

Sanctions, Rewards and Regime Types

BYUNGWON WOO

Oakland University, Department of Political Science
2200 N. Squirrel Rd
Rochester, MI 48309
woo@oakland.edu

DANIEL VERDIER

The Ohio State University, Department of Political Science
2140 Derby Hall, 154 N. Oval Mall
Columbus, OH 43210-1373
verdier.2@polisci.osu.edu

Abstract

Should democracies be rewarded and autocracies punished, or should it be the reverse? This is an important question for U.S. presidents who regularly find themselves wanting to alter the behavior of foreign governments favorable to U.S. interests. Existing studies on economic sanctions and rewards provide an uneasy answer to that sanctions are more effective toward democracies and rewards work better toward autocracies, suggesting democracies need to be punished while autocracies need to be rewarded. We revisit the issue of regime type and incentive form by building a game theoretical model focusing on domestic political dynamics in a Target country. The theoretical model yields the claim that intermediate regimes stand out from democracies and dictatorships by being less responsive to sanctions and, consequently, by attracting more rewards. Empirical analysis supports the claims and shows that democracies and dictatorships are more responsive to foreign aid and economic sanctions than autocracies.

1 Introduction

Should democracies be rewarded and autocracies punished, or should it be the reverse? This is an important question for U.S. presidents who regularly find themselves wanting to alter the behavior of foreign governments in a way that is favorable to U.S. interests, for instance, curbing Jewish settlements in the West Bank and containing nuclear proliferation in Iran and North Korea. In a world where she could have her wish, there is little doubt that the citizen of an established democracy would have a preference for proffering positive incentives to democracies while reserving sanction threats for nondemocracies. But as much as one would yearn to do what feels right, concerns for efficiency may command a different tack. As we shall see, a large number of political scientists who have done research on sanction threats have concluded that such threats are more effective when used against democracies than nondemocracies. And although rewards have not received as much attention as sanctions, at least one study concludes that democracies are less sensitive to promises of reward than nondemocracies are. Against popular wish, current science claims that democracies should be punished and nondemocracies rewarded—Israel should be threatened while Iran and North Korea should be engaged.

We revisit the issue of regime type and incentive form to conclude that neither the popular nor the scientific belief have it quite right. We find that both sanctions and rewards work better with democracies than nondemocracies, with the notable exception of dictatorial regimes, which, we argue, behave no differently from democracies. Instead, we find a U-shaped relation between regime and incentive. It is intermediate regimes like Iran, regimes that are neither quite as democratic as Israel nor as absolutely autocratic as North Korea that are the least responsive to either type of incentive.

In connection with the regime type question, we raise the question of which incentive, between sanction threat and promise of reward, is the more efficient of the two. Current research, qualitative exclusively, on the relative usefulness of sanction and rewards leads to the conclusion that sanction threats and promises of reward are most efficient when

used simultaneously. Implicit is the idea that the two forms of incentives are functional substitutes—raising one a little bit allows to save on the other by an equivalent extent. We argue and show that sanctions and rewards are not necessarily substitutes. They are so when used to coerce regime types located at the extremes of the accountability spectrum—democracies and dictatorships—but not against intermediate regimes. Against intermediate regimes, including a sanction in a reward package has the counterintuitive effect of diluting the marginal effect of the reward.

The reason why sanction threats do not work against intermediate regimes is that they trigger so-called “rallies round the flag” among the targeted populations of such regimes, making compliance with the sanction impossible. Although democracies and dictatorships do experience similar rallies in response to sanction threats, the policy consequences in such cases are immaterial for they do not affect compliance. As a result, we argue that, in equilibrium, sanction threats are more likely to be used along with promises of rewards against democracies and dictatorships than against intermediate regimes.

What makes our argument unique and the propositions that follow from it unusual is that we formalize the consequences that sanctions and rewards have on the domestic make-up of the targeted country. We argue that sanctions elicit rallies round the flag among groups who expect to benefit from the sanction and thus oppose compliance. In contrast, rewards elicit so-called “fifth-column” effects among those who expect to benefit from the reward and thus support compliance. Furthermore, rewards elicit “extortion,” that is, a preemptive investment in wrongdoing on the part of the target government with an aim to raise the ante. It is only by modeling the three effects together—the rally, the fifth, and the extortion effects—that we may be able to determine under what circumstances sanctions, rewards, or a mix of the two constitute the most appropriate incentive.

The theoretical model yields the claim that intermediate regimes stand out from democracies and dictatorships by being less responsive to sanctions and, consequently, by attracting more rewards. Two variations on this claim are alternatively tested on existing datasets on

aid and sanctions. The empirical evidence supports the claim.

The rest of the paper is organized as follows. We review the existing literature on sanctions and aid, in which we locate the present research. We then offer the argument, which we implement by means of several game-theoretic models, one per regime type, between a hypothetical Sender and a hypothetical Target. We solve the games, derive predictions and proceed to test them on two existing datasets, the OECD dataset on foreign aid and the sanction dataset put together by Hufbauer et al. (2007). The article ends with a discussion and expansion of the main findings.

2 Literature

Works on international incentives and regime types bear on two subjects: sanctions and foreign aid. The literature on the interaction between sanctions and regime type displays a remarkable degree of coherence. It mostly builds on the early work of Kaempfer and Lowenberg (1988), who argue that, in order to be effective, sanctions must threaten costs on groups and individuals who are in a position of power. Following Bueno de Mesquita et al's (1999) characterization of such groups as larger in democracies than in nondemocracies, the field argues that sanctions affect the two types of regime differently. Accountable to a broad base of support, leaders in democracies enjoy fewer opportunities to resist economic hardship than autocrats, who can often insulate themselves and their immediate supporters from retribution by the victims of sanctions. As a result, sanctions cause more antigovernment protests in democratic than in autocratic regimes (Allen 2008), lead authoritarian incumbents to increase repression to suppress dissent (Wood 2008, Peksen and Drury 2008), while making democracies more likely to comply than nondemocracies, with compliance measured by sanction duration (Bolks and Al-Sowayel 2000, Allen 2005). In the same vein, comprehensive trade sanctions are not as effective against autocrats than "smart" sanctions, which directly target the groups and individuals who are close to the leader (Brooks 2002, Lektzian and Souva 2007). Also, a change in leadership is unlikely to have any impact on sanction

duration in democracies, where the newly elected leader must rely on the same median voter as did the incumbent. In contrast, leadership turnover in autocracy is more likely to elevate a different group to primacy, with corresponding impact on sanction duration (McGillivray and Stam 2004).¹

While the sanction and regime type literature portrays democracies as more responsive to external pressure than nondemocracies when pressure takes the form of a sanction, the current aid literature reverses this finding. In the only study that examines the relationship between regime type and aid effectiveness, Lai and Morey (2006) find that democracies are less sensitive to promises of aid than nondemocracies. The finding reflects the fact, they argue, that autocrats rely on rents distributed to a limited number of cronies to establish their authority, whereas democratic leaders must cater to the broad public—one dollar of aid has a bigger marginal impact in nondemocracy than in democracy. Taken together, these two strands of literature point to a lack of equivalence between sanction threats and aid promises: sanction threats have a stronger impact on broad-based governments, whereas aid promises have a stronger impact on elitist governments.

The above literature on international incentives presents several limitations. First, the utility rationale with respect to rewards, the idea that nondemocratic rulers have a greater utility for rewards than democratic rulers, is debatable. Second, the constraint rationale with respect to sanctions—the idea that autocrats can insulate themselves from retribution—fails to take into account the fact that autocrats often are residual claimants, with a vested interest in pursuing the best policies for the most productive sections of their society as long as these policies are no menace to their rule.

Third, and most importantly, despite the fact that a sanction is a reward with a negative sign and vice versa, there is no common theoretical framework integrating these a priori interchangeable incentives. We seek to remedy this asymmetry by investigating in greater details the way rulers respond to international incentives. To that effect, we integrate within our analysis two mechanisms identified by the qualitative literature on sanctions: the rally-

round-the-flag effect and the fifth-column effect. A rally round the flag occurs whenever a sanction threat arouses a nationalist response from potential winners from the sanction, making compliance with the sanction threat more difficult. Conversely, a fifth-column effect occurs when it is those who expect to be hurt who mobilize in support of compliance (Galtung 1967, Selden 1999, Rowe 2001, Nincic 2005). Usually, Selden (1999, 21) argues, the rally effect will prevail over the fifth-column effect, because gainers from sanctions prevail over losers for the same reasons that protectionists typically prevail over free traders. Pro-compliance fifth columns are more likely to prevail in response to promises of reward, which, unlike sanction threats, are not coercive and trigger no rally (Galtung 1967, Baldwin 1971, Long 1996).

This does not mean that reward promises are better than sanction threats. Many of the same authors also point to a limitation of positive incentives—their vulnerability to extortion. It is the idea that offering rewards to the targeted country for quitting wrongdoing will lead this country to engage in more wrongdoing in the hope of obtaining larger rewards (Baldwin 1971, Bernauer 1999, 167, Haass and O’Sullivan 2000). A partial and analytically uneasy consensus seems to have jelled around the notion that sanction threats and promises of reward are most efficient when used simultaneously (Amini 1997, Dorussen and Mo 1999, Cortright and Lopez 2000, Haass and O’Sullivan 2000).

We integrate these various components—sanctions and rewards, the rally and fifth-column effects, and, in relation to the latter, the possibility of extortion—in one generic sanction model featuring a Target country with two groups, one that gains from sanctions but loses from rewards, and another that loses from sanctions but gains from rewards. Each incentive, depending on the circumstances, provides the ruler with the opportunity to extend political tenure. In keeping with the literature, we argue that democratic leaders are less able to seize that opportunity than nondemocratic leaders. Basically, we assume that the identity of the supporting coalition is exogenous in democracies, while endogenous in nondemocracies; autocrats have a greater freedom to play off one group against the other.

We see several advantages to our approach in relation to existing ones. First, it is more

comprehensive, combining insights from diverse strands of the sanction and aid literatures into one unified model. Second, it offers precise mechanisms by which sanctions and rewards are transformed into policy outcomes. Last, it yields a result that is counterintuitive and testable. Splitting the regime continuum into three categories—democracy, limited autocracy, and dictatorship—, among the three, it is the intermediate form, limited autocracy, that stands out. Rulers in democracy and dictatorship are more responsive to sanctions than rulers in limited autocracy, who are mostly responsive to rewards.

3 The Strategic Use of External Incentives

In this section, we further develop the idea, already central in the literature, that, in response to external incentives, nondemocratic leaders are better able to seize an opportunity to extend tenure than their democratic counterparts.

We begin by assuming that the target country, the one that the sender country seeks to bribe or coerce into changing its behavior, is made of two coalitions and one leader. One coalition is trade or aid oriented, while the other is introverted. Typical political economy models distinguish between, on the one side, export-oriented constituencies (merchants, exporters, services) and, on the other side, protectionist sectors (home-based industry, agriculture) (Galtung 1967, Rowe 2001, Selden 1999, Nincic 2005). In such an economy, an embargo threatens to undercut the relative wealth and power of the export sectors while enriching and empowering the protectionist sectors. In contrast, more trade promises to tilt the balance in favor of the export sectors. Like trade, aid may also split society into two coalitions, one made up of constituencies that thrive on aid (military, urban populations) and, over time, become dependent on that aid for their well-being, and another coalition who is chaffing under the political dominance of the former group and opposing both the donor's aid and demands. We refer to the two coalitions of the target country as export oriented and protectionist for short.

An external incentive, whether in the form of a reward or a sanction, once implemented,

may alter the domestic balance between coalitions. A sanction, for instance, weakens the export coalition while strengthening the protectionist coalition. This is a particular instance of the so-called "rally round the flag" effect, which occurs whenever a sanction threat arouses a defiant response within the target's government or population, making compliance with the sanction threat more difficult (Galtung 1967). Conversely, a reward (more trade, more aid) strengthens the export coalition at the political expense of the protectionist coalition, enacting an instance of the so-called "fifth-column effect," making compliance with the sender's demand easier (Selden 1999).

The target government chooses between compliance and defiance according to its support coalition. For instance, complying has the support of the export coalition, but not that of the protectionist coalition. Vice versa, defying has the support of the protectionist coalition, but not of the export coalition.

The important point, however, is that the target government's supporting coalition, whether export or protectionist, is not predetermined, but may be chosen by the leader strategically, in response to the external incentive. This will happen if the leader is not bound to a particular coalition. In such a case, the leader has the option of changing side if, by doing so, he is able to improve his tenure prospect. For instance, if at the time when the sender offers the incentive, say a sanction, the leader has the support of the export coalition, he has a choice between, on the one hand, sticking to the export coalition and complying with the demand and, on the other hand, shifting coalition and denying the demand with the hope of cashing on the rally-round-the-flag-effect caused by the sanction. Conversely, it is possible to imagine a case in which a leader who is initially endorsing the interests of the protectionist faction would contemplate switching allegiance to the export faction in response to the promise of a fat reward in exchange for compliance with the senders' demand, as the implementation of such a reward would enrich the export coalition as in the scenario of the fifth-column effect.

Whether the leader of the target country enjoys sufficient margin of maneuver to jump

boat in response to an external incentive is not for us to say, but must be assessed empirically. In some cases, leaders come from, or are so identified with, the coalition that they represent that the idea of them changing side, especially once in power, seems ludicrous. Such cases are mostly identified with parliamentary democracies, where the government is led by the leader of the party that won the most seats at the last elections. In contrast, there are other cases where the leader of the government enjoys enough autonomy vis-a-vis any particular faction that he can pick and choose the faction to mobilize in accordance with the policy he feels most advisable to pursue. A famous historical example is Prussian chancellor Bismarck's historical switch in the mid 1870s from a free trade policy that was supported in Parliament by a coalition of Liberals along with Socialists and Catholics to a policy of protection that had the support of the Conservative junkers and industrialists.

A more recent instance of similar reversal occurred in 2005 Iran, when Supreme Leader Ayatollah Khomeini decided to terminate the reformist experiment under president Mohammad Khatami and throw his support behind the antiwestern government of Mahmoud Ahmadinejad. Although Bismarck was not reacting to a sanction threat—only to a world recession—Khomeini was. Four rounds of UN sanctions have since then helped consolidate the regime realignment from the cosmopolitan reformists to the nationalist hard-liners. The sanctions provided the new government with the means and rationale to build up the political and economic power of a para-military organization—the Revolutionary Guards—an organization which, today, controls the country's strategic missile forces, with ties to companies in oil, construction, telecommunications, and weapons manufacture as well as black market enterprises smuggling embargoed products, alcohol and nuclear fuel in particular.

Although an empirical question, a look at the political institutions may help us address the question of whether a leader enjoys enough autonomy to try to capitalize on the expected change in equilibrium among domestic groups caused by the implementation of the external incentive. There is a positive correlation between autonomy and autocracy. In democracies, elections force leaders to credibly identify with a party or a movement. They change their

policy orientation at the risk of tarnishing their reputation, typically losing the support of their party without for all that gaining that of the opposition party. In contrast, the absence of free elections in autocracies enables a leader to experiment with policies that, at the time of their adoption, may not be popular with her support base. An autocratic ruler who would expect these presently unpopular policies to eventually be successful and elicit support later from other groups in the polity would have a rationale to pursue these policies in the first place. Therefore, in keeping with the existing literature, we hold the democracy-autocracy divide as a plausible proxy for the distribution of political autonomy.

At the very end of the democracy-autocracy spectrum stands the absolute autocrat (dictator for short), a type for whom domestic support and tenure-maximization are of no immediate concern. Although no dictator is ever absolute in that extreme sense, some come very close. And as they do, they should not place much weight on the expected side-effects of external incentives such as the rally round the flag and the fifth column. Instead, a dictator should privilege two dimensions: first, the actual demand per se—what are its costs and benefits; second, unlike rulers who depend on coalitional rivalry to stay in power, dictators have an interest in maximizing the welfare of all groups in society, for they stand in the position of residual claimant of their subjects' output—they have a monopoly over property rights (Findlay 1990, Barzel 2000).

The regime typology—democracy-autocracy-dictatorship—is crucial to our understanding of the coalition realignment that may or may not take place in response to an external incentive in the form of a sanction threat or the promise of a reward. The rally-round-the-flag and fifth-column effects can be felt in any type of regime, but they have significant consequences for rulers that enjoy enough autonomy to strategically choose between compliance and defiance with a demand on the basis of the external incentives that are threatened or promised along. This condition eliminates democratic leaders from that list, for they do not enjoy enough autonomy from their support base. While in regimes where no support of any coalition is ever necessary for the government to stay in power, rallies and fifth column

effects are also inconsequential.

We do not mean to say that rallies never occur in democracies—they do. Anti-Americanism has become a powerful tool for mobilizing support around a government. But the difference is that such a rally, when it happens, has limited policy consequences. It does not lead to a radical change in the country’s policy preference, but merely consolidate the anti-American orientation of its dominant coalition—its effects are redundant. The same is true of a dictatorship; a dictator with a preference for defiance may take advantage of the sanction threat to stage a rally around his rule, with no meaningful impact on his policy preference. Likewise, fifth-column effects are unlikely to bear any meaningful influence in democracies or dictatorships.

The above discussion has presented the mechanisms by which a sender’s incentive affects the strategic calculation of a target government. It has also led to the proposition that leaders in intermediate autocratic regimes have both the freedom and rationale for the leader to cash on the expected effects of a sanction or a reward; democratic leaders do not enjoy the freedom, while dictators do not have a rationale.

It remains to determine whether a sender can better incentivize a target whose regime type is known in advance by threatening a sanction than by promising a reward. The question is complex enough for its solution to require the construction of a multivariate model.

4 Model

4.1 Payoffs

We assume that Target and Sender are competing for a good of total worth $Z \in [0, \infty)$. (Target is a “he”, Sender a “she.”) Target moves first by claiming $z \leq Z$. Target’s claim z represents an investment in a behavior that is deemed delinquent by Sender; z could, for instance, be the share of a territory of total size Z that Sender considers to be hers.

Sender moves second by offering a trade or aid incentive package with the intention to coerce Target into either giving up z altogether or keeping it, depending of what works

better for Sender. The incentive package has two components: a reward t and a sanction s , both positive. Sanction s is bounded upward, $s \leq S$ so as to rule out the option of threatening Armageddon. Promising the moon is not an option either, but this possibility is endogenously ruled out by Sender's maximization. Then the target government responds by either complying with the demand (giving up z) or defying (keeping z). We assume that threats and promises are enforceable.²

Sender's utility function is that of a unitary actor with no particular a-priori preference for reward or sanction. Sender merely finds either kind of incentive costly to implement: a rise in aid hurts taxpayers, whereas a drop hurts international lobbies, while a rise in trade hurts domestic producers, whereas a drop hurts exporters. There is always a group of discontented producers who punishes the government. Formally, Sender maximizes

$$U = Z - z - \xi_1 t - \xi_2 s, \quad (1)$$

with respect to sanction s and reward t . ξ_2 and ξ_1 , both strictly positive, are the marginal costs of implementing the sanction and the reward respectively.

Within the target country, two coalitions compete on the basis of relative wealth initially set to $1 - p$ for the free trade side, to p for the protectionist side, with $p \in [0, 1]$. Moreover, the free trade coalition benefits from reward t (more trade or more aid) but is hurt by sanction s (less trade or less aid) whereas it is the opposite for the protectionist coalition. As a result, the free trade coalition earns $W^{FT}|_{z,s,t,C} = 1 - p + \delta_1 t$ if its government complies and $W^{FT}|_{z,s,t,D} = 1 - p - \delta_2 s$ if its government defies whereas the protectionist coalition earns $W^P|_{z,s,t,C} = p - \delta_1 t$ in the case of compliance and $W^P|_{z,s,t,D} = p + \delta_2 s$ in the case of defiance. δ_1 and δ_2 capture the propensity of the regime to respond respectively to a positive and negative incentive.

The target government's payoff function shows two components. First, Target government's payoff is a positive function of the aggregate wealth of its supporting coalition

$g(W^{i^*})$, with $i^* = FT, P$ identifying the supporting coalition. Since this function can take about any form as long as it is positive, we simply write that Target government maximizes the aggregate wealth of its supporting coalition: $g(W^{i^*}) = W^{i^*}$. Second, as mentioned earlier, Target government benefits from investing in the delinquent behavior z , with an expected benefit of bz and at a cost of cz^2 (b is the marginal gain and c a component of the marginal cost) A generic payoff function for Target government can be written as

$$V = W^{i^*} + bz - cz^2, \text{ with } i^* \text{ the supporting coalition} \quad (2)$$

The first component of the payoff varies with the nature of the regime. Consider the case of the limited autocracy. Target government's payoff is a positive function of the aggregate wealth of the coalition that supports it *ex post*, that is, the free trade coalition in case of compliance and the protectionist coalition in case of defiance. Hence, compliance yields $V_{aut|z,s,t,C} = W^{FT}|_{z,s,t,C} - cz^2$ or, after substitution, $1 - p + \delta_1 t - cz^2$. Conversely, defiance yields $V_{aut|z,s,t,D} = W^P|_{z,s,t,D} + bz - cz^2$ or, after substitution, $p + \delta_2 s + bz - cz^2$. The non-investment payoff is $1 - p$ if the free traders are dominant *ex ante* and p otherwise.

In a democracy, the government values what its *ex ante* supporting coalition values. This is the result of defining democracy as *ex ante* accountable. If the free trade coalition is in power at the outset, that is, $p < 1/2$, complying yields $V_{dem}^{FT}|_{z,s,t} = 1 - p + \delta_1 t - cz^2$, while defying yields $1 - p - \delta_2 s + bz - cz^2$. If it is the protectionist coalition that is in power, i.e., $p > 1/2$, complying yields $p - \delta_1 t - cz^2$, while defying yields $V_{dem}^P|_{z,s,t} = p + \delta_2 s + bz - cz^2$. The non-investment payoff is $1 - p$ or p depending on the *ex ante* dominant coalition.

In a dictatorship, the dictator values the aggregate wealth of the country. Like a unitary actor, he benefits from an aid increase, $V_{dic}|_{z,t,s,C} = 1 + \delta_1 t - cz^2$, and loses from a cut, $V_{dic}|_{z,s,t,D} = 1 - \delta_2 s + bz - cz^2$. The non-investment payoff is 1.³

4.2 Target Type and Extortion

A positive incentive, argues the literature, invites extortion on the part of the target. Extortion implies that Target is investing z at cost cz^2 for no other reason than to extract a reward from Sender. Extortion is made possible by the fact that Sender is not aware of the actual purpose of the investment. Sender's ignorance is modeled by positing two possible types of Target government, randomly drawn from the set $\Theta_T = \{0, b\}$ featuring two types, a "security" type "S" with marginal gain $b > 0$ and an "extortionist" type "E" with marginal gain $b = 0$. The labels refer to the situation in which investment z enhances the security of one type but has no intrinsic value for the other type. Nature draws the security type with probability h and the extortionist type with probability $1 - h$. Target knows its type, but Sender only knows the probability distribution.⁴

4.3 Tree, Strategies, and Equilibrium

We are now ready to provide a formal definition of the strategies and draw the tree (Figure 1).⁵ A strategy for Sender in this game is the mapping $\{\sigma_1 = (I, NI), z \in \zeta = R^+\} \rightarrow \{t \in T = R^+, s \in S = R^+\}$, specifying for an investment decision and each z value the values of t and s . A strategy for Target is, at first, the mapping $\Theta_T \rightarrow \{\sigma_1, \zeta\}$ specifying for each type whether to invest or not, and, if the decision to invest is made, then the value of z and the mapping $\Theta_T \times \{\sigma_1, \zeta\} \times \{T, S\} \rightarrow (c, d)$ specifying for each type, the decision to invest, each choice of z , and in response to all possible sender's proposals the decision whether to comply or defy.

We further deconstructed the target's choice of investment z into two steps: a first in which target chooses whether or not to invest in the delinquent behavior. If he chooses not to invest ($z = 0$), the game is over—Sender cannot offer an incentive. Only if he goes ahead with the investment ($z \in [\varepsilon, Z]$), with ε and Z respectively the smallest observable and the largest possible investment in the delinquent behavior ($\varepsilon, Z > 0$), does he get to choose the actual value of z and can Sender respond.

[FIGURE 1 HERE]

We denote Sender's posterior belief about Target type by the conditional probability $q(b|z) = \Pr(b > 0|z)$; q is Sender's updated belief, after having observed z , that target is of the security type. The equilibrium concept utilized is the Perfect Bayesian (PBE), which requires posterior beliefs to be calculated using Bayes' rule and each strategy to maximize expected utility given these beliefs and other players' strategies. We use the trembling hand refinement to pin down actions and beliefs that fall off the equilibrium path and eliminate a few eccentric equilibria.

4.4 Solution

We solve the game for each kind of regime, generating four propositions, which, together, show that limited autocracies are treated less harshly than either democracies or dictatorships. We start with the case of democracy with a free trade supporting coalition.

Proposition 1 (*Free Trade Democracy equilibrium*) *There are three PBEs: (1) If $S > \bar{S}$, there is a **pooling on the extortionist** type's preference for not investing. Off the equilibrium path, Sender offers the same incentives as in (2); $q = h$.*

(2) *If $0 < S < \bar{S}$ and $h > \hat{q}$, there is a **pooling on the security** type's preference for investing $z_S^* = \begin{cases} \frac{b}{2c} \text{ if } Z > \frac{b}{2c} \\ Z \text{ if } Z < \frac{b}{2c} \end{cases}$, with Sender offering $t_S^* = \frac{bz_S^* - \delta_2 S}{\delta_1}$ and $s^* = S$, while Target complies; $q = h$.*

(3) *If $0 < S < \bar{S}$ and $h < \hat{q}$, there is a **semi-separating** in which the security type invests z_S^* while the extortionist type mimics him with probability $g^* = \frac{h}{1-h} \frac{z_S^*(\delta_1 - b\xi_1) + s^*(\delta_1\xi_2 + \delta_2\xi_1)}{\xi_1(bz_S^* - \delta_2 s^*)}$ and does not invest with probability $1 - g^*$. Upon seeing an investment, Sender offers t_S^* and $s^* (= S)$ with probability $r^* = \frac{cz_S^{*2}}{bz_S^* - \delta_2 s^*}$ but $t_E^* = 0$ and s^* with probability $1 - r^*$. The security type complies in response to t_S^* but defies in response to t_E^* ; the extortionist type always complies. $q = \hat{q} \equiv \xi_1 \frac{t_S^*}{z_S^* + \xi_2 s^*}$.*

With $\bar{S} = \frac{bz_S^ - cz_S^{*2}}{\delta_2}$. All results assume $\delta_1 > b\xi_1$.*

The proof of proposition 1 is offered in the appendix and graphed in Figure 2 along parameters S and Z on the horizontal and vertical axes respectively.⁶ We extensively develop the intuition behind the results, because the same reasoning underlies all the other proofs.

[FIGURE 2 HERE]

The target that is of the security type (thereafter, Security) is willing to invest in the delinquent behavior provided that Sender does not threaten to implement a tough sanction. It is in Sender's interest, however, to threaten the highest sanction possible because sanction threats, unlike promises of reward, are costless, for they need not be acted upon in case of success. We thus expect Sender to avail himself of the full extent of sanction available, that is, $s^* = S$. Now, S may not be enough of an incentive to elicit Target's compliance, forcing Sender in such a case to supplement the sanction threat with a reward promise t^* . Sanction and reward, therefore, complement each other, with t^* being a reverse function of S .

Given that Sender is always threatening the full value of S to elicit compliance, there will be values of S that are so high that Security will find himself better off not investing in the first place (when his compliance payoff is inferior to his do-nothing payoff). The area in which this obtains lays right of the curve drawn in Figure 2 as $S = \bar{S}$.

Alternatively, one might easily imagine that when the maximum sanction allowed is too low, it would be Sender who would rather do nothing. There are parametric specifications, indeed, for which there exists a second delimiting S -curve, this time close to the origin, below which Sender remains inactive; it is just that the specification of marginals that we opted for in Figure 2 and throughout the paper—we posited $\delta_2 > b\xi_1$ —places this case out of the positive range.

To fully characterize the equilibria graphed in Figure 2, we need to determine the extortionist type's best reply to both Sender's and Security's moves. Note, first, that, were the game one of complete information, Extortionist would never invest because, short of a positive incentive, which Sender, in such a case, would have no reason to offer, not investing

would always be better than investing. But incomplete information gives Extortionist the option within the interval $[0, \bar{S}]$ of improving upon his reservation value by mimicking Security. Hence, there exists a *pooling on Security equilibrium* in that interval. Sender gets both types to comply by mixing bribe and sanction. An extortion rent is paid to Extortionist, because the bribe is calculated to buy Security's compliance, a compliance that is expensive because Security values the investment (his b is positive). Extortionist, who has no value for the investment (his b is zero) and would thus comply in exchange for nothing, extorts the same bribe.

The scheme works fine provided that Sender believes that she is facing a security type with a high probability (h is high). However, if Nature failed to stack the deck with enough security types (h is low), then Sender has a cheaper alternative at her disposal, one that could wreak havoc with Extortionist's plan: she could tailor the incentive, not to the costly security type, but to the cheaper extortionist type, giving just enough for Extortionist to comply, thereby canceling the rent component of the incentive. Note that it would make business sense for Sender to act like this because the cheaper incentive would more than offset the occasional cases of defiance suffered from the rare security types. But for Extortionist, Sender's counterstrategy would mean the end of extortion.

Unless, of course, Extortionist is smart enough to make his presence scarcer than Nature did initially. This is the essence of a semi-separating strategy. Basically, Extortionist randomizes the decision to invest, investing with probability g and not investing with probability $1 - g$, in such a way that, upon getting the opportunity to play, Sender believes that she is facing a security type with a high probability, irrespective of the low initial draw of h . By pumping back up Sender's posterior belief q to a level that is high enough, Extortionist is able to bootstrap his payoff to that of the munificent pooling equilibrium, with two caveats. First, he receives the pooling payoff only g of the time. Second, the semi-separating equilibrium cannot hold unless Extortionist, himself, is indifferent between investing and not investing, for if he found investing better, then he would invest with certainty, unraveling

his own randomization strategy. To help him commit to his mixed strategy, Sender, in turn, must randomize between offering the high incentive that it takes to get Security to comply and the low incentive that suffices to get Extortionist to comply so as to bring the latter's payoff down enough to make him indifferent between investing and not investing.

Left of cutpoint \bar{S} , consequently, we have a pooling on Security equilibrium if Nature chose a high frequency of such types in the first place and a semi-separating equilibrium otherwise. Both equilibria involve extortion, in the sense that the state of uncertainty in which Sender finds herself forces her to be more generous than if she knew the identity of her protagonist. The difference between the two equilibria is that, in the pooling equilibrium, Sender is paying a rent (she pays a higher transfer on average than it would take to elicit compliance) and Extortionist cashing it. In the semi-separating equilibrium, Sender is still paying a rent, but Extortionist is not making a real profit, for the transfer just covers his reservation value for not investing.⁷

The case of democracy with a protectionist coalition yields starkly different results in terms of the kinds of incentives that are offered by Sender:

Proposition 2 (*Protectionist Democracy equilibrium*) *There is one **separating** PBE in which the security type invests and defies and the extortionist type does not invest, while Sender in either case offers no incentives.*

The third regime, dictatorship, is very similar to the first case of democracy, in which free traders are dominant. This similarity is easily read off Target's payoffs in Table 2, where the only difference is the first component of each payoff, $1 - p$ for democratic government, 1 for the dictator. As a result, the solution is very similar, with the minor exception of Target's payoffs. Proposition 1 and Figure 2 apply equally well to free trade democracy and dictatorship.⁸

We last solve the game for the limited autocracy regime. The results significantly differ from all preceding ones. The size of the supporting coalition, which, in democracy, is ex-

ogenously given and, in dictatorship, irrelevant, changes endogenously in limited autocracy. The initial size of each coalition thus plays a central role in the results of Proposition 3.

Proposition 3 (*Limited Autocracy Equilibrium*) *There are four perfect-Bayesian-Nash equilibria: (1) If $p < \underline{p}$, there is a **pooling on the extortionist** type's preference for not investing. Off the equilibrium path, Sender offers $s = t = 0$; $q = h$.*

(2) *If $\underline{p} < p < \bar{p}$ and $h > \hat{h}$, there is a **pooling on the security** type's preference for investing $z_S^* = \begin{cases} \frac{b}{2c} \text{ if } Z > \frac{b}{2c} \\ Z \text{ if } Z < \frac{b}{2c} \end{cases}$, with Sender offering $t_S^* = \frac{2p+bz-1}{\delta_1}$ and $s^* = 0$, while Target complies; $q = h$.*

(3) *If $\underline{p} < p < \bar{p}$ and $h \leq \hat{h}$, there is a **semi-separating** in which the security type invests z_S^* while the extortionist type mimics him with probability $g^* = \begin{cases} h \frac{\delta_1 z_S^* - \xi_1 (2p + bz_S^* - 1)}{(1-h)\xi_1 bz_S^*} \text{ if } \frac{p}{2} \geq 1 \\ h \frac{\delta_1 z_S^* - \xi_1 (2p + bz_S^* - 1)}{(1-h)\xi_1 (p + bz_S^* - 1)} \text{ if } \frac{p}{2} < 1 \end{cases}$ and does not invest with probability $1 - g^*$. Upon seeing an investment, Sender offers t_S^* and $s^* (= 0)$ with probability $r^* = \begin{cases} \frac{cz_S^*}{b} \text{ if } \frac{p}{2} \geq 1 \\ \frac{1-2p+cz_S^{*2}}{bz_S^*} \text{ if } \frac{p}{2} < 1 \end{cases}$ but $t_E^* = \begin{cases} \frac{2p-1}{\delta_1} \text{ if } \frac{p}{2} \geq 1 \\ 0 \text{ if } \frac{p}{2} < 1 \end{cases}$ and s^* with probability $1 - r^*$. The security type complies in response to t_S^* but defies in response to t_E^* ; the extortionist type always complies. $q = \hat{h}$.*

(4) *If $p > \bar{p}$, there is a **separating** equilibrium in which the security type invests z_S^* , while the extortionist type does not invest. Upon observing the investment, Sender offers nothing and the security type defies; $q = 1$.*

With $\underline{p} = \frac{1}{2}(1 - bz_S^ + cz_S^{*2})$, $\bar{p} = \frac{1}{2\xi_1}(\xi_1 + z_S^*\delta_1 - bz_S^*\xi_1)$, and $\hat{h} = \frac{\xi_1(t_S^* - t_E^*)}{z_S^* - \xi_1 t_E^*}$. All results assume $\delta_1 > b\xi_1$.*

Proposition 3 is proven in the appendix and graphed in Figure 3.⁹ The horizontal axis represents the initial strength p of the protectionist coalition, while the vertical axis represents the maximum investment Z . The reason for changing the parameter on the horizontal axis is that sanctions play no role in limited autocracy because a sanction threat may risk a perverse rally round the flag. The initial size of the respective coalitions, in contrast, plays a determinant role in the results.

On the very left, the free traders form a robust majority: p is low (below the lower cutpoint \underline{p}). Target government, irrespective of type, is already siding with the free traders and will keep doing so in the future. There is no point in making the objectionable investment in the first place. Target invests nothing and the game is over. This is a case where the sender relies on a majority that is favorable to maintaining open trade relations between the two economies to do her bidding. There is no reward and thus no extortion.

[FIGURE 3 HERE]

On the opposite side of the spectrum, the protectionists are a robust majority: p is high (above the upper cutpoint \bar{p}). Sender cannot profitably engineer a shift to the free trade coalition by offering a carrot, for it would be too costly, costlier than doing nothing. As a result, the two types go their separate routes. Security, who is wired to benefit from the investment in the delinquent behavior, does invest and, absent any incentive from Sender, then defies. Extortionist, who, in contrast to Security, has no use for the investment in the first place other than to extract a rent from Sender, anticipating that no incentive will be proffered, shuns from investing.

Squeezed between these two cutpoints are the extortion equilibria, in which Security steadfastly invests in the delinquent behavior while Extortionist mimics, systematically or randomly, Security's investment, hoping to fool Sender into buying him out of that investment for the same reward than Sender is paying to Security. The extortion equilibria are twofold: a pooling on Security equilibrium for high values of h and a semi-separating equilibrium for lower values of h , according to a logic that is identical to that developed in Proposition 1 and need not be repeated here.

The two extortion equilibria feature a remarkable case of fifth-column effect: this effect occurs in the range where the protectionists are in power ex ante ($p > \frac{1}{2}$) and for values of p below the second cutpoint ($p < \bar{p}$). In this area, by means of a positive transfer, Sender is able to engineer within the domestic politics of the target a power realignment away from the protectionist coalition toward the free trade coalition, with the latter being supportive

of compliance with Sender's demand. This equilibrium makes a powerful case in theory for a pure engagement policy, even though the existence of a rent makes the engagement policy second-best as far as Sender is concerned.

The starkest result, however, is that sanction threats are not part of any equilibrium solution, even though a sanction would, in other circumstances, make it unnecessary for Sender to pay a rent. It is indeed a property of principal-agent games that no rent needs to be paid to an agent enjoying private information in order to induce that agent to act in the principal's interest provided that this principal enjoys enough room to punish the agent. By means of a sanction threat, Sender should typically be able to implement a screening strategy by which he would lure Extortionist into complying while forcing Security to defy. In our game, the rally round the flag interferes with the freedom to punish, with the consequence that even such a screening strategy, with or without rent, provides Sender with no optimal course of action.

Sanctions are never used, either because they are large enough to cause a rally effect, thereby causing defiance, an outcome that hurts Sender, or because they are small enough not to cause a rally effect, but come at a cost nevertheless, which Sender would rather not pay. The cost is twofold: direct ($-\xi_1 s$), like any other incentive, and indirect as well, in the form of a higher compliance transfer¹⁰. The indirect cost reflects the idea that a sanction in this particular case is not the functional substitute of a reward but in effect cancels out the effect of the reward, thereby calling for a higher reward to extract the same level of compliance. In sum, large sanctions encourage rather than deter defiance, whereas small sanctions do nothing except drain Sender's budget. Only bribes are used, because they reinforce the fifth-column coalition (free traders) whose interests are aligned with Sender's interest in extracting compliance.

The no-sanction result is robust to any kind of variation in the two marginals δ_1 and δ_2 , which measure the propensity of the regime to respond respectively to a positive and negative incentive by means of a coalition realignment, as long as these marginals are greater than

zero. The results are also robust to variations in ξ_1 and ξ_2 , the respective marginal costs of the positive and negative incentives for Sender. The fifth-column effect, in contrast, exists only if $\delta_1 > b\xi_1$.

5 Predictions

The model predicts a differential use of rewards and sanctions across regime types. We focus on the extortion equilibria, the only ones in which Sender confronts Target with a set of incentives. These equilibria exist in the context of dictatorship, limited autocracy, and democracy controlled by a free trade majority; there is no such equilibrium in a democracy controlled by a protectionist majority. Focusing on the first three, limited autocracy stands apart from dictatorship and free trade democracy in two ways. First, only a positive incentive is offered in limited autocracy, whereas sanctions are also threatened in the other two regimes. Second, the positive incentive is higher in limited autocracy ($t_{aut}^* = \frac{2p+bz_S^*-1}{\delta_1}$) than in the other two regimes ($t_{dic,dem}^* = \frac{bz_S^*-\delta_2S}{\delta_1}$); this is easily seen by setting the value of p in t_{aut}^* to its average value of one half. These two differences are captured in claims 1 and 2.

Claim 1 *Limited autocracies are offered high positive incentives only; dictatorships and democracies are offered moderate positive and negative incentives.*

Claim 2 *All countries comply in response to positive incentives, yet (1) dictatorship and democracy comply in response to modest positive incentives, whereas (2) limited autocracies comply in response to high positive incentives only.*

If the same amount of positive incentives provided, limited autocracies should comply less than its two more extreme cousins.

Given the difficulty of approximating “high” and “modest” incentives for a different Target, we derive a corollary to claim 2 that democracies and dictatorships respond to positive incentives more strongly than limited autocracies do.

Both claims point in the same direction, that limited autocracies are treated less harshly than democracies or dictatorships whenever they are offered, or subjected to, incentives designed to have them quit behaving in a way that is deemed hostile by another country.

Although in equilibrium Sender should not threaten a sanction toward a limited autocracy, in practice they do (Hufbauer et al. 2007). This discrepancy between model and reality suggests that the model is leaving out some aspects of reality. Missing is the modeling of domestic politics for the Sender similar to that prescribed for the target.¹¹ If the sanctioning country, like the target, had a preference for sanctioning, results would look different. For instance, Kaempfer and Lowenberg (1988) model the sender as composed of two groups, one of them protectionist and lobbying for sanctions, while Baldwin (1971, 34) stresses voters' general dislike to reward criminal action : "When the North Koreans seized the Pueblo, it was "unthinkable" that President Johnson should offer to buy it back." There is no doubt that a more realistic model would have to incorporate domestic political constraints on the Sender's side of the kind these authors refer to. Nevertheless, our model is not devoid of interest for all that, but it allows us to offer the following corollary based on Target's behavior off the equilibrium path:

Claim 3 *If a sanction is imposed, dictatorships and democracies should comply, whereas limited autocracies should not.*

6 Empirical Test

We test our claims on a foreign-aid dataset and an economic-sanctions dataset. On the foreign-aid dataset from the OECD, we test the claim that while all regime types are more likely to comply when positive incentives are offered, limited autocracies only comply in response to high positive incentives (claim 2). On the economic-sanctions dataset by Hufbauer, Shott, Elliott, and Oegg (2007), we test the claim 3 that limited autocracies have the worst compliance record. We focus on the cases where the United States is involved as a Sender by analyzing American foreign aid and American unilateral sanctions cases. This focus on

the U.S. cases allows us to examine the claims while controlling for the factors associated with diverse or multiple senders.

Due to data limitation, we cannot test claim 1 regarding the type of incentive offered. In the most ideal research setting, we would have a dataset containing observations each of which recording an incident of delinquent behavior, a positive or negative incentive being offered, and an ensuing outcome. Unfortunately, such a comprehensive dataset neither is readily available nor practical to assemble. As a best possible alternative, we focus on testing claims 2 and 3—the claims on how a Target responds when positive or negative incentives are extended.

6.1 Compliance in response to a Positive Incentive (claim 2)

The most comprehensive cross-sectional, time-series aid data are available for U.S. Official Development Assistance (ODA).¹² This dataset covers bilateral flows from the United States to most developing countries from 1960 to 2009. U.S. aid data are relevant because Washington links aid allocation to voting by members of the United Nations General Assembly (Wang 1999, Dreher et al. 2008). In this first test, the U.S. assumes the role of sender and each aid recipient that of potential target country.

Building on the existing empirical studies linking bilateral and multilateral aid to UNGA voting behavior, we measure target’s compliance — the dependent variable — as voting with the United States at the UNGA. The main independent variable is an abrupt increase in foreign aid. We create *Posincen20* and *Posincen40*, each equal to one if annual American foreign aid allocated to a country increases by more than the specified number of millions of dollars (20 or 40) and to zero otherwise. While these cut points are arbitrary and are in absolute dollar terms (as opposed to percentage terms), there are reasons that we think these are useful. First, annual aid amount for a country does not fluctuate much from year to year and the distribution of aid amount approximate bipolar distribution, clustering around 0 or large amount. Within observations with large amount, there is still significant variation,

raising possibility of few outliers skewing the results. Thus we decide to have a dummy variable for an abrupt aid increase. These abrupt increases are rather rare to be found less than 5% of all observations. Second, the U.S. has a limit in the annual budget for foreign aid, thus thinking from the sender's perspective, it makes more sense to consider the abrupt increase in aid in absolute dollar term. If we measures an abrupt increase in percentage of GDP increase, it will likely lead to underrepresentation of larger economies. Third, thinking from a leader of an aid recipient country's perspective, we think that considering aid increase in absolute dollar term makes more sense. We focus on aid commitment as opposed to aid disbursement as our theoretical model focuses on reward promises. We expect the coefficient to be positive across all regime types, but the substantive effect to be stronger in democracies and dictatorships than in limited autocracies.

The independent variable is the regime of the target country, as measured in the Polity IV score (Jagers and Gurr 1995). We translate the 20 point scale score to three categories: *Democracy* for scores 6 or higher; *Dictatorship*, for scores -7 or below; and *Limited Autocracy* for scores in between -7 and 6.

We control for national capability, regime type, and GDP, three variables that several authors have found to predict UNGA voting (Dreher et al. 2008, Dreher and Sturm 2006). We use COW national capability score, *Capability* to capture national capability (Singer, Bremer, and Stuckey 1972, Singer 1987) and the *Polity IV* score in its unmodified -10+10 scale format.

The assembled dataset is both time-series and cross-sectional. We use error correction models for their many advantages. The ECM is a dynamic model that estimates the rate at which the dependent variable will change after a change in independent variables (De Boef and Keele 2008). In comparison to more widely used Autoregressive Distributive Lag models, the main advantages of an ECM is that it provides closer ties to the theory presented in the game theoretic model.¹³ Since our model implies a dynamic process between incentive offered by a Sender and responding compliance of a Target, an ECM is particularly useful.

In addition, interpretation of estimation results, both immediate and longer term effects, are more intuitive in ECMs. Lastly, the ECM does not require making arbitrary assumptions about the lag structure in a typical time-series models. We use the most general version of the error correction model without imposing any restriction with fixed effects. The estimated equation can be generically expressed as follows:

$$\Delta Y_t = \alpha_1 Y_{t-1} + \beta_0 \Delta X_t + \beta_1 X_{t-1} + \epsilon_t$$

The change of the dependent variable is a function of the lagged dependent variable, the change in the independent variables, the lagged independent variables, and a stochastic term. We try both inclusion and exclusion of some idiosyncratic cases, such as Israel, Iraq, Afghanistan, and Pakistan and find that inclusion of these cases make little substantive changes. Including all independent variables shrinks the time span of the data to 1984-2001. We ran two sets of models, one with *Posincen40* the other with *Posincen20*. Because the results were consistent to each other, we only report the latter (Table 2). The first column from the left lists variables, while the second reports the coefficients of the independent variables for all observations. To test the claim that the impact of incentives on compliance is distributed across regimes according to a trichotomous pattern, we run three additional tests, each limited to the observations belonging to one regime type at a time—dictatorship, limited autocracy, and democracy. Given the full equivalence between models with interaction terms and models for separate regime categories, we choose to use models for each regime type and compare statistical significance and substantive effects between regime types. This allows us to take full advantage of the benefits of error correction models without complicating the model — interacting a differential and a lag of an independent variable with a lag of the dependent variable.

[TABLE 2 HERE]

Our results confirm the accepted finding that changes in positive incentives have an

impact on UN voting. The coefficients for the pooled model are statistically significant at the $p < 0.01$ level. Substantively, the result suggests that, when aid is increased by 20 millions dollars or more, one can expect the UN voting alignment score to increase by 0.056 points in that year. This is no small change given that the UN voting congruence variable varies only between zero and one. In addition to this “immediate” effect, one may also calculate a long term effect of the increase in aid. The total long-run effect distributed over time is equal to about 0.1 point. The long run multiplier, defined as the total effect of an independent variable on the dependent variable, can be calculated by dividing the coefficient of the lagged independent variable by that of the lagged dependent variable (Keele and DeBoef 2008:191).

More importantly, our results confirm claim 2 that the impact of positive incentives is stronger in dictatorships and democracies than in limited autocracies. The coefficients for the positive incentive variable are statistically significant when the observations are limited to each type of regime, but both the immediate and long-term substantive effects are stronger in dictatorships and democracies than in limited autocracies. The immediate effect, which can be read right off of the table as the coefficient of independent variable is about 0.094 in dictatorships, 0.054 in democracies, but only 0.043 in limited autocracies. While the immediate effects in democracies and limited autocracies look more similar to each other than to those in dictatorships, the total substantive effects occurring over the following years show more similarities between dictatorships and democracies. The long-run multiplier measuring the total effect is 0.125 in dictatorships, and 0.118 in democracies, but only 0.05 in limited autocracies. Overall, the the total effects are thus larger in dictatorships and democracies than in limited autocracies. Most of the total effect (0.094 out of 0.125) in dictatorships are immediate compared to that in democracies (only 0.054 out of 0.118 is immediate) where the effect is born out more in the following years. The total effect is much smaller (0.05) and immediate (0.043 out of 0.05) in limited autocracies.

Our result departs from that of Lai and Morey (2006). They found that non-democracies

are more likely to comply than nondemocracies, while we find that democracies and dictatorships are more likely to comply than limited autocracies. The different results do not merely reflect different ways of measuring aid increase—we use a sharp increase of foreign aid over a set threshold while they use foreign aid as a percentage of a recipient country’s GDP—since we were able to reproduce their findings using our measure. Instead, the difference lies in the slicing of the regime type: when dichotomized, the effect of foreign aid is, in accordance with Lai and Morey’s result, stronger for non-democracies than for democracies; but when trichotomized, the effect is, in accordance with our theory, very strong for dictatorships and democracies yet less strong for intermediate regimes.

Among the control variables, only Capability seems to affect a country’s UN voting behavior. In accordance with prior findings, power makes a country less likely to vote with the United States. This effect seems to be especially strong for limited autocracies. Neither the Polity IV variable nor the GDP variable show any significant impact on UN voting.

In sum, the results that we report here largely support existing hypotheses on the linear relationship between a positive incentive, operationalized as a sharp increase in U.S. aid, and recipient’s compliance, operationalized as UN voting congruence with the United States. Yet, even though all countries are more compliant when offered a positive incentive, some are more so than others: dictatorships and democracies are more responsive than limited autocracies.

6.2 Compliance in response to Economic Sanctions (claim 3)

The second leg of our empirical analysis examines how the three types of regimes respond to economic sanctions. If a sanction is imposed, we claim that dictatorships and democracies should comply, whereas limited autocracies should not. This is slightly different from the consensus in the literature, which views democracies as more responsive than dictatorships and limited autocracies lumped together (Bolks and Al-Sowayel 2000, Allen 2005, Hufbauer et al. 2007).

We test our predictions on the recently updated Economic Sanctions Reconsidered dataset

by Hufbauer et al. (2007). We only consider unilateral sanctions by the U.S., dismissing sanctions involving multiple states and international organizations because the model assumes a unitary sender and we wish to remain consistent with the first test.

We derive two versions of the dependent variable from the compliance *Score* in the Hufbauer et al. (2007) dataset. Score is coded on a 16-point scale and created by multiplying two 4-point scale variables—the result of the episode and the contribution of the economic sanction to that result. Thus, when an economic sanction decisively brings a successful outcome, Score is equal to 16. In the very opposite case, it is equal to unity. We create two different dependent variables using Score. First, we create *Logged Score* by taking natural log of Score to cancel out the multiplying effect, which exaggerates the higher end of the scale. Second, we create *Success*, a binary variable coding any value above nine as success and any below nine as failure.

The main independent variable is our trichotomous regime type variable. We expect that a democratic or a dictatorial regime are positively related to sanction success while an autocratic regime reduces the likelihood of sanction success.

We include as controls the variables that have been found most relevant by the literature. The first is the sanction cost, found by Dashti-Gibson et al. (1997) and Hufbauer et al. (2007) to increase the likelihood of sanction success. The dataset offers two cost variables, one measuring the economic cost to the sender, the other that to the target. We only include the *Sender Cost* in the analysis, because available data for the cost to the target are clustered around zero, with a few observations significantly larger than zero, thus raising the risk of letting a few outliers drive the results. The Sender Cost ranges from 1 (little effect on sender) to 3 (major loss to sender). An additional variable, *Prior Relations*, controls for whether the two countries are friends or foes. Drezner (1999) argues that sanctions work better against friends than foes. The variable ranges from 1 (antagonistic) to 3 (cordial). Van Bergeijk (n.d.) and Hufbauer et al. (2007) found that trade interdependence between sender and target increases the likelihood of success. We include the *Tradelink* variable in our analysis.

Last, we control for economic health and political stability of a target. Dashti-Gibson et al. (1997) found that when a target is economically healthy and stable, sanctions are less likely to be successful. The *Stability* variable ranges from 1 (distress) to 3 (strong and stable).

A look at the summary statistics (not provided) indicate that only 28 percent of 71 U.S. unilateral sanctions are coded as successful. Of all cases, 41 percent are against dictatorships, 21 percent against democracies, and the residual, 38 percent, against limited autocracies.

Estimation results are reported in Table 5. We estimate two different models according to the dependent variable of use: a logit model on Success and an OLS model on Logged Score. Overall, the two regime variables bear the correct sign, are statistically significant and the results are robust to inclusion and exclusion of the control variables.

[TABLE 3 HERE]

The theoretical model predicts off the equilibrium path that compliance in response to sanctions is non linear in regime type. Because the rally-round-the-flag effect is more likely in a limited autocracy, we expect that sanctions against dictatorships and democracies are more likely to succeed than sanctions against limited autocracies. We find strong support for the claim. The logit model (with dichotomous Success as the dependent variable) exhibits positive and statistically significant coefficients for Democracy and Dictatorship, 2.492 and 1.273 respectively. Since the excluded regime type among the three regimes is limited autocracy, the results can be interpreted as saying that sanctions against democracies and dictatorships are more likely to succeed than against limited autocracies. The same pattern is observable in the OLS model.

To get a sense of the substantive impact of regime type, we use the results of the logit model to calculate the predicted probability of sanction success while holding other variables at their respective means. When a target is limited autocracy, the predicted probability of sanction success is a mere 9 percent; when dictatorship, it significantly increases to 25 percent; when democracy, it further increases to 50 percent.

The results confirm existing findings that sanctions are more likely to work against democracies than non-democracies (Bolks and Al-Sowayel 2000, Allen 2005, Hufbauer et al. 2007). Furthermore, the results conform with our prediction that dictatorships should behave like democracies more than limited autocracies. The discrepancy between our result and the standard result in the literature reflects the fact that the literature combines two distinct types of regimes (autocracy and dictatorship) together. If one only includes Democracy in the model (thus examining the effect of democracy vs. lumped up nondemocracies), one gets the literature's positive and statistically significant coefficient for the Democracy variable.

7 Conclusion

What incentives work best against what regimes? The literature on regime form and incentive type has it that sanctions work better against accountable forms of government whereas rewards work better when directed to unaccountable ones. This is because unaccountability allows autocrats to brush off the pain and dissent caused by the sanction while accountability implies that democratic leaders have little use for a mere handful of bribes.

We disagree with this clear-cut dichotomy, arguing instead that autocratic regimes with a measure of accountability that falls in between ex-ante accountable democracies and downright unaccountable dictatorships, actually are accountable to their populations or subset thereof, if not at all times, like democracies, at least in the medium or long run—ex post. More importantly, this intermediate situation provides them with the opportunity to ride external incentives to their advantage, engineering a rally round the flag in response to a sanction or a fifth column in response to a reward. This makes them *sui generis* and actually less responsive to either sanctions or rewards than the other two types of regime. The reason is that in both democracies and dictatorships, rewards and sanctions are functional substitutes: the greater the sanction can the sender inflict, the lower the reward needs the sender afford. In contrast, in limited autocracies, where sanctions risk backfiring, rewards and sanctions work at cross purposes: the higher the sanction is, the higher the reward needs

to be to cancel out the risk of rally round the flag that is caused by the sanction. As a result, a dollar of reward elicits less compliance in an intermediate regime than in a democracy or a dictatorship—sanctions are unusable and rewards less effective.

Therefore, it is intermediate regimes like Iran, regimes that are neither quite democratic like Israel nor absolutely autocratic like North Korea, that are the least responsive to either type of incentive. Only the promise of large rewards is likely to work, putting Washington in a difficult policy position because promising a large sum of rewards toward a country like Iran is bound to be politically unpalatable.

Our empirical research confirms that the difference between perfectly accountable and perfectly unaccountable leaders, the difference that is emphasized in the literature, is less important than it seems when intermediate forms of regime are analyzed separately. The source of the variation is caused by the intermediate group, which responds to incentives in a form that is opposite to the other two forms of regime. It is only by lumping dictators with limited autocrats that one gets the common result that compliance is a linear function of accountability in the face of incentives.

The non-linear relation between regime type and compliance questions the rationale that has been given in the literature to account for the different responses offered to external incentives by democracies and non-democracies. The currently-held rationale that rulers in democratic regimes enjoy less freedom of maneuver than rulers in non-democratic regimes may be necessary to explain why democracies are strongly responsive to incentives but is insufficient to explain why dictators in particular are equally responsive to the same incentives. Key to the non-linear result is the idea that dictators need not worry about tenure but, instead, can enjoy the advantages that come with the status of residual claimant to their subjects' output, a status that make them responsive to the overall welfare of society.

Notes

¹One should note the existence of two works extending the democratic peace argument to economic sanctions (Lektzian and Souva 2003, Cox and Drury 2007). We do not include them here because they focus on the joint effect of regime type in Sender and Target, whereas we limit our investigation to the Target.

²This assumption is potentially problematic as one could argue that Target may not believe that Sender would act upon threats and promises costly to him if he had to. Nevertheless, we assume perfect credibility. Credibility does not result from the way the present game is played, but it does result from the way the larger, unmodelled game would be played. Sender is engaged in subsequent sanction games, involving other targets one at a time. He has an interest in establishing a reputation as credible sanctioner and the only way of doing so is by delivering on the threats and promises that he makes to any target. This is a standard result in reputation games of imperfect information; see Kreps and Wilson (1981).

³Target government's payoffs are gathered in Table 1 and included in Supporting Information.

⁴Not all extortion models require incomplete information. The present one does because we vest all negotiating power in the Sender, who makes a take-it-or-leave-it offer to Target. It would be unnecessary if Target was making the offer, as such is the case in the mafia and corruption models by Polinsky and Shavell 2001, Schlicht 1996, and Bueno De Mesquita and Hafer 2008.

⁵The game tree is presented in Figure 1 and included in Supporting Information.

⁶The appendix and Figure 2 are included in Supporting Information.

⁷We still refer to the semiseparating as an extortion equilibrium because the indifference result is an approximation, merely required for mathematical tractability. In real life, unlike game theory, it takes more than plain indifference to get an individual to follow one's own wishes.

⁸The formal proof is included in the Appendix included in Supporting Information.

⁹The appendix and Figure 3 are included in Supporting Information.

¹⁰Raising sanction s by one unit means having to raise the compliance transfer t by δ_2/δ_1 .

¹¹Incorporating the dynamics of domestic politics in the sender, in addition to those in the target, will offer more realistic claims, but it also adds technical complexity to the models that are already quite complex. Given the trade-off, we decide to leave out the domestic politics of the sender.

¹²DAC Statistics is available at www.oecd.org/dac/stats.

¹³Keele and Deboef (2008) demonstrate the equivalence of the two models and highlight many advantages of error correction models.

REFERENCES

- Alesina, Alberto and David Dollar. 2000. "Who Gives Foreign Aid to Whom and Why?" *Journal of Economic Growth* 5(1): 33-63.
- Allen, Susan Hannah. 2005 "The Determinants of Economic Sanctions Success and Failure." *International Interactions* 31(2): 117-138.
- Allen, Susan Hannah. 2008. "Political Institutions and Constrained Response to Economic Sanctions." *Foreign Policy Analysis* 4: 255-74.
- Amini, Gitty M. 1997. "A Larger Role for Positive Sanctions in Cases of Compellence." Working Paper No. 12, Center for International Relations, University of California at LA.
- Baldwin, D. 1971. "The Power of Positive Sanctions." *World Politics* 24(1): 19-38.
- Barzel, Yoram. 2000. "Property Rights and the Evolution of the State." *Economics of Governance* 1, 1: 25-51.
- Bennett, D. Scott, and Allan Stam. 2000. "EUGene: A Conceptual Manual." *International Interactions* 26:179-204.
- Bernauer, Thomas. 1999. "Positive Incentives in Nuclear Proliferation and Beyond." In *The Politics of Positive Incentives in Arms Control*, edited by Thomas Bernauer and Dieter Ruloff, pp. 157-191. Columbia: University of South Carolina Press.
- Brooks, Risa. 2002. "Sanctions and Regime Type: What Works, and When?" *Security Studies* 11(4): 1-50.
- Bueno de Mesquita, Bruce, James Morrow, Randolph Siverson and Alastair Smith. 1999. "An Institutional Explanation of the Democratic Peace." *American Political Science Review* 93(4): 791-807.
- Bueno De Mesquita, Ethan and Catherine Hafer. 2008. "Public Protection or Private Extortion?" *Economics and Politics* 20(1): 1-32.
- Bolks, Sean M. and Dina Al-Sowayel. 2000. "How Long Do Economic Sanctions Last? Examining the Sanctioning Process through Duration." *Political Research quarterly* 53(2): 241-65.

Cortright, D. and G.A. Lopez. 2000. *The Sanctions Decade: Assessing UN Strategies in the 1990s*. Boulder, CO: Lynne Rienner.

Cox, Dan G. and A. Cooper Drury. 2006. "Democratic Sanctions: Connecting the Democratic Peace and Economic Sanctions." *Journal of Peace Research* 43(6): 709-722.

Dashti-Gibson, J., P. Davis and B. Radcliff. 1997. "On the Determinants of the Success of Economic Sanctions: An Empirical Analysis." *American Journal of Political Science* 41(2): 608-18.

Dorussen, Han & Jongryn Mo, 1999. "Sanctions and Incentives." Paper presented at the 1999 Annual Meeting of the American Political Science Association, Atlanta, GA, 2-5 September, 1999.

Dreher, Axel. and Jan-Egbert Sturm. 2006. "Do IMF and World Bank Influence Voting in the UN General Assembly?" KOF Working papers 06-137, KOF Swiss Economic Institute, ETH Zurich.

Dreher, Axel., Rainer Thiele and Peter Nunnenkamp. 2008. "Does US Aid Buy UN General Assembly Votes? A Disaggregated Analysis" *Public Choice* 136(1): 139-164.

Drezner, Daniel W. 1999. *The Sanctions Paradox: Economic Statecraft and International Relations*. New York: Cambridge University Press.

Findlay, Ronald. 1990. "The New Political Economy: Its Explanatory Power for LDCs." *Economics and Politics* 2, 2: 193-221.

Galtung, Johan. 1967. "On the Effects of International Economic Sanctions: With Examples from the Case of Rhodesia." *World Politics* 19(3): 378-416.

King, Gary, Michael Tomz, and Jason Wittenberg. 2000, "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44(2): 341-355.

Haas, Richard N. and Meghan L. O'Sullivan. 2000. "Terms of Engagement: Alternatives to Punitive Policies." *Survival* 42(2)

Hufbauer, C. Gary. et al. 2007. *Economic Sanctions Reconsidered*. 3rd Edition. Peterson

Institute for International Economics.

Jagers, K. and T. R. Gurr. 1995. "Tracking democracy's Third Wave with the Polity III Data." *Journal of Peace Research* 32:469-82.

Kaempfer, William H. and Anton D. Lowenberg. 1988. "The Theory of International Economic Sanctions: A Public Choice Perspective." *American Economic Review* 78(4): 786-793.

Keele, Luke. and Suzanna Deboef. 2008 "Taking Time Seriously: Dynamic Regression." *American Journal of Political Science* 52(1): 184-200.

Kilby, Christopher. 2006. "Donor Influence in Multilateral Development Banks: the Case of the Asian Development Bank." *Review of International Organizations* 1(2):173-95.

Kilby, Christopher. 2009. "Donor influence in international financial institutions: Deciphering what alignment measures measure." Working paper.

Kreps, David M. and Wilson, Robert. 1981. "Reputation and Imperfect Information." *Journal of Economic Theory* 27: 253-79.

Lai, Brian. and Daniel S. Morey. 2006. "Impact of Regime Type on the Influence of U.S. Foreign Aid." *Foreign Policy Analysis* 2: 385-404.

Leeds, Brett Ashley, Jeffrey M. Ritter, Sara McLaughlin Mitchell, and Andrew G. Long. 2002. "Alliance Treaty Obligations and Provisions, 1815-1944." *International Interactions* 28: 237-260.

Lektzian, David, and Mark Souva. 2003. "The Economic Peace Between Democracies: Economic Sanctions and Domestic Institutions." *Journal of Peace Research* 40(6): 641-60.

Lektzian, David, and Mark Souva. 2007. "An Institutional Theory of Sanction Onset and Success." *Journal of Conflict Resolution* 51(6): 848-71.

McGillivray, Fiona, and Allan C. Stam. 2004. "Political Institutions, Coercive Diplomacy, and the Duration of Economic Sanctions." *Journal of Conflict Resolution* 48(2): 154-72.

Nincic, Miroslav. 2005. *Renegade Regimes: Confronting Deviant Behavior in World*

Politics. New York: Columbia.

Peksen, Dursun, and A. Cooper Drury. 2008. "Coercive or Corrosive: The Negative Impact of Economic Sanctions on Democracy." Working paper.

Polinsky, A. Mitchell and Steven Shavell. 2001. "Corruption and Optimal Law Enforcement." *Journal of Public Economics* 81: 1-24.

Rowe, David M. 2001. *Manipulating the Market: Economic Sanctions, Institutional Change, and the Politics of White Rhodesia*. Ann Arbor: The University of Michigan Press.

Schlicht, Ekkehart. 1996. "Exploiting the Coase Mechanism: The Extortion Problem." *Kyklos* 49, 3:319-330.

Selden, Zachary. 1999. *Economic Sanctions as Instruments of American Foreign Policy*. Westport, Conn.: Praeger.

Singer, J. David, Stuart Bremer, and John Stuckey. 1972. "Capability Distribution, Uncertainty, and Major Power War, 1820-1965." in Bruce Russett (ed) *Peace, War, and Numbers* Beverly Hills: Sage, 19-48.

Singer, J. David. 1987. "Reconstructing the Correlates of War Dataset on Material Capabilities of States, 1816-1985." *International Interactions* 14: 115-32.

Smith, Adam. 1976. *The Wealth of Nations. Cannan edition [1904]*. Chicago: University of Chicago Press.

Van Bergeijk, Peter A.G. nd. "Economic sanctions; Why do they fail; why do they succeed." in W. van Genugten and G.A. de Groot (eds), *United Nations Sanctions*, Intersentia, Antwerpen, (1999) pp. 97-112.

Voeten, Erik and Adis Merdzanovic, "United Nations General Assembly Voting Data" [hdl:1902.1/12379 UNF:3:Hpf6qOkDdzzvXF9m66yLTg==](https://hdl.handle.net/1902.1/12379_UNF:3:Hpf6qOkDdzzvXF9m66yLTg==)

Wang, T. Y. 1999. "U.S. Foreign Aid and UN Voting: An Analysis of Important Votes." *International Studies Quarterly* 43: 199-210.

Wood, Reed M. 2008. "'A Hand upon the Throat of the Nation': Economic Sanctions and State Repression, 1976-2001." *International Studies Quarterly* 52: 489-513.

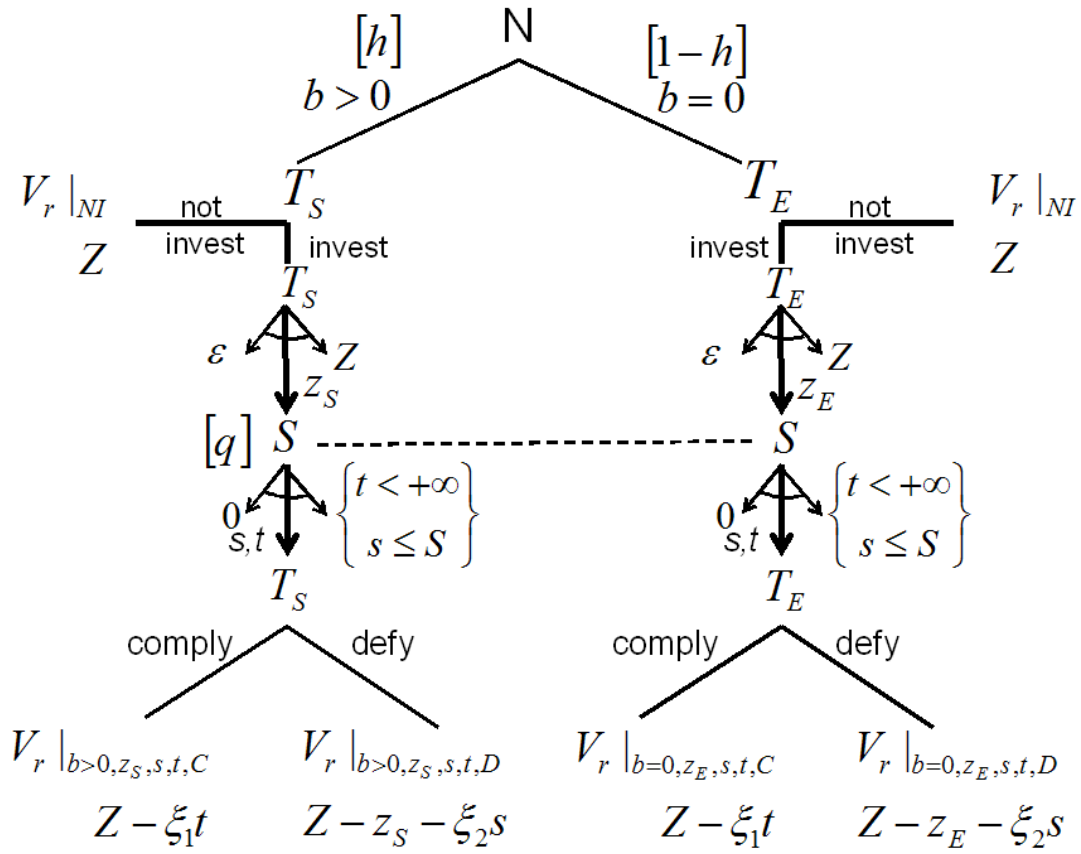


Figure 1: Game Tree, with $V_r \in \{_{dem}^{FT, P}, aut, dic\}$

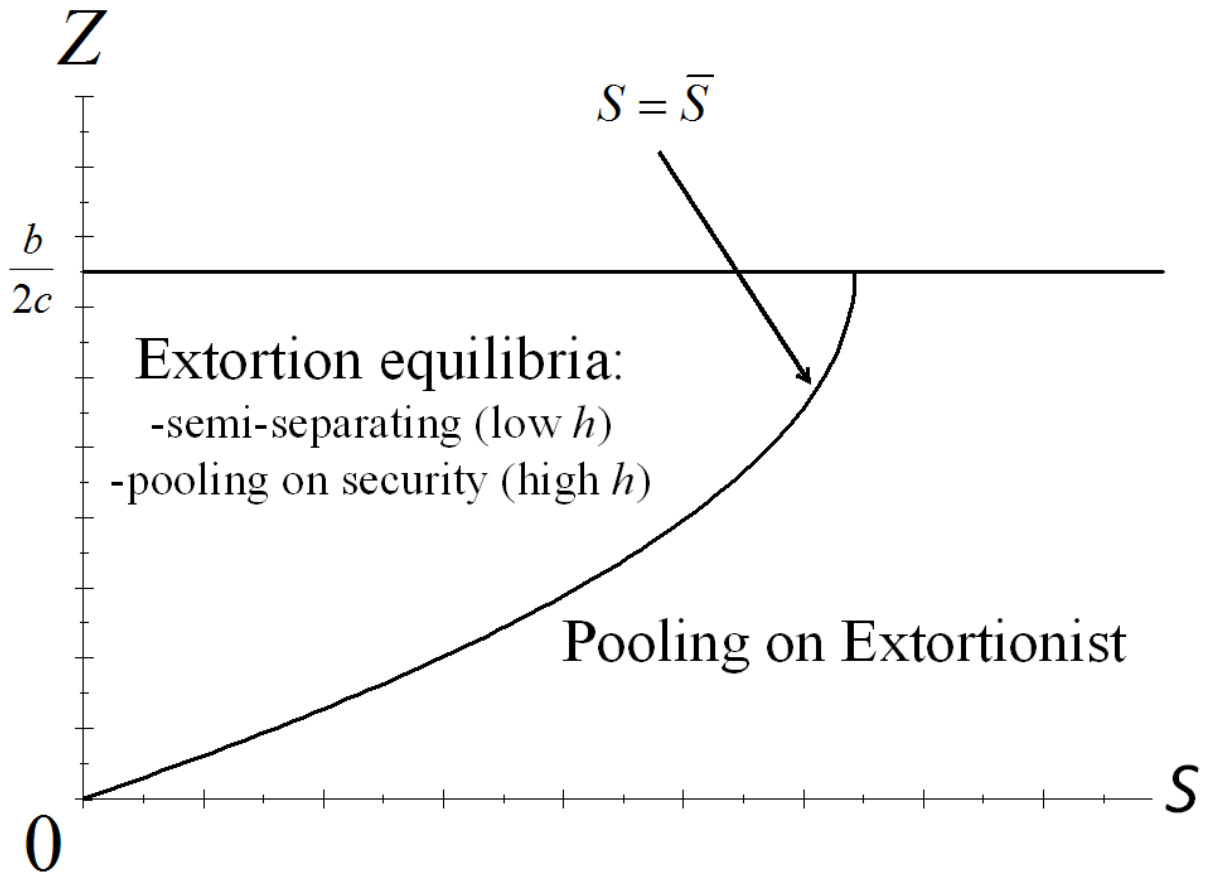


Figure 2: Solution for Free Trade Democracy

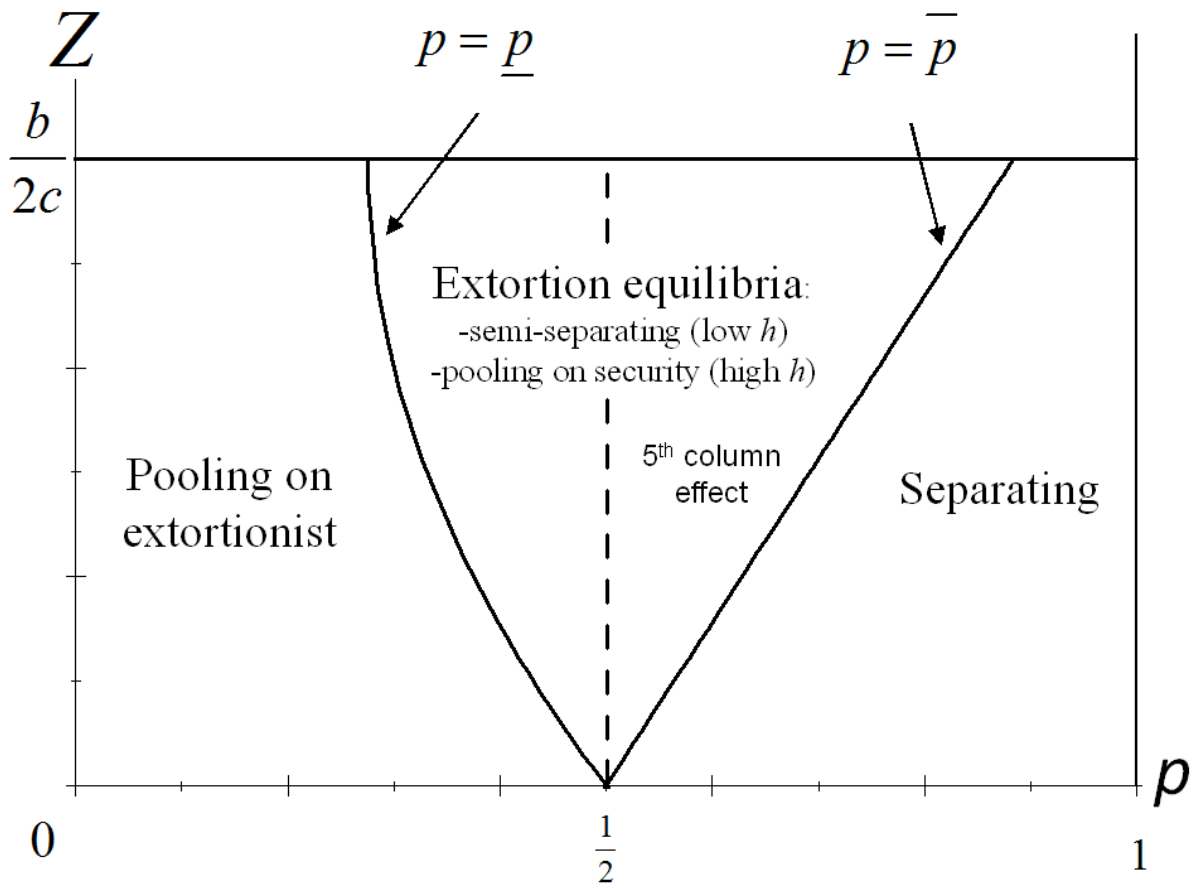


Figure 3: Solution for Limited Autocracy

Table 2: Change in the UN Voting: ECM Model				
Variable	All	Dictatorship	Autocracy	Democracy
Lagged UN Vote	-.647 ** (.024)	-.902 ** (.046)	-.717 ** (.044)	-.609 ** (.044)
Change of POS.INCEN20	.056 ** (.013)	.094 ** (.034)	.043 ** (.019)	.054 ** (.018)
Lagged POS.INCEN20	.065 ** (.019)	.113 ** (.054)	.036 (.029)	.072 ** (.026)
Change of Polity	-.002 (.002)	.003 (.023)	-.004 (.003)	.000 (.011)
Lagged Polity	-.001 (.001)	.009 (.022)	-.005 ** (.002)	-.000 (.010)
Change in Capability	-27.189 ** (9.240)	-28.922 ** (14.168)	-37.315 (21.672)	-28.173 (17.372)
Lagged Capability	-9.975 ** (3.565)	-9.429 (5.474)	-31.006 ** (14.099)	-5.558 (5.748)
Change in GDPPC	-.007 (.007)	-.012 (.012)	.014 (.014)	-.002 (.014)
Lagged GDPPC	-.007 (.004)	.005 (.009)	-.007 (.004)	-.005 (.006)
Constant	.370 ** (.021)	.521 ** (.184)	.403 ** (.037)	.366 ** (.089)
<i>N</i>	1871	568	668	635
<i>Group</i>	118	67	81	60
R^2	.10	.16	.09	.14
σ_u	.142	.208	.198	.106
σ_e	.155	.183	.127	.132
ρ	.457	.563	.707	.390
Notes: Standard Errors in Parentheses. ** : $p < .05$				

Table 3: Success of Economic Sanctions: Logit and Regression Models		
Variable	Success	Log of Score
Democracy	2.492 ** (.906)	.466 ** (.233)
Dictatorship	1.273 ** (.736)	.344 ** (.185)
Prior Relations	.211 (.612)	.086 (.179)
Sender Cost	-.433 (.632)	-.142 (.132)
Tradelink	.009 (.029)	.010 (.008)
Stability	-1.221 (.628)	-.191 (.130)
Constant	.273 (1.723)	1.718 ** (.493)
N	71	71
$R^2(Pseudo)$.23	.20
$Prob. > \chi^2/F$.007	.007
Notes: Robust Standard Errors in Parentheses. ** : $p < .10$		