nccr trade regulation
swiss national centre of competence in research

Working Paper No 2013/31

Targeting Autocrats: Economic Sanctions and Regime Change

Manuel Oechslin

Abstract

When it comes to international economic sanctions, the most frequent goal is regime change and democratization. Yet, use and consequences of such sanctions are little understood. Past experiences suggest that they are often unsuccessful. Moreover, paradoxically, targeted regimes tend to respond by implementing policies which severely amplify the sanctions' harmful effects. This paper offers a political-economy model which provides an intuitive explanation for these observations. Autocratic regimes lower the supply of public goods to reduce private-sector productivity and hence the resources of potential challengers. As a result, sanctions-induced challenges become less likely - and the sanctions episode may end in failure.

JEL classification: F51; O11; O19

Keywords: Economic sanctions; regime change; democratization; public goods; exile

NCCR TRADE WORKING PAPERS are preliminary documents posted on the NCCR Trade Regulation website (<www.nccr-trade.org>) and widely circulated to stimulate discussion and critical comment. These papers have not been formally edited. Citations should refer to a "NCCR Trade Working Paper", with appropriate reference made to the author(s). Research for this paper was funded by the Swiss National Science Foundation under a grant to the National Centre of Competence in Research on Trade Regulation and the European Commission, based at the World Trade Institute of the University of Bern, Switzerland.

FNSNF
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

Die Nationalen Forschungsschwerpunkte (NFS) sind ein Förderinstrument des Schweizerischen Nationalfonds.
Les Pôles de recherche nationaux (PRN) sont un instrument d'encouragement du Fonds national suisse.
The National Centres of Competence in Research (NCCR) are a research instrument of the Swiss National Science Fondation.

Targeting Autocrats: Economic Sanctions and Regime Change*

Manuel Oechslin,
Tilburg University†

January 31, 2013

Abstract

When it comes to international economic sanctions, the most frequent goal is *regime change and democratization*. Yet, use and consequences of such sanctions are little understood. Past experiences suggest that they are often unsuccessful. Moreover, paradoxically, targeted regimes tend to respond by implementing policies which severely amplify the sanctions' harmful effects. This paper offers a political-economy model which provides an intuitive explanation for these observations. Autocratic regimes lower the supply of public goods to reduce private-sector productivity and hence the resources of potential challengers. As a result, sanctions-induced challenges become less likely – and the sanctions episode may end in failure.

JEL classification: F51; O11; O19

Keywords: Economic sanctions; regime change; democratization; public goods; exile

*I am grateful to Johannes Binswanger, Reto Föllmi, Ragnar Torvik, and seminar participants at the University of Cambridge, the University of Lausanne, Tilburg University, the University of Vienna, and the NEUDC Conference in Boston for helpful comments. Support from the NCCR Trade Regulation is acknowledged.

†Department of Economics, PO Box 90153, 5000 LE Tilburg, The Netherlands; phone: +31 13 4668089; fax: +31 13 4663042, email: m.oechslin@uvt.nl.

1 Introduction

Over the course of the 20th century, international economic sanctions have become an increasingly important foreign policy tool. Since the outbreak of World War I, there has been a total of 187 sanctions episodes, about 66 of which started after the collapse of the Soviet empire (Hufbauer et al., 2007; Tables 1A.1 and 1A.2). Economic sanctions usually combine restrictions on international trade and investment and are generally viewed as an instrument to induce specific changes in a target country. In practice, sanctioning states have indicated a variety of goals but the most frequent by far is to promote democratization by pushing autocratic (or even despotic) regimes out of power (46% of the cases in the 1914–2000 period).

Yet, despite their frequent use, our knowledge about how economic sanctions might foster *regime change and democratization* is very limited. There is a general notion that, as Mack and Khan (2000) put it, "the pain inflicted by sanctions on citizens of target states will cause them to pressure their government into making the changes demanded by the sanctioning body." But very little analytical work has actually been devoted to the exact channels through which sanctions are supposed to promote democratization. As a result, our understanding of the factors determining the likelihoods of success and failure is highly incomplete. It is the purpose of the present paper to make some progress in this regard by building a political-economy model which reflects some basic features of a typical target country.

A closer look at the history of economic sanctions aiming at regime change and democratization corroborates the view that a better understanding of their use and consequences is required. In particular, past experiences with such sanctions offer a number of observations that are puzzling. One of these observations is that targeted regimes hardly try to dampen the negative effects on the economy; targeted regimes rather tend to respond by pursing policies which severely compound the sanctions-induced hardship inflicted on the general population. For instance, as I will discuss in the following section, the governments of Haiti (in the 1991-1994 period) and Iraq (in the 1990-2003 period) hampered economic activity by aggressively cutting the supply of public services or by directly preventing people from pursuing their businesses.[1] Referring to the situation in Iraq, Mueller and Mueller (1999, p. 49) even note that "the country's political leadership sometimes seems more interested in maximizing the nation's suffering (...) than in relieving it." Similar evidence comes from other target countries: In

[1] Cross-sectional data suggests that reducing the supply of public goods and services is a widely observed reaction: According to evidence presented in Escriba-Folch and Wright (2010; Figure 2), personal rulers usually respond to external pressure by strongly reducing productive public spending (e.g., on public employment or capital goods) as a fraction of the GDP or the public budget.

February 2002, Zimbabwe's ruler responded to the threat of sanctions by "promising" to accelerate and broaden the implementation of his disastrous economic policies (*The Economist*; "Zimbabwe on the brink" and "The friendlessness of Robert Mugabe", February 19 and March 21, 2002). Clearly, such behavior requires explanation.

A second puzzling observation concerns the use of sanctions to promote regime change and democratization. It appears that many of these sanctions episodes ended prematurely, i.e., the sanctions were lifted after a couple of years even though the intended goal had not been achieved. According to Hufbauer et al. (2007), there was a total of 57 such episodes in the 1914–2000 period.[2] Yet, only 12 of them (or 21%) can be judged as at least "partly successful" – which means that the stated goal was at least "partly" achieved and that the sanctions' contribution was at least "substantial".[3] In 37 cases (or 65%), the sanctions were lifted although regime change and democratization was not even partly accomplished (after six years on average). Eight episodes were still ongoing in 2000. Again, this observation requires explanation: Why might a sanctioning body (e.g., the UN Security Council) suddenly abandon a long-standing sanctions regime that has not yet paid off?

The political-economy model I am proposing to look into these issues rests on three simple element. First, consistent with the focus on regime change and democratization, I consider an autocratic target country, i.e., a country where the government has substantial leeway to implement its preferred policies but also to divert public resources for its own benefit. Second, the state plays an important role in the private sector of the economy: By providing public goods and services, the government can affect the productivity of private firms and hence the citizens' incomes. Third, challenging the regime in order to promote a transition to democracy comes at an economic cost: During periods of power transitions the public sector is paralyzed so that the economy as a whole becomes less productive.

It turns out that this simple framework offers an intuitive explanation for both of the observations outlined above. At the heart of this explanation is that the incumbent elite may use the supply of public goods as a tool of defense, an aspect that has so far been neglected in the literature. To see how this works, consider a country that has just been placed under a sanctions

[2] A regime is classified as autocratic if the Polity score (Marshall and Jaggers, 2008) is zero or lower. Note that in 23 cases sanctions to promote regime change were imposed on fairly democratic states with a Polity score above zero. I do not look at these cases since I am interested in the impact of sanctions in autocracies.

[3] Sanctions imposed to promote goals other than "regime change and democratization" seem to be more successful on average. Again relying on the work by Hufbauer et al. (2007), the historical success rate among these remaining 124 episode is (a still modest) 36%. The observation that sanctions are frequently used despite their modest success rate is sometimes called the "Sanctions Paradox" (see Drezner, 1999).

regime. As intended, the application of economic pressure makes a switch to democracy more rewarding to the ordinary citizens. The sanctions regime thus renders a previously reluctant citizenry more inclined to revolt. Yet, this is exactly what the ruling elite wants to prevent as the consequences of a revolt would be more damaging than an orderly exit – which, however, requires time to organize. To buy this time, the elite has to discourage any challenges from the citizenry by making challenges more costly – and a straightforward way to do so is to decrease the supply of public goods: With a lower level of public goods, the citizens' incomes are lower and, as a consequence, a given cost associated with a revolt translates into a bigger loss in terms of instantaneous utility. The evidence I am going to discuss below suggests that the Haitian and Iraqi governments were quite successful in applying such a defense strategy. In particular, it appears that the Iraqi government's strategy to make the middle class struggle was instrumental in preventing serious challenges to its rule.

But not only the ordinary citizens fare badly. The situation for the ruling elite deteriorates as well because the economy generates less output due to the sanctions and the elite's own defense strategy. So it is looking for an exit that guarantees survival and income when out of power. Yet, as the quality of exile opportunities varies over time, it may be optimal for the elite to hold out for a while. This incentive to wait is the reason for why the sanctions may fail: While waiting, the elite has to "starve" the population continuously – which entails the possibility of losing all potential exile opportunities (because, e.g., the political cost to the leadership of potential host countries turns prohibitive). When this situation arises, the only option left is to hold out indefinitely. Observing this change in the elite's constraints, the sanctioning body lifts the sanctions regime as it does not want to harm the general population without improving the prospects of regime change and democratization.

Besides addressing use and effects of economic sanctions, the present framework also speaks to a number of related research questions. On the on hand, it offers a fresh perspective on the widespread phenomenon of underinvestment in public goods. Here, such underinvestment is not due to the lack of economic incentives (as, e.g., in Acemoglu, 2005; Oechslin, 2010); underinvestment is rather used as a defense tool that reduces the resources of potential challengers. On the other hand, by highlighting the role of secure exile destinations, the model contributes to a discussion in international relations or law (e.g., Goldsmith, 2003) on how to deal with ex dictators who have committed atrocities while in power.

In economics, the existing literature on international sanctions is relatively small. The seminal work by Schelling (1960, 1965) discusses fundamental principles governing the use and effectiveness of threats and promises designed to achieve a wide range of goals, among them

foreign policy goals. More closely related, Kaempfer and Lowenberg (1992) analyze the role of sanctions in helping the domestic opposition to overcome free rider problems associated with a revolt. Similarly, starting from Wintrobe's (1990) dictatorship model, Kaempfer et al. (2004) focus on the political impact of sanctions placed on autocracies to alter certain undesirable policies.[4] These latter two contributions connect with the present paper as they deal with the impact of sanctions on the political struggle in the target country. Yet, none of them explicitly deals with regime change and democratization nor does any of them address why threatened autocrats may find it optimal to compound the hardship inflicted by the sanctions. In terms of methods, there is a close link to work by Eaton and Engers (1992, 1999) who also rely on models of repeated games to explore the use and effectiveness of sanctions. Moreover, emphasizing incomplete information, the second contribution also suggests an explanation for why unsuccessful sanctions episodes may emerge in equilibrium. However, both contributions take the involved states as the actors and do not explicitly model the political struggle (or the economic environment) in the target country.

There is further a huge literature on economic sanctions in political science and international relations. The part of this literature most closely related to the present paper focuses on how political institutions determine the impact of economic sanctions on regime stability. For instance, Marinov (2005) shows empirically that sanctions are more likely to destabilize democratic governments than non-democratic ones (relatedly, Lektzian and Souva, 2007, find that sanctions against a democratic government are more likely to lead to policy compliance). Escriba-Folch and Wright (2010), on the other hand, provide evidence suggesting that the extent to which sanctions destabilize a non-democratic government depends on the type of authoritarian rule. Although related, it is clear that the present paper differs significantly from these contributions, both in terms of methods and focus.

The rest of this paper is organized as follows. Section 2 reviews anecdotal evidence documenting how threatened regimes responded in an extremely harmful manner to the imposition of sanctions. Section 3 introduces the basic model, which is then solved in Section 4. In Section 5, I present a straightforward extension of the basic framework that entails the emergence of unsuccessful sanctions episodes in equilibrium. Section 6 discusses the model's implications for related research questions. Section 7, finally, concludes.

[4] In an earlier paper, Kaempfer and Lowenberg (1988) propose a theory suggesting that sanctions might not be imposed to secure policy changes abroad but to benefit pressure groups within the sanctioning country.

2 Anecdotal Evidence from Haiti and Iraq

This section reviews some anecdotal evidence from two recent target countries, Haiti (1991–1994) and Iraq (1990–2003). Besides being well-studied cases, these two sanctions episodes have in common their prime purpose, namely ousting a highly autocratic regime in order to reinstate (Haiti) or establish (Iraq) democracy. In each case, before outlining the regime's policy responses, I will briefly review some facts about the imposed sanctions.

Haiti (1991–1994). In 1990, Haiti saw for the first time fair and democratic presidential elections. The vote produced a clear winner, Jean-Bertrand Aristide, who was sworn into office in February 1991 (see, e.g., Werleigh, 1995, for details). The new president promised to eliminate power and privileges of the old political elites. Yet, this process came to an abrupt halt when a military coup ousted Aristide after just seven months in office. In response, the Organization of American States, including the United States, imposed economic sanctions. Initially, these non-mandatory sanctions were targeted at the Haitian government but subsequently became more comprehensive and included severe limitations on most imports and exports from Haiti. Eventually, in June 1993, the United Nations imposed mandatory sanctions, comprising inter alia an oil and arms embargo. However, even with UN backing, the sanctions did not achieve their goal. In the end, it was a US military operation that restored democracy.

During the time the sanctions were in place, economic activity declined significantly (according to the World Bank's WDI, by about 10% p.a.) and the state almost ceased to function: As noted by Elizabeth Gibbons (1999, p. 31), then UNICEF's representative in Haiti, the new constitutional government "found a thoroughly debilitated, atrophied state structure" when taking over in October 1994. Moreover, Gibbons points out that "the state infrastructure was even more dilapidated than it had been in 1991". It is clear that the sanctions directly challenged the state's ability to perform its functions.[5] However, the rulers' own behavior significantly contributed to this decline. Evidence for the authorities' destructive role comes, inter alia, from the agriculture and infrastructure sectors: Werleigh (1995, pp. 166–7), for instance, points out that the authorities systematically destroyed part of the agricultural infrastructure and prevented technicians from offering their services to the farmers. In addition, by arbitrarily expropriating the farmers' returns, they "wiped away any incentive to invest in

[5] For instance, Gibbons (1999, p. 19) argues that the petrol embargo forced schools to operate only three days a week because teachers could not afford transportation to work. Some observers, however, are not convinced that direct effect of the embargo was substantial. They point out that embargo was only weakly enforced, partially lifted from time to time, and excluded the necessities of life (see, e.g., Werleigh, 1995, pp. 164–5).

economic activity." There is also evidence that hundreds of thousands of skilled farmers were forced to abandon their working places, causing a further reduction in agricultural production. Similarly, Gibbons and Garfield (1999, p. 1501) report that the authorities severely hampered the functioning of local water management committees and, as a consequence, compounded the supply problems stemming from the embargo.

Similar examples of harmful government policies can be found in the public health sector. In 1991, after decades under kleptocratic rule, the public health system had already been in a poor state (obviously, the Aristide government could not improve things in just seven months). Yet, even when compared to the dire standards of preceding autocratic regimes, the policies during the sanctions period were extremely poor. In particular, there is evidence that the new rulers deliberately weakened the battered health system even further: Gibbons (1999, p. 49), for instance, reports that there were occasions when the the authorities deprived the entire system of a broad set of vital supplies by impounding medicine, humanitarian fuel, etc.; further examples include the hasty abandonment of vital immunization programs. More generally, Gibbons (1999, p. 50) concludes that the government's (and the military's) policies in the public health sector were extremely harmful, thereby "directly contributing to the violations of the population's basic right to life and health." This assessment is supported by other local observers, among them Farmer et al. (2003, p. 420), who note that the "sharp fall in the quality and coverage of [health] services" was largely due to the striking "absence of commitment to public health on the part of the Haitian army."

In sum, the evidence suggests that government policies during the sanctions period were even poorer than those pursued during the three decades (1957–1986) of personal rule by the Duvalier family. There is no doubt that, by contributing to the collapse of public services and the country's infrastructure, Haiti's military and the various de facto governments in office between 1991 and 1994 considerably deepened the hardship caused by the sanctions.

Iraq (1990–2003). The sanctions against Iraq were imposed by the United Nations Security Council in August 1990 in response to the Iraqi invasion of Kuwait. The sanctions regime was comprehensive but did not include medicines, health equipment or foodstuffs (see, e.g., Reuther, 1995, for details). Moreover, in August 1991, the Security Council authorized the sale of oil in order to pay for the import of humanitarian goods. However, this opportunity was not taken up by the Iraqi government before 1996. The sanctions against Iraq had two goals, impairing the country's military potential and promoting regime change and democratization. While the first goal was probably achieved, the second clearly was not. Once again, it eventually took a

US-led military operation to oust the regime.

When it comes to government policies during the sanctions episode, there are many parallels to the Haitian case. With a comprehensive embargo in place (and the infrastructure damaged by the Gulf War), it was clearly difficult to keep public services going. Yet, many observers emphasize that the government actively contributed to worsening the situation. Reuther (1995, pp. 126, 130), for instance, reports that in large parts of the country the Iraqi regime did not even try to maintain the basic functions of the state or to repair the damaged infrastructure. Instead, the rulers deliberately deprived a substantial fraction of the population of government services and participation in the economy. About 3.1 million people in the three northern governorates were even subjected to a "total internal embargo" which came on top of the international (i.e., "external") sanctions. This "double embargo", jointly imposed by the international community and the Iraqi government, triggered a collapse of economic activity in the north of the country and – as a consequence – led to years of internal conflict between different rival Kurdish factions.[6] Clearly, this ongoing infighting substantially weakened the capacity of the Kurds to challenge the incumbent government in Baghdad.

As in the Haitian case, the public health sector provides further examples of destructive government behavior. For instance, in an article for the *New York Times* ("Were Sanctions Right?", July 27, 2003), David Rieff reports that – although there were more than enough drugs for the political elites – the authorities deprived the population of medical supplies so that ordinary citizens "had been subjected to two sets of sanctions, those of the United Nations and those of Saddam Hussein himself."[7] Such observations are also mirrored in a remark by Hans von Sponeck, then the UN coordinator for humanitarian assistance, who – according to the above-mentioned newspaper article – highlighted that "local repression and international sanctions became brothers-in-arms in their quest to punish the Iraqi people for something they had not done." Anecdotal evidence of this type may also constitute the basis of Mueller and Mueller's (1999, p. 49) conclusion that, after the imposition of sanctions, the Iraqi rulers were more interested in maximizing the nation's suffering than in relieving it.

Gibbons (1999, p. 39), finally, argues that the hardship caused by both the sanctions and the government's harmful policies decisively weakened the population's ability to challenge the Iraqi regime. This claim is supported by another report from Iraq in the *New York Times* ("As Hussein Builds, His People Struggle To Live", January 31, 1998) which cites a diplomat

[6] In addition to strangulating the economy, the regime cut off the region from food aid (Clawson, 1993, pp. 40-1). More generally, it seems that food supplies were diverted away from those who opposed the regime.

[7] According to Garfield (1999), Iraq had invested heavily in health services in the 15 years prior to the embargo and maintained an advanced public health system before the sanctions hit in 1991.

as saying that "if any sector of society outside the military might have formed a political opposition, the Iraqi middle class would have been the only hope." Yet, so the diplomat continues, "it has now been totally destroyed." What the diplomat meant was that most members of the middle class had to assume two or even three jobs to support their families; yet, in spit of these huge efforts, many families were not able to survive without food rations from government. Obviously, this daily struggle to make ends meet, combined with the dependence on the government, left little room for forming or joining an opposition movement.

Further observations. The patterns observed in Haiti and Iraq are not unique. Similar evidence comes from other target countries, notably Zimbabwe. As mentioned earlier, Zimbabwe's ruler responded to the threat of sanction in 2002 by announcing policies that were likely to ruin the country even faster. Moreover, Meredith (2007) presents casual evidence suggesting that the Zimbabwean regime systematically strangulated areas that had voted for the opposition so that starvation and death were likely to occur along party lines. Finally, it seems that the strategy of creating economic hardship in order to keep power is sometimes applied even in absence of any sanctions (which is also consistent with the model below). In a recent article on Egypt, *The Economist* ("An end or a beginning?", February 5, 2011) concludes that persistent poverty helped to prop up the regime. One basis for this conclusion is an observation by a young Cairo lawyer: "People survive on a day-to-day basis. They can't go for long without a daily wage and daily bread, so they can't afford to make trouble."

3 The Model

Agents, preferences, and economic activity. I consider an infinite-horizon economy in discrete time. The society starts out with two different players, the ruling elite (E) and the citizenry (N). Both groups derive utility from consumption of a non-storable final output good (which is also the numéraire). Preferences are given by the intertemporal utility function

$$U_{i,t} = E_t \left\{ \sum_{s=0}^{\infty} \beta^s \ln(c_{i,t+s}) \right\}, \tag{1}$$

where $c_{i,t}$ refers to consumption by player $i \in \{E, N\}$ in period t and $\beta \in (0,1)$ denotes the discount factor. Note that the instantaneous utility function is logarithmic, implying that the marginal utility rises as the level of consumption falls.[8]

[8] Logarithmic instantaneous utility is chosen for convenience. Other concave utility functions (e.g., the standard CRRA formulation $u(c) = c^{1-\theta}/(1-\theta)$, $\theta > 0$) would lead to qualitatively similar results.

The final good is produced by the citizenry only. Specifically, the citizenry has access to a technology which generates a net income (i.e., output minus cost of inputs) of

$$Y_t = A_t G_t \tag{2}$$

units of the final good. The first factor in (2), the productivity parameter A_t, is taken to reflect the "availability" of crucial foreign input factors. It also serves as the channel through which economic sanctions affect the domestic economy. More precisely, I assume that the imposition of trade sanctions increases the cost of foreign inputs and hence decreases their use – which is mirrored in a lower net income.[9] The second factor, G_t, refers to the level of the public good provided by the government. It captures in a simple way that the state plays an important role in promoting private economic activity by, for instance, building and maintaining infrastructure, upholding law and order, or enforcing private contracts. Note that G_t reflects the level of the public good at the time production takes place, i.e., towards the end of the period. As I will describe below, this level may be lower than the one provided initially as a result of damages associated with political turmoil.

Policy choices and the supply of the public good. In each period t, two policy variables have to be determined. First, there has to be a decision on the tax rate on the citizenry's income. The tax rate is denoted by $\tau_t \in [0, \tau^m]$, where $\tau^m < 1$. Limiting the government's ability to tax is necessary to generate interesting implications. Yet, imposing a maximum tax rate is just a reduced-form way of modeling more realistic limits to taxation which prevent the government from fully appropriating private incomes.[10]

Figure 1 here

The second policy choice is the supply of the public good, $X_t \in [0, X^m]$, where $0 < X^m < \infty$. The associated per-unit cost (in terms of the final good) is given by ϕA_t, where $\phi < 1$. An intuitive way of looking at this cost is to suppose that it reflects the number of government officials employed to produce the public good. From this perspective, the cost can be interpreted as the public wage bill which moves in lockstep with private-sector incomes. The assumption of a maximum supply, on the other hand, implies in a straightforward manner that there

[9] Note that (2) can be derived from a production function of the form $q = ((1-\alpha)^{-(1-\alpha)}\alpha^{-\alpha})(m)^{\alpha}(G)^{1-\alpha}$, $0 < \alpha < 1$, where m is the quantity of the foreign input (whose price, f, depends positively on the sanctions). The optimal choice of m leads to $Y = f^{-\alpha/(1-\alpha)}G = AG$, where A is a decreasing function of f.

[10] In reality, raising taxes is costly, and disproportionally so for high tax rates (as, e.g., modeled in Acemoglu and Robinson, 2001). Thus, the revenue-maximizing rate (which is an interpretation of τ^m) lies below 100%.

are decreasing returns in the production of the public good. The relationship between public expenses and the level of the public good is illustrated in Figure 1.

As mentioned above, the level of the public good available to the citizenry, G_t, deviates from the one initially supplied by the government, X_t, in times of political turmoil. More specifically, the relationship between X_t and G_t is given by

$$G_t = \max\{X_t - \eta_t \chi, 0\}, \tag{3}$$

where $\eta_t \in \{0, 1\}$ is an indicator variable that takes on the value 1 if the elite exits the economy and χ refers to the magnitude of the associated reduction. As we shall see, the exit of the elite may be the result of a popular revolt or, alternatively, due to a voluntary decission to flee the country. However, no matter what the exact reason is, assuming that the public good suffers in times of political turmoil is obvious. Myriad examples suggests that – when protesters clash with the regime or the regime abandons power abruptly – roads are blocked, law and order collapses, and enforcing contracts turns difficult. It is further natural to assume that, as implied by (3), the relative size of the reduction is larger if the supply of the public good is lower: An underdeveloped traffic infrastructure, for instance, means that one blocked road may be sufficient to cause gridlock; similarly, an understaffed judiciary is highly vulnerable to absenteeism (due to political turmoil). Regarding the magnitude of χ, I assume

$$\beta X^m < \chi < X^m. \tag{R1}$$

The lower bound is a non-crucial assumption that simplifies the exposition of the analysis (Footnote 14 discusses what happens if it is relaxed). The upper bound rules out that the exit of the elite implies zero output by assumption (which would be an uninteresting case).

Finally, as to the relationship between the social benefit and cost of the public good, it is clear that $X_t = X^m$ maximizes the social surplus as the marginal cost of providing the public good, ϕA_t, is only a fraction ϕ of the marginal impact on the aggregate output, A_t. Yet, as will be discussed below, the elite is not interested in the social surplus but in the government budget surplus (which it can appropriate when in power). The marginal impact of X_t on the budget surplus is $\tau_t A_t - \phi A_t$, where $\tau_t \in [0, \tau^m]$. In this regard, I impose

$$\tau^m > \phi. \tag{R2}$$

Parameter restriction (R2) ensures that, from an economic perspective, the elite prefers the same level of the public good as the citizenry does. Any deviation from X^m will thus be rooted in the fact that the elite uses the supply of the public good as a defense tool.

Political regimes and the transition of political power. There are two political regimes, dictatorship and democracy. In what follows, I denote the political state by $S_t \in \{R, D\}$, where D stands for democracy. Under dictatorship, economic policies are set by the elite. Moreover, with the exception of the period of its (possible) exit, the elite can appropriate and consume any fraction of the government budget surplus it likes. Democracy, on the other hand, means that economic policies are determined by the citizenry.

The economy starts as a dictatorship ($S_0 = R$). However, as long as $S_t = R$, the ruling elite's power is continuously threatened as the citizenry may revolt in any single period. The citizenry's decision in this respect is denoted by $\rho_t \in \{0, 1\}$, with 1 indicating a revolt. If the citizenry decides in favor of a revolt, democracy will be *irreversibly* established in the following period (i.e., $S_{t+1} = S_{t+2} = \cdots = D$). Moreover, in this case, the elite is immediately ousted. The result of such a forced and quick exit is that the elite definitively loses all sources of income (or even faces injuries or death) so that $U_{E,t} \to -\infty$.

A revolt is not the only road to democracy, however, as the elite may seize an opportunity to voluntarily leave the country for exile abroad. Yet, because political circumstances in potential host countries may be in flux, the existence of such an opportunity is not assured but only emerges with an exogenous probability $p \in (0, 1]$ in each period. The state variable in this regard is denote by $F_t \in \{0, 1\}$, with 1 meaning existence. If the elite seizes an existing opportunity, which is indicated by $\sigma_t = 1$, where $\sigma_t \in \{0, 1\}$, democracy will again be irreversibly established in the following period. Moreover, as of the current period, the elite's recurrent income is given by $\omega > 0$ so that $U_{E,t} = \ln(\omega)(1 - \beta)^{-1}$.[11] If the elite prefers to stay, though, the political state remains unchanged (i.e., $S_t = S_{t+1} = R$).

It is now clear that the "exit" variable introduced above is defined by $\eta_t \equiv \max\{\rho_t, \sigma_t\}$. The variable summarizes whether period t lays the foundation for a transition to democracy in $t + 1$: We have $S_t = S_{t+1} = R$ if $\eta_t = 0$ and $S_{t+1} = S_{t+2} = \cdots = D$ otherwise.

Economic sanctions. Economic sanctions affect the targeted economy by depleting the productivity of the domestic firms, A_t. I consider the case of continuous sanctions and assume that $A_t \in (0, 1]$, where the upper bound corresponds to the "natural" productivity level, i.e., the level in a sanctions-free environment. So a lower level of A_t corresponds to a tougher

[11] Relinquishing political power does not necessarily imply exile. The "retired" elite could also find arrangements that would allow it to stay (probably involving house arrest for some members). Both variants are actually present among the 12 successful sanctions episodes mentioned in the introduction (according to the Archigos dataset, available at http://www.rochester.edu/college/faculty/hgoemans/Archigos.2.9-August.pdf). So $\ln(\omega)(1 - \beta)^{-1}$ could also be interpreted as the lifetime utility under such an arrangement.

sanctions regime. In reality, a negative impact on productivity is typically achieved by imposing restrictions on international trade, especially the denial of critical imports (see, e.g., Hufbauer et al., 2007, pp. 44-5). It is obvious, though, that the import of critical input factors can never be perfectly prohibited (due to the possibility of smuggling or because some trading partners refuse to impose sanctions, for instance). In practice, the denial of critical imports means that these goods become more expensive but are still available. As a result, they are used in lower quantities so that productivity is impaired.[12]

In each period, the productivity level A_t is determined by a sanctioning body (e.g., the United Nations Security Council or the United States Congress). Consistent with the focus of this paper, the dominant goal of the sanctioning body is to promote regime change and democratization, i.e., a transition from state R to D.[13] Yet, the sanctioning body is not completely indifferent towards the fate of the citizens but cares about their well-being. In formal terms, the body's preferences are reflected by the instantaneous utility function

$$z_t = \Pr[S_{t+1} = D] + \epsilon\, c_{N,t}, \tag{4}$$

where ϵ is a constant that is strictly positive but arbitrarily close to zero. Moreover, we can think of the intertemporal utility function as the (infinite) sum of the discounted instantaneous utilities. Specification (4) essentially implies lexicographic preferences with respect to the two goals democratization and well-being of the citizenry: These preferences will induce the sanctioning body to take all measures necessary to maximize the chances of democratization in each period. Yet, should there be more than one way to do so, they make the sanctioning body prefer the option that is least damaging for the general population (and hence guarantee the uniqueness of the equilibrium).

Equilibrium concept and time-line. The focus is on the (pure strategy) Markov Perfect Equilibrium (MPE), where strategies depend only on the state of the system and prior actions within the same period. In the present setup, the state of the system is represented by the political state, $S_t \in \{R, D\}$, and the state reflecting exile opportunities, $F_t \in \{0, 1\}$.

Regarding the time-line, suppose first that $S_t = R$. Then, the timing of events is as follows: *First*, the sanctioning body opts for a certain level of A_t (by choosing the intensity of the trade sanctions accordingly). *Second,* the elite determines the policy vector, $\Pi_t = (\tau_t, X_t)$. *Third*,

[12] In the context of Footnote 9, $A = 1$ follows from $f = 1$, where 1 is assumed to be the undistorted price of the foreign input. $A = 0$, in contrast, is associated with $f \to \infty$ (i.e., the unavailability of the foreign input).

[13] For the body, regime change and democratization may be an end in itself (as, e.g., in the case of Haiti) and/or a means to promote other goals (like ending a belligerent regime as, e.g., in the case of Iraq).

the citizenry decides on whether or not to revolt against the elite, ρ_t. If $\rho_t = 1$, no further decisions are taken but the elite is forced to exit the game immediately, Π_t is implemented, the payoffs materialize, and the period ends. Otherwise, if $\rho_t = 0$, there is a *fourth* event: The elite learns F_t and, provided that $F_t = 1$, decides on leaving the country for exile abroad, σ_t. Finally, Π_t is implemented, the payoffs materialize, and the period ends.

Suppose now that $S_t = D$. Then, there will no longer be any decisions by the sanctioning body (because its goal has been achieved irreversibly) or by the elite (as it is no longer part of the game). The only actor that still takes decisions is the citizenry which determines the policy vector $\Pi_t = (\tau_t, X_t)$ at the beginning of the period. These policies are then immediately implemented, the payoffs materialize, and the period ends.

4 Analysis

This section derives the politico-economic equilibrium, focusing first on the choices under democracy before turning to the equilibrium strategies under dictatorship.

4.1 Equilibrium under Democracy

Suppose that $S_t = D$ so that productivity is at its maximum level ($A_t = 1$) and the elite is no longer part of the game. Under these circumstances, the budget constraint of the public sector is given by $\tau_t X_t \geq \phi X_t$. It is clear that this constraint must hold with equality: Imposing a tax rate higher than necessary to finance the public good is suboptimal. Therefore, in any single period, the citizenry prefers $\tau_t = \phi$, which is also feasible because of restriction (R2). The current level of consumption by the citizenry is thus given by

$$c_{N,t} = (1 - \phi)X_t.$$

It is further obvious that maximizing the above expression requires $X_t = X^m$ so that $\Pi_t = (\phi, X^m)$. Finally, since switching back to dictatorship is impossible, identical policies will be implemented in all future periods $t + 1$, $t + 2$, $\cdots$. As a result, once the political state has switched to D, the uniform level of lifetime utility incurred by the citizenry is

$$V(D) = \frac{\ln\left((1 - \phi))X^m\right)}{1 - \beta}. \tag{5}$$

Note that $V(D)$ is the highest lifetime utility the citizenry can achieve because, in each period, it consumes the full social surplus (which, in turn, is at its maximum level).

4.2 Equilibrium under Dictatorship

This subsection describes the equilibrium under dictatorship. I first derive the MPE for any given (exogenous) sanctions intensity. After that, I will provide the characterization of the full politico-economic equilibrium (which entails the endogenous determination of A).

4.2.1 Exogenous Sanctions Intensity

Under the assumption of an exogenous sanctions intensity, two different types of equilibria can be shown to exist. These equilibria are called "stable dictatorship" and "destabilized dictatorship". The sanctions intensity determines which one prevails.

Stable dictatorship. Suppose that $(\tau^m - \phi)AX^e > \omega$, where X^e is defined below. Then, there exists a unique MPE. In this equilibrium, we have $\Pi_t = (\tau^m, X^e)$ and $\rho_t = \sigma_t = 0$ (irrespective of the realization of F_t) at each $t \geq 0$. The equilibrium supply of the public good, X^e, is defined to be the maximum value of X in the range $[0, X^m]$ that satisfies

$$\left(\frac{(1-\phi)X^m}{(1-\tau^m)A}\right)^{\frac{\beta}{1-\beta}} \leq \frac{X^{\frac{1}{1-\beta}}}{X-\chi}. \tag{6}$$

Figure 2 illustrates inequality (6) by plotting the left-hand side (LHS) and right-hand side (RHS) separately as functions of X. Both panels of the figure suggest that X^e is a non-decreasing function of A. In Panel *a.*, the relationship is strictly increasing for all $A \in (0, 1]$. In Panel *b.*, there exists an $A^m < 1$ so that X^e strictly increases in A if $A < A^m$ and $X^e = X^m$ for all $A \geq A^m$.[14] Thus, under a parameter constellation that invokes the situation pictured in Panel *a.*, the supply of the public good is below its maximum even in absence of any sanctions (which is not true if Panel *b.* shows the relevant situation). The positive association between A and X^e suggests that the negative impact of economic sanctions on the overall output is compounded by the elite's harmful response: AX^e does not only fall because of the sanctions' direct impact but also because the elite lowers the supply of the public good.

Figure 2 here

One way to establish this MPE is to show that there are no one-stage deviations from the above actions that improve one of the actors' payoff (see, e.g., Acemoglu, 2009, Appendix C).

[14] It is easy to check that Panel *b.* is relevant if $((1-\phi)/(1-\tau^m))^{\beta/(1-\beta)} \leq X^m/(X^m - \chi)$. Note further that $\beta X^m \leq \chi$ imposed in (R1) implies that the RHS is non-increasing at $X = X^m$. This feature is not crucial, however: If the RHS were increasing, X^e would still be non-decreasing in A but might be discontinuous at some point. Allowing for this would complicate the analysis without generating any further insights.

To do so, we have to go backwards through a given period t. So suppose that the elite has just learnt that $F_t = 1$ and thus has to decide on whether or not to leave the country voluntarily. Depending on its decision, $\sigma_t \in \{0, 1\}$, the elite's current consumption level is

$$c_{E,t} = (1 - \sigma_t)(\tau^m - \phi)AX^e + \sigma_t \omega. \tag{7}$$

Yet, assuming that the described equilibrium actions are implemented from $t + 1$ onwards, $\sigma_t = 0$ means that the elite will continue to consume $(\tau^m - \phi)AX^e$ in all future periods. On the other hand, $\sigma_t = 1$ would imply a consumption level of ω in all future periods. As a result, since $(\tau^m - \phi)AX^e > \omega$, staying is the optimal choice.

Consider now the citizenry's decision on whether or not to oust the elite. Observe first that the citizenry's current level of consumption is $c_{N,t} = (1 - \tau_t)A_t G_t$, where G_t is given by (3). So, clearly, a revolt reduces $c_{N,t}$ and therefore the citizenry's instantaneous utility. There are also positive effects of a revolt, however: A revolt promises the maximum lifetime utility as of tomorrow. So when deciding on $\rho_t \in \{0, 1\}$, the citizenry compares the associated short-run cost to the long-run benefit – and decides against a revolt if the former exceeds the latter. In formal terms, the citizenry prefers $\rho_t = 0$ if and only if

$$\ln\left((1 - \tau_t)A_t X_t\right) + \beta E_t\{V(R, F_{t+1})\} \geq \ln\left((1 - \tau_t)A_t(X_t - \chi)\right) + \beta V(D),$$

where the left-hand side is the lifetime utility associated with $\rho_t = 0$ and $V(R, F_{t+1})$ denotes the continuation value associated with dictatorship. Rearranging terms yields

$$\beta\left(V(D) - E_t\{V(R, F_{t+1})\}\right) \leq \ln\left(X_t/(X_t - \chi)\right), \tag{8}$$

i.e., an expression that depends neither on the current tax rate, τ_t, nor on the the current productivity level, A_t. As a result, both variables are irrelevant when it comes to comparing the costs and benefits of a revolt in period t.[15] To check condition (8), remember that $V(D)$ is given by (5). Moreover, note that – given that the assumed equilibrium actions are implemented in all subsequent periods – all future realizations of F are irrelevant so that $E_t\{V(R, F_{t+1})\}$ is simply given by $\ln\left((1 - \tau^m)AX^e\right)(1 - \beta)^{-1}$. Taking these continuation values into account, the decisive condition for $\rho_t = 0$ turns into

$$\left(\frac{(1 - \phi)X^m}{(1 - \tau^m)AX^e}\right)^{\frac{\beta}{1-\beta}} \leq \frac{X_t}{X_t - \chi}. \tag{9}$$

[15] Although the two variables affect the absolute size of the loss in terms of current consumption, they are irrelevant for the magnitude of the relative loss – and it is the relative loss that determines the fall in instantaneous utility and therefore the short-run cost of a revolt.

Moving one step backwards, it remains to analyze the elite's policy choice in t. It is clear that – also in period t – the elite will opt for the maximum tax rate because τ_t has no impact on ρ_t and, other things equal, τ^m maximizes the current government budget surplus (and hence $c_{E,t}$). With respect to the choice of X_t, we have to check two requirements. First, condition (9) must be satisfied if $X_t = X^e$. Second, the elite must find it optimal to choose $X_t = X^e$. As to the first requirement, it is easy to check that $X_t = X^e$ turns condition (9) into condition (6) – which obviously holds for X^e. With respect to the second requirement, we have to distinguish the two cases $X^e = X^m$ and $X^e < X^m$. If $X^e = X^m$, the government budget surplus (and hence $c_{E,t}$) is maximized and so there is no incentive whatsoever to deviate. If $X^e < X^m$, condition (9) holds with equality; as a result, $X_t > X^e$ would induce a revolt – and hence will never be implemented as a revolt implies $U_{E,t} \to -\infty$.

To summarize, given that the actions described initially are implemented in all future periods $t+1$, $t+2$, $\cdots$, they will also be implemented in period t because there are no payoff-improving one-stage deviations. As a result, the existence of the equilibrium is established (while uniqueness will be discussed below).

Destabilized dictatorship. Suppose now that $(\tau^m - \phi)AX^e \leq \omega$, where X^e is again the maximum value of $X \in [0, X^m]$ satisfying (6). Then, there exists a unique MPE. In this equilibrium, we have $\Pi_t = (\tau^m, X^e)$, $\rho_t = 0$, and $\sigma_t = F_t$ at each $t \geq 0$ for which $S_t = R$ holds. So, in contrast to the stable-dictatorship case, this equilibrium entails that the elite leaves the country at the first opportunity. In all other respects, however, the two equilibria have similar properties. In particular, X^e is again determined by condition (6) so that the positive association between A and X^e remains unchanged.

The way to establish this MPE is similar to the one chosen above. So let's again start with the elite's decision on σ_t (after it has learnt that $F_t = 1$). Assuming that the described equilibrium actions are implemented from $t+1$ onwards, the level of consumption in exile, ω, is never lower than the one achieved when in power, $(\tau^m - \phi)AX^e$. It is therefore clear that, as soon as the opportunity emerges, exit is the preferred option.

Yet, the elite can only decide on leaving the country voluntarily if it has not been ousted by the citizenry at the previous stage. The decisive condition in this regard is now given by

$$\begin{aligned} & (1-p)\left[\ln\left((1-\tau_t)A_tX_t\right) + \beta E_t\left\{V(R, F_{t+1})\right\}\right] + p\left[\ln\left((1-\tau_t)A_t(X_t-\chi)\right) + \beta V(D)\right] \\ \geq \quad & \ln\left((1-\tau_t)A_t(X_t-\chi)\right) + \beta V(D), \end{aligned}$$

where the left-hand side is the expected lifetime utility associated with $\rho_t = 0$. Here, in

contrast to the stable-dictatorship case, abstaining from a revolt does not rule out a transition to democracy (and the materialization of the associated cost). Because the elite will exit the game if circumstances permit later on (which happens with probability p), a transition to democracy is still a possibility. Rearranging terms yields

$$\beta\left(V(D) - E_t\{V(R, F_{t+1})\}\right) \leq \ln\left(X_t/(X_t - \chi)\right), \tag{10}$$

i.e., a condition that is similar to (8) and hence also independent of τ_t and A_t. As was the case above, in order to check condition (10), we need an explicit expression for $E_t\left\{V(R, F_{t+1})\right\}$. In the appendix, I show that – given that the assumed equilibrium actions are implemented in all future periods $t+1$, $t+2$, $\cdots$ – this expression is

$$E_t\left\{V(R, F_{t+1})\right\} = \frac{(1-p)\ln\left((1-\tau^m)AX^e\right) + p\ln\left((1-\tau^m)A(X^e - \chi)\right) + \beta p V(D)}{1 - \beta(1-p)}. \tag{11}$$

Taking (5) and (11) into account, the decisive condition for $\rho_t = 0$ (i.e., the equivalent to condition 9) in the destabilized-dictatorship equilibrium turns into

$$\left(\frac{(1-\phi)X^m}{(1-\tau^m)A(X^e)^{1-p}(X^e - \chi)^p}\right)^{\frac{\beta}{1-\beta(1-p)}} \leq \frac{X_t}{X_t - \chi}. \tag{12}$$

The final step is again the analysis of the elite's policy choice. The argumentation with respect to the tax rate is unchanged and does not require reiteration. Similarly, one can show the optimality of $X_t = X^e$ along the lines discussed above. Imposing $X_t = X^e$ makes condition (12) equivalent to condition (6) so that the choice of X^e in period t does not induce a revolt. Moreover, as is the case in the stable-dictatorship regime, the elite will not implement a value higher than X^e because this would either be impossible (if $X^e = X^m$) or because this would lead to a revolt (if $X^e < X^m$) – which completes the proof of existence.

With respect to uniqueness, note that there are no MPE which entail a revolt along the equilibrium path. The reason is that – in anticipation of a revolt – the elite would always find a one-stage deviation that increases its payoff. Consider the situation in $t = 0$: Instead of choosing Π_0 in a way that would lead to $\rho_0 = 1$, the elite has the option to push X_0 sufficiently close to χ in order to prevent a revolt – and would certainly do so since a revolt is the worst possible outcome. It is further clear that equilibria in which the citizenry stays calm cannot entail policies different from $\Pi = (\tau^m, X^e)$. Consider first the tax rate. An equilibrium rate less than τ^m would again be associated with payoff-improving one-stage deviations on the part of the elite: In any given period t, raising the tax rate to τ^m would increase the elite's expected current consumption without affecting the citizenry's decision on ρ_t. Similarly, notwithstanding the relationship between F and σ, values of $X \in [0, X^m]$ other than X^e would

create opportunities for profitable one-stage deviations, either because a lower value of X_t would be required to avoid a revolt or because a higher value could be chosen without inducing a revolt. It is thus clear that the two equilibria described above are the only ones (and that the relationship between $(\tau^m - \phi)AX^e$ and ω determines which one prevails).

Summary and discussion. To summarize the findings so far, which is done in Proposition 1 below, it is convenient to define a critical productivity level A^c so that

$$(\tau^m - \phi)A^c X^e(A^c) = \omega, \tag{13}$$

where X^e is defined as above and hence a non-decreasing function of A.

Proposition 1 *Assume that (R1) and (R2) hold and suppose that the economy starts as a dictatorship ($S_0 = R$). Then, depending on the level of A (i.e., the intensity of economic sanctions), one of the two following Markov Perfect Equilibria prevails:*

- *Stable dictatorship: If $A > A^c$, there is a unique MPE where $\Pi_t = (\tau^m, X^e)$ and $\rho_t = \sigma_t = 0$ (irrespective of the realization of F_t) at each $t \geq 0$.*
- *Destabilized dictatorship: If $A \leq A^c$, there is a unique MPE where $\Pi_t = (\tau^m, X^e)$, $\rho_t = 0$, and $\sigma_t = F_t$ at each $t \geq 0$ for which $S_t = R$ holds.*

Proposition 1 highlights two important implications of this model. First, the sanctions must be sufficiently harmful to promote regime change and democratization.[16] Second, even in the destabilized-dictatorship equilibrium, the citizenry never revolts against the elite but waits for the elite to exit voluntarily. A third implication, illustrated by Figure 3, is that the supply of the public good decreases as the sanctions intensity increases (i.e., as A falls).

Figure 3 here

These three implications are closely related to each other. In order to make the citizenry willing to bear the cost of a revolt, a switch to democracy (and the associated end of the sanctions regime) must be sufficiently rewarding. A revolt, however, is exactly what the elite must avoid as a forced removal from power would lead to the lowest possible continuation value. So the elite tries to do whatever it takes to prevent a revolt – which, however, does not rule out voluntary exit (should such an exit prove feasible and beneficial). The supply of

[16] This statement requires a qualification. It is possible that $A^c \geq 1$, which would mean that the destabilized-dictatorship equilibrium prevails even in the absence of any sanctions (i.e., even if $A = 1$). Without loss of generality, the remainder of this paper focuses on parameter constellations for which $A^c < 1$ holds.

the public good, finally, is the tool of defense the elite may resort to. Clearly, facing economic sanctions, the elite must increase the cost of a revolt in order to stay unopposed – and it does so by lowering the supply of the public good. A lower level of the public good means that the citizenry earns less income and thus has a higher marginal valuation of consumption. As a result, any further loss of income induced by a revolt translates into a bigger cost in terms of lower instantaneous utility.

It is finally interesting to note that the elite's defense strategy works despite the fact that it also increases the benefit of a revolt (as it implies an even lower supply of the public good under dictatorship). It must thus be the case that the increase in the cost of a revolt outweighs the rise in the benefit. The main reason is the shape of the instantaneous utility function: Since instantaneous utility is concave, reducing the supply of the public good increases the utility cost of a revolt by a larger amount than it widens the difference between the levels of the instantaneous utility under dictatorship and democracy.

4.2.2 Endogenous Sanctions Intensity

The politico-economic equilibrium. It remains to characterize the full politico-economic equilibrium which entails the the endogenous determination of the sanctions intensity. With A chosen by the sanctioning body in every period, there exists a unique MPE that involves the use of sanctions. In line with the sanctioning body's goal, this MPE is a destabilized-dictatorship equilibrium with $A_t = A^c$ at each t for which $S_t = R$ holds.

To establish this equilibrium, one has to show that – given that the assumed equilibrium actions are implemented in all future periods $t+1$, $t+2$, $\cdots$ – the sanctioning body does not have any incentive to deviate from A^c in period t.[17] To see that no such deviation incentive exists, remember that the current productivity level, A_t, is irrelevant for the citizenry's decision on ρ_t (inequality 12); remember further that, independently of A_t, the elite prefers the highest possible levels of τ_t and X_t (which means $\tau_t = \tau^m$ and $X_t = X^e(A^c)$, as $X_t > X^e(A^c)$ would imply a revolt). So the sanctioning body's choice of A_t has no impact on Π_t or ρ_t (provided that $A_{t+1} = A_{t+2} = \cdots = A^c$). Yet, a deviation from A^c may have an impact on the elite's exit decision. In particular, if the sanctioning body chose $A_t > A^c$, the elite would stay irrespective of the realization of F_t (as it is exactly indifferent between staying and leaving if $A_t = A^c$). As a result, such a move would imply $\Pr[S_{t+1} = D] = 0$ and hence a fall in the sanctioning

[17] Note that the destabilized-dictatorship equilibrium discussed in Subsection 4.2.1 was exactly established under the assumption of a time-invariant A (as of period $t+1$). As a result, the trade-offs considered by the elite or the citizenry when deciding on their actions in period t are unchanged.

body's instantaneous utility z_t (remember that ϵ is arbitrarily small). On the other hand, if the sanctioning body chose $A_t < A^c$, the probability of a democratic transition would be unchanged but the citizenry's current level of consumption would fall. Again, this would mean a lower instantaneous utility for the sanctioning body. So the conclusion is that there are no payoff-increasing deviations from $A_t = A^c$.

As for uniqueness, note that levels other than A^c cannot be part of a MPE involving the use of sanctions. A higher level would imply a stable-dictatorship equilibrium and hence $\Pr[S_{t+1} = D] = 0$ at each t. As a result, the sanctioning body could lift its payoff by opting for $A_t = 1$. On the other hand, a level lower than A^c would be associated with the destabilized-dictatorship equilibrium so that $\Pr[S_{t+1} = D] = p$ at each t. In this case, the sanctioning body could lift its payoff by opting for, e.g., $A_t = A^c$ as such a deviation would not affect the probability of democratization but lift the citizenry's current level of consumption.

Finally, note that – besides this destabilized-dictatorship regime – there exists a second MPE which, however, does not involve the use of sanctions. As can be easily checked, this second MPE is a stable-dictatorship equilibrium with $A_t = 1$ at each t.[18]

Summary and discussion. The characterization of the politico-economic equilibrium is now complete. Proposition 2 summarizes the findings derived in Subsections 4.2.1 and 4.2.2.

Proposition 2 *Assume that (R1) and (R2) hold and suppose that the economy starts as a dictatorship ($S_0 = R$). Then, there exists a unique MPE that involves the use of sanctions. This MPE is a destabilized-dictatorship equilibrium (as described in Proposition 1) that entails $A_t = A^c$ at each t for which $S_t = R$ holds, where A^c is defined by (13). The probability of regime change and democratization is thus given by $p \leq 1$ in every period.*

Looking at Proposition 2, one finding is particularly noteworthy: Although the sanctioning body could bring economic activity almost to a full stop, and although the citizenry could oust the elite in every period, none of these actually happens. Instead, the applied sanctions intensity just makes the elite's life in power unpleasant so that it leaves for exile voluntarily at the first opportunity. This pattern reflects that the citizenry is taken hostage by the elite: Controlling the supply of the public good, the latter determines how much resources the citizenry can generate – and hence the citizenry's willingness to revolt. So, as the sanctions intensity

[18] Note further that the proof existence of the destabilized-dictatorship equilibrium relies on the fact that $p > 0$ (if p were zero, choosing $A_t = 1$ over $A_t = A^c$ would increase the sanctioning body's payoff). So, if $p = 0$, the stable-dictatorship equilibrium (with $A_t = 1$ at each t) would be the unique equilibrium.

rises, the elite simply reduces the citizenry's resources to make sure that the incentives to revolt remain insufficient. Anticipating this defense strategy, the sanctioning body must content itself with promoting a voluntarily exit of the elite.

Such a voluntary exit, however, requires time as a suitable destination is not always available. So the sanctioning body may be forced to apply a certain sanctions intensity over an extended period of time (an issue at which the following section looks at). As can be derived from equation (13) and the definition of X^e, the critical sanctions intensity is decreasing in ω and increasing in χ. The intuition is straightforward in the case of ω: If a life in exile is comfortable, a relatively small fall in the rents from holding power at home is sufficient to make the elite leave. A high χ, on the other hand, means that a small decrease in the supply of the public good has a relatively large discouraging impact on the citizenry. So, for a given sanctions intensity, the supply of the public good and the elite's income are higher – which means that the sanctions must be more intense to make the elite leave.

5 Unsuccessful Sanctions

The successful use of international sanctions, as has just been discussed, requires a consistent and potentially long-standing application of a certain sanctions intensity. This section explores difficulties that may be associated with this requirement. In particular, I demonstrate that the need for patience may be one reason for why sanctions to promote regime change and democratization are often lifted before being successful.

5.1 A Modified Setup

The simple extension of the baseline model presented in this section rests on the notion that the elite's exit options may deteriorate over time. This is an obvious idea. To defend itself, the elite has to "starve"the population. Yet, the consequences of this defense strategy – while being felt by the local population immediately – may not be perceived for some time outside the target country (e.g., because there are usually many more trouble spots that compete for international media coverage). Yet, as soon as it becomes widely known that the elite takes an active part in hampering economic activity, the political leadership of potential host countries may become reluctant to offer exile: As the reputation of the threatened elite gets (even more) damaged, the political cost of offering exile may turn prohibitively high.[19] Similarly, the longer

[19] Consider Zimbabwe, a country that has been facing sanctions since 2002 and whose population suffers under disastrous economic policies. There is no doubt that South Africa, to which many elite members have close

the elite starves the population, the more difficult it may get to find an arrangement that would allow the elite to stay in the country after having relinquished power.

A simple way of introducing this idea into the present setup is to modify the state variable reflecting exile opportunities. In particular, assume that there are now three possible states: $F_t \in \{0, 1, 2\}$, where 2 is an absorptive state, reflected in the assumption $\Pr[F_t = 2|F_{t-1} = 2] = 1$. Put differently, once F has switched to 2, exile opportunities will no longer emerge. Suppose further that, given $F_{t-1} \neq 2$, F switches to this absorptive state with an exogenous probability q in each period; the probabilities of $F_t = 0$ and $F_t = 1$ are given by $(1-q)(1-p)$ and $(1-q)p$, respectively. No other modifications are introduced.

5.2 Analysis

To describe the politico-economic equilibrium in this modified setup, I have to distinguish the two cases $F_{t-1} = 2$ and $F_{t-1} \neq 2$.

All exile opportunities lost ($F_{t-1} = 2$). If F has reached its absorptive state, neither the current period t nor any future period will offer the elite an opportunity to leave. As a result, the extended version of the model is identical to the baseline version with $p = 0$ – which means that the only MPE is a stable-dictatorship equilibrium with $A_t = 1$ at each t (see Subsection 4.2.2, in particular Footnote 18). Reflecting this stationary environment, the value functions of the citizenry and the elite are simply given by, respectively,

$$V(R, 2) = \frac{\ln\left((1-\tau^m))X^m\right)}{1-\beta} \quad \text{and} \quad W(R, 2) = \frac{\ln\left((\tau^m - \phi))X^m\right)}{1-\beta},$$

where – without loss of generality – the formulations of the two value functions assume $X^e(1) = X^m$. Intuitively, if the elite does not have any option to leave, economic sanctions are impotent as the elite takes all measures necessary to prevent a revolt. Recognizing this fact, the sanctioning body lifts any pre-existing sanctions as it does not want to harm the citizenry without improving the prospects of regime change and democratization.

Upcoming exile opportunities ($F_{t-1} \neq 2$). As long as exit options arise with a positive probability, the only politico-economic equilibrium involving the use of sanctions is similar to

ties, would be a natural exile option. Yet, the country is turning slowly into a place hostile to the Zimbabwean elite. In 2008, powerful South African trade unions asked their members to never ever serve Mugabe anywhere (*Reuters*, "South Africa unions calls for isolation of Mugabe", June 24, 2008). More recently, South Africa's High Court ruled that the authorities must prosecute members of the Zimbabwean elite suspected of having committed egregious crimes (*New York Times*, "A Landmark Ruling in South Africa", May 14, 2012).

the one in the baseline model. In particular, we have a destabilized-dictatorship equilibrium with $A_t = A^c$ at each t for which $F_{t-1} \neq 2$ and $S_t = R$ hold. However, there are some slight changes to the definitions of X^e and A^c. Specifically, X^e is now defined to be the maximum value of X in the range $[0, X^m]$ that satisfies the inequality

$$\left(\frac{[(1-\phi)X^m]^{(1-\beta(1-q))/(1-\beta)}}{(1-\tau^m)A[(1-\tau^m)X^m]^{\beta q/(1-\beta)}}\right)^{\frac{\beta}{1-\beta(1-q)}} \leq \frac{X^{\frac{1+\beta q}{1-\beta(1-q)}}}{X-\chi}, \tag{14}$$

i.e., an expression that has the same properties as inequality (6). Most importantly, X^e is still a non-decreasing function of A, as a figure similar to Figure 2 would immediately reveal. The modified A^c, on the other hand, is defined by

$$(\tau^m - \phi)A^c X^e(A^c) = \omega\left(\frac{\omega}{(\tau^m-\phi)X^m}\right)^{\frac{\beta^2 q}{(1-\beta)(1+\beta q)}}, \tag{15}$$

where the second factor on the right-hand side is less than one (see Footnote 16).

The way to establish existence and uniqueness of this equilibrium is similar to the one chosen in the previous section. The exposition here is therefore brief and focuses mainly on the derivations of X^e and A^c. I start again with the elite's decision on whether or not to exit (provided that $F_t = 1$). This decision is now more involved. When deciding on σ_t, the elite can no longer just compare the period incomes when in power and facing sanctions, $(\tau^m - \phi)AX^e$, or when in exile, ω. It also takes into account that the sanctions may be lifted when it is still in power. Sticking to the assumption that the elite leaves if holding out is not strictly preferred, the appendix derives that $(\tau^m - \phi)A^c X^e(A^c)$ must be determined by

$$\begin{aligned} \ln(\omega) + \beta\frac{\ln(\omega)}{1-\beta} &= \ln((\tau^m-\phi)A^c X^e(A^c)) \\ &\quad +\beta\left[(1-q)\frac{\ln\omega}{1-\beta} + q\left(\ln((\tau^m-\phi)A^c X^e(A^c)) + \beta W(R,2)\right)\right], \end{aligned} \tag{16}$$

from which equation (15) follows immediately. The left-hand side of (16) gives the value associated with $\sigma_t = 1$ (if $F_t = 1$). The right-hand side is the expected value if the elite decided to stay (in spite of $F_t = 1$). In this case, the expected continuation value reflects that F will take on the value 2 with probability q in the following period.

Moving one step backwards, the decisive condition for the absence of a revolt ($\rho_t = 0$) is

$$\begin{aligned} & (1-\tilde{p})\left[\ln\left((1-\tau_t)A_t X_t\right) + \beta E_t\left\{V(R, F_{t+1})\right\}\right] + \tilde{p}\left[\ln\left((1-\tau_t)A_t(X_t-\chi)\right) + \beta V(D)\right] \\ \geq\ & \ln\left((1-\tau_t)A_t(X_t-\chi)\right) + \beta V(D), \end{aligned}$$

where $\tilde{p} \equiv (1-q)p$. Simplifying the above expression leads again to condition (10). An expression for $E_t\left\{V(R, F_{t+1})\right\}$ is derived in the appendix, and the use of this expression in

(10) implies that the citizenry refrains from revolting if

$$\left(\frac{[(1-\phi)X^m]^{(1-\beta(1-q))/(1-\beta)}}{(1-\tau^m)A(X^e)^{1-\tilde{p}}(X^e-\chi)^{\tilde{p}}[(1-\tau^m)X^m]^{\beta q/(1-\beta)}}\right)^{\frac{\beta}{1-\beta(1-q)(1-p)}} \leq \frac{X_t}{X_t-\chi}. \qquad (17)$$

Condition (17), which parallels condition (12), is again the basis for the analysis of the elite's policy choice. In this regard, note that one can use exactly the same arguments as in Section 4 to establish that – also in this case – the elite has no incentives to deviate from the policy vector $\Pi_t = (\tau^m, X^e)$.[20] Similarly, the absence of any deviation incentives on the part of the sanctioning body can be shown along the lines discussed in Section 4.

Summary and discussion. Proposition 3 summarizes the key implications of the extended version of the model:

Proposition 3 *Consider the extended version of the model and assume that (R1) and (R2) hold. Moreover, suppose that the economy starts as a dictatorship ($S_0 = R$) whose elite may still receive exile opportunities ($F_{-1} \neq 2$). Then, the MPE is characterized as follows:*

- *As long as $F_{t-1} \neq 2$ (and $S_t = R$), there exists a unique MPE that involves the use of sanctions. This MPE is a destabilized-dictatorship equilibrium where $A_t = A^c$, $\Pi_t = (\tau^m, X^e)$, $\rho_t = 0$, and $\sigma_t = 1$ iff $F_t = 1$ (with the definitions of A^c and X^e given in 5.2). The probability of democratization is given by $(1-q)p < 1$ in every period.*

- *As soon as $F_{t-1} = 2$ (and $S_t = R$), there exists a unique MPE. This equilibrium is a stable-dictatorship equilibrium where $A_t = 1$, $\Pi_t = (\tau^m, X^m)$, and $\rho_t = \sigma_t = 0$. The chances of democratization are therefore zero.*

A comparison of Propositions 2 and 3 immediately reveals the key difference between the two versions of the model. While the baseline model predicts that the sanctions are in place until democracy is established, the extended version allows for thier abolishment even if the central policy goal, regime change and democratization, has not been achieved: A sanctions regime is lifted as soon as it turns ineffective, even if the ruling elite is still in power. Put differently, in the extended version of the model, the sanctioning body makes use of sanctions although it anticipates that they will fail with a certain probability.

A second difference between the two versions becomes apparent when one compares equations (13) and (15). The two equations determine the aggregate output, $A^c X^e(A^c)$, if the economy is subject to sanctions. As the right-hand side of (15) is smaller than that of (13),

[20] In particular, note that τ_t does not enter condition (17) and that $X_t = X^e$ turns (17) into (14).

the output in the extended version of the model must be lower. The need for more damaging sanctions in the latter case is due to the fact that there is a positive probability that the sanctions will be lifted even if the elite holds on to power. So, other things equal, the expected continuation value associated with holding out is higher – which requires the sanctioning body to inflict greater harm. Note further that the right-hand side of (15) is decreasing in q. So the equilibrium intensity of a sanctions regime is the higher the less likely it is to succeed.

6 Discussion

Besides examining the use and effects of sanctions, the present framework contributes to related areas of research. Specifically, it offers a novel perspective on the widespread phenomenon of underinvestment in public goods. In the context of non-democratic politics, the literature emphasizes that underinvestment may be the result of the fact that self-interested elites do not consider the economy-wide benefits of such investments; instead, the elites only consider their private rewards – which are considerably lower if the state has a low capacity to tax (e.g., Acemoglu, 2005) or if the incumbent elites have a short time-horizon in office (e.g., Oechslin, 2010). In the present framework, underinvestment in public goods has a completely different root: It is a defense strategy. In fact, here, the ruling elite would prefer to implement the maximum level of the public good. But it may not be advisable to do so even in absence of any sanctions (see Figure 2*a*.) because a better supply of public goods also means higher earnings for the general population. Higher earning, on the other hand, imply that a given fall in incomes – which is the inevitable result of a revolt – is less harmful (i.e., leads to a lower fall in instantaneous utility from consumption). Hence we get the seemingly paradoxical result that providing more public goods makes the citizens more inclined to revolt against a self-interested elite. The elite therefore keeps the supply of public goods, and hence the general income level, low in order to ensure that – in the words of the lawyer quoted at the end of Section 2 – "... [people] can't afford to make trouble".

By emphasizing the role of exile options, this paper may further contribute to a literature in political science and law on how to deal with ex dictators who have committed atrocities while in power. Part of this literature argues that a systematic prosecution (e.g., by the International Criminal Court, ICC) of such ex-dictators is problematic because – anticipating prosecution after leaving power – sitting dictators may be induced to hold on to power at all cost (e.g., Goldsmith, 2003). Although there may be many arguments in favor of prosecution, the present framework supports the significance of this concern in the context of sanctions

aiming at regime change and democratization. If a more systematic prosecution translated into a higher probability of a complete loss of exile destinations (i.e., into a higher q), more sanctions episodes would end in failure. Moreover, a higher q would imply that active sanctions regimes must have a higher intensity to make a dictator seize an upcoming exile opportunity. As a result, the general population would suffer from even harsher economic conditions under an active sanctions regime. From the perspective of this model, only measures that improve – rather than impair – exile prospects can benefit the general population. Examples would be arrangements that accelerate the emergence of a secure exile destination or arrangements that enhance the elite's welfare in exile (i.e., measures that raise p or ω).

7 Conclusions

This paper develops a political-economy model to study the use and impact of international economic sanctions aiming at regime change and democratization. The model suggests that, when threatened by such sanctions, a dictatorial regime may use the supply of public goods and services as a tool of defense. The intuition is straightforward. As intended, the imposition of sanctions makes a previously reluctant citizenry more inclined to revolt. Thus, to prevent an immediate ouster, the elite has to increase the cost of a revolt – and it can do so by reducing the supply of public goods. A lower supply means lower incomes for the citizenry and hence more strain (i.e., a steeper fall in utility) associated with a revolt's destructive effects. This defense strategy remains in place until the elite leaves voluntarily, catching an appropriate exile opportunity. Yet, a sanctions regime does not necessarily end in success. Waiting entails the possibility that exile opportunities disappear altogether. In this case, holding out indefinitely becomes the only option – and the sanctioning body gives in.

The present analysis offers a coherent perspective on past experiences with sanctions imposed to promote regime change and democratization. It suggests an explanation for why targeted regimes – far from trying to mitigate the consequences for the general population – respond by taking measures which severely amplify the sanctions' direct effects. Moreover, the model is able to match the typical course of a sanctions episode: In many cases, the sanctions were kept in place for a number of years but eventually abandoned although the desired result had not been achieved. Although the model's implications are consistent with substantial anecdotal evidence, it is undisputed that there may be other channels – which are ignored here – that could play a role in explaining the observed patterns. For instance, targeted regimes may compound the economic impact for propaganda purposes (i.e., to highlight how "inhumane"

the sanctions are). Or a sanctions regime may be lifted "prematurely" because the stated goal (i.e., regime change and democratization) was not the true one. I consider the exploration of such channels to be an interesting area of future research.

It is finally worthwhile to point out that the model contributes to related fields of research. On the one hand, it offers a novel explanation for the widespread phenomenon of under-investment in public goods. In the present framework, such under-investment is not the result of a lack of economic incentives on the part of the ruling group. In fact, under-investing elites would prefer to spend more on productive public goods – but cannot do so without risking a revolt. On the other hand, the model may be relevant for the debate on the effects of prosecuting "retired" dictators for atrocities committed while in power. The spotlight here is on a possible negative effect. More aggressive prosecution could mean that secure exile options dwindle more quickly. If so, a sanctioning body will apply a higher sanctions intensity and thus hurt the general population even more. A more explicit application of the present framework to these areas of research is another obvious avenue for future work.

References

[1] Acemoglu, Daron (2005); "Politics and Economics in Weak and Strong States," *Journal of Monetary Economics*, 52(7), 1199-1226.

[2] Acemoglu, Daron (2009); *Introduction to Modern Economic Growth*. Princeton: Princeton University Press.

[3] Acemoglu, Daron and James Robinson (2001); "A Theory of Political Transitions", *American Economic Review*, 91(4), 938-963.

[4] Clawson, Patrick (1993); *How has Saddam Hussein Survived? Economic Sanction, 1990-1993*. Washington, DC: National Defense University.

[5] Drezner, Daniel (1999); *The Sanctions Paradox. Economic Statecraft and International Relations*. Cambridge, UK: Cambridge University Press.

[6] Eaton, Jonathan and Maxim Engers (1992); "Sanctions", *Journal of Political Economy*, 100(5), 899-928.

[7] Eaton, Jonathan and Maxim Engers (1999); "Sanctions: Some Simple Analytics", *American Economic Review, P&P*, 89(2), 409-414.

[8] Escriba-Folch, Abel and Joseph Wright (2010); "Dealing with Tyranny: International Sanctions and the Survival of Authoritarian Rulers", *International Studies Quarterly*, 54(2), 335-359.

[9] Farmer, Paul, Mary Smith, and Patrice Nevil (2003); "Unjust Embargo of Aid for Haiti", *The Lancet*, 361, 420-423.

[10] Garfield, Richard (1999); *The Impact of Economic Sanctions on Health and Well-being*. London: Overseas Development Institute.

[11] Gibbons, Elizabeth (1999); *Sanctions in Haiti: Human Rights and Democracy under Assault*. Westport, CT: Praeger Publishers.

[12] Gibbons, Elizabeth and Richard Garfield (1999); "The Impact of Economic Sanctions on Health and Human Rights in Haiti, 1991-1994," *American Journal of Public Health*, 89(10), 1499-1504.

[13] Goldsmith, Jack (2003); "The Self-Defeating International Criminal Court", *University of Chicago Law Review*, 70(1), 89-104.

[14] Hufbauer, Gary Clyde, Jeffrey Schott, Kimberly Ann Elliott, and Barbara Oegg (2007); *Economic Sanctions Reconsidered* (3rd ed.). Washington, DC: Peterson Institute for International Economics.

[15] Kaempfer, William and Anton Lowenberg (1988); "The Theory of International Economic Sanctions: A Public Choice Approach", *American Economic Review*, 78(4), 786-793.

[16] Kaempfer, William and Anton Lowenberg (1992); "Using Threshold Models to Explain International Relations", *Public Choice*, 73(4), 419-443.

[17] Kaempfer, William, Anton Lowenberg, and William Mertens (2004); "International Economic Sanctions Against a Dictator", *Economics and Politics*, 16(1), 29-51.

[18] Lektzian, David and Mark Souva (2007); "An Institutional Theory of Sanctions Onset and Success", *Journal of Conflict Resolution*, 51(6), 848-871.

[19] Mack, Andrew and Asif Khan (2000); "The Efficacy of U.N. Sanctions", *Security Dialogue*, 31(3), 279-292.

[20] Marinov, Nikolay (2005); "Do Economic Sanctions Destabilize Country Leaders?", *American Journal of Political Science*, 49(3), 564-576.

[21] Marshall, Monty and Keith Jaggers (2008); *Polity IV Project: Political Regime Characteristics and Transitions, 1800-2008*. Electronic resource.

[22] Meredith, Martin (2007); *Mugabe: Power, Plunder, and the Struggle for Zimbabwe*. New York: Public Affairs.

[23] Mueller, John and Karl Mueller (1999); "Sanctions of Mass Destruction", *Foreign Affairs*, 78(3), 43-53.

[24] Oechslin, Manuel (2010); "Government Revenues and Economic Growth in Weakly Institutionalised States", *Economic Journal*, 120(June), 631-650.

[25] Reuther, David (1995); "UN Sanctions Against Iraq", in David Cortright and George Lopez, eds., *Economic Sanctions: Panacea or Peacebuilding in a Post-Cold War Word?* Boulder, CO: Westview Press.

[26] Schelling, Thomas (1960); *The Strategy of Conflict*. Cambridge: Harvard University Press.

[27] Schelling, Thomas (1965); *Arms and Influence*. New Haven: Yale University Press.

[28] Werleigh, Claudette Antoine (1995); "The Use of Sanctions in Haiti: Assessing the Economic Realities", in David Cortright and George Lopez, eds., *Economic Sanctions: Panacea or Peacebuilding in a Post-Cold War Word?* Boulder, CO: Westview Press.

[29] Wintrobe, Ronald (1990); "The Tinpot and the Totalitarian: An Economic Theory of Dictatorship", *American Political Science Review*, 84(3), 849-872.

Appendix: Derivations

Derivations Section 4. To derive equation (11), assume $S_{t+1} = R$. Then, the value of the citizenry's lifetime utility at the end of the priod $t+1$, i.e., after the decision on σ_{t+1}, is

$$V(R, F_{t+1}) = \begin{cases} \ln((1-\tau^m)AX^e) + \beta E_{t+1}\{V(R, F_{t+2})\} & : \quad F_{t+1} = 0 \\ \ln((1-\tau^m)A(X^e - \chi)) + \beta V(D) & : \quad F_{t+1} = 1 \end{cases}. \tag{18}$$

According to the law of iterated expectations, the expected value of $V(R, F_{t+1})$, formed at the end of period t, can be written as $E_t\{V(R, F_{t+1})\} = (1-p)E_t\{V(R,0)\} + pE_t\{V(R,1)\}$. Taking into account equation (18), this expression turns into

$$\begin{aligned} E_t\{V(R, F_{t+1})\} &= (1-p)\left[\ln((1-\tau^m)AX^e) + \beta E_t\{E_{t+1}\{V(R, F_{t+2})\}\}\right] \\ &\quad + p\left[\ln((1-\tau^m)A(X^e - \chi)) + \beta V(D)\right]. \end{aligned}$$

Note further that $E_t\{E_{t+1}\{V(R, F_{t+2})\}\} = E_{t+1}\{V(R, F_{t+2})\}$; moreover, it is clear that $E_{t+1}\{V(R, F_{t+2})\} = E_t\{V(R, F_{t+1})\}$. Using these two facts in the above expression and solving for $E_t\{V(R, F_{t+1})\}$ results in equation (11).

Derivations Section 5. I start with the derivation of equation (16). As the elite prefers to exit the game if the continuation values associated with staying ($\sigma_t = 0$) or leaving ($\sigma_t = 1$) are exactly identical, $(\tau^m - \phi)A^c X^e(A^c)$ is determined by

$$\ln(\omega) + \beta\frac{\ln(\omega)}{1-\beta} = \ln((\tau^m - \phi)A^c X^e(A^c)) + \beta E_t\{W(R, F_{t+1})\}. \tag{19}$$

To make further progress, one needs to know the explicit functional form of $E_t\{W(R, F_{t+1})\}$. To find it, note that – given $S_{t+1} = R$ – the elite's lifetime utility at the end of period $t+1$ is

$$W(R, F_{t+1}) = \begin{cases} \ln((\tau^m - \phi)A^c X^e) + \beta E_{t+1}\{W(R, F_{t+2})\} & : \quad F_{t+1} = 0 \\ \ln(\omega) + \beta\ln(\omega)(1-\beta)^{-1} & : \quad F_{t+1} = 1 \\ \ln((\tau^m - \phi)A^c X^e) + \beta W(R, 2) & : \quad F_{t+1} = 2 \end{cases}. \tag{20}$$

According to the law of iterated expectations, the expected value of $W(R, F_{t+1})$, formed at the end of period t, can be written as $E_t\{W(R, F_{t+1})\} = (1-q)[(1-p)E_t\{W(R,0)\} + pE_t\{V(R,1)\}] + qW(R,2)$. Taking into account equation (20) and observing the fact that $E_t\{W(R, F_{t+1})\} = E_t\{E_{t+1}\{W(R, F_{t+2})\}\}$, this expected value turns into

$$\begin{aligned} E_t\{W(R, F_{t+1})\} &= (1-q)(1-p)\left[\ln((\tau^m - \phi)A^c X^e) + \beta E_t\{W(R, F_{t+1})\}\right] \\ &\quad + (1-q)p\left[\ln(\omega) + \beta\ln(\omega)(1-\beta)^{-1}\right] \\ &\quad + q\left[\ln((\tau^m - \phi)A^c X^e) + \beta W(R,2)\right], \end{aligned}$$

i.e., an expression that can be solved explicitly for the expected value of $W(R, F_{t+1})$. Applying the definition $\tilde{p} \equiv (1-q)p$, one obtains

$$\begin{aligned} E_t\left\{W(R,F_{t+1})\right\} &= \frac{(1-\tilde{p})\ln((\tau^m-\phi)A^cX^e)+\tilde{p}\ln(\omega)}{1-\beta(1-q)(1-p)} \\ &\quad +\frac{\beta\tilde{p}\ln(\omega)(1-\beta)^{-1}+\beta qW(R,2)}{1-\beta(1-q)(1-p)}. \end{aligned}$$

Finally, using this expression in equation (19) and rearranging terms yields equation (16).

I now turn to the derivation of $E_t\{V(R, F_{t+1})\}$. The approach is similar to the one chosen above. Given $S_{t+1} = R$, the citizenry's lifetime utility at the end of period $t+1$ is

$$V(R,F_{t+1}) = \begin{cases} \ln((1-\tau^m)AX^e)+\beta E_{t+1}\{V(R,F_{t+2})\} & : \quad F_{t+1}=0 \\ \ln((1-\tau^m)A(X^e-\chi))+\beta V(D) & : \quad F_{t+1}=1 \\ \ln((1-\tau^m)AX^e)+\beta V(R,2) & : \quad F_{t+1}=2 \end{cases}. \tag{21}$$

Relying again on the law of iterated expectations, the expected value of $V(R, F_{t+1})$ can be written as $E_t\{V(R,F_{t+1})\} = (1-q)[(1-p)E_t\{V(R,0)\}+pE_t\{V(R,1)\}]+qV(R,2)$. Moreover, taking into account equation (21) and the fact that $E_t\{E_{t+1}\{V(R,F_{t+2})\}\}$ is identical to $E_t\{V(R,F_{t+1})\}$, this expected value turns into

$$\begin{aligned} E_t\{V(R,F_{t+1})\} &= (1-q)(1-p)\left[\ln((1-\tau^m)AX^e)+\beta E_t\{V(R,F_{t+1})\}\right] \\ &\quad +(1-q)p\left[\ln((1-\tau^m)A(X^e-\chi))+\beta V(D)\right] \\ &\quad +q\left[\ln((1-\tau^m)AX^e)+\beta V(R,2)\right]. \end{aligned}$$

Finally, solving for $E_t\{V(R,F_{t+1})\}$ yields the expression referred to in Section 5:

$$\begin{aligned} E_t\left\{V(R,F_{t+1})\right\} &= \frac{(1-\tilde{p})\ln\left((1-\tau^m)AX^e\right)+\tilde{p}\ln\left((1-\tau^m)A(X^e-\chi)\right)}{1-\beta(1-q)(1-p)} \\ &\quad +\frac{\beta\tilde{p}V(D)+\beta qV(R,2)}{1-\beta(1-q)(1-p)}. \end{aligned}$$