Sanctions as Instruments of Regime Change
Online Appendix

Andrew J. Coe and Stephanie N. Kang

Section 1 provides proofs of the propositions. Section 2 provides additional evidence on the South Africa case.

1 Proofs of Propositions

Lemma 1. *Let $\overline{\sigma}$ be such that $c(\overline{\sigma}) = \phi$. If S threatens sanctions of a severity higher than $\overline{\sigma}$, she will never impose them, and so R and T will ignore her demand. If S threatens sanctions of a severity no more than $\overline{\sigma}$, she will impose them if her demand is not met.*

Proof. Consider any terminal subgame of the game, in which S decides whether to impose sanctions. The sole effect of imposing sanctions for S is to reduce her payoff up to that subgame by $c(\sigma)$, while the sole effect of not imposing sanctions is to reduce her payoff up to that point by either 0 (if her demand was met) or ϕ (if it was not). If $\sigma < \overline{\sigma}$, then because $c'(\sigma) > 0$, it must be that $0 < c(\sigma) < \phi$ and subgame perfection implies that S will impose sanctions if and only if her demand was not met in the prior play.

If instead $\sigma > \overline{\sigma}$, then it must be that $c(\sigma) > \phi > 0$ and subgame perfection implies that S will not impose the sanctions in equilibrium. Backward induction implies that T sets $x = 0$ if he wins an internal political contest (IPC) while R would set $x = r$ if victorious, so that the continuation values of an IPC for the two players are $V^T_{IPC} = p(1-r) + 1 - p - \gamma_T$ and $V^R_{IPC} = p + (1-p)(1-r) - \gamma_R$. If instead T makes an offer x that R accepts, their continuation values are $V_x^T = 1 - x$ and $V_x^R = 1 - |r - x|$. Subgame perfection requires R to accept any $x > pr - \gamma_R$ and reject any $x < pr - \gamma_R$, so that T's optimal offer is $x = \max\{pr - \gamma_R, 0\}$. Since neither player's equilibrium strategy depends on d, both ignore S's demand. $\square$

Proposition 1. *An internal political contest occurs if and only if* $d \geq p \max\{r, d\} + (1 - p)\sigma + \gamma_T$, $p\sigma \geq \gamma_R + \gamma_T + 2p \max\{d - r, 0\}$, *and* $\sigma \leq \overline{\sigma}$. *If R wins, she will set policy* $x = \max\{r, d\}$ *and no sanctions will be imposed. If T wins, he will set policy* $x = 0$ *and sanctions will be imposed.*

Proof. If $\sigma > \overline{\sigma}$ or $d = 0$, then the proof of Lemma 1 establishes that no IPC will occur, so assume $\sigma \leq \overline{\sigma}$ and $d > 0$ for the remainder of the proof.

There are three cases to deal with depending on how large σ is relative to $d - r$ and d. We ignore the edge cases $\sigma = d$ and $\sigma = d - r$. In those cases, two equilibria exist: one entails the behavior we demonstrate for a greater σ, the other that for a lower σ.

Case 1: $\sigma < d - r < d$. After winning an IPC, T would choose $x = 0$ and R would set $x = r$. Then the values of an IPC are $p(1 - r) + (1 - p)(1 - \sigma) - \gamma_T$ for T and $p(1 - \sigma) + (1 - p)(1 - r) - \gamma_R$ for R. Then it is easily shown that in the bargaining prior to an IPC, R and T would strictly prefer accepting and offering some $x > pr - p\sigma - \gamma_R$ respectively to an IPC, so that no IPC can occur in equilibrium.

Case 2: $\sigma > d$, then T would set $x = d$ and R would set $x = \max\{r, d\}$ after winning an IPC. Then the values of an IPC are $p(1 - \max\{r, d\}) + (1 - p)(1 - d) - \gamma_T$ for T and $p(1 - \max\{d - r, 0\}) + (1 - p)(1 - |r - d|) - \gamma_R$ for R. If $d \geq r$, then clearly R and T would strictly prefer accepting and offering $x = d$ respectively, since this saves both their respective cost of an IPC while yielding the same policy. If instead $d < r$, then it is easily shown that R and T would strictly prefer respectively accepting and offering some $x > pr + (1-p)d - \gamma_R$ to an IPC, so that no IPC can occur in equilibrium.

Case 3: $d - r < \sigma < d$, then T will set $x = 0$ and R will set $x = \max\{r, d\}$ after winning an IPC. Then the values of an IPC are $p(1 - \max\{r, d\}) + (1 - p)(1 - \sigma) - \gamma_T$ for T and $p(1 - \max\{r, d\} + r) + (1 - p)(1 - r) - \gamma_R$ for R. In bargaining prior to an IPC, the minimum x which R could possibly accept in equilibrium is $x_{min} \equiv p(2r - \max\{r, d\}) - \gamma_R$.

First suppose $d \leq x_{min}$, which implies $x_{min} \geq 0$. Observe that the only possible optimal offers for T are x_{min} and any $x < x_{min}$, if $x_{min} > 0$. Any offer $x \geq x_{min}$ will be accepted by R and not sanctioned by S; the best of these for T is clearly $x = x_{min}$. Any offer below x_{min} will lead to an IPC with the values above. Hence T chooses effectively between an IPC and offering $x_{min} = pr - \gamma_R$ (since $d \leq x_{min} \leq pr - \gamma_R < r$). The latter would yield a value of $1 - pr + \gamma_R$ for T and so would be strictly preferred. Thus no IPC can occur in equilibrium.

Now suppose instead that $d > x_{min}$. Observe that the only possible optimal offers for T are d, $\max\{0, x_{min}\}$, and any $x < x_{min}$ if $x_{min} > 0$. Any offer x that is at least d would be accepted by R and not sanctioned by S, yielding a continuation value for T of $1 - x$; of these, $x = d$ is strictly best for T. Any offer in $[x_{min}, d)$ would be accepted but sanctioned, giving a value for T of $1 - x - \sigma$; of these, $x = x_{min}$ is strictly best for T. Any offer less than x_{min} would lead to an IPC with the value above.

The first inequality in the proposition is equivalent to the condition that T would strictly prefer an IPC to yielding to S's demand, and the second inequality is equivalent to the condition that T would strictly prefer an IPC to satisfying R while spurning S's demand. Hence, in equilibrium T will spurn S's demand and induce an IPC with an unsatisfactory offer to R if and only if both inequalities hold.

This proves that the inequalities in the proposition are necessary for an IPC to occur in equilibrium. Those inequalities are sufficient when combined with the conditions $d - r < \sigma < d$ and $d > x_{min} > 0$ used in the case arguments above. We proceed to show that these latter conditions are not necessary because they are implied by the inequalities in the proposition.

From the first inequality in the proposition, we have $d \geq pd + (1 - p)\sigma + \gamma_T > \sigma$, and also

$d \geq pr + (1-p)\sigma + \gamma_T > pr - \gamma_R \geq x_{min}$. From the second inequality in the proposition, we have $p\sigma \geq 2p(d-r) \Rightarrow \sigma > d-r$. This leaves only $x_{min} > 0$. To see this, first suppose $d \geq r$. Then the second inequality can be rearranged to $2pr \geq 2pd - p\sigma + \gamma_R + \gamma_T$. Because the first inequality implies $d > \sigma$, we have $2pr > pd + \gamma_R + \gamma_T \Rightarrow x_{min} = 2pr - pd - \gamma_R > \gamma_T \geq 0$. Now suppose $d < r$. Because the first inequality implies $d > \sigma$, we have $r > \sigma$, so that the second inequality implies $pr > \gamma_R + \gamma_T$ and thus $x_{min} > 0$. □

Proposition 2. *Suppose $r = 1$. If $p\overline{\sigma} \geq \gamma_R + \gamma_T$ and $pk > (1-p)\left[\overline{\sigma} + c\left(\frac{\gamma_R+\gamma_T}{p}\right)\right] + \gamma_S + \gamma_T$, then S will make a demand that she knows will be spurned and threaten sanctions of severity $\sigma = \frac{\gamma_R+\gamma_T}{p}$, and an internal political contest will occur. Otherwise, S will demand $d = \max\{\overline{\sigma}, \min\{p - \gamma_R + \overline{\sigma}, p + (1-p)\overline{\sigma} + \gamma_T\}\}$ and threaten sanctions of severity $\overline{\sigma}$, T will yield to S's demand, and no internal political contest or sanctions will occur.*

Proof. For any σ, d S chooses, either an IPC will result or T will make an offer that R accepts. In order to determine S's best option, we compute the optimal choice for S among those that will lead to an IPC, and compare the value for S of this to that of the optimal choice that will lead to an accepted offer.

S's best choice among those that will lead to an IPC is immediate from Proposition 2: $\sigma = \frac{\gamma_R+\gamma_T}{p}$, which must be no greater than $\overline{\sigma}$ to cause an IPC, and $d \geq p + (1-p)\sigma + \gamma_T$, which yield a continuation value for S of $V^S_{IPC} = p + (1-p)\left[-c\left(\frac{\gamma_R+\gamma_T}{p}\right) - k\right] - \gamma_S$. Any σ or d lower than these will, by Proposition 1, not lead to an IPC. Any σ higher than this that still leads to an IPC will change S's value only by increasing the cost of the sanctions S will have to impose if T wins the IPC, making S strictly worse off.

We will partition the possible choices by S that do not lead to an IPC and compute the optimal choice within each partition.

Case 1: $\sigma > \overline{\sigma}$: By the proof of Lemma 1, T offers $\max\{p - \gamma_R, 0\}$ and R accepts, giving S a value $\max\{p - \gamma_R, 0\} - k$.

Case 2: $d \leq \sigma \leq \overline{\sigma}$: This implies that $d < p + (1-p)\sigma + \gamma_T$, so by Proposition 1, no IPC will occur. Observe that T would set $x = d$ after winning an IPC, since instead setting $x = 0$ would result in suffering the cost of sanctions σ at a gain of $d \leq \sigma$. R's reservation value is thus $V'^R_{IPC} = p + (1-p)d - \gamma_R$ and T's offer will be at least this. If $V'^R_{IPC} < d$, then T would strictly prefer to offer d, since offering less would result in suffering sanctions at a cost σ for a gain of at most d. Thus T will make an offer of $x^* = \max\{V'^R_{IPC}, d\}$, giving a value for S of $x^* - k$. The choice that maximizes this value is $d = \sigma = \overline{\sigma}$, yielding $\max\{p + (1-p)\overline{\sigma} - \gamma_R, \overline{\sigma}\} - k$ for S. This is strictly better for S than the previous partition.

Case 3: $\sigma \leq \overline{\sigma}$, $\sigma < d < \min\{p + \sigma - \gamma_R, p + (1-p)\sigma + \gamma_T\}$: Observe that this partition only exists if $\sigma < p + \sigma - \gamma_R \Leftrightarrow p - \gamma_R > 0$. The second term in the minimum implies that the first inequality in Proposition 1 does not hold, ruling out an IPC. Because $d > \sigma$, T would set $x = 0$ after winning an IPC so that R's reservation value is $V^R_{IPC} = p - \gamma_R$. If $d \leq p - \gamma_R$, R's minimum also satisfies S and so is clearly T's optimal offer. If $d > p - \gamma_R$, offering R's minimum would give T $1 - p + \gamma_R - \sigma$ while offering d would give T $1 - d$, which by definition of the partition is larger, so that offering d is optimal. Hence S's optimal choice is to set $\sigma = \overline{\sigma}$ and $d^* = \min\{p + \overline{\sigma} - \gamma_R, p + (1-p)\overline{\sigma} + \gamma_T\}$, giving her a value of $d^* - k$.

When this partition exists, then both terms in the minimum in S's value here exceed both terms in the maximum in S's value in the previous case, so that the best of this partition is strictly better for S than anything in the previous partitions.

Case 4: $\sigma \leq \overline{\sigma}$, $\sigma < d$, $d > \min\{p + \sigma - \gamma_R, p + (1-p)\sigma + \gamma_T\}$: The definition of this partition implies that the first two of the three inequalities in Proposition 1 hold, so for an IPC not to occur, it must be that the the last inequality in that proposition does not hold: $\sigma < \frac{\gamma_R + \gamma_T}{p}$. This in turn implies that the first term in the minimum defining this partition is smaller, so that $d > p + \sigma - \gamma_R$. Because $\sigma < d$, T would set $x = 0$ after winning an IPC and R's reservation value is $V^R_{IPC} = p - \gamma_R$. If T offered d, he would receive a value of $1 - d$, while offering R's minimum (or 0, if R's minimum is not feasible) would give him $1 - p + \gamma_R - \sigma$

(or $1-\sigma$ if R's minimum is infeasible), which is larger and so strictly preferred. Hence S would receive a value of $\max\{p-\gamma_R,0\}-c(\sigma)-k$, which is strictly worse for S than the best of any previous partition.

Collecting the analysis, if $p-\gamma_R\leq 0$, it is not possible for S to incite an IPC, and the optimal choice for S is $d=\sigma=\overline{\sigma}$, yielding a value of $\max\{p+(1-p)\overline{\sigma}-\gamma_R,\overline{\sigma}\}-k$. If instead $p-\gamma_R>0$, then the optimal choice for S that does not incite an IPC is $\sigma=\overline{\sigma}$, $d=d^*$, giving her a value of $\min\{p+\overline{\sigma}-\gamma_R,p+(1-p)\overline{\sigma}+\gamma_T\}-k$. Comparing this value to that of the optimal IPC-inciting choice from above, and recalling that by Proposition 1 inciting an IPC is only feasible if $p\overline{\sigma}\geq\gamma_R+\gamma_T$, yields the conditions in the proposition. The requirement for an IPC to be possible that $p-\gamma_R>0$ is implied by the first condition in the proposition, so that those conditions are sufficient. □

Proposition 3. *If R's ideal policy is close enough to S's (r close enough to 1), then S becomes more likely to use sanctions to incite regime change rather than policy change as the cost of coexisting with T (k) rises, the costs of an internal political contest (γ_R, γ_T, γ_S) fall, R's ideal policy moves closer to S's (r rises), and the probability of R winning (p) rises.*

Proof. We proceed to calculate the best possible choice for S of a d, σ among pairs that will not incite an IPC. We then examine how the feasibility of inciting an IPC varies in the parameters, and compute the best possible choice among strategies that will incite an IPC when any exist. We then compare the continuation values of the two to obtain the desired comparative statics. We partition S's possible strategies into subsets where the effect of S's strategy on subsequent behavior by T and R is clear.

Case 1: $\sigma>\overline{\sigma}$: Similarly to Proposition 2, S gets a value of $\max\{pr-\gamma_R,0\}-k$.

Case 2: $d\leq\sigma\leq\overline{\sigma}$: Similarly to Proposition 2, S gets a value of $\max\{pr+(1-p)\overline{\sigma}-\gamma_R,\overline{\sigma}\}-k$, strictly better than the previous partition.

Case 3: $\sigma\leq\overline{\sigma}$, $\sigma<d-r$: Observe that T would set $x=0$ and R would set $x=r$ after

winning an IPC, since for both the value of this outweighs the value of meeting S's demand and avoiding sanctions. R's expected value from an IPC is thus $V_{IPC}^{R} = p(1-\sigma) + (1-p)(1-r) - \gamma_R$, and the minimum offer x needed to satisfy R is given by $1 - |r - x| = V_{IPC}^{R}$, so that $x = pr - p\sigma - \gamma_R$. It is easy to check that T strictly prefers making this offer (or 0, if this offer is infeasible) to any other, so that S gets $\max\{pr - p\sigma - \gamma_R, 0\} - c(\sigma) - k$, strictly worse than the previous partition.

Case 4: $\sigma \leq \overline{\sigma}$, $d - r \leq \sigma < d \leq p\,[2r - \max\{r, d\}] - \gamma_R$: Observe that in this and all subsequent partitions, T would set $x = 0$ after winning an IPC, while R would choose $x = \max\{r, d\}$. The minimum offer that R would accept in equilibrium is $x_{min} \equiv p\,[2r - \max\{r, d\}] - \gamma_R$. The last inequality defining this partition means that offering x_{min} is feasible and would also meet S's demand (since $x_{min} \geq d \geq 0$). It also implies that $d \leq pr - \gamma_R < r$, so that $x_{min} = pr - \gamma_R$. It is easy to show that T would strictly prefer making this offer to any other, yielding $pr - \gamma_R - k$ for S, strictly worse than partition 2 when the current partition exists.

Case 5: $\sigma \leq \overline{\sigma}$, $d - r \leq \sigma < d$, $p\,[2r - \max\{r, d\}] - \gamma_R < d$
$\leq \min\{p\,[2r - \max\{r, d\}] + \sigma - \gamma_R, p\max\{r, d\} + (1-p)\sigma + \gamma_T\}$: The second-to-last inequality implies a policy of x_{min} would not meet S's demand, so that in equilibrium there are only three possible optimal offers for T: d, which satisfies both S and R; x_{min}, which satisfies R but leads to imposed sanctions; and anything less than x_{min}, which leads to an IPC. The first term in the minimum implies that T would prefer offering d to x_{min}, while the second term implies T would prefer d to an IPC. Hence in equilibrium d is offered and accepted, so that S gets a value of $d - k$.

Throughout the remainder of the proof, we will interpret "R's ideal policy is close enough to S's" to mean that $r > \overline{\sigma} + \frac{\gamma_T}{1-p}$. This implies that $pr + (1-p)\sigma + \gamma_T \leq r$ for any $\sigma \leq \overline{\sigma}$, which in turn means the inequality that distinguishes this partition can be simplified to $pr - \gamma_R < d \leq \min\{pr + \sigma - \gamma_R, pr + (1-p)\sigma + \gamma_T\}$. Since both terms in the minimum rise in σ and are independent of d, the optimal strategy for S within this partition must be $\sigma = \overline{\sigma}$ and $d = \min\{pr + \overline{\sigma} - \gamma_R, pr + (1-p)\overline{\sigma} + \gamma_T\}$.

Case 6: $\sigma \leq \overline{\sigma}$, $d - r \leq \sigma < d$, $p\sigma \leq \gamma_R + \gamma_T - 2p[r - \max\{r, d\}]$, $d > \min\{p[2r - \max\{r,d\}] + \sigma - \gamma_R, p\max\{r,d\} + (1-p)\sigma + \gamma_T\}$: Note that the second-to-last inequality implies that the first term in the minimum binds. The last inequality then implies that T would strictly prefer offering x_{min} and suffering sanctions to yielding to S's demand, while the second-to-last implies that T would strictly prefer offering x_{min} (or 0, if $x_{min} < 0$) and suffering sanctions to a lesser offer that would be rejected by R. Hence in equilibrium $\max\{x_{min}, 0\}$ is offered and accepted. By the last inequality, this does not meet S's demand, so that S has to impose sanctions and receives a value of $\max\{x_{min}, 0\} - c(\sigma) - k \leq \max\{pr - \gamma_R, 0\} - c(\sigma) - k$, strictly worse than partition 2.

Case 7: $\sigma \leq \overline{\sigma}$, $d - r \leq \sigma < d$, $p\sigma > \gamma_R + \gamma_T - 2p[r - \max\{r, d\}]$, $d > \min\{p[2r - \max\{r,d\}] + \sigma - \gamma_R, p\max\{r,d\} + (1-p)\sigma + \gamma_T\}$: Note that the second-to-last inequality implies that the binding term in the minimum is the second one. Then the conditions of Proposition 1 are met and this partition results in an IPC, yielding a value for S of $p\max\{r,d\} + (1-p)[-k - c(\sigma)] - \gamma_S$.

From the arguments above, it is clear that either partition 2 or 5 contains the best non-IPC-inciting strategy for S, and that partition 7 contains the best IPC-inciting strategy for S. Observe that which of 2 or 5 contains the optimum non-IPC-inciting strategy for S is determined by whether $pr - \gamma_R$ is positive or not. If $pr - \gamma_R < 0$, then partition 5 does not exist since it requires $d \leq pr + \sigma - \gamma_R < \sigma$, but also $d > \sigma$, so it must be that the best of partition 2 is optimal overall. If $pr - \gamma_R > 0$, then the first term in the minimum for partition 5 clearly exceeds both terms in the maximum defining partition 2. The second term in partition 5's minimum clearly exceeds the second term in partition 2's maximum, and also exceeds the first term since by assumption $r > \overline{\sigma}$. Hence the optimal non-IPC-inciting strategy for S is $\sigma^*_{non} \equiv \overline{\sigma}$ and:

$$d^*_{non} \equiv \begin{cases} \overline{\sigma} & \text{if } pr - \gamma_R \leq 0 \\ \min\{pr + \overline{\sigma} - \gamma_R, pr + (1-p)\overline{\sigma} + \gamma_T\} & \text{if } pr - \gamma_R > 0 \end{cases}$$

Now consider the feasibility for S of instead inciting an IPC. Using our assumption that $r > \overline{\sigma} + \frac{\gamma_T}{1-p}$, we have that $pr + (1-p)\sigma + \gamma_T < r$ for any $\sigma \leq \overline{\sigma}$, which implies that the first condition for an IPC to occur from Proposition 1 can always be satisfied by choosing $d = r$ and $\sigma \leq \overline{\sigma}$. With $d = r$, the second condition from Proposition 1 can be written $p\sigma > \gamma_R + \gamma_T$. Combining this with the third and final condition that $\sigma \leq \overline{\sigma}$, it is possible in equilibrium for S to incite an IPC if and only if $p\overline{\sigma} \geq \gamma_R + \gamma_T$. Clearly, the feasibility of inciting an IPC rises in p and falls in γ_R and γ_T, and does not depend on k, γ_S, or r, subject to our assumption that r is high enough.

Next we compute the optimal IPC-inciting strategy for S, given that such a strategy is feasible. From the above argument, if any strategy will incite an IPC, then setting $d = r$ and $\sigma = \frac{\gamma_R + \gamma_T}{p}$ will do so, so the question is simply whether S would do better to demand more than r when trying to incite an IPC. Given that $d^*_{IPC} \geq r$, the first condition for an IPC to occur from Proposition 1 is automatically satisfied since $d > pd + (1-p)\sigma + \gamma_T \Leftrightarrow d > \sigma + \frac{\gamma_T}{1-p}$, which by assumption is less than r. From partition 7, the continuation value for S will be $pd + (1-p)\left[-k - c(\sigma)\right] - \gamma_S$. Since this value rises in d but falls in σ, the optimal strategy must entail the second feasibility condition being just met, or $p\sigma = \gamma_R + \gamma_T + 2p(d - r)$. Substituting this for σ in S's continuation value and differentiating with respect to d, we obtain the first-order condition $c'\left(\frac{\gamma_R + \gamma_T}{p} + 2(d^* - r)\right) = \frac{p}{2(1-p)}$.

The third and final feasibility condition for an IPC to occur from Proposition 1 requires $\sigma \leq \overline{\sigma}$. The argument inside $c'(\cdot)$ above will be equal to this upper bound on σ when $d^* = r + \frac{\overline{\sigma}}{2} - \frac{\gamma_R + \gamma_T}{2p}$, so that this is an upper bound on the optimal demand. Using this and the first-order condition above, we have $\sigma^*_{IPC} = \frac{\gamma_R+\gamma_T}{p} + 2(d^*_{IPC} - r)$ and:

$$d^*_{IPC} = \begin{cases} r & \text{if } \frac{p}{2(1-p)} \leq c'\left(\frac{\gamma_R+\gamma_T}{p}\right) \\ c'\left(\frac{\gamma_R+\gamma_T}{p} + 2(d^*_{IPC} - r)\right) = \frac{p}{2(1-p)} & \text{if } c'\left(\frac{\gamma_R+\gamma_T}{p}\right) < \frac{p}{2(1-p)} \leq c'\left(\overline{\sigma}\right) \\ r + \frac{\overline{\sigma}}{2} - \frac{\gamma_R+\gamma_T}{2p} & \text{if } c'\left(\overline{\sigma}\right) < \frac{p}{2(1-p)} \end{cases}$$

Finally, we can compare the continuation values of the best non-IPC-inciting and best IPC-inciting strategies for S, given that the latter is feasible. We showed above that feasibility entails $p\overline{\sigma} \geq \gamma_R + \gamma_T$. Since $r > \overline{\sigma}$ by assumption, this inequality also implies that $pr > \gamma_R$. When inciting an IPC is feasible, we can use these to simplify the optimal non-IPC-inciting strategy and write its continuation value as:

$$V^S_{non} = d^*_{non} - k = pr + (1-p)\overline{\sigma} + \gamma_T - k$$

$$V^S_{IPC} = pd^*_{IPC} + (1-p)\left[-k - c\left(\sigma^*_{IPC}\right)\right] - \gamma_S$$

Clearly, inciting an IPC becomes more attractive to S as k rises and γ_S falls. An increase in either γ_R or γ_T lowers d^*_{IPC} or leaves it unchanged in all the cases in its definition, and shifts the constraints defining those cases in a way that also decreases d^*_{IPC} or leaves it unchanged, and raises σ^*_{IPC} even holding d^*_{IPC} constant, so that V^S_{IPC} must decline. By contrast, V^S_{non} clearly rises in γ_T and does not depend on γ_R, so that inciting an IPC becomes more attractive to S as γ_R or γ_T fall.

Next consider an increase in r. We can see that the derivative of d^*_{IPC} with respect to r is 1, while that of σ^*_{IPC} is 0 accounting for its dependence on d^*_{IPC}. Hence V^S_{IPC} rises in r at a rate p, just like V^S_{non}. However, recall that k is treated as a function of r, and assumed to be increasing in r in the model setup. Hence, as r grows, inciting an IPC becomes more attractive to S.

Lastly consider an increase in p. Differentiating V^S_{IPC} with respect to p, we have:

$$\begin{aligned}\frac{\partial V^S_{IPC}}{\partial p} &= d^*_{IPC} + p\frac{\partial d^*_{IPC}}{\partial p} + k + c\left(\sigma^*_{IPC}\right) - (1-p)c'\left(\sigma^*_{IPC}\right)\frac{\partial \sigma^*_{IPC}}{\partial p} \\ &= d^*_{IPC} + p\frac{\partial d^*_{IPC}}{\partial p} + k + c\left(\sigma^*_{IPC}\right) - (1-p)c'\left(\sigma^*_{IPC}\right)\left[-\frac{\gamma_R+\gamma_T}{p^2} + 2\frac{\partial d^*_{IPC}}{\partial p}\right]\end{aligned}$$

Observe that d^*_{IPC} cannot fall: in every case in its definition, it increases or is unchanged, and the constraints shift in favor of higher cases. At $d^*_{IPC} = r$, $\frac{\partial d^*_{IPC}}{\partial p} = 0$, and every term in the above expression is non-negative, so that $\frac{\partial V^S_{IPC}}{\partial p} > d^*_{IPC} = r > r - \overline{\sigma} = \frac{\partial V^S_{non}}{\partial p}$. If instead $d^*_{IPC} > r$, then from its definition we must have $c'\left(\sigma^*_{IPC}\right) \leq \frac{p}{2(1-p)}$, implying:

$$\begin{aligned}\frac{\partial V^S_{IPC}}{\partial p} &\geq d^*_{IPC} + p\frac{\partial d^*_{IPC}}{\partial p} + k + c\left(\sigma^*_{IPC}\right) - (1-p)\frac{p}{2(1-p)}\left[-\frac{\gamma_R+\gamma_T}{p^2} + 2\frac{\partial d^*_{IPC}}{\partial p}\right] \\ &= d^*_{IPC} + k + c\left(\sigma^*_{IPC}\right) + \frac{\gamma_R+\gamma_T}{2p} > r - \overline{\sigma} = \frac{\partial V^S_{non}}{\partial p}\end{aligned}$$

Hence inciting an IPC becomes more attractive to S as p rises.

Since the attractiveness of inciting an IPC rises in p, r, and k and falls in γ_S, γ_R, and γ_T, and its feasibility rises in in p, falls in γ_R and γ_T, and does not depend on the other variables, we have established the claims in the proposition. □

2 Additional Evidence on South Africa Sanctions

Here we show that the behavior and expectations of the US government, the South African government (henceforth "SAG"), and the white South African opposition matched the corresponding equilibrium (coercive sanctions from 1977 to 1985, incitement sanctions from 1986 to 1991) prescribed by our model. We use primary sources from Burton (2016) for the Carter Administration and the account of Thomson (2008), which is based on the limited documents available but also interviews with the policy's architects, for the Reagan administration. These are cited as “S[document number]” and “T[page number” for brevity. Because incitement sanctions were imposed by Congress over Reagan’s opposition, we employ transcripts of Congressional hearings on South Africa, cited according to their document code, e.g., “HRG-1980-FOA-0041”. For South Africa’s perspective, we draw on Giliomee (2012), which is based on interviews with the key leaders in South Africa, cited as “G[page number]”.

Under our theory, coercive sanctions should involve a modest demand, rather than the more extensive demand the sender would make under incitement sanctions, and the sender should intend to replace the government only in the latter. The sender should expect the incumbent government to grant the modest but not the extensive demand. We should observe bargaining between sender and incumbent over the modest demand, versus an internal political contest over whether to meet the extensive demand. Sanctions should cause the incumbent (under coercion) or replacement (if one comes to power under incitement) government to grant the sender’s demand and only then should sanctions be dropped.

President Carter demanded only “a progressive transformation” (S268) involving “intermediate steps [that] would be acceptable to [SAG Prime Minister] Vorster [. . .] toward the liberalization of South African society” (S267). The administration was “abandoning the concept of majority rule”, its ideal policy, expressly because “That phrase strikes terror among the [governing] South Africans” and “we [. . .] don’t want to turn over South Africa to the Reds” (S267). Carter clarified that “sequential progress should be condoned. Let Vorster tell you what they *will* do & how long it will take. [. . .] Don’t set our requirements so high as to obviate any cooperation” (S274). Clearly, the US strategy was to formulate a modest demand that SAG would grant, rather than a more extensive demand that SAG would reject and that might lead to its replacement. Bargaining over exactly what steps the US wished to see and which would be acceptable to SAG followed (S269, S271, S273, S274, S276, S278), and sanctions were imposed as SAG initially responded unsatisfactorily (S284, S285, S313, S314, S324) even as haggling continued (S317, S321, S323). The Reagan administration pursued a similar strategy, though described more delicately as "seeking to [...] bolster those committed to evolutionary change" and "to foster conditions in which all South Africans can more fully share and participate in the economy and political process" but "neither to destabilize South Africa nor align ourselves with apartheid policies that are repugnant to us" (HRG-1981-FOA-0076, 11–12). SAG implemented liberalizing reforms from 1979 through 1984 (T113), in response to which the US gradually relaxed its sanctions (T113–118), even as bargaining continued (T122-124).

No internal political contest over whether to grant or spurn the US demand occurred. Vorster called a snap election soon after Carter's demand and his National Party campaigned on *both* opposition to foreign interference *and* constitutional reform proposals to extend political participation to "Indians" and "Coloreds" (Midlane, 1979, 374, 376–377), exactly the kind of liberalizing step the US had demanded. All the other major parties criticized the sanctions, and most also supported constitutional reform, with the sole exception receiving only 3.3% of the vote (Midlane, 1979, 379–382). Vorster and the National Party won "the largest majority in South African history and almost certainly the support of the overwhelming majority of English as well as Afrikaans speaking South Africans" (Midlane, 1979, 385). Vorster was succeeded by P.W. Botha, a similarly right-wing politician who had nonetheless chaired the cabinet committee that developed the reform proposals (Stultz, 1984, 362). In the next election, Botha and the National Party again campaigned on both opposition to foreign interference and also liberalizing reforms to the constitution and labor laws, and won another landslide (Lemon, 1982, 511–516, 524).

By contrast, with the Comprehensive Anti-Apartheid Act of 1986, the US Congress forthrightly declared its goal "to bring an end to apartheid in South Africa and lead to the establishment of a nonracial, democratic form of government." Eschewing the vague demand for liberalizing steps, Congress set precise requirements that SAG had to meet for sanctions to be lifted: release all political prisoners; end the state of emergency used to suppress protest; unban political parties and allow free association, speech, and political participation; repeal the major apartheid laws; and agree to negotiate with black leaders.[1] It did so knowing that it was "not self-evident that that government, at least those who are in charge now, are prepared really to go into negotiations for power-sharing with black leaders" (HRG-1985-FOR-0007, 307). More bluntly, one congressman asked "what's the difference between the current government and the brown shirts [right-wing conservatives opposed to any reform of apartheid]? How did P.W. Botha rise to power? He came out of the same brown shirt movement [. . .] They both stand for the same thing" (HRG-1986-BFU-0017, 36–37). Clearly, the Botha government was not expected to concede these more extensive demands. Congress instead wished to "strengthen the hand of those white South Africans who are arguing for change" and recognized "that there are responsible white leaders [. . .] who don't take any solace at all in continuing the policies" of apartheid (HRG-1985-BHU-0010, 40, 50–51, and similar on 74). The intent was to incite "the [white] South African business community" and "English-speaking South African interests that have also adopted a policy which challenges the Government's rigid system of apartheid" as well as "white South Africans in the Government that would like to do more than they are doing" to "[do] things that they know they have to do"; implicitly, to force a change in the government (HRG-1985-FOR-0007, 282, 317–318). As one Senator explained, "by putting pressure on those who recognize the need for change, we may contribute to the sort of rapid dismantlement of apartheid [. . .] The white South African community is sharply divided. Those white South Africans in the business community who are most receptive to real change are likely to suffer if strong economic sanctions are adopted. We should give them another compelling reason to throw themselves behind the reform movement" (HRG-1986-BHU-0021, 2, similar on 3).

As our theory predicts, an internal political contest began once it was clear stronger sanctions would be imposed. National Party elites polarized between a "securocrats" faction led

[1]See sections 4 and 311 of Public Law 99-440.

by incumbent State President Botha that opposed further reforms, and a faction, eventually to be led by F.W. De Klerk, that favored reforms that might satisfy the US (G182–194). "The race for PW Botha's successor started immediately after [...] the Rubicon speech" (G283) in which Botha publicly spurned the new US demands (G194–201). Sanctions led several economic elites and major civil society organizations to align themselves with the reform faction (G251–252, 254, 269, 272), which made several failed attempts to force Botha into agreeing to further reforms (G256–261, 270–272), even as Botha shifted control over key government policies from the reform faction to the securocrats (G262–266). Botha won the 1987 election soon after sanctions were imposed, but this time his campaign was based on the need to secure the country from an escalating war against Communists in Angola and the African National Congress, rather than on reform or Western sanctions (G267–269, 282).

But the contest was not over. Botha was replaced as leader of the National Party in a caucus election that De Klerk won (G277) on a platform of further reform (G280). De Klerk immediately consolidated power within the party (G290–293, 295) and "reformulat[ed] the party's policy programme," so that the "NP's manifesto for the election [...] came across as a liberal democratic manifesto" (G295), after which "the simmering conflict between Botha as president and De Klerk as party leader came to a head," and Botha's pro-reform cabinet members forced him to resign (G277, 278). "Once Botha had resigned," the NP "project[ed] the party as 'under new management', with a new leader and new policies. It worked," though the National Party won by its smallest margin since 1958 (G296). "For De Klerk the election sent a strong message that most voters [...] wanted a new leader with fresh ideas" for ending apartheid, and De Klerk pledged drastic reform in his victory speech (G297). He proceeded to remove much of the power of the securocrats over government policy and implement reforms (G301–312) that were designed to ensure they satisfied the US. "de Klerk was following a set formula agreed with officials in the United Kingdom and the United States" to meet the demands of the US Congress, after which the US quickly ended its sanctions (T161).

References

Burton, Myra F., ed. 2016. *Foreign Relations of the United States, 1977–1980, Volume XVI, Southern Africa.* Washington, DC: United States Government Printing Office.

Giliomee, Hermann. 2012. *The Last Afrikaner Leaders: A Supreme Test of Power.* Charlottesville: University of Virginia Press.

Lemon, Anthony. 1982. "Issues and Campaigns in the South African General Election of 1981." *African Affairs* 81(325):511–526.

Midlane, Matthew. 1979. "The South African General Election of 1977." *African Affairs* 78(312):371–387.

Stultz, Newell M. 1984. "Interpreting Constitutional Change in South Africa." *The Journal of Modern African Studies* 22(3):353–379.

Thomson, Alex. 2008. *U.S. Foreign Policy Towards Apartheid South Africa, 1948-1994.* New York: Palgrave Macmillan.