vote for the rich, the bureaucrats obtain a wage equal to $(1 - q_0) A^L / q_0$, whereas if they vote for the poor, they obtain a wage of \hat{w}^D and public with provision of \hat{G}^D for one period (while the state is inefficient), and subsequently a payoff equal to the payoff of the poor under an efficient state.

Lemma 19 *Suppose that condition (36) holds. Then, the rich prefer to set $I_t = 0$ for all t if the following condition is satisfied*

$$\left[(1 - \check{\tau}^D) A^H + \check{G}^D \right] < [1 - p(\pi(\check{\tau}^E))] A^H + p(\pi(\check{\tau}^E)) (A^H - A^L) \quad (37)$$

where $\check{\tau}^D$ and \check{G}^D are defined in Lemma 17, and $\check{\tau}^E$ is given by (35).

Proof. It is immediate that (37) is sufficient to ensure that the rich prefer to be in power with an inefficient state, set the tax rate $\check{\tau}^E$, evade taxes and pay bribes equal to A^L with probability $p(\pi(\check{\tau}^E))$ to living under democracy with taxes and public good provision given by $\check{\tau}^D$ and \check{G}^D as in Lemma 17. ■

Condition (37) states that the payoff to the rich when the state is efficient (and the poor are in power) is lower than the expected payoff that they get when the state is inefficient (and they are in power). The latter payoff reflects the following facts: only the poor pay taxes, tax payers are inspected with probability $p(\pi(\check{\tau}^E)) = \check{\tau}^E$, and the rich offer a bribe equal to A^L to the bureaucrat inspecting them.

The following proposition characterizes the equilibrium with bureaucratic corruption. Since its proof follows that of Proposition 3 closely, it is omitted.

Proposition 8 *Consider the political environment with emerging democracy. Then, if conditions (36) and (37) hold, the unique MPE is one in which the rich elite choose $I_t = 0$ in the initial period and for all t thereafter, the rich party R always remains in power and the following policies are implemented at all dates:*

$$\begin{aligned} w_t &= \tilde{w}^E \equiv \frac{1 - q_0}{q_0} A^L, \quad X_t = \pi(\check{\tau}^E), \\ G_t &= \tilde{G}^E \equiv 0, \quad \text{and } \tau_t = \check{\tau}^E, \end{aligned}$$

where $\check{\tau}^E$ is given by (35). Moreover, only the poor pay taxes, while the rich evade taxes and pay a bribe equal to $b = A^L$ when inspected.

If, on the other hand, one or both of conditions (36) and (37) hold with the reverse inequality, the unique MPE involves $I_t = 1$ for all t , and for all $t \geq 1$, $d_t = P$ and the unique policy vector is

$$\begin{aligned} w_t &= \check{w}^D \equiv (1 - \check{\tau}^D) A^L, \quad X_t = \pi(\check{\tau}^D), \\ G_t &= \check{G}^D \equiv (1 - n) \check{\tau}^D A^H + [n - \pi(\check{\tau}^D)] \check{\tau}^D A^L - (1 - \check{\tau}^D) A^L \pi(\check{\tau}^D) - K, \end{aligned}$$

where $\check{\tau}^D$ and \check{G}^D are defined in Lemma 17.

Proof. The first part of the proposition follows from Lemmas 16-19. The only part that remains to be proved is that when one or both of conditions (36) and (37) hold with the reverse inequality, the poor will be in power. To see this, note that these conditions (with the reverse inequality) are sufficient for the rich to prefer democracy to setting up an inefficient state and evading taxes. Moreover, as before, if the state is efficient (i.e., $I_{t-1} = 1$), the poor will be in power. Therefore, we only have to show that the rich elite would not prefer an inefficient state and no tax evasion. This is straightforward since to prevent tax evasion by themselves, the rich would have to set a higher tax rate than $\check{\tau}^E$, since the “no tax evasion constraint” for the rich under $I_{t-1} = 0$ is

$$p(\pi(\tau)) \max \left\{ A^H - \frac{q_0}{1 - q_0} w, 0 \right\} + [1 - p(\pi(\tau))] A^H \leq (1 - \tau) A^H.$$

At $\check{\tau}^E$, this constraint is violated (since it is satisfied as equality for A^L). Thus, this constraint will be satisfied at some tax rate $\tau' > \check{\tau}^E$, which would give a per-period utility of $(1 - \tau') A^H$ to rich agents, which is strictly less than $p(\pi(\check{\tau}^E)) \max \{ A^H - q_0 w / (1 - q_0), 0 \} + [1 - p(\pi(\check{\tau}^E))] A^H$. Therefore, the rich are always better off evading taxes when in power. This establishes that conditions (36) and (37) holding with reverse inequality are sufficient for the equilibrium with the poor in power to emerge. ■

The most interesting result in Proposition 8 is that, when they are able to capture democratic politics, the rich do not pay any taxes at all. Instead, they (sometimes) pay bribes equal to the tax burden on poor agents, A^L . This implies that patronage politics turns *de jure* proportional taxation into a *de facto* regressive one. In other words, when the rich elite are able to set up an inefficient state and receive the support of bureaucrats, they are not only able to limit redistribution and public good provision, but they are also able to shift most of the burden of taxation to the poor. Consequently, the tax rate faced by the poor may be higher when corruption is possible than in the baseline model where both rich and poor pay taxes.

6 An Empirical Implication and Some Evidence

A distinctive empirical implication of our model is that democracies where relative wages of bureaucrats are high should provide fewer public goods. This is because, all else equal, bureaucrats are paid higher relative wages when the elite use patronage politics to limit redistribution and public good provision. In contrast, a naive intuition may suggest that relative wages of bureaucrats and public good provision should be correlated positively, either because when there is greater provision of public goods, more activities are entrusted to bureaucrats and they need to be paid more, or because countries with a greater willingness to tax will spend more both on public employment and on public good provision.²¹

We next look at the cross-country correlation between the relative wages of bureaucrats and public good provision among democracies. Our measure of the relative wage of bureaucrats is average wage of public-sector employees relative to GDP per capita from World Bank for 1991-2000. Our main measure of public good provision is total (central) government expenditure as a fraction of GDP for 1990-1998, and we also look at social services and welfare spending as a fraction of GDP as an alternative dependent variable.²² Both of these variables are from the IMF’s International Financial Statistics (see the details in the Appendix). To focus on democracies, we limit the sample to countries with an average Polity score greater than or equal to 5 over the period 1990-1998, which corresponds to “stable democracies” (see Persson and Tabellini, 2003). Our baseline sample contains 51 observations. Figure 1 shows the correlation between the relative wages of bureaucrats and government expenditure share of GDP. A strong negative relationship is visible in the figure, with most European countries having lower relative wages for bureaucrats and government expenditures than in Latin American

²¹An alternative intuition may be that, with a fixed government budget, higher public sector wages would force the government to reduce the rest of public good expenditures. In practice, there is considerable variation in the level of government budgets, and we will see below that same results apply with a measure of spending on social services and welfare.

²²We choose the total government expenditure as our main measure, both because we have more observations on this variable and also because the alternative, social services and welfare spending share of GDP, is heavily influenced by the age structure of the population. See below.

and Asian countries (there are few African countries in our sample).

The regression corresponding to Figure 1 is shown in column 1 of Table 1, with the robust standard errors in parentheses. The correlation between relative bureaucratic wages and government expenditure share of GDP is statistically highly significant, with a t-statistic of approximately 5. Column 2 of the table controls for GDP per capita. Richer countries spend more on public goods and income per capita is also correlated with relative bureaucratic wages. This regression shows that log income per capita is indeed significant, but the relationship between relative bureaucratic wages and government expenditures remains strong (the coefficient declines from -4.96 to -3.63, which continues to be significant at less than 1%). Column 3 also controls for the Polity democracy score, which is insignificant and has little effect on the coefficient of the relative wage of bureaucrats. Figure 2 shows the conditional correlation between relative bureaucratic wage and government expenditure share of GDP corresponding to column 3 of Table 1. The same negative relationship as in Figure 1 is again visible.

Column 4 controls for the age structure of the population, in particular the fraction of the population between the ages of 15-64 and the fraction over the age of 65. We expect the age structure of the population to have a direct effect on Social Security spending and thus also on total government expenditures. The results in column 4 show that controlling for the age structure variables significantly reduces both the coefficient estimate of the relative bureaucratic wage and the standard errors. The coefficient estimate is now -1.82, with a standard error of 0.85, which is still significant at 5%.

Columns 5-8 repeat the same regressions using social services and welfare spending as a percentage of GDP as the dependent variable. The results are similar to those for government expenditure and typically stronger, except when we control for the age structure variables. In particular, in columns 5-8, relative bureaucratic wage is significant at less than 1%. Once we include the controls for the age structure of the population, however, the relationship between the relative bureaucratic wage and social services and welfare spending is no longer significant; the coefficient estimate declines significantly and the standard error doubles. This result might reflect the fact that social services and welfare spending are closely related to the age structure of the population and there is little cross-sectional variation left once we control for the age structure variables.

In addition to the results shown in Table 1, we have also experimented with including “semi-stable” democracies (those with Polity scores between 0 and 5). The results are similar but slightly weaker. The results are also similar when we construct the sample using Freedom House measures of political and civil rights. We also checked the robustness of the results to various other controls. The results are broadly similar when we control for the legal origin of the country, for parliamentary versus presidential systems, for majoritarian versus proportional democracies, and for the age of democracy. Nevertheless, the results are significantly weakened

or disappear when we control for a full set of continent dummies. This is not entirely surprising, since, as Figures 1 and 2 show, the results reflect the contrast of European countries to Latin American and Asian countries.

Overall, it appears that there is a significant negative relationship between government expenditure and the relative wages of bureaucrats, which becomes weaker when we control for the age structure of the population and for continent dummies. While this cross-country correlation is not as robust as we would like it to be, it is nonetheless encouraging for our approach, since the negative relationship between relative wages of bureaucrats and government expenditure is a counter-intuitive implication of our model and a naive intuition would have suggested the opposite relationship between these two variables.

7 Concluding Remarks

Inefficiencies in the bureaucratic organization of the state are often viewed as an important factor in retarding economic development. Many sociological accounts of comparative development emphasize the role of state capacity (or lack thereof) in explaining why some societies are able to industrialize and modernize (e.g., Evans, 1995, Migdal, 1988). In addition, inefficient state organizations appear to coincide with limited amounts of public good provision and redistribution towards the poor. Existing approaches do not address the question of why certain societies choose or end up with such inefficient organizations and do not clarify the relationship between inefficient state organizations and limited redistribution.

We presented a simple theory of the emergence and persistence of inefficient states, in which the organization of the public bureaucracy is manipulated by the rich elite in order to influence redistributive politics. In particular, by instituting an inefficient state structure, the elite are able to use patronage and capture democratic politics. This enables them to limit the extent of redistribution and public good provision. Captured democracies not only limit redistribution, but also create a number of major distortions: the structure of the state is inefficient, there is too little public good provision and there may be overemployment of bureaucrats.

We also showed that an inefficient state creates its own constituency and tends to persist over time. Intuitively, an inefficient state structure creates more rents for bureaucrats than would an efficient state structure. When the median (poor) agent comes to power in democracy, he will reform the structure of the state to make it more efficient so that the higher taxes can be collected at lower cost (especially in terms of lower rents for bureaucrats). Anticipating this, when the organization of the state is inefficient, bureaucrats support the rich, who set lower taxes but pay high wages to bureaucrats. In order to generate enough political support, the coalition of the rich and the bureaucrats may not only choose an inefficient organization of the state, but they may further expand the size of bureaucracy so as to gain additional votes.

The model shows that an equilibrium with an inefficient state is more likely when there is greater income inequality and when democratic taxes are anticipated to be higher. An interesting implication of this result is that inequality and redistribution may be negatively correlated because higher inequality makes the capture of democratic politics more likely.

The pattern of elite control in democracy based on patronage politics and the emergence of an inefficient state organization bears some resemblance to the inefficient bureaucratic structures in a number of countries. In addition to these case studies, we provided cross-country correlations consistent with a distinctive implication of our model, that among democracies there should be a negative relationship between the relative wages of state employees and the amount of public good provision.

The general message from our analysis is that “not all democracies are created equal”; while some democracies will adopt policies that redistribute to poorer segments of the society, others may become captured by traditional elites. These captured democracies not only choose low levels of redistribution, but, as part of their political rationale for survival, they also typically create a range of inefficiencies. Our model suggests that these inefficiencies might be related to the relatively poor performance of a number of democracies in Latin America and Asia.²³

Analyses of the effect of such policies on economic growth and investigations of other methods via which the rich may limit the amount of redistribution in democratic politics are interesting areas for future work. Another interesting area for further study is a more careful empirical analysis of the relationship between the variation in the extent of government expenditure, relative wages of state employees and potential elite capture of democratic politics.

²³Another potential political factor in the poor economic performances of Latin American democracies is “populism”. Why some countries pursue populist policies is beyond the scope of the current paper. Nevertheless, it may be conjectured that the political environment may be more conducive to populism when the majority of the population fare relatively badly under democracy (see Acemoglu, 2007) and the type of democratic capture studied in this paper is likely to limit the benefits of democracy for the majority of the population.

Appendix A: Omitted Proofs

7.1 Proofs of Lemmas 1, 2, 5, 6, 7, 8 and 9

Proof of Lemma 1. Write the payoff of an individual producer $j \notin \mathcal{X}_t$ at time t when the tax rate is τ_t and the size of (effort-exerting) bureaucracy is x_t as

$$V_t^j = \max \left\{ (1 - \tau_t) A^j, (1 - p(x_t)) A^j \right\} + G_t + \beta V_{t+1}^j(\sigma^*),$$

where A^j is the productivity of this individual, and σ^* is the optimal policy, so that $\beta V_{t+1}^j(\sigma^*)$ is the discounted optimal continuation value for the individual. The max incorporates two terms. The first, $(1 - \tau_t) A^j$, is what the individual will consume if he pays a fraction τ_t of his income in taxes. The second, $(1 - p(x_t)) A^j$, is his expected consumption when evading taxes. In particular, in this case, the individual takes home his full productivity A^j with probability $(1 - p(x_t))$, but is caught and loses all his current income with probability $p(x_t)$. Limited liability implies $c_t^j \geq 0$ and the current behavior has no effect on the continuation value $\beta V_{t+1}^j(\sigma^*)$ given the anonymity assumption. This expression immediately establishes that the max term will pick tax evasion, i.e., $z_t^j = 0$, if $p(x_t) < \tau_t$, as claimed in the lemma. ■

Proof of Lemma 2. Write the payoff of a bureaucrat $j \in \mathcal{X}_t$ at time t when the wage rate is w_t and the detection probability is q_t as

$$V_t^j = \max \left\{ w_t - h, (1 - q_t) w_t \right\} + G_t + \beta V_{t+1}^j(\sigma^*),$$

where σ^* is the optimal policy, so that $\beta V_{t+1}^j(\sigma^*)$ is the discounted optimal continuation value for the individual. The max operator incorporates two terms representing the payoff to exerting effort and receiving the wage for sure, $w_t - h$, and the payoff to shirking. Since, by assumption, bureaucrats cannot be fired for shirking and limited liability makes sure that $c_t^j \geq 0$, the payoff to shirking is $(1 - q_t) w_t$. Whenever $w_t < h/q_t$, the max operator will pick the second term, so that we have $e_t^j = 0$ for all $j \in \mathcal{X}_t$ as claimed in the lemma. ■

Proof of Lemma 5. Given that $I_t = 1$, the solution to the equivalent of program (12) in the proof of Proposition 2 for party R involves choosing the policy vector $w_t = (1 - \tau^N) A^L + h$, $X_t = \pi(\tau^N)$, $G_t = G^N \equiv 0$. ■

Proof of Lemma 6. The policy vector in (13) is the optimal policy of the citizens in permanent democracy (Proposition 1). Now suppose that party P is in power at time t , and suppose that it chooses the policy vector specified in the lemma. Since this includes $I_t = 1$, the following period, we start with $I_t = 1$ as part of the payoff-relevant state vector. Suppose that σ^* is such that $v(I = 1 | B) = P$. Then party P wins the majority at time $t+1$. Alternatively suppose that $v(I = 1 | B) \neq P$, but $X < n - 1/2$. Then, party P again wins the majority at time $t+1$. In both cases, repeating this argument for the next period shows that party P keeps power at all dates and establishes the lemma.

To complete the proof we only need to rule out the case where $v(I = 1 | B) = R$ and $X \geq n - 1/2$ (the proof to eliminate the case where bureaucrats randomize between the two parties in a way to bring party R to power is identical). Since $v(I = 1 | B) = R$ and $I = 0$ is costly for the rich (recall program (12) in the proof of Proposition 2), party R will choose $I_t = 1$. Then from Lemma 5,

$$w_t = (1 - \tau^N) A^L + h, \quad X_t = \pi(\tau^N), \quad G_t = G^N \equiv 0.$$

This implies that the utility of the bureaucrat is the same as a poor producer. Then denoting the utility of a bureaucrat supporting party d by $V^B(d)$, we have

$$\begin{aligned} V^B(R) &= (1 - \tau^N) A^L + \beta V^j(\sigma^*) \\ &< (1 - \tau^D) A^L + G^D + \beta V^j(\sigma^*) \\ &= V^B(P), \end{aligned}$$

where the inequality follows from the fact that the last term is the maximal utility of a poor agent. Since this is also the utility that a bureaucrat will receive when party P is in power, $v(I = 1 | B) = R$ cannot be a best response, completing the proof of the lemma. ■

Proof of Lemma 7. That any party, when in power and inheriting $I_{t-1} = 0$, will choose $w_t = h/q_0$ follows immediately from Lemma 3 (otherwise, the investment in infrastructure K cannot be financed and there will be zero production). The fact that party R will choose $G_t = G^E \equiv 0$ follows immediately from the program in (12) after imposing $w_t = h/q_0$. That party P will choose \hat{G}^D as in (14) follows from the maximization problem (10) with the additional restriction that $w_t = h/q_0$. ■

Proof of Lemma 8. Lemma 6 establishes that $I_{t-1} = 0$ is necessary. Now suppose that $I_{t-1} = 0$ and consider the scenario in which party R chooses $I_t = 0$ and $X_t \geq X_{t-1}$ (so that no current bureaucrat will be fired). Consider the case in which individual $j \in \mathcal{X}_t$ is pivotal and chooses $v_t^j = R$ in all future periods. Then, his net per-period payoff will be $w_t - h = (1 - q_0)h/q_0$, and give him a lifetime utility of

$$V_t^j = \frac{1}{1 - \beta} \frac{(1 - q_0)h}{q_0}. \quad (38)$$

In contrast, if $j \in \mathcal{X}_t$ were to choose $v_t^j = P$ when pivotal, his value would be

$$V_t^j = \frac{h}{q_0} - h + \hat{G}^D + \beta \hat{V}_{t+1}^j. \quad (39)$$

where \hat{V}_{t+1}^j is the continuation value when party P is in power from then on, given by

$$\hat{V}_{t+1}^j = \frac{1}{1 - \beta} [(1 - \tau^D) A^L + G^D].$$

This last expression incorporates the fact that if the poor are in power, they reform the bureaucracy, setting $I_t = 1$, and that $I = 1$ is an absorbing state.

The comparison of (38) and (39) gives (15)—as a weak inequality—as a necessary condition for bureaucrats to support party R when they are pivotal. Condition (16) is also necessary since, if it were violated, bureaucrats would not be pivotal and party R would receive less than half of the votes even with all of bureaucrats voting $v_t^j = R$. This argument establishes that both (15) and (16) are necessary. Moreover, (15)—as a strict inequality—and (16) are also sufficient to ensure $d_t = R$, since when both of these conditions hold, it is a weakly dominant strategy for bureaucrats to vote for party R whenever $I_{t-1} = 0$ and the coalition of bureaucrats and the rich have a majority. ■

Proof of Lemma 9. Suppose that bureaucrats play $v(I = 0 | B) = R$ (that is, they will vote for party R whenever the state is inefficient). Under the rule of party P , the per period return of the rich is $(1 - \tau^D) A^H + G^D$. When $\tau^E \geq \hat{\tau}^E$, party R can remain in power by choosing $I = 0$ and obtain the per period return $(1 - \tau^E) A^H$, which establishes the first part (19).

For the second part, note that party R can always choose its myopic optimum when in power. This will give a representative rich agent utility

$$V^R = (1 - \tau^E) A^H + \frac{\beta}{1 - \beta} [(1 - \tau^D) A^H + G^D].$$

Here $(1 - \tau^E) A^H$ is current consumption, and $\beta [(1 - \tau^D) A^H + G^D] / (1 - \beta)$ is the continuation value, which follows from the observation that since, by assumption, $\tau^E < \hat{\tau}^E$, we have $n - 1/2 > \pi(\tau^E)$ and thus party R will lose the election at the next date. Then Lemma 6 implies that party P will win all elections in all future dates. Alternatively, party R can choose $X = n - 1/2$ and guarantee to be in power forever, but at the expense of taxing the rich at the higher rate $\hat{\tau}^E$. This will give a representative rich agent utility

$$\hat{V}^R = \frac{(1 - \hat{\tau}^E) A^H}{1 - \beta}.$$

Comparison of \hat{V}^R with V^R in the previous expression gives the second part of (19). ■

7.2 Proof of Proposition 5

Consider changes in inequality that keep mean pre-tax income, $Y = (1 - n)A^H + nA^L$, constant. This implies the following simple relationship between the pre-tax incomes of rich and poor agents:

$$A^H = \frac{Y - nA^L}{1 - n}. \quad (40)$$

To prove the desired result, we need to show that (15) and (19) in Lemmas 8 and 9 are more likely to hold when there is greater inequality, i.e., when A^L is lower (and A^H is given by (40)).

Let us rewrite condition (15) as

$$\frac{h}{q_0} \geq \frac{1}{1 - q_0} \left[(1 - \tau^D) A^L + G^D + \frac{1 - \beta}{\beta} \hat{G}^D \right] \equiv \underline{\theta}. \quad (41)$$

Next, consider condition (19) in Lemma 9. Suppose first that $\tau^E \geq \hat{\tau}^E$, i.e., when $X = \pi(\tau^E)$ (recall that more generally $X = \max\{\pi(\tau^E), n - 1/2\}$). Then, combining the government budget constraint (17) with (40) gives

$$\tau^E = \frac{\pi(\tau^E)}{Y - \pi(\tau^E)A^L} \frac{h}{q_0} + \frac{K}{Y - \pi(\tau^E)A^L}. \quad (42)$$

Substituting for τ^E from this expression, condition (19) can be rewritten as

$$\frac{h}{q_0} < \left(\tau^D - \frac{G^D}{A^H} \right) \frac{Y - \pi(\tau^E)A^L}{\pi(\tau^E)} - \frac{K}{\pi(\tau^E)} \equiv \bar{\theta}. \quad (43)$$

Instead, when $\tau^E < \hat{\tau}^E$, the size of the bureaucracy is $X = n - 1/2 \equiv \lambda$. Solving for $\hat{\tau}^E$ from (18) and (40) as

$$\hat{\tau}^E = \frac{\lambda h/q_0 + K}{Y - \lambda A^L}, \quad (44)$$

the relevant part of condition (19), $(1 - \hat{\tau}^E)A^H > (1 - \beta)(1 - \tau^E)A^H + \beta[(1 - \tau^D)A^H + G^D]$, can be expressed as

$$\frac{h}{q_0} < \frac{1}{\lambda} \{ [(1 - \beta)\tau^E + \beta(\tau^D - G^D/A^H)] (Y - \lambda A^L) - K \} \equiv \bar{\theta}^*. \quad (45)$$

These three expressions define $\underline{\theta}$, $\bar{\theta}$ and $\bar{\theta}^*$. Now summarizing our analysis, an inefficient state will be created under two different scenarios:

1. if $X = \pi(\tau^E)$ and if conditions (41) and (43) are satisfied, which requires

$$\underline{\theta} \leq h/q_0 < \bar{\theta};$$

2. if $X = n - 1/2 \equiv \lambda$ and if conditions (41) and (45) are satisfied, which requires

$$\underline{\theta} \leq h/q_0 < \bar{\theta}^*.$$

We will prove that higher inequality makes the inefficient state equilibrium more likely by showing that the upper thresholds ($\bar{\theta}$ in case 1 and $\bar{\theta}^*$ in case 2) are increasing and the lower threshold, $\underline{\theta}$, is decreasing in the level of inequality—these naturally imply that the intervals $\Delta \equiv \bar{\theta} - \underline{\theta}$ and $\Delta^* \equiv \bar{\theta}^* - \underline{\theta}$ increase with income inequality. We will also show that an increase in inequality does not cause a switch from 1 to 2 or vice versa in a way to make the inefficient state less likely.

We first establish an intermediate result:

Claim 1 We have

$$\frac{\partial \tau^D}{\partial A^L} = -\frac{1 + \pi'(\tau^D)}{\pi''(\tau^D)(A^L + h)} < 0, \quad (46)$$

and

$$\frac{\partial \tau^E}{\partial A^L} = \frac{\tau^E \pi(\tau^E)}{Y - \pi'(\tau^E)(\tau^E A^L + h/q_0) - \pi(\tau^E)A^L} > 0. \quad (47)$$

Proof. The first-order condition of program (9) for an interior τ^D is

$$\frac{\partial G^D}{\partial \tau} = A^L. \quad (48)$$

Using (40), the equilibrium level of the public good (8) provided by the poor is

$$G^D = \tau^D Y - K - \pi(\tau^D)(A^L + h). \quad (49)$$

The first-order condition (48) therefore becomes

$$Y - \pi'(\tau^D)(A^L + h) - A^L = 0. \quad (50)$$

The solution for τ^D is always positive since $K > 0$ needs to be financed. Moreover, the assumption that $\tau^D < 1 - \pi(\tau^D) < 1$ ensures that $\tau^D < 1$. Differentiating (50) gives (46).

Next, differentiating the government budget constraint (17) and using (40) gives (47), where the denominator is positive since τ^E is always to the left of the peak of the Laffer curve. ■

Next, given the definition of $\underline{\theta}$ in (41) we obtain that

$$\lim_{\beta \rightarrow 1} \frac{\partial \underline{\theta}}{\partial A^L} = \frac{1}{1 - q_0} [1 - \tau^D - \pi(\tau^D)] > 0 \quad (51)$$

where the expression for the limit uses the fact that $\partial \hat{G}^D(A^L) / \partial A^L$ exists, is finite and is independent on β , and the inequality again follows from the assumption that $\tau^D < 1 - \pi(\tau^D) < 1$. This inequality implies that for sufficiently high β , the inefficient state becomes more attractive to bureaucrats as the level of inequality increases.

We now show that higher inequality, represented by a decrease in A^L with A^H given by (40), makes the inefficient state also more attractive to the rich by increasing $\bar{\theta}$ and $\bar{\theta}^*$.

Consider two cases:

Case 1: $X = \pi(\tau^E)$.

From (49), we have

$$\tau^D Y = G^D + K + \pi(\tau^D)(A^L + h). \quad (52)$$

Substituting (52) into (43) and some algebra gives

$$\bar{\theta} = \frac{n}{\pi(\tau^E)} \frac{A^H - A^L}{A^H} G^D + \frac{A^L}{A^H} G^D + \frac{\pi(\tau^D)}{\pi(\tau^E)} (A^L + h) - \tau^D A^L. \quad (53)$$

Next, note that

$$\frac{\partial [(A^H - A^L) / A^H]}{\partial A^L} = -\frac{\partial [A^L / A^H]}{\partial A^L} = -\left[\frac{1}{A^H} + \frac{n}{1 - n} \frac{A^L}{(A^H)^2} \right] < 0$$

$$\frac{\partial [1 / \pi(\tau^E)]}{\partial A^L} = -\frac{\pi'(\tau^E)}{[\pi(\tau^E)]^2} \frac{\partial \tau^E}{\partial A^L} < 0$$

$$\frac{\partial [\pi(\tau^D) / \pi(\tau^E)]}{\partial A^L} = \frac{\pi'(\tau^D)}{\pi(\tau^E)} \frac{\partial \tau^D}{\partial A^L} - \frac{\pi(\tau^D) \pi'(\tau^E)}{[\pi(\tau^E)]^2} \frac{\partial \tau^E}{\partial A^L} < 0$$

and from (49), using (50),

$$\frac{\partial G^D}{\partial A^L} = \frac{\partial \tau^D}{\partial A^L} A^L - \pi(\tau^D) < 0. \quad (54)$$

Differentiating (53), in turn, gives

$$\begin{aligned} \frac{\partial \bar{\theta}}{\partial A^L} &= -\frac{n\pi'(\tau^E)}{[\pi(\tau^E)]^2} \frac{\partial \tau^E}{\partial A^L} \frac{A^H - A^L}{A^H} G^D - \frac{n}{\pi(\tau^E)} \left[\frac{1}{A^H} + \frac{n}{1-n} \frac{A^L}{(A^H)^2} \right] G^D + \\ &+ \frac{n}{\pi(\tau^E)} \frac{A^H - A^L}{A^H} \left[\frac{\partial \tau^D}{\partial A^L} A^L - \pi(\tau^D) \right] + \left[\frac{1}{A^H} + \frac{n}{1-n} \frac{A^L}{(A^H)^2} \right] G^D + \\ &+ \frac{A^L}{A^H} \left[\frac{\partial \tau^D}{\partial A^L} A^L - \pi(\tau^D) \right] + \left[\frac{\pi'(\tau^D)}{\pi(\tau^E)} \frac{\partial \tau^D}{\partial A^L} - \frac{\pi(\tau^D)\pi'(\tau^E)}{[\pi(\tau^E)]^2} \frac{\partial \tau^E}{\partial A^L} \right] (A^L + h) + \\ &+ \frac{\pi(\tau^D)}{\pi(\tau^E)} - \frac{\partial \tau^D}{\partial A^L} A^L - \tau^D, \end{aligned}$$

which can be rewritten as

$$\begin{aligned} \frac{\partial \bar{\theta}}{\partial A^L} &= -\frac{n\pi'(\tau^E)}{[\pi(\tau^E)]^2} \frac{\partial \tau^E}{\partial A^L} \frac{A^H - A^L}{A^H} G^D - \left[\frac{n}{\pi(\tau^E)} - 1 \right] \left[\frac{1}{A^H} + \frac{n}{1-n} \frac{A^L}{(A^H)^2} \right] G^D + \quad (55) \\ &+ \frac{\partial \tau^D}{\partial A^L} A^L \left[\frac{n}{\pi(\tau^E)} - 1 \right] \frac{A^H - A^L}{A^H} - \pi(\tau^D) \left[\frac{n}{\pi(\tau^E)} \frac{A^H - A^L}{A^H} + \frac{A^L}{A^H} \right] - \tau^D + \\ &+ \left[\frac{\pi'(\tau^D)}{\pi(\tau^E)} \frac{\partial \tau^D}{\partial A^L} - \frac{\pi(\tau^D)\pi'(\tau^E)}{[\pi(\tau^E)]^2} \frac{\partial \tau^E}{\partial A^L} \right] (A^L + h) \\ &+ \frac{\pi(\tau^D)}{\pi(\tau^E)}. \end{aligned}$$

Since $\partial \tau^E / \partial A^L > 0$, $\partial \tau^D / \partial A^L < 0$ and $n > \pi(\tau^E)$, all the terms in (55) except the last line, $\pi(\tau^D) / \pi(\tau^E)$, are negative. Furthermore, we have

$$\frac{\pi'(\tau^D)}{\pi(\tau^E)} \frac{\partial \tau^D}{\partial A^L} (A^L + h) + \frac{\pi(\tau^D)}{\pi(\tau^E)} = \frac{1}{\pi(\tau^E)} \left[-\pi'(\tau^D) \frac{1 + \pi'(\tau^D)}{\pi''(\tau^D)} + \pi(\tau^D) \right], \quad (56)$$

where the right and side of (56) is obtained using the expression for $\partial \tau^D / \partial A^L$ in (46). Now, the log-concavity of $\pi(\tau)$ implies that $[\pi'(\tau)]^2 > \pi''(\tau)\pi(\tau)$ and is sufficient to ensure that (56) is negative. This implies that all the terms in (55) including $\pi(\tau^D) / \pi(\tau^E)$ are negative, and therefore $\partial \bar{\theta} / \partial A^L < 0$ as desired. Note that this conclusion holds irrespective of the value of β .

Case 2: $X = n - 1/2 \equiv \lambda$.

The proof parallels that of case 1. From (45), using (52) and the fact that $Y - \lambda A^L = (1-n)A^H + (1/2)A^L$, we obtain

$$\bar{\theta}^* = \frac{1}{\lambda} \left\{ (1-\beta)\tau^E(Y - \lambda A^L) + \beta \left[\pi(\tau^D)(A^L + h) + K - \lambda \tau^D A^L + G^D - \frac{(1-n)A^H + (1/2)A^L}{A^H} G^D \right] - K \right\} \quad (57)$$

Note that

$$\frac{\partial [(1-n)A^H + (1/2)A^L] / A^H}{\partial A^L} = \frac{Y}{2(1-n)(A^H)^2}. \quad (58)$$

Differentiating (57) and using (54), (58) and the fact that $\partial \tau^E / \partial A^L$ exists and is independent of

β , we have

$$\begin{aligned} \lim_{\beta \rightarrow 1} \frac{\partial \bar{\theta}^*}{\partial A^L} &= \frac{1}{\lambda} \left\{ \pi(\tau^D) + \pi'(\tau^D) \frac{\partial \tau^D}{\partial A^L} (A^L + h) - \lambda \tau^D - \lambda \frac{\partial \tau^D}{\partial A^L} A^L + \frac{\partial \tau^D}{\partial A^L} A^L - \pi(\tau^D) + \right. \\ &\quad \left. - \frac{Y}{2(1-n)(A^H)^2} G^D - \frac{(1-n)A^H + (1/2)A^L}{A^H} \left[\frac{\partial \tau^D}{\partial A^L} A^L - \pi(\tau^D) \right] \right\}, \end{aligned}$$

which can be rewritten as

$$\begin{aligned} \lim_{\beta \rightarrow 1} \frac{\partial \bar{\theta}^*}{\partial A^L} &= \frac{1}{\lambda} \left\{ \pi(\tau^D) + \pi'(\tau^D) \frac{\partial \tau^D}{\partial A^L} (A^L + h) - \lambda \tau^D - \frac{Y}{2(1-n)(A^H)^2} G^D + \right. \\ &\quad \left. + \frac{1}{2} \frac{A^H - A^L}{A^H} \frac{\partial \tau^D}{\partial A^L} A^L - \left(n - \frac{1}{2} \frac{A^L}{A^H} \right) \pi(\tau^D) \right\}. \end{aligned} \quad (59)$$

Since $\partial \tau^D / \partial A^L < 0$, $\pi'(\tau^D) > 0$, $n > 1/2$ and $A^L < A^H$, all terms in (59) other than $\pi(\tau^D)$ are negative. Therefore, a sufficient condition to ensure that (59) is negative is

$$\pi'(\tau^D) \frac{\partial \tau^D}{\partial A^L} (A^L + h) + \pi(\tau^D) = -\pi'(\tau^D) \frac{1 + \pi'(\tau^D)}{\pi''(\tau^D)} + \pi(\tau^D) < 0, \quad (60)$$

where the right hand side of (60) is obtained using the expression for $\partial \tau^D / \partial A^L$ in (46). This condition is equivalent to (56) and the log-concavity of $\pi(\tau)$ is sufficient to ensure it. This establishes that $\partial \bar{\theta}^* / \partial A^L$ is negative for sufficiently high β as desired.

The proof so far has established that the lower threshold θ declines as inequality increases and that the upper thresholds $\bar{\theta}$ and $\bar{\theta}^*$ increase as inequality increases. To complete the proof of the proposition, we need to ensure that there would be no switch from the wider to the smaller interval, which could happen if we have a switch from $\tau^E \geq \hat{\tau}^E$ to $\tau^E < \hat{\tau}^E$, or vice versa. However, as $\beta \rightarrow 1$, we have that (19) is equivalent to

$$\begin{aligned} \text{if } \tau^E \geq \hat{\tau}^E &\Leftrightarrow X = \max\{\pi(\tau^E), n - 1/2\} = \pi(\tau^E) && \text{and } (\tau^D - \tau^E) A^H > G^D \\ \text{if } \tau^E < \hat{\tau}^E &\Leftrightarrow X = \max\{\pi(\tau^E), n - 1/2\} = n - 1/2 \equiv \lambda && \text{and } (\tau^D - \hat{\tau}^E) A^H > G^D \end{aligned}$$

so that for β sufficiently large, at the point of a possible switch, $\tau^E = \hat{\tau}^E$, we have $\bar{\theta} = \bar{\theta}^*$. This completes the proof. ■

7.3 Proof of Lemma 16

Suppose that the bureaucratic wage is given by \tilde{w}^E in (31). Then the incentive compatibility constraint (26) of the bureaucrats inspecting low-skill producers is satisfied even when the producers offer a bribe as large as their income A^L . Holding τ fixed, a decrease in w from \tilde{w}^E will allow bureaucrats to accept bribes, and thus reduce government revenues to zero. Therefore, it cannot be optimal. Increasing w is also not beneficial for the rich.

Condition (29), on the other hand, ensures that the poor choose to pay taxes. Holding w fixed at \tilde{w}^E , increasing taxes would induce the poor not to pay taxes and is therefore not beneficial. Reducing taxes is also not beneficial. Given (31) and (32), it is also straightforward to verify that the utility of the rich is decreasing in G , so that this variable is set equal to zero.

This argument shows that (31) and (32) gives a stationary point of the optimization problem of the rich, since the rich will not find it beneficial to change either one of x or w by itself. To complete the proof, we need to show that it is also not beneficial to change x and w simultaneously. We will do this by showing that the payoff function of the rich is strictly quasi-concave. Consider the problem of the

rich if they do not pay taxes. Clearly, constraints (29) and (30) in equilibrium hold as equalities. We can thus solve out for the tax rate from the government budget constraint as

$$\tau = \frac{xw + K + G}{(n - x)A^L}.$$

Substituting this expression into (29), we obtain

$$w = \frac{K + G}{[q_0/(1 - q_0)](n - x)p(x) - x}, \quad (61)$$

provided that $w \geq (1 - q_0)A^L/q_0$ (which will be true in equilibrium). Now, substituting (61) in the objective function of the rich (28), and observing that this is maximized at $G = 0$, we can represent the problem of the rich as the following single dimensional maximization problem

$$\max_x U(x) = A^H - \frac{K}{(n - x) - (1 - q_0)x/q_0p(x)}.$$

If this problem is strictly quasi-concave, it must have a unique solution. Corresponding to this unique x , there will be unique levels of τ and w , since these variables are defined uniquely by the previous equalities. To check that this function is indeed strictly quasi-concave, note that

$$U'(x) = -\frac{1 + (1 - q_0)/q_0p(x) - (1 - q_0)[p'(x)x]/q_0[p(x)]^2}{[(n - x) - (1 - q_0)x/q_0p(x)]^2}K.$$

For $U(x)$ to be strictly quasi concave, it is sufficient that its second derivative is negative when $U'(x) = 0$. For this, it is sufficient for $-\left\{[p(x)]^2 + (1 - q_0)p(x)/q_0 - (1 - q_0)[p'(x)x]/q_0\right\}$ to be strictly decreasing in x . It can be easily verified that this is always the case since $p(x)$ is increasing and concave in x . This completes the proof that (31) and (32) is optimal for the rich when they prefer not to pay taxes. ■

Appendix B: Data Sources and Definitions

Our dataset builds on the cross-country dataset compiled by Persson and Tabellini (2003) (henceforth PT). Our sample of “stable democracies” consists of countries with an average Polity score greater than or equal to 5 over the period 1990-1998. We also report results with a sample containing all democracies (defined as countries with average Polity score greater than 0 over the period 1990-1998). With the exception of the average government wage to per capita GDP (which comes from the World Bank), all our variables are from PT’s dataset. These variables are the following:

Central government expenditures as a percentage of GDP: constructed using the item Government Finance-Expenditures in the IFS, divided by GDP at current prices and multiplied by 100. Source: IMF-IFS CD-Rom and IMF-IFS Yearbook.

Consolidated central government expenditures on social services and welfare as percentage of GDP: from the GFS Yearbook, divided by GDP and multiplied by 100. Source: IMF-GFS Yearbook 2000 and IMF-IFS CD-Rom.

Log GDP per capita: per capita real GDP defined as real GDP per capita in constant dollars (chain index) expressed in international prices, base year 1985. Data through 1992 are taken from the Penn World Table 5.6, while data on the period 1993-98 are computed from data taken from the World Development Indicators, the World Bank. These later observations are computed on the basis of the latest observation available from the Penn Word Tables and the growth rates of GDP per capita in the subsequent years computed from the series of GDP at market prices (in constant 1995 U.S. dollars) and population, from the World Development Indicators.

Sources: Penn World Tables - mark 5.6 (PWT), available on <http://datacentre2.chass.utoronto.ca/pwt/docs/topic.html>.

The World Bank’s World Development Indicators, available on <http://www.worldbank.org>.

Polity: score for democracy, computed by subtracting the AUTOC score from the DEMOC score, and ranging from +10 (strongly democratic) to -10 (strongly autocratic). **AUTOC (DEMOC)** is the index of autocracy (democracy), derived from codings of the competitiveness of political participation, the regulation of participation, the openness and competitiveness of executive recruitment, and constraints on the chief executive.

Source: Polity IV Project (<http://www.cidcm.umd.edu/inscr/polity/index.htm>).

Age structure variables: percentage of population between the ages of 15 and 64 in the total population and percentage of population over the age of 65 in the total population.

Source: World Development Indicators CD-Rom 1999.

Average government wage relative to per capita GDP: mean value of the average government wage to per capita GDP between 1991 and 2000. It is computed as the average of the two data points available for the periods 1991-95 and 1996-2000. When data for one of the two periods are not available, only the available time period is used. The variable is calculated by dividing the average government wage by the GDP per capita figure. The average government wage is calculated as the total central government wage bill divided by the number of employees in total central government. The total central government wage bill is the sum of wages and salaries paid to central government employees, including armed forces personnel. The number of employees in total central government is the sum of total civilian central government and the Armed Forces.

Source: Schiavo-Campo, de Tommaso and Mukherjee (1997), Table A-3, <http://www-wds.worldbank.org/external/default>.

References

- Acemoglu, Daron (2005) "Politics and Economics in Weak and Strong States," *Journal of Monetary Economics*, 52, 1199-1226.
- Acemoglu, Daron (2007) "An Economic Model of Populism," in progress.
- Acemoglu, Daron and James A. Robinson (2001) "Inefficient Redistribution," *American Political Science Review*, 95, 649-662.
- Acemoglu, Daron and James A. Robinson (2006a) *Economic Origins of Dictatorship and Democracy*, New York; Cambridge University Press.
- Acemoglu, Daron and James A. Robinson (2006b) "Persistence of Power, Elites and Institutions," NBER Working Paper, No. 12108.
- Acemoglu, Daron and Thierry Verdier (1998) "Property Rights, Corruption and the Allocation of Talent: A General Equilibrium Approach," *Economic Journal*, 108, 1381-1403.
- Acemoglu, Daron and Thierry Verdier (2000) "The Choice Between Market Failures and Corruption," *American Economic Review*, 90(1), 194-211.
- Aghion, Philippe and Patrick Bolton (1990) "Government Domestic Debt and the Risk of Default: A Political-Economic Model of the Strategic Role of Debt," in *Capital Markets and Debt Management*, R. Dornbusch and M. Draghi, eds., Cambridge, MA: MIT Press.
- Alesina, Alberto (1988) "Credibility and Political Convergence in a Two-Party System with Rational Voters," *American Economic Review*, 78(4), 796-805.
- Alford, Robert R. and Roger Friedland (1985) *Powers of Theory. Capitalism, the State and Democracy*, New York; Cambridge University Press.
- Austen-Smith, David (1987) "Interest Groups, Campaign Contributions, and Probabilistic Voting," *Public Choice*, 54, 123-139.
- Baron, David P. (1994) "Electoral Competition with Informed and Uninformed Voters," *American Political Science Review*, 88, 33-47.
- Barro, Robert (1999) *Determinants of Economic Growth: A Cross-country Empirical Study*, MIT Press, Cambridge MA.
- Beck, Linda J. (1997) "Senegal's "Patrimonial Democrats": Incremental Reform and the Obstacles to the Consolidation of Democracy," *Canadian Journal of African Studies/Revue Canadienne des Études Africaines*, 31(1), pp. 1-31.
- Becker, Gary S. and Casey Mulligan (2003) "Deadweight Costs and the Size of Government," *Journal of Law and Economics*, 46, 293-340.
- Bénabou, Roland (2000) "Unequal Societies: Income Distribution and the Social Contract," *American Economic Review*, 90(1), 96-129.
- Besley, Timothy and Stephen Coate (1997) "An Economic Model of Representative Democracy," *Quarterly Journal of Economics*, 108, 85-114.

- Boone, Catherine (1990) "State Power and Economic Crisis in Senegal," *Comparative Politics*, 22(3), 341-357.
- Centeno, Miguel (2002) *Blood and Debt: War and Nation-State in Latin America*, University Park, PA.; Pennsylvania State University Press.
- Coate, Stephen and Stephen Morris (1995) "On the Form of Transfers to Special Interests," *Journal of Political Economy*, 103(6), 1210-1235.
- Coate, Stephen and Stephen Morris (1999) "Policy Persistence," *American Economic Review*, 89(5), 1327-1336.
- Cukierman, Alex, Sebastian Edwards and Guido Tabellini (1992) "Seigniorage and Political Instability," *American Economic Review*, 82(3), 537-555.
- Debs, Alexandre (2006) "The Wheel of Fortune: Agency Problems in Dictatorships," MIT mimeo.
- Dixit, Avinash K. (2002) "Incentives and Organizations in the Public Sector: An Interpretative Review," *Journal of Human Resources*, 37(4), 696-727.
- Dixit, Avinash K. and John Londregan (1998) "Ideology, Tactics and Efficiency in Redistributive Politics," *Quarterly Journal of Economics*, 109(2), 497-529.
- Dixit, Avinash K., Gene Grossman and Elhanan Helpman (1997) "Common Agency and Coordination: General Theory and Application to Government Policy Making," *Journal of Political Economy*, 105(4), 752-769.
- Egorov, Georgy and Konstantin Sonin (2005) "Dictators and Their Viziers: Agency Problems in Dictatorships," New Economic School, Moscow, mimeo.
- Epstein, Stepan (2000) *Freedom and Growth. The Rise of States and Markets in Europe, 1300-1750*, London; Routledge.
- Etzioni-Halevy, Eva (1983) *Bureaucracy and Democracy*, London; Routledge.
- Evans, Peter B. (1989) "Predatory, Developmental, and Other Apparatuses: A Comparative Political Economy Perspective on the Third World State," *Sociological Forum*, 4(4), 561-587.
- Evans, Peter B. (1992) "The State as Problem and Solution: Predation, Embedded Autonomy, and Structural Change," in Stephan Haggard and Robert R. Kaufman (eds.), *The Politics of Economic Adjustment: International Constraints, Distributive Conflicts and the State*, Princeton; Princeton University Press.
- Evans, Peter B. (1995) *Embedded Autonomy*, Princeton; Princeton University Press.
- Fiorina, Morris P. and Roger G. Noll (1978) "Voters, Legislators and Bureaucracy: Institutional Design in the Public Sector," *American Economic Review*, 68(2), 256-260.
- Fudenberg, Drew and Jean Tirole (1991) *Game Theory*, Cambridge; MIT Press.
- Gay, Robert (1990) "Popular Incorporation and Prospects for Democracy: Some Implications of the Brazilian Case," *Theory and Society*, 19(4), 447-463.

Geddes, Barbara (1991) "A Game Theoretic Model of Reform in Latin American Democracies," *American Political Science Review*, 85(2), 371-92.

Glazer, Amihai (1989) "Politics and the Choice of Durability," *American Economic Review*, 79(5), 1207-1213.

Golden, Miriam A. (2003) "Electoral Connections: The Effects of the Personal Vote on Political Patronage, Bureaucracy and Legislation in Postwar Italy," *British Journal of Political Science*, 33, pp. 189-212.

Gomes, Armando and Philippe Jehiel (2005) "Dynamic Processes of Social and Economic Interactions: On the Persistence of Inefficiencies," *Journal of Political Economy*, 113(3), 626-667.

Gramsci, Antonio (1971) *Selections from the Prison Notebooks*, New York; International Publishers.

Grossman, Gene M. and Elhanan Helpman (1996) "Electoral Competition and Special Interest Politics," *Review of Economic Studies*, 63(2), 265-286.

Hassler, John, José V. Rodríguez Mora, Kjetil Storesletten and Fabrizio Zilibotti (2003) "The Survival of the Welfare State," *American Economic Review*, 93(1), 87-112.

Herbst, Jeffery I. (2000) *States and Power in Africa: Comparative Lessons in Authority and Control*, Princeton NJ; Princeton University Press.

Kohli, Atul (2004) *State-Directed Development: Political Power and Industrialization in the Global Periphery*, New York; Cambridge University Press.

Levi, Margaret (1989) *Of Rule and Revenue*, Berkeley and Los Angeles CA; University of California Press.

Lizzeri, Alessandro and Nicola Persico (2001) "Provision of Public Goods under Alternative Electoral Systems" *American Economic Review*, 91(1), 225-239.

Migdal, Joel (1988) *Strong Societies and Weak States: State-Society Relations and State Capabilities in the Third World*, Princeton; Princeton University Press.

Miliband, Ralph (1969) *State in Capitalist Society*, New York, Basic Books.

Osborne, Martin J. and Al. Slivinsky (1996) "A Model of Political Competition with Citizen-Candidates," *Quarterly Journal of Economics*, 111(1), 65-96.

Perotti, Roberto (1996) "Growth, Income Distribution, and Democracy: What the Data Say," *Journal of Economic Growth*, 1(2), 149-87.

Persson, Torsten, Gérard Roland and Guido Tabellini (2000) "Comparative Politics and Public Finance," *Journal of Political Economy*, 108(6), 1121-1161.

Persson, Torsten and Guido Tabellini (2003) *The Economic Effect of Constitutions*, Cambridge MA; The MIT Press.

Persson, Torsten and Lars E. O. Svensson (1989) "Why a Stubborn Conservative would Run a Deficit: Policy with Time-Inconsistent Preferences," *Quarterly Journal of Economics*,

104(2), 325-345.

Poulantzas, Nicos (1978) *Political Power and Social Classes*, London, Verso.

Rauch, James and Peter B. Evans (2000) "Bureaucratic Structure and Bureaucratic Performance in Less Developed Countries," *Journal of Public Economics*, 75, 49-71.

Robinson, James A. (2001) "When is a State Predatory?," Unpublished manuscript, Harvard University.

Robinson, James A. and Ragnar Torvik (2005) "White Elephants," *Journal of Public Economics*, 89, 197-210.

Rodrik, Dani (1995) "The Political Economy of Trade Policy," in G. Grossman and K. Rogoff (eds.), *Handbook of International Economics*, vol. 3, Amsterdam; North-Holland.

Roett, Riordan (1999) *Brazil: Politics in a Patrimonial Society*, Westport, CT: Praeger.

Rothstein, Bo and Eric M. Uslaner (2005) "All for All: Equality, Corruption, and Social Trust," *World Politics*, 58(1), 41-72.

Saint-Paul, Gilles (1996) "Exploring the Political Economy of Labour Market Institutions," *Economic Policy*, 11(23), 265-315.

Schiavo-Campo, Salvatore, Giulio de Tommaso and Amitabh Mukherjee (1997) "An International Statistical Survey of Government Employment and Wages," The World Bank, Policy Research Working Paper 1806.

Shleifer, Andrei and Robert W. Vishny (1994) "Politicians and Firms," *Quarterly Journal of Economics*, 109(4), 995-1025.

Tabellini, Guido and Alberto Alesina (1990) "Voting on the Budget Deficit," *American Economic Review*, 80(1), 37-49.

Therborn, Göran (1980) *What Does the Ruling Class Do When it Rules? State Apparatuses and State Power under Feudalism, Capitalism and Socialism*, London; Verso.

Ticchi, Davide and Andrea Vindigni (2005) "Endogenous Constitutions," University Pompeu Fabra, Working Paper 896.

Tilly, Charles (1990) *Coercion, Capital and European States, AD 990-1990*, Cambridge MA; Blackwell.

Van Rijckeghem, Caroline and Beatrice Weder (2001) "Bureaucratic Corruption and the Rate of Temptation: Do Wages in the Civil Service Affect Corruption, and by How Much?," *Journal of Development Economics*, 65(2), 307-331.

Weyland, Kurt (1996) "Obstacles to Social Reform in Brazil's New Democracy," *Comparative Politics*, 29(1), 1-22.

Wilson, John D. (1990) "Are Efficiency Improvements in Government Transfer Policies Self-Defeating in Political Equilibrium," *Economics and Politics*, 2, 241-258.

World Bank (2004) *Doing Business in the World: Understanding Regulation*, World Bank, Oxford University Press.

Table 1. The Relationship Between Bureaucratic Wage and Government Expenditures

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent variable is government expenditures as a percentage of GDP				Dependent variable is government expenditures on social services and welfare as a percentage of GDP			
Average government wage to per capita GDP	-4.96 (1.01)	-3.63 (1.07)	-3.66 (1.03)	-1.82 (0.85)	-3.91 (0.65)	-2.85 (0.56)	-2.84 (0.56)	-0.80 (1.08)
Log of per capita real GDP		4.19 (1.63)	2.04 (2.18)	-0.90 (2.29)		4.02 (0.9)	2.91 (1.27)	0.24 (1.13)
Polity			1.7 (1.05)	0.17 (1.11)			0.84 (0.54)	0.01 (0.33)
Percentage of population between 15 and 64 years old				-0.69 (0.37)				-0.24 (0.16)
Percentage of population over 65 years old				1.95 (0.28)				1.28 (0.22)
R ²	0.23	0.31	0.34	0.57	0.3	0.5	0.52	0.77
Number of observations	51	51	51	51	46	46	46	46

Notes: Robust standard errors in parentheses. The sample consists of countries with an average polity index for the period 1990-98 greater than or equal to 5. The dependent variable in columns (1)-(4) is central government expenditures as a percentage of GDP for the period 1990-98. The dependent variable in columns (5)-(8) is consolidated central government expenditures on social services and welfare as a percentage of GDP for the period 1990-98. Average government wage to per capita GDP is the ratio of the total central government wage bill to the number of central government employees for the period 1991-2000. The log of per capita real GDP is the natural log of the real GDP per capita in constant 1985 dollars, averaged for the period 1992-98. All data are from Persson and Tabellini (2003) except the average government wage to per capita GDP, which are from the World Bank.

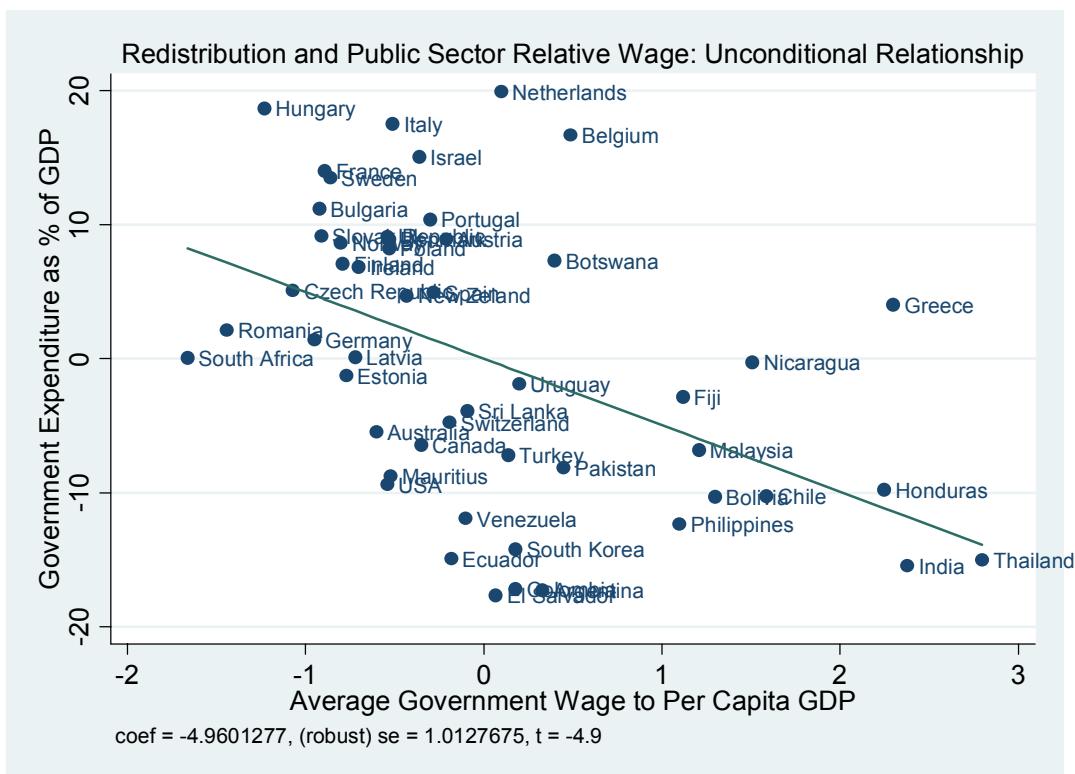


Figure 1. The figure reports the fitted values of the unconditional relationship between average government wage to per capita GDP and central government expenditures as percentage of GDP for the sample of countries with an average polity index for the period 1990-98 greater than or equal to 5 (see text for details).

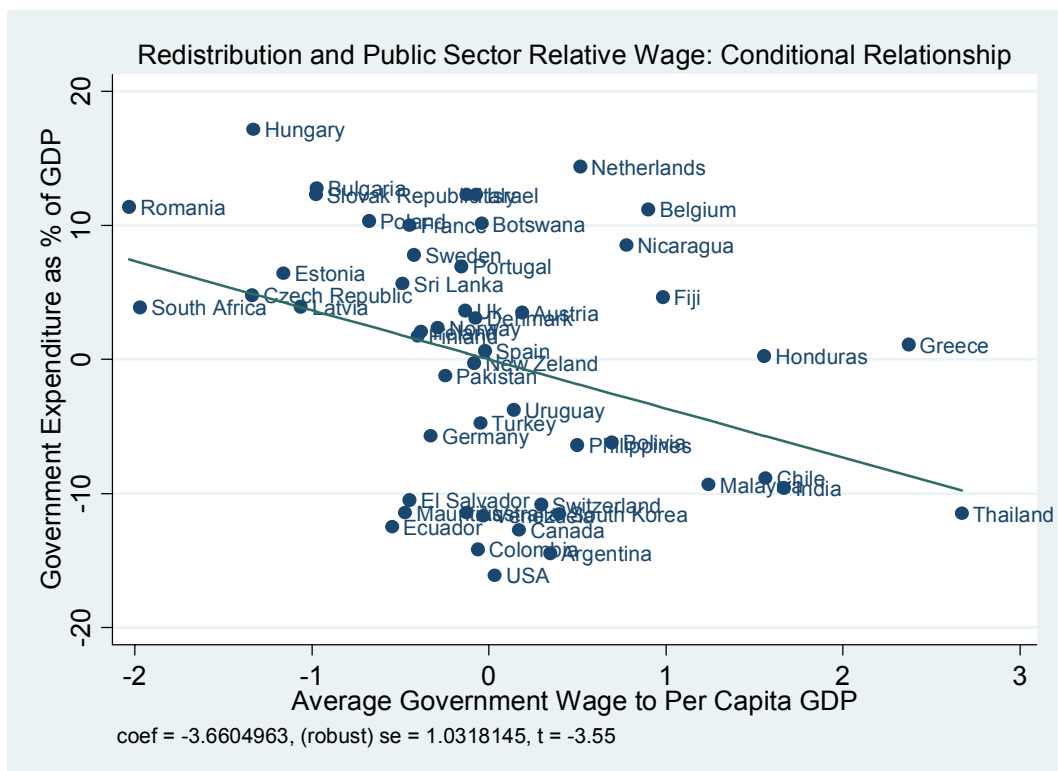


Figure 2. The figure reports the fitted values of the conditional relationship between average government wage to per capita GDP and government expenditures as percentage of GDP for the sample of countries with an average polity index for the period 1990-98 greater than or equal to 5. The control variables are the log of per capita real GDP and the average polity index (see text for details).