Costly Economic Sanctions as Bargaining Tactics:
A Game Theoretic and Empirical Analysis

by

Catherine C. Langlois*, Associate Professor

McDonough School of Business, Georgetown University

Langlois@msb.edu

and

Jean-Pierre P. Langlois, Professor

Department of Mathematics, San Francisco State University

Langlois@math.sfsu.edu

November 30, 2005

*The authors, considering their contribution to be equal, have listed their names alphabetically.

Costly Economic Sanctions as Bargaining Tactics:
A Game Theoretic and Empirical Analysis

Abstract

We place sanctioning behavior in the context of rational bargaining in continuous time, drawing out, as a result, a new relationship between the cost of sanctions and the duration of sanctioning episodes. We interpret sanctioning as an attempt to wear down a target into acquiescence, without specifying a date for resolution. Instead, acceptance by either side of the other side's terms could happen at any time during the sanctioning episode. For this to rationally occur, both sides must face a constant balance between accepting or waiting further, a condition we call countervailing. Countervailing strategies are shown to form perfect equilibria whether the players are fully informed or not. If adopted, such strategies must leave recognizable traces in the data since their theoretical predictions run counter to inferences made in many empirical studies: it is the cost incurred *by the sanctioner* that must correlate with the *target*'s chances of acquiescence, *not* the cost incurred by the target. Using a classic sanctions data base we find empirical support for the predictions of our countervailing model.

Costly Economic Sanctions as Bargaining Tactics: A Game Theoretic and Empirical Analysis

"If a man knocks at a door and says that he will stab himself on the porch unless given $10, he is more likely to get the $10 if his eyes are bloodshot."[1] Schelling's characteristically vivid imagery, meant to illustrate the power of self-immolation in bargaining situations, deserves further development. What if the man at the door engages in some slow blood letting activity on his person, while wedging your door open to the icy winter storm. Will you give in to his request before he decides to give up on you? And what motivates you to transfer part of your wealth to the stranger? The icy wind, or the strange evidence of resolve unfolding at your doorstep? Sanctioning activity by states is rarely analyzed in the context of bargaining theory.[2] Yet, Schelling's stranger behaves like a sanctioning state: he is willing to incur, and to inflict, costs to compel the target to move away from the *status quo* to a situation that he favors. And the target presumably loses utility from acquiescence. If this were not so, a self-interested target would have abandoned the *status quo* of its own accord, without the prompting of a demand by the sanctioner.

Sanctioning activity is most often viewed "as a modality for defending standards of behavior," (Doxey, 1996, p.9). As such, sanctions are a punishment for behavior that does not conform to international norms. As Daoudi and Dajani note, sanctions express the sender's view on "morality and justice" (Daoudi and Dajani, 1983, p.161). If sanctions are simply just reprisals, they should be imposed at minimum cost to the sender, since sender costs would have no strategic value. But sanctions can also be viewed as a bargaining tactic designed to compel the other side to acquiesce to the sanctioner's demand. Then costs incurred by both parties have strategic value because they determine the terms of a waiting game in which each side hopes for the other's early capitulation. *Status quo*, demand, costs, and delays are then inextricably linked by the calculus of rational bargaining, and this perspective sheds new light on the very problematique of sanctioning behavior.

The purpose of this article is to place sanctioning behavior in the context of rational bargaining in continuous time, and this requires a novel modeling effort. Sanctioning is an attempt to wear down the target into acquiescence through the continuous application of costly pressure. However, we argue, no precise date is set for the target's acceptance although the very imposition of sanctions is intended to set the stage for a possible acquiescence at any time. But if acceptance of the sanctioner's terms is to be a rational choice at all times, it must be neither better nor worse, for the target, than waiting further. This perpetual balance between the payoffs of ending or continuing the struggle cuts both ways: rational resistance by the target must involve the expectation that the sender could give up and remove the self-inflicted costs of sanctions at any time. So, it must be neither better nor worse for the sender than to continue imposing them. We call "countervailing" those strategies that maintain a constant balance between the payoffs of ending and continuing the confrontation, and show, in a continuous time bargaining framework, that countervailing strategies form a perfect equilibrium *whether the protagonists are fully informed or not.*

A countervailing model of behavior yields new insight into the nature of sanctioning, and the relationship between costs and the duration of sanctions. Indeed, if the protagonists countervail each other, it is the cost *incurred by the sanctioner* that determines the *target*'s chances of acquiescence, *not* the cost incurred by the target. This insight runs counter to inferences made in many empirical studies of sanctioning data. Yet, guided by our theoretical approach, we find empirical support for the predictions of our model using the classic sanctions data base developed by Hufbauer, Schott and Elliott (1990a, 1990b). As such, we propose that bargaining theory, when set in continuous time, offers new, empirically relevant, insight into the way sanctions work.

1. The Sanctions Debate: A Brief Literature Review

The sanctions debate centers around their usefulness. Pape argues that "sanctions have little independent usefulness for the pursuit of non-economic goals," (Pape, 1997, p.93), and reexamines the data coding of Hufbauer, Schott and Elliott (1990a,b). But as Drezner (2000) points out, Pape looks for a direct and exclusive causal link between sanctioning and outcome, ignoring the possibly complex causal chain within which sanctions play their role. In fact, Drezner argues, most attempts to test the effect of economic sanctions using multivariate analysis fail to bolster the case for sanctions because the very techniques used "are limited in their ability to control for interaction effects" Drezner (2000, p.220). Thus authors such as Dashti, Davis and Radcliff (1997) focus on the domestic politics of the target and the resulting ability to bear the costs of sanctions, Hart (2000) emphasizes the signaling role that sanctions play, while Drezner (1998,1999) examines sanctions success with the green eyeshades of the protagonists' expectations of future conflict. Yet Drezner's (2000) own Boolean analysis of causality suggests that, in Russia's use of economic coercion against the newly independent states of the former Soviet Union, sanctions must impose high costs on the target to prompt concessions *if* conflict expectations are high.

The merits of Boolean analysis notwithstanding, it remains an a-theoretical approach to data analysis and, as such, does not inform on the process by which sanctions impact the target state's policies, or the conditions under which sanctions are imposed at all. In response to the empirical claim that sanctions usually fail to coerce the target, and are costly to senders, authors such as Tsebelis (1991), Eaton and Engers (1992, 1999), and Smith (1996) develop game theoretic models that focus on sender motivation to impose sanctions and target incentive to resist.[3] These models highlight the strategic importance of sanction threats and the role of imperfect information in the actual implementation of sanctioning. To conclude that sanctions do not "work" on the basis of observed sanctioning episodes may well understate the importance of sanctions as a tool of statecraft because "the most successful episodes are likely to end before sanctions are imposed," Drezner (2003, p.654).

Indeed, upon examination of data on sanction threats, Drezner (2003) finds that targets made significant concessions more frequently under the threat of sanctions than once sanctions are actually imposed.

While academics write about the doubtful impact of sanctions, policy makers continue to defend them. As Jesse Helms puts it: "Take away sanctions and how can the United States deal with terrorists, proliferators, and genocidal dictators? Our options would be empty talk or sending in the Marines," (Helms, 1999, p.5). The practical point of view is then, as Baldwin (2000) would argue, one of choice. Sanctions may be costly but more effective than "empty talk". And they probably cost less than "sending in the Marines." But costs also have strategic value. As Baldwin points out, the willingness to incur costs is "widely regarded as a standard indicator of one's resolve" (Baldwin, 1985, p.107). By analyzing sanctions in the context of bargaining, the strategic importance of costs becomes clear. Wagner (1988) and Morgan and Schwebach (1997) draw upon bargaining theory to analyze sanctions. Wagner pictures the protagonists as associating demands to the sharing of the gains from trade. His focus is on the Nash bargain reached under the threat of costly sanctions. Morgan and Schwebach associate a probability of acceptance by each side to all possible bargains and maximize the joint probability of acceptance to determine the outcome of bargaining. Higher costs increase the probability of acceptance of all possible bargains. Thus, sanctions "affect little the nature of any concessions but the pace of concessions is accelerated," (Morgan and Schwebach, 1997, p.35).

For Morgan and Scwebach (1997), and indeed most authors who appeal to the bargaining literature when discussing sanctions, costs are exogenous. They represent, conceptually, the inevitable burden of time wasted without agreement. Costs are not part of the strategic effort to coerce the other side to acquiesce. We propose that rational sanctioning imposes costs on both parties to hold each to the value of their preferred position. Thus, a rational sanctioner holds the target to the utility of the demand that he is making. This, we will show, is the *only* way to ensure that the target could accept the demand with some probability *at all times* during the sanctioning episode. The result of our analysis is a relationship between the *status quo,* the demand, costs to sender and target and the duration of sanctions. The empirical testing of our model requires that we estimate a hazard model. McGillivray and Stam note that "only a fraction of the large body of research on sanctions has looked at why some coercion attempts last for decades, but others last only a few months," (McGillivray and Stam, 2004, p.157). Our work therefore contributes to the recent efforts, by authors such as Mc Gillivray and Stam (2004), Dorussen and Mo (2001) and Bolks and Al-Sowayel (2000), to explain sanctions duration.[4]

2. Rational Sanctions, Conceptually

2.1 The Decision Variables

Conventional wisdom attributes the effectiveness of sanctions to the costs they impose on the target, and the theoretical challenge is to explain sanctions if they are also costly to the sender.[5] Yet there is another way to conceptualize the role of costly sanctions. Sanctioning is an attempt to gain the target's acquiescence through the continuous application of costly pressure without setting a precise date for the target's acceptance of the sanctioner's terms. The imposition of sanctions is then meant to set the stage for likely acquiescence at any point in time, while the actual timing of acquiescence is left to the other side. In fact, if the duration of sanctioning episodes is to have any strategy value, the timing of demands, imposition of sanctions, acquiescence of the target or backdown of the sender should be decision variables. This requires modeling sanctioning behavior in continuous time, but highlights the fact that sanctions impose costly pressure on a continuous basis.

If sanctioning creates a situation where the target can rationally resist or accept the other side's terms at any time, then acceptance must be neither better nor worse for the target than waiting further. If it was strictly better at one such time, acceptance would have already occurred at least some instant before. And if it was strictly worse, acceptance would be irrational and could not be expected at all at that time. Therefore the target's expected utility of waiting must, at all times during the sanctioning episode, match the payoff of immediate acceptance. However, rational resistance by the target, even if it is probabilistic, must incorporate the expectation that the sender, with some probability, could give up and remove the sanctions that are costly to itself. But if it is rational for the sender to give up at any point in time, he must also be indifferent between accepting the target's position, and continuing to apply sanctions in the hope that the target will acquiesce to its demand. In other words, at all times during the sanctioning episode, each side's expected utility of waiting must match the payoff of immediate acceptance of the other side's terms. Strategies that achieve this balance are called countervailing.

We formalize bilateral countervailing behavior in a continuous time bargaining framework, and we show that it results in a perfect equilibrium *whether the protagonists are fully informed or not*.[6] Moreover, the protagonists can *only* ensure that the other side will accept the terms offered with some probability, at any time, *if* they adopt a countervailing strategy. Indeed, threats could induce resolution at discrete dates, but it is sanctions, engineered to hold the other side to the utility of your bargaining position, that can give agreement a *persistent* chance. This, we argue, is a possible justification for the actual implementation of costly sanctions. And it is one that has been overlooked in the literature. We begin by formalizing the argument in general terms assuming, first, that there is complete information.

Let two states, the target state i and the sender state j, bargain in continuous time.[7] In the current *status quo* SQ state i implements policy that is distasteful to state j. State j can then make a demand D of state i. Both demand D and *status quo* SQ are assumed to be on the Pareto frontier so that acceptance by either side requires no further bargaining to divide a remaining surplus. If it were accepted, the demand D would improves j's utility at the expense of i's. In utility terms $U_i(\mathrm{D}) < U_i(\mathrm{SQ})$ and $U_j(\mathrm{SQ}) < U_j(\mathrm{D})$.

For now, let us assume that i sticks to the *status quo* as its bargaining position, refusing to budge in the face of sender state j's demand. State j can then attempt to compel i

to accept its demand by implementing costly sanctions. Even if i can somehow mitigate their effects by using countermeasures, the game would move to a sanction point S that brings i's utility down. And since j couldn't be hoping for i to accept a demand that would force its utility further down from S, it must be the case that $U_i(\text{S}) < U_i(\text{D}) < U_i(\text{SQ})$.

If the sender j's utility were to increase or remain stable as a result of sanctioning, $U_j(\text{S}) \geq U_j(\text{SQ})$, there would be little theoretical challenge: indeed, under these circumstances, sender j would be better off sanctioning the target than not, and would never accept a return to the *status quo*, while i could only do better by acquiescing to the sender's demand as soon as possible since $U_i(\text{S}) < U_i(\text{D})$. The theoretical challenge arises when $U_j(\text{S}) < U_j(\text{SQ})$ since, now, if j's sanctioning behavior is to be credible, it must balance the possibility that the target will accept the sender's demand with the target's incentive to resist, maintaining the *status quo*. This last situation is pictured in Figure 1, in utility terms:

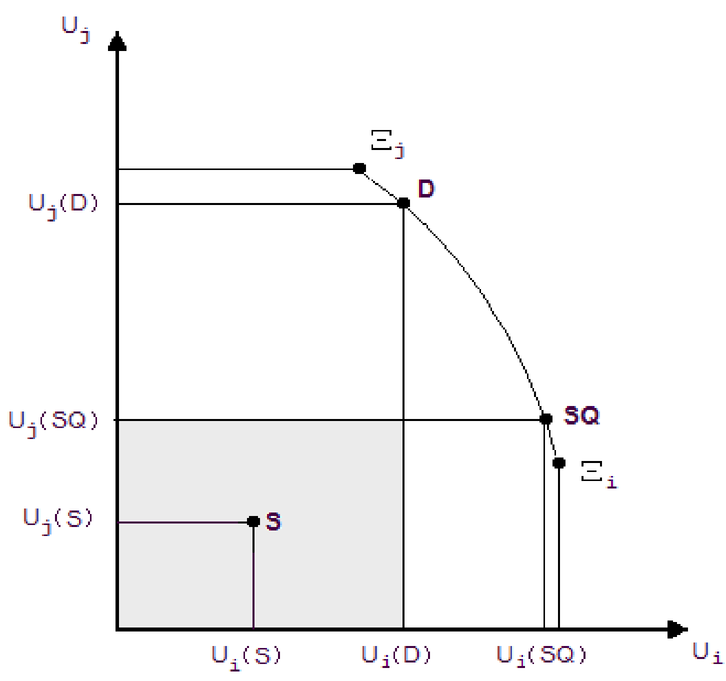

Figure 1: Sanctioning an Uncompromising Target

Ξ_i and Ξ_j in Figure 1 are extreme points on the Pareto frontier that will be important references if the players have incomplete information about each other's types.

2.2 An Example of Countervailing Behavior

We now outline how the two sides can countervail each other and how this determines a probability of acceptance of the other side's preferred position for each player. For i to be countervailed, its utility for accepting j's demand D must exactly match its expected utility of enduring the sanction S with some chance of prevailing. But since the sanction is worse than the demand, this can only hold if prevailing means that j eventually gives up, letting the *status quo* SQ stand. To understand the terms of this equation we need to define expected utility in the time continuum. Any future development at time t is naturally discounted by a factor $e^{-r_i t}$ where the magnitude of $r_i > 0$ is a measure of i's impatience. If the state of the game at time t is denoted $Y(t)$ then i should derive discounted utility $e^{-r_i t} U_i(Y(t))dt$ within an infinitesimal interval of length dt at time t. The *expected* utility of the developments described by $Y(t)$ is therefore:

$$E_i = \int_0^\infty r_i e^{-r_i t} U_i(Y(t))dt \qquad (1)$$

The normalizing factor r_i ensures that expectations are comparable to the utility of instant results: accepting D, for instance, should yield $U_i(\text{D}) = \int_0^\infty r_i e^{-r_i t} U_i(\text{D})dt$, or the utility of D, $U_i(\text{D})$, for all foreseeable future. If some settlement X (SQ or D in the current discussion) is reached at time s after continually implementing a fixed sanction S, then (1) yields:

$$E_i = (1 - e^{-r_i s})U_i(\text{S}) + e^{-r_i s} U_i(X) \qquad (2)$$

Of course, s is not necessarily known for sure, and neither is the nature of X since either side can accept the other's terms. To understand this issue we need to give further structure to the players' behavior. But what we are about to assume for this conceptual discussion will *not* be necessary for the general sanctioning model of Section 3. Instead, the behavior that we now describe will emerge as subgame perfect equilibrium behavior in the general model.

Assume that each side uses as acceptance strategy an independent random variable with exponential distribution. Formally, let $\phi_j(s) = e^{-\mu_j s}$ ($\mu_j \geq 0$) be the probability that j will *not* have accepted by time s, and similarly for i. The probability that j accepts at date s within an interval ds is then $-\phi_i(s)d\phi_j(s) = \mu_j e^{-(\mu_i+\mu_j)s}ds$ and similarly for i. The expected payoff resulting from such behavior applied to (2) then reads[8]

$$E_i = \frac{r_i U_i(\text{S}) + \mu_i U_i(\text{D}) + \mu_j U_i(\text{SQ})}{r_i + \mu_i + \mu_j}$$

In fact, E_i is i's expected payoff at *all* times as long as nothing changes in the two sides' relative bargaining positions and acceptance strategies.[9] So, i is countervailed if $E_i = U_i(\text{D})$ at all times which yields by cancellation of the μ_i terms:

$$\mu_j = \lambda_i = r_i \frac{U_i(\text{D}) - U_i(\text{S})}{U_i(\text{SQ}) - U_i(\text{D})} \text{ and } \phi_j(s) = e^{-\lambda_i s} \qquad (3)$$

Examining the drivers of sanctioner j's probability of acceptance λ_i, we note in the numerator the difference $(U_i(\text{D}) - U_i(\text{S}))$ which is the cost *to the target* of failing to acquiesce to the sanctioner's demand, suffering the consequences of sanctions instead. The higher that cost, the higher the probability that the *sanctioner* will back down and accept the *status quo*, because the sanctioner is holding the target to the utility of the demand made. Target state i currently suffers the cost of sanctions and will continue to do so until it accepts j's demand or j backs down and accepts the *status quo*. Since accepting j's demand would bring i's utility up to the expected level it is held to by j, that expectation can only be realized, given current sanctions, if there is a probability that state j will back down, and that probability must increase with the costs imposed. The difference between the target's utility at the *status quo* and the utility it would have if it acquiesced to j's demand, $U_i(\text{SQ}) - U_i(\text{D})$, also plays a role in the probability that j will back down. The bigger the gap between what j wants to hold i to, and i's utility at the *status quo*, the less likely j will be to give in, backing down and accepting the *status quo* which is i's current bargaining position.

If i adopts the *status quo* as its bargaining position, it must hold state j to an *expected* utility of $U_j(\mathrm{SQ})$ while it is currently experiencing the utility of the sanctioning point with some hope of prevailing. A symmetric calculation for $E_j = U_j(\mathrm{SQ})$ yields:

$$\phi_i(s) = e^{-\lambda_j s} \quad \text{with} \quad \lambda_j = r_j \frac{U_j(\mathrm{SQ}) - U_j(\mathrm{S})}{U_j(\mathrm{D}) - U_j(\mathrm{SQ})} \tag{4}$$

The parameter λ_j and the corresponding probability that target i will acquiesce to the sender's demand at any point in time during the sanctioning episode, increases with the costs *borne by the sanctioner.* Thus, increased costs to the sanctioner actually increase the chances that the sanctioning episode will be successful. This counterintuitive proposition results from a rational approach to sanctions. And a theory driven analysis of the widely used Hufbauer, Schott and Elliot (1990a,b) data base supports the prediction of our model.

2.3 Sanctions and Target Counteroffers

If the *status quo ante* is the target's bargaining position, the costs that determine the target's acceptance of the sender's demand are real costs, actually incurred by the sender as a result of sanctioning. But the target could also make a counteroffer in the face of the sender's demand. If the target does make a counteroffer, it is the cost of sanctions relative to the counteroffer position that will matter.[10] The target's probability of acceptance is then no longer determined only by actual costs but by opportunity costs as well. Indeed, the sender could have small actual costs but a large opportunity cost of maintaining the sanctions in the face of a generous counteroffer by the target and this would increase the target's likelihood of acceptance of the sender's demand. To see this, consider counteroffer O as represented in Figure 2 below:

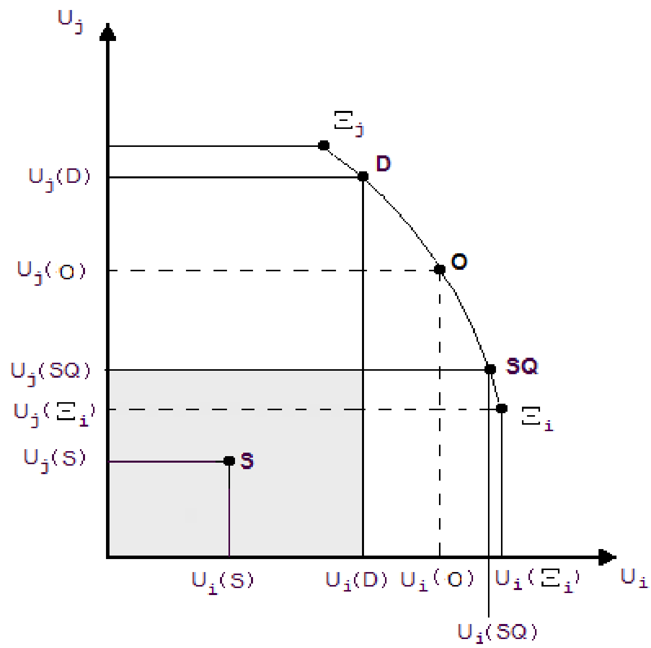

Figure 2: Sanctioning with Counteroffers

The bargaining process is conciliatory if state i makes a counteroffer O. By doing so the gap between j's demand and i's offer narrows and this increases the likelihood that the target will accept the sender's demand since probabilities of acceptance are now expressed with reference to bargaining position O rather than the initial *status quo* SQ. Indeed, the probability that target i has not accepted the sender's demand by date s now reads:

$$\phi_i(s) = e^{-\lambda_j s} \quad \text{with} \quad \lambda_j = r_j \frac{U_j(\text{O}) - U_j(\text{S})}{U_j(\text{D}) - U_j(\text{O})} \qquad \text{(5a)}$$

Since $U_j(\text{D}) - U_j(\text{O}) < U_j(\text{D}) - U_j(\text{SQ})$ and $U_j(\text{O}) - U_j(\text{S}) > U_j(\text{SQ}) - U_j(\text{S})$ parameter λ_j, and corresponding probability with which the target acquiesces to the sender's demand, increase unambiguously. But the sender's willingness to accept the target's counteroffer also increases. Indeed, while the target is making a compromise counteroffer of O, the sender is still holding the target to the expected utility of his demand D. As costly sanctions are currently imposed, the sender must be more likely to accept counteroffer O, than the *status quo*, SQ, since $U_i(\text{O}) < U_i(\text{SQ})$. So, from the moment at which a counteroffer is on the table, expected delay to resolution should decrease. This is of course reflected in sender j's probability of acceptance of the target's counteroffer O since:

$$\phi_j(s) = e^{-\lambda_i s} \quad \text{with} \quad \lambda_i = r_i \frac{U_i(\text{D}) - U_i(\text{S})}{U_i(\text{O}) - U_i(\text{D})} \qquad \text{(5b)}$$

However, this does not mean that sanctioning episodes that end in compromise should necessarily be shorter than sanctioning episodes that do not. The theory does not predict a relationship between a string of compromise offers and sanctions duration. Indeed, compromise offers come to the table after some time without agreement. Overall, a process that involves compromise along the way could well last longer than one which ends with the sender accepting the *status quo*.

The target may not be conciliatory when faced with sanctions and could, instead, harden its position moving to extreme position Ξ_i. The target would then be moving to the worst position possible for the sender on the Pareto frontier and would now hold the sender to its utility. Such a move, as pictured in Figure 2, above, reduces the probability that the target will acquiesce to the sender's demand. Indeed:

$$\phi_i(s) = e^{-\lambda_j s} \text{ where } \lambda_j = r_j \frac{U_j(\Xi_i) - U_j(\text{S})}{U_j(\text{D}) - U_j(\Xi_i)} < r_j \frac{U_j(\text{SQ}) - U_j(\text{S})}{U_j(\text{D}) - U_j(\text{SQ})} \qquad \text{(6)}$$

It turns out that the adoption of the extreme points Ξ_i and Ξ_j as bargaining positions is necessary, in perfect Bayesian equilibrium, if sender and target are unsure of each other's type. Only when the uncertainty is resolved can the players adopt positions SQ, D, or O that are not extreme points on the Pareto frontier. We now illustrate these strategies further by developing a game theoretic sanctioning model in continuous time, that also accommodates incomplete information.

3. A Rational Sanctioning Model

3.1 The Payoff Structure and Discounted Objective

Where sanctioning is at stake, bargaining is not about the sharing of a surplus. Rather, it is about the joint adoption of an alternative *status quo* under the threat, or the actual implementation, of sanctions. Player moves fall into three categories: (1) the formulation of offers and counteroffers that represent one or the other player's proposed choices on the issues involved; (2) acceptance of the other side's last offer that is assumed to end the game, and establish a new *status quo* for all foreseeable future; (3) the implementation of sanctions and possible countermeasures while disagreement lasts. In order to let the *timing* of decisions be truly strategic, without restrictions on when and whose turn it is to speak, it is necessary to frame the problem in continuous time.

To fix the ideas, we will develop a simple model of sanctions but the methodology applies to more general models. Consider a sender state that disagrees with the nature and level of political freedom in the target state. Political freedom in the target state is represented by a variable $y_{i,1} \in [0,1]$, with $y_{i,1} = 0$ indicating complete lack of freedom while $y_{i,1} = 1$ indicates full political freedom. To punish the target for failing to redress the situation, the sender state j can impose economic sanctions at levels that vary along a continuum represented by $y_j \in [0,1]$. If $y_j = 0$, no sanctions are imposed. If $y_j = 1$ the sender state imposes maximum possible sanctioning. In response to these sanctions, the target state can implement countermeasures, such as third party supply agreements, at a level measured by variable $y_{i,2} \in [0,1]$. Let SQ $= (\, y_{i,1}, y_{i,2},\, y_j) = (0,0,0)$ be the initial *status quo* that target state i hopes to preserve. Sender state j wants to see more political freedom prevail in state i and formulates its demand accordingly. Ideally, j hopes for full freedom and could make the demand D $= (y_{i,1}, y_{i,2_i}, y_j) = (1,0,0)$.

In order to pressure i, j may impose sanctions up to the maximum $y_j = 1$. This would move the parties to $Y = (y_{i,1}, y_{i,2},\, y_j) = (0,0,1)$ if i makes no concession and does not adopt countermeasures But the target's response to the imposition of sanctions affects its utility $U_i(y_{i,1}, y_{i,2}, y_j)$ which decreases with $y_{i,1}$ and y_j, and increases with $y_{i,2}$ when sanctions are imposed. The following functional forms are possible representations of the two sides preferences but the existence results obtained in Theorems 1 and 2 in no way depend on this specific form:

$$U_i(y_{i,1}, y_{i,2},\, y_j) = 1 - y_{i,1} + c_i(y_{i,2} - 2)y_j \tag{7}$$

where $c_i > 1$ to reflect i's preference for lower political freedom and lower sanctions, while the sender's utility could be represented as follows:

$$U_j(y_{i,1}, y_{i,2},\, y_j) = y_{i,1} + c_j(y_j - 2)y_{i,2} \tag{8}$$

where $c_j > 0$ to reflect j's preference for higher political freedom and lower sanctions. Maximal imposition of sanctions by the sender, $y_j = 1$, leads the target's rational adoption of countermeasures $y_{i,2} = 1$ and the definition of a sanctions state S $= (y_{i,1}, y_{i,2_i}, y_j) = (0, 1, 1)$ which forms a Nash equilibrium of the "one-shot" game.[11] At S, $U_i(\text{S}) = 1 - c_i < 0$, and $U_j(\text{S}) = -c_j < 0$.

It is well known that constant play of the Nash equilibrium of the constituent game provides a subgame perfect equilibrium (SPE) in a repeated game through time, with discounting of the future. Moreover, if the shadow of the future is long enough, standard trigger schemes based on the threat of reversion to the Nash equilibrium provide SPEs that support a more cooperative solution. But the question here is a bit more complex: all cooperative points P $= (y_{i,1}, 0, 0)$ with $y_{i,1} \in [0, 1]$ ranging from SQ to D yield non-negative utilities for both sides and are therefore candidates for such a trigger scheme. The *status quo* SQ is one such point, and it is i's favorite as well as the starting position of the game. So, how can j rationally pressure i into accepting any move, especially a move to D? The issue is not that D is equally sustainable in SPE. It is, instead, that a move to D would disturb a current situation, which is sustainable in equilibrium, to advantage j at i's expense. Our question is then: "can sender j engineer a move to D in subgame perfect equilibrium by imposing the costly sanctions of point S?" The puzzle arises from the twin facts that the move from SQ to S is detrimental to *both* sides rather than just to i, and that any Pareto optimal point P is a potential bargaining outcome.[12]

To investigate these issues, we must clearly define the players' strategies and objectives. In general, Let $Y(t) = (y_i(t), y_j(t)) \in \mathcal{Y} = \mathcal{Y}_i \times \mathcal{Y}_j$ represent the players' choices as they stand at date t, where $\mathcal{Y}_i$ and $\mathcal{Y}_j$ will denote the players' respective action spaces. In the above example $\mathcal{Y}_i \times \mathcal{Y}_j = [0,1]^2 \times [0,1]$. Let us also denote by $X_i(t) \in \mathcal{Y}$ and $X_j(t) \in \mathcal{Y}$ the two sides' respective current offers of settlement at date t, possibly SQ and D. Each side's utility is a flow through time that depends on the current state of the game $Y(t)$ as it varies according to the players' strategies. As the game proceeds through continuous time t, changes occur in $Y(t)$ that reflect the players' possible decisions. Discontinuities in $Y(t)$, meaning sudden jumps from one choice to another, will be called "events".[13] As the game unfolds, all developments are recorded and come to form the history h^t (at time t) upon which the players plan their future moves according to their respective strategies. The resulting game, starting at time t, is called a subgame in which the planned moves must be optimal, regardless of the prior developments, for the strategies to form a SPE. The players' objectives must therefore be formulated in any such subgame.

A possible *future* evolution of the game, following a history h^t, will be called a path and it is a map $\sigma^t : \tau \to \left(Y^t(\tau), X_i^t(\tau), X_j^t(\tau)\right)$ from $[0,\infty)$ into $\mathcal{Y} \times \mathcal{Y} \times \mathcal{Y}$ that, to any additional delay τ associates the intended state $Y^t(\tau)$ and the intended offers $X_i^t(\tau)$ and $X_j^t(\tau)$ by i and j respectively, at time $(t+\tau)$. If one side accepts the other's offer, call it X, after some delay s then $Y^t(\tau) \equiv X$ for all $\tau \geq s$. The players' strategy profile $\Psi = (\psi_i, \psi_j)$ therefore associates to any prior history h^t an intended path σ^t. If $Y^t(\tau)$ denotes the expected state of the game at future time $(t+\tau)$, it is valued $r_i U_i(Y^t(\tau))d\tau$ by i within an infinitesimal period $d\tau$ and it is discounted by $e^{-r_i\tau}$ when viewed from time t. At time t, i's expected utility if path σ^t is followed therefore reads

$$E_i(\sigma^t) = \int_0^s r_i e^{-r_i\tau} U_i(Y^t(\tau))d\tau + e^{-r_i s} U_i(X) \tag{9}$$

The two sides' respective strategies define offers, counteroffers, conditions of acceptance as well as current sanctions and countermeasures. Therefore, they result in a path σ^t, and associated expected utility $E_i(\sigma^t)$. When considering path σ^t the timing of each decision is as important as its magnitude, especially that to accept the other side's terms denoted X in formula (9).[14]

Along any given path σ^t, let $F_i^t(s)$ denote the cumulative probability that i accepts j's current offer (and $F_j^t(s)$ for j) by date $(t+s)$ according to the strategies induced by history h^t. Further denote by $H_{ij}^t(s) = 1 - F_i^t(s) - F_j^t(s)$ the probability that neither side accepts by that time and by $Y^t(s)$ the state on the path before acceptance. We show in Lemma 1 in appendix that Player i's expected utility for path σ^t can be then written:

$$E_i(\sigma^t) = \int_{[0,\infty)} e^{-r_i s} dG_{ij}^t(s) \qquad (10)$$

with $dG_{ij}^t(s) = U_i\big(X_j^t(s)\big)dF_i^t(s) + U_i(X_i^t(s))dF_j^t(s) + r_i U_i(Y^t(s))H_{ij}^t(s)ds$

In words, $dG_{ij}^t(s)$ is player i's expected utility within the infinitesimal instant of length ds at time $(t+s)$. It is the weighted average of infinitesimal utilities associated to all possible events at $(t+s)$, namely that i accepts j's offer, that j accepts i's offer, or that neither player accepts the prior offer made by the other side. The weights are the probabilities of these events. This corresponds to the standard definition of an expected utility but applies to the utility flow at each instant. The justification for this formulation requires some precise definitions and mathematical work that are presented in appendix and lead to Lemma 1. In the above example, we can set $X_j^t(s) \equiv$ SQ, $X_i^t(s) \equiv$ D, and $Y^t(s) \equiv$ S. In the most general case, strategies can involve probabilistic choices of one path or another at event times. Since the probability $\mathbb{P}^t(\sigma^t)$ of each path σ^t would clearly result from such strategies one would use the expectation $\mathbb{E}_i^t = \sum_{\sigma^t} \mathbb{P}^t(\sigma^t) E_i(\sigma^t)$ as objective function. But it will be enough to work with individual paths σ^t for all our existence results and data analysis.

3.4 A Subgame Perfect Equilibrium

We assume that the game always begins with an existing status quo SQ which will remain i's (the target) position until he makes a counteroffer. And the game only starts when j (the sender) expresses a demand denoted D. Either side can change its position at any time as long as the new position suggested keeps the players on the Pareto frontier $\mathcal{P}$.[15] We also assume that neither player ever offers more than the other requests, and changes in the offers made do not occur with infinite frequency. The two sides' last offers $X_i(t)$ and $X_j(t)$ are therefore always well defined. Formally, an extreme offer Ξ_i for player i is such that all Pareto optimal X satisfy $U_i(X) \leq U_i(\Xi_i)$. Extreme offer Ξ_i is one that represents the best possible outcome for player i, and therefore the worst outcome for his opponent by Pareto optimality, and symmetrically for Ξ_j.

While disagreement lasts, the two sides also implement a current state $Y(t)$ which can be anywhere in the action space $\mathcal{Y}$. $\mathrm{S} = (y_i^{\mathrm{S}}, y_j^{\mathrm{S}})$ is a costly sanction state if for all $y_i \in \mathcal{Y}_i$ and $y_j \in \mathcal{Y}_j$[16]

$$U_i(y_i, y_j^{\mathrm{S}}) < U_i(\Xi_j) \quad \text{and} \qquad U_j(y_i^{\mathrm{S}}, y_j) < U_j(\Xi_i) \tag{11}$$

In the example above $\Xi_i = \mathrm{SQ}$ and $\Xi_j = \mathrm{D}$ are extreme offers, and the Nash equilibrium is a costly sanction state since $U_i(y_{i,1}, y_{i,2}, 1) \leq U_i(\mathrm{S}) = 1 - c_i < 0 = U_i(\mathrm{D})$ and $U_j(0, 1, y_j) \leq U_j(\mathrm{S}) = -c_j < 0 = U_j(\mathrm{SQ})$.

Player j adopts a countervailing strategy towards i, if he makes i indifferent at all times between accepting j's last offer or continuing the game. Formally, it must be true at any time t that:

$$E_i(\sigma^t) = U_i(X_j(t)) \tag{12}$$

We will also say that, by choosing a strategy that achieves (12) for any of i's strategies, j countervails i. In the continuous time bargaining game that we have defined, it turns out that countervailing strategies determine subgame perfect equilibria whether the players are fully informed or not.

If the players did *not* adopt a countervailing strategy, there would necessarily be stretches of time where the probability that one side or the other will accept the other's position is nil. Once costly sanctions are imposed, this could only occur if one side expected the other to soften his stance at some point in the future. Focussing on the target's expectations, the target would then have to anticipate a lowering of the sanctioner's demand, or an increase in the probability that the sanctioner back down letting him remain at the *status quo*, unpunished. But then, why would a sanctioner state suffer costly sanctions only to let the target know that things will improve later.[17] Instead, one would expect states to impose costly sanctions in the hope that this will prompt the other side to comply with the sanctioner's demand or at least make a conciliatory move. But, in the very nature of sanctions, persistent pressure is on. There is no deadline for relief, and there is none for resolution either. In fact resolution is presumably expected at any point in time, once sanctions are imposed. Only countervailing strategies provide rational protagonists with a persistent chance that the disagreement will end. If sanctioning is a rational activity for sender states, countervailing behavior should be empirically relevant. We now exhibit countervailing strategies in our continuous time sanction game and state our first main result.

Consider the following strategy ψ_j for j:

Offer constantly some $X_j \in \mathcal{P}$ and play constantly the choice y_j^{S} of a costly sanction state S. After any history h^t wait for a further delay of at least s with probability

$$\phi_j^t(s) = e^{-\lambda_i s} \text{ and } \qquad \lambda_i = r_i \frac{U_i(X_j) - U_i(y_i^t(s), y_j^{\mathrm{S}})}{U_i(X_i^t(s)) - U_i(X_j)} \qquad (13)$$

but accept instantly if $X_i(t) = X_j$.[18] We have

***Theorem 1*:** The strategy ψ_j is countervailing for player i so that for all times t formula (12) holds. Moreover, the strategy profile $\Psi = (\psi_i, \psi_j)$ (with ψ_i defined symmetrically) forms a SPE.

The proof is given in appendix.

3.5 A Perfect Bayesian Equilibrium

We now assume incomplete information on both sides of the dispute. So $i, l \in \mathcal{I}$ and $j, k \in \mathcal{J}$ are *types* of target and sender players with various utilities and impatience factors on each side. In our example this merely means that the coefficients c_i, r_i (and c_j, r_j) vary from one type to the other. In general we assume finitely many types. Most importantly, we assume incentive compatibility between the types as follows:

• All types on each side have the same incentives, meaning that all types of that side have the same preference ordering over outcomes.

• The extreme offers denoted $\Xi_{\mathcal{I}}$ and $\Xi_{\mathcal{J}}$ for each side are common to all types.

• All types share a same Nash equilibrium NE that is a costly sanction state as defined in (11).

Both sides have beliefs about the other side's type that evolve with time according to developments. So, $b_k(t) \geq 0$ such that $\sum_{k \in \mathcal{J}} b_k(t) = 1$ are the beliefs about side $\mathcal{J}$ (and symmetrically for $\mathcal{I}$) and they are updated as the game unfolds. However, by incentive compatibility, offers as well as choices in current state $Y(t)$ are pooling, and only acceptance can be separating.[19] The evolution of play, when the protagonists are unsure of the other's type, relies on the definition of a countervailing strategy aimed at an opponent type assumed to have the highest probability of acceptance of the offer made under this strategy. More precisely, let $i \in \mathcal{I}$ be such that

$$\lambda_i = \max_{l \in \mathcal{I}} \{\lambda_l = r_l \frac{U_l(\Xi_{\mathcal{I}}) - U_l(\text{NE})}{U_l(\Xi_{\mathcal{J}}) - U_l(\Xi_{\mathcal{I}})}; b_l(t) > 0\} \qquad (14)$$

and λ_j defined symmetrically.[20]

In our construction of a PBE i and j direct their strategies to the other's type that is "most likely to accept" at that time, while all other types with positive beliefs are assumed "sure to wait", meaning that if $k \in \mathcal{J} - \{j\}$ and $b_k(t) > 0$ then $\phi_k^t(s) \equiv 1$.[21] As a result, side $\mathcal{J}$ as a whole can be expected by all types of side $\mathcal{I}$ to wait further with probability

$$\phi_{\mathcal{J}}^t(s) = \sum_{k \in \mathcal{J}} b_k(t) \phi_k^t(s) = 1 - b_j(t) + b_j(t)\phi_j^t(s) \qquad (15)$$

In the incomplete information case, it is this $\phi_{\mathcal{J}}^t$ that impacts all types on side $\mathcal{I}$ in their calculus of balancing the chances of the other side's acceptance with their own acceptance and the costs of waiting further. Not surprisingly, countervailing of i holds when

$$\phi_{\mathcal{J}}^t(s) = e^{-\lambda_i s} \qquad (16)$$

with λ_i defined above. And this choice will result in all $l \in \mathcal{I} - \{i\}$ such that $b_l(t) > 0$ to prefer waiting with certainty. Meanwhile, beliefs about each player $k \in \mathcal{J}$, from any time t on, must evolve continuously in time according to Bayes' Law

$$b_k(t+s) = \frac{b_k(t)\phi_k^t(s)}{\phi_{\mathcal{J}}^t(s)} \tag{17}$$

For all $k \in \mathcal{J} - \{j\}$ with $b_k(t) > 0$, and therefore $\phi_k^t(s) \equiv 1$, it is clear from (17) that $b_k(t+s)$ must be increasing with s while beliefs $b_j(t+s)$ must correspondingly decrease. Conceptually, since j is the only type likely to accept, a continued lack of acceptance on side $\mathcal{J}$ makes it less and less likely that j is the actual opponent. Indeed, (15) and (16) together result in:

$$\phi_j^t(s) = \frac{e^{-\lambda_i s} + b_j(t) - 1}{b_j(t)} \tag{18}$$

which is not only decreasing with s but actually reaches zero in finite time $\theta = -\frac{ln(1-b_j(t))}{\lambda_i}$. But clearly, $b_j(t+\theta) = 0$ by (18) for $k = j$. If that happens j is eliminated by side $\mathcal{I}$ from its countervailing effort and is replaced by the next candidate that then maximizes the corresponding λ_k's. Indeed, λ_i (and therefore i) itself may change while j is still being countervailed by side $\mathcal{I}$ if belief $b_i(t+s)$ happen to fall to zero for $s < \theta$.

Consider therefore the strategy profile Ψ we just outlined (see the formal Definition 1 in appendix) where the two sides make the extreme offers $\Xi_{\mathcal{I}}$ and $\Xi_{\mathcal{J}}$, play the Nash equilibrium NE as sanction state, and wait with probabilities determined by (15) and (16). We then have

***Theorem 2*:** The strategy profile Ψ, together with beliefs (17), forms a PBE.

The theorem is proved in appendix. When the players do not have full information, they can adopt a sequence of countervailing strategies, aimed at the currently assumed type. And countervailing behavior takes a form that is very similar to that of the perfect information case. The main difference is that, until each side has found out about the other's exact type, the offers and demands must remain extreme. Moreover, the sanctions point must be a Nash equilibrium of the constituent game. Interestingly, this holds even if information is *almost* complete. In Harsanyi's perspective that complete information is only an ideal limit of the incomplete information case, this is important: the slightest doubt about the other side's type lead to uncompromising and hurtful behavior.[22] However, incomplete information does not necessarily add to delays that are already present in the complete information case since the λ_i parameters of the slightly perturbed types may be higher than that of the true type, thus implying faster expected acceptance.

Nevertheless, whether the players are fully informed or not, the probability that the *sender* accepts the target's position will still increase with the sanctions cost suffered *by the target,* while the target will be more likely to acquiesce to the sender's demand when the sanctions costs suffered *by the sender* increase. This results from (14) and (16) and it is the basic proposition that we test on the sanctions data.

4. Empirical Evidence

4.1 The Data

We seek to validate the predictions of our model using a data base constructed by Daniel Drezner.[23] Drezner (1998,1999) tests his conflict expectations model of sanctions on a subset of Hufbauer, Schott and Elliott's sanctions data base. As his model "addresses situations where a sender country uses economic diplomacy to coerce a target country into a policy concession," (Drezner, 1999, p.102), he eliminates sanctions cases with no clearly identified sender, and cases where the sanctions are actually embargoes designed to weaken the target militarily. Moreover, because Drezner's model predicts that the difference between sender and target costs will impact the outcome of sanctions, he develops a cardinal measure of sender costs expressed as a percentage of the sender country's GDP.[24] As sender costs are critical in our model, and we also seek to explain the use of economic tools to coerce a target state to move from the current *status quo*, the data base constructed by Drezner is ideal for our purposes.

The rational sanctioning strategy that we describe requires that the sender impose costs on the target, and experience no gain from sanctioning activity. Examination of the costs incurred by the sender in Drezner's data base reveals that both target and sender suffered costs as a result of sanctioning in only 51 of 114 cases. We refer to these as the costly sanctions cases. For the remaining 63 cases, the sender actually gained from sanctions activity in 53 cases.[25] In another 10 cases, the target incurred no costs.[26] We refer to this set of 63 cases as the cost free sanctions cases. Explaining sanctions outcome or duration in these 63 cases requires a different model from ours. In fact, for these cases, we would expect the relationships predicted by our model to fail. Indeed, if the sender *gains* from imposing costly sanctions on the target, then the target should acquiesce to the sender's demand immediately, if sanctioned, unless the utility lost from doing so outweighs the costs suffered as a result of sanctioning. In such cases, the target's acquiescence is no longer positively linked to the costs suffered by the sender as our model predicts. And if the sender gains from imposing sanctions, he should not accept the target's compromise offer unless that offer provides him with greater utility than the imposition of sanctions. We would then observe compromise offers either being accepted immediately, or forever turned down. Sender acceptance of the compromise should then depend on the extent of its gains from sanctioning, regardless of the costs imposed on the target. But this is contrary to what our model predicts since sender acceptance depends positively on target costs.

The above discussion suggests that a "one-size-fits-all" approach to the empirical analysis of sanctions data may be misguided because it imposes the same set of drivers and hypothesized effects on all cases. We will indeed find that the variables that explain duration in those cases where sanctions are costly to both target and sender, fail to explain duration when the sender gains from imposing sanctions. And in the 51 costly sanctioning cases, it is also necessary to distinguish between two sets of sanctions episodes: those that end with target acquiescence to the sender's demands and those that do not. In 14 of the 51 costly sanctions cases, the target acquiesced to the sender's demand. Of the 37 remaining sanctioning episodes, 6 are ongoing,[27] and 31 have been resolved with the sender accepting a compromise offer from the target, or acquiescing to the *status quo*. If the target acquiesced, then the observed duration of the sanctioning episode should reflect the *target's* probability of acquiescence. If a compromise was reached or the status quo prevailed, the length of the sanctioning episode should reflect the drivers of the *sender's* probability of acceptance. In other words, given observed outcomes, we should not expect all sanctioning episodes to be related to the same set of explanatory variables, or to be related to them in the same way if some are common drivers of all sanctioning experiences. Again a "one-size-fits-all" approach could veil the true drivers of the sanctioning episode. We will therefore analyze the cases in which the target acquiesced separately from the cases in which the sender accepted a compromise or the *status quo*. In fact, Dashi-Gibson, Davis and Radcliff (1997) also proceed similarly separating out all cases in which destabilization of the target government is the sender's goal.

4.2 Empirical Results

4.2.3 The Predictions of the Costly Sanctions Model

Our model predicts the probabilities $\phi_i^t(s)$ and $\phi_j^t(s)$, defined at date t, that states i and j will not have accepted each other's bargaining positions by date $(t+s)$. In the language of survival analysis, these are survival functions (see Cleves, Gould and Gutierrez, 2004). In the cases we are interested in, survival is that of the sanctions. Indeed consider $\phi_i^t(s)$ and set $t=0$ for simplicity. If T marks the end a sanctioning episode with state i accepting j's offer, then:

$$\phi_i^t(s) = e^{-\lambda_j s} = \text{Probability}(\text{T} > s) = \text{Survival function}$$

The hazard function $h_i(s)$ that is typically estimated in survival analysis is defined as follows:

$$h_i(s) = \frac{f_i(s)}{\phi_i^t(s)} \text{ where}$$

$$f_i(s) = \frac{d}{ds}(1 - \phi_i^t(s)) = \frac{d}{ds}(1 - e^{-\lambda_j s}) = \lambda_j e^{-\lambda_j s}$$

It follows that hazard rate $h_i(s) = \lambda_j$.

Consider the cases in which the target i acquiesces to the sender's demand. The appropriate hazard rate given this realization is:

$$h_i(s) = \lambda_j = r_j \frac{U_j(\text{SQ or O}) - U_j(\text{S})}{U_j(\text{D}) - U_j(\text{SQ or O})}$$

where O is a possible counteroffer by the target. We therefore expect an estimation of the hazard function to increase with sanctions costs to the sender, a proxy for the utility gap between the sanctions point and the *status quo* or counteroffer

For those cases that end in acceptance of a compromise outcome or acceptance of the *status quo* by the sender, the appropriate hazard rate is:[28]

$$h_j(s) = \lambda_i = r_i \frac{U_i(\text{D}) - U_i(\text{S})}{U_i(\text{SQ or O}) - U_i(\text{D})}$$

Now higher costs to the target, a proxy for the utility gap between the sanctions point and the demand by the sender, increase the probability that the sanctions episode will end.

4.2.4 The Estimated Hazard Rates

We estimated Weibull hazard functions to test the relationship between cardinal estimates of costs, CTARGET and CSENDER and the probability that the sanctions episode will end. Indeed, if imperfect information is present there is no reason to believe that the hazard rate will be constant. To fully explain the duration data, we introduced standard control variables such as ASSIST, a dummy variable taking on the value of 1 if the target received assistance during the sanctions episode, ALLY, which codes the prior relationship between sender and target on a scale of 1 to 3 with a 3 indicating a cordial relationship, COOPERATION an index of international cooperation with the sender, and OUTCOME which codes the policy outcome associated to the sanctions episode weighted by the importance of the issue.[29] While we also introduced variables such as the relative power of the sender, the trade linkage between target and sender, or the target's health and stability, none of these turned out to be significant and did not improve our estimation of the hazard function upon examination of a log likelihood ratio test. All estimations were performed using STATA 8.

Table 1 presents estimation of the hazard function for the 14 costly sanctions cases in which the target acquiesced to the sender's demand.[30] We also provide estimation of a hazard function using the same explanatory variables on the corresponding 19 cost free sanctions cases.[31] It is important to note, at the outset, that we do *not* attempt to explain the duration of sanctions episodes if the sender does *not* incur costs as a result of sanctioning. Our empirical estimations for these cases are only meant to illustrate that, when the conditions of our model are not met, the data does not support the model's predictions either.

Table 1: Estimated Hazard when Target Acquiesces to Sender Demand

The Target Acquiesces to the Sender's Demand				
	Costly Sanctions Cases		***Cost Free Sanctions Cases***	
Explanatory Variables	Estimated Coefficient	t-statistic	Estimated Coefficient	t-statistic
CSENDER	0.9995**	2.10	– 3.7160	– 0.11
CTARGET	– 0.5254**	– 3.54	0.1306	0.26
ALLY	4.3734**	3.38	2.8304**	4.37
COOPERATION	– 0.8049*	– 1.95	– 2.8555	– 0.85
Constant	– 11.874**	– 3.28	– 6.4142*	– 1.72
Duration dependence parameter, p	3.3967**	4.35	2.4954**	5.30
Number of cases	14		19	
log Likelihood	– 6.1637**		– 13.9945**	
*p < 0.10, **p < 0.05 are two tailed significance levels.				

When the target acquiesces to the sender's demand, our model predicts that higher *sender* costs will increase the probability that the target will accept the sender's terms, ending the sanctions episode. The coefficient on CSENDER is indeed positive and significant for the costly sanctions cases of Table 1. Our estimations of the hazard rate also indicates that higher target costs decrease the probability that the target will acquiesce. This makes sense in the context of our model because the costs imposed on the target increase with the gap between the sender's demand and the *status quo*. The higher that gap, the less likely the target is to acquiesce to the sender's demand. Interestingly, the signs on CTARGET and CSENDER are reversed when the same relationship is estimated using the costless sanctions data, but the coefficients are not significant.

Our estimations also indicate that, in the costly sanctions cases, the target is more likely to acquiesce to the sender's demand if the sender and the target are allies (the coefficient on ALLY is positive and significant). This is compatible with Drezner's (1998, 1999) argument about future expectations of conflict. Drezner anticipates that "*ceteris paribus*, targets will concede more to allies than adversaries." An allied target should then also concede to the sender's demand with higher probability than an adversarial target. Our estimates suggest that, in those cases that end with target acquiescence, international cooperation is associated to longer sanctioning episodes (the coefficient on COOPERATION is negative and significant). Interestingly, Martin (1992) finds that the sender's willingness to absorb sanctioning costs is strongly associated to international cooperation. Here we find, indeed, that international cooperation is a significant factor in explaining the probability that the target will acquiesce to the sender's demand when the sender incurs costs. The fact that international cooperation is related to longer sanctions episodes is not incompatible with Martin's findings. In fact, Bolks and Al-Sowayel (2000), find that increased external hostility to the target during the sanctioning episode increases duration because it favors the development of countermeasures. International cooperation with the sender is indicative of increased hostility to the target.

In those cases where the target does not acquiesce to the sender's demand, either the sender backs down and accepts the *status quo*, or a compromise is reached and accepted by the sender. An estimated hazard function for the costly sanctions cases that end with compromise or a return to the *status quo* is presented in Table 2.[32] In some of these cases, the sanctioning episode was not over. They were therefore coded as right censored.[33] We estimate the hazard function found relevant to the costly sanctions cases, using the cost free sanctions cases. Again, we find that, when the conditions of our model are not met, the data does not support the model's predictions either.

Table 2: Estimated Hazard when Sanctions End with Compromise or Return to the *Status Quo*

Sanctions End with Compromise or Return to the *Status Quo*				
	Costly Sanctions Cases		***Cost Free Sanctions Cases***	
Explanatory Variables	Estimated Coefficient	t-statistic	Estimated Coefficient	t-statistic
CTARGET	0.4537**	2.73	0.06391	0.80
OUTCOME	– 0.2655**	– 2.00	– 0.0582	– 0.49
ASSIST	1.4528**	2.83	– 0.6856*	– 1.65
Constant	– 1.6041**	– 3.40	– 1.7685**	– 4.25
Duration dependence parameter, p	1.1704**	7.27	1.1438**	8.46
Number of cases	36		44	
log Likelihood	– 50.5587**		– 61.3933**	
*p < 0.10, **p < 0.5 are significance levels on two tailed tests.				

As our model predicts, the probability with which the sender agrees to a compromise or accepts the *status quo*, ending the sanctions, increases with the costs imposed *on the target.* Indeed, for the costly sanctions cases, the coefficient on CTARGET is positive and significant. It is not significant for the cost free cases. Our hazard estimation also indicates that the sender is more likely to accept a compromise or the *status quo* if the target state receives assistance during the sanctioning episode (the coefficient on ASSIST is positive). We also find that the hazard rate decreases with OUTCOME. OUTCOME adjusts Hufbauer, Schott and Elliott's coded assessment of the policy result for issue salience,[34] and is indicative of conciliatory movement on the part of the target. However, the compromise in many cases is reached at the end of the sanctioning episode without a formal counteroffer being on the table. As a result, the duration data does not reflect how long a compromise offer has been on the table. Consequently, we cannot expect a compromise outcome to be associated to an increased probability of acceptance by the sender *during* the sanctions episode. In fact, the conciliatory stance may actually occur after sanctions end, as in the US-

South Africa sanctions episode over nuclear safeguards. President Reagan authorized nuclear related exports, effectively ending sanctions, before South Africa began indicating that it would consider signing the nonproliferation treaty in the late 1980's.

We therefore find that the costly sanctions cases that end in compromise tend to last longer than those that do not. This is in line with Dashti-Gibson, Davis and Radcliff's (1997) finding that sanctions "success" is negatively correlated to duration. The longer the sanctions episode, the less likely are the protagonists to come to agreement. The same relationship between the extent of compromise and duration emerges if compromise solutions simply take longer to reach, as our estimations suggest. Such a relationship between duration and outcome is also proposed by Miyagawa (1992) who argues that, if the sanctioning episode is not quickly resolved with target acquiescence, time hardens the target's resolve. Our empirical estimate would support such an argument: hardening of the target's resolve would make compromise less likely, and compromise outcomes would take longer to reach.

Conclusion

Economic sanctions are often described as having two strikes against them: they are costly to the sanctioner, and they take time to achieve their goal, if they succeed at all. We argue in this paper that these are, instead, necessary characteristics of rational sanctioning. Sanctioning is an attempt to wear the target down into acquiescence through the constant application of costly pressure. But since no precise date is set for the target's acceptance, sanctions must be setting the stage for possible acquiescence at any time during the sanctioning episode. If acceptance of the sanctioner's terms is to be rational at all times, it must be neither better nor worse for the target than waiting further. Indeed, this logic cuts both ways: continued resistance by the target implies an expectation that the sender could give up its effort at any time. And this can only be rational for the sender if it is neither better nor worse than persisting in the imposition of sanctions. Strategies that maintain this balance between the payoffs of ending and continuing the struggle are referred to as countervailing.

We show that countervailing strategies form perfect equilibria whether the protagonists are fully informed or not. Indeed, the incomplete information case is hardly different from the complete information case, and need not yield any additional delay. However incomplete information, even if it is negligible, requires that the two sides adopt uncompromising bargaining positions and costly sanctioning behavior. Countervailing behavior, whether adopted by fully informed protagonists or not, must leave distinctive traces in the data. Indeed, if it is adopted, it is the costs *incurred by the sanctioner* that correlate with the *target*'s chances of acquiescence, *not* the costs incurred by the target. This insight runs counter to inferences made in many empirical studies of sanctioning data. Yet we find empirical support for the predictions of our model using the classic sanctions data base developed by Hufbauer, Schott and Elliott (1990a, 1990b).

Appendix

Formally, define the statement $A_i^t =$ "i successfully accepts j's offer *before* time t". Then, the cumulative distribution introduced in the text reads $F_i^t(s) = \mathbb{P}(A_i^{t+s}|h^t)$ the probability of A_i^{t+s} conditional in the prior history h^t.[35] It is important to note that F_i^t is an expression of the *two* sides' strategies since i could only accept if j hasn't yet done so. In general, a distribution function is only assumed non-decreasing and taking its values in $[0, 1]$. In particular, F_i^t can admit discontinuities that are interpreted as discrete probabilities of acceptance at the corresponding times. But an issue arises when both sides *simultaneously attempt* to accept each other's *distinct* offers. In order to avoid setting arbitration rules that would be debatable we simply assume that such attempts fail. But common discontinuities in F_i^t and F_j^t should be resolved accordingly.[36] However, we observe that if at least one side (say i) uses a *continuous* survival function ϕ_i^t then the probability of simultaneous acceptance attempts is nil. And since our existence results only involve continuous ϕ_i^t's we do not need to account for that difficulty. The expected utility of i accepting j's offer at "some future time" conditional in the prior history h^t thus reads:

$$\mathbb{E}_i(A_i^\infty) = \int_{[0,\infty)} \left(\int_0^s r_i e^{-r_i \tau} U_i(Y^t(\tau)) d\tau + U_i(X_j^t(s)) \right) dF_i^t(s) \qquad \text{(A1)}$$

While the term $\int_0^s r_i e^{-r_i \tau} U_i(Y^t(\tau)) d\tau$ is continuous in s, $U_i(X_j^t(s))$ is not necessarily so. And common discontinuities between $U_i(X_j^t(s))$ and $F_i^t(s)$ can raise some definition issues in (A1). But it is reasonable to assume that U_i is continuous and to restrict j to strategies such that X_j^t has only isolated discontinuities. In that case $U_i(X_j^t(s))$ is measurable and (A1) is well defined.[37]

$\mathbb{E}_i(A_j^\infty)$ as well as $\mathbb{E}_i(\neg(A_i^\infty \vee A_j^\infty))$ are similarly defined. We have

***Lemma 1*:** Formula (10) in the text defines i's objective at any time t.

***Proof*:** With the above definitions $E_i(\sigma^t) = \mathbb{E}_i(A_i^\infty) + \mathbb{E}_i(A_j^\infty) + \mathbb{E}_i(\neg(A_i^\infty \vee A_j^\infty))$. The two terms involving the factor $\int_0^s r_i e^{-r_i \tau} U_i(Y^t(\tau)) d\tau$ sum up to

$$\int_{[0,\infty)} \left(\int_0^s r_i e^{-r_i \tau} U_i(Y^t(\tau)) d\tau \right) \left(dF_i^t(s) + dF_j^t(s) \right) \qquad \text{(A2)}$$

$$= \Big(- H_{ij}^t(s)\int_0^s r_i e^{-r_i\tau} U_i(Y^t(\tau))d\tau\Big)\Big]_0^\infty + \int_0^\infty r_i e^{-r_i\tau} U_i(Y^t(\tau))H_{ij}^t(\tau)d\tau$$

after integrating by parts (which holds since $F_i^t(s)$ is monotonous and $\int_0^s r_i e^{-r_i\tau} U_i(Y^t(\tau))d\tau$ is continuous in s). But the first term on the right in (A2) reduces to $-\mathbb{E}_i(\neg(A_i^\infty \vee A_j^\infty))$. Adding up all three above terms in $E_i(\sigma^t)$ yields (10). □

***Lemma 2*:** At any time t, for any delay $\theta > 0$, and for any strategy for either side

$$E_i(\sigma^t) = \int_{[0,\theta)} e^{-r_i s} dG_{ij}^t(s) + e^{-r_i\theta} H_{ij}^t(\theta) E_i(\sigma^{t+\theta}) \qquad \text{(A3)}$$

***Proof*:** Since $F_i^t(s)$ and $F_j^t(s)$ are conditional probabilities

$$dF_i^t(\theta + s) = H_{ij}^t(\theta) dF_i^{t+\theta}(s)$$

and symmetrically for j. Similarly $H_{ij}^t(\theta + s) = H_{ij}^t(\theta) H_{ij}^{t+\theta}(s)$. So, since there can be no left-hand side discontinuity in F_i^t (or F_j^t or H^t) at θ

$$E_i(\sigma^t) = \int_{[0,\theta)} e^{-r_i s} dG_{ij}^t(s) + \int_{[\theta,\infty)} e^{-r_i s} dG_{ij}^t(s)$$
$$= \int_{[0,\theta)} e^{-r_i s} dG_{ij}^t(s) + e^{-r_i\theta} H_{ij}^t(\theta) \int_{[0,\infty)} e^{-r_i s} dG_{ij}^{t+\theta}(s) \qquad \square$$

Equation (A3) in Lemma 2 is often called a "Bellman equation". It simplifies the optimization of $E_i(\sigma^t)$ into that of the given integral over $[0,\theta)$ and the choice of an optimal θ assuming that $E_i(\sigma^{t+\theta})$ is similarly optimized. In formula (13) in the text, define formally the statement $a_j^t =$ "j has issued an acceptance before time t". Then let j's survival function be $\phi_j^t(s) = 1 - \mathbb{P}(a_j^{t+s} | h^t)$.

***Lemma 3*:** Assume that, for all $s \in [0,\tau)$, the players choose constant $X_i^t(s) \equiv X_i$, $X_j^t(s) \equiv X_j$, and $Y^t(s) \equiv Y$ and that j uses $\phi_j^t(s) = e^{-\lambda_i s}$ with $\lambda_i = r_i \frac{U_i(X_j) - U_i(Y)}{U_i(X_i) - U_i(X_j)}$. Then any survival strategy $\varphi_i^t(s)$ for player i and for $s \in [0,\tau)$ yields

$$\int_{[0,\tau)} e^{-r_i s} dG_{ij}^t(s) = \big(1 - e^{-r_i\tau} H_{ij}^t(\tau)\big) U_i(X_j) \qquad \text{(A4)}$$

***Proof*:** Since $dF_j^t(s) = -\varphi_i^t(s) d\phi_j^t(s)$ and similarly for $dF_i^t(s)$

$$dG_{ij}^t(s) = e^{-\lambda_i s}\Big(- U_i(X_j) d\varphi_i^t(s) + (\lambda_i U_i(X_i) + r_i U_i(Y))\varphi_i^t(s) ds\Big)$$

But integration by parts yields[38]

$$\int_{[0,\tau)} e^{-(r_i+\lambda_i)s} d\varphi_i^t(s) = e^{-(r_i+\lambda_i)s} \varphi_i^t(s)\Big]_0^{\tau^-} + (r_i + \lambda_i) \int_{[0,\tau)} e^{-(r_i+\lambda_i)s} \varphi_i^t(s) ds$$

and using the definition of λ_i

$$\int_{[0,\tau)} e^{-r_i s} dG_{ij}^t(s) = \big(1 - e^{-(r_i+\lambda_i)\tau}\varphi_i^t(\tau^-)\big)U_i(X_j)$$

But $e^{-\lambda_i\tau}\varphi_i^t(\tau^-) = H_{ij}^t(\tau)$ by left-continuity at τ. □

***Lemma 4*:** Assume that $(\theta_n)_{n\in\mathbb{N}}$ is the sequence of successive event times. If it is finite, complement it by an arbitrary sequence increasing to infinity. Then

$$\lim_{n\to\infty} e^{-r_i\theta_n} H_{ij}^t(\theta_n) = 0 \qquad \text{(A5)}$$

***Proof*:** If $\lim_{n\to\infty}\theta_n = \infty$ this is obvious. Otherwise the increasing and bounded sequence θ_n must converge to some limit Θ. But the history at time Θ would contain infinitely many events and must, by assumption, have probability $H_{ij}^t(\Theta) = 0$. And by left-continuity $\lim_{n\to\infty} H_{ij}^t(\theta_n) = 0$. □

***Proof of Theorem 1*:** Assume that j behaves according to ψ_j. If at time t i accepts X_j then $\sigma^t \equiv (X_j, X_j, X_j)$ and $E_i(\sigma^t) = U_i(X_j)$. Otherwise, for any path σ^t, any strategy η_i for i entails an increasing sequence of future event times $(t+\theta_n)_{n\in\mathbb{N}}$, when i adjusts his strategic variables, and corresponding delays $\tau_n = \theta_n - \theta_{n-1}$ (with $\theta_0 = 0$).[39] By Lemma 2 written with $\tau = \tau_1$ and replacing according to Lemma 3 in (A3)

$$\begin{aligned} E_i(\sigma^t) &= \big(1 - e^{-r_i\tau_1} H_{ij}^t(\tau_1)\big)U_i(X_j) + e^{-r_i\tau_1} H_{ij}^t(\tau_1) E_i(\sigma^{t+\tau_1}) \\ &= \big(1 - e^{-r_i\theta_1} H_{ij}^t(\theta_1)\big)U_i(X_j) + e^{-r_i\theta_1} H_{ij}^t(\theta_1) E_i(\sigma^{t+\theta_1}) \end{aligned}$$

And iterating, since $H_{ij}^t(\theta_1) H_{ij}^{t+\theta_1}(\tau_2) = H_{ij}^t(\theta_2)$, yields

$$\begin{aligned} E_i(\sigma^t) &= \big(1 - e^{-r_i\theta_2} H_{ij}^t(\theta_2)\big)U_i(X_j) + e^{-r_i\theta_2} H_{ij}^t(\theta_2) E_i(\sigma^{t+\theta_2}) \\ &= .. = \big(1 - e^{-r_i\theta_n} H_{ij}^t(\theta_n)\big)U_i(X_j) + e^{-r_i\theta_n} H_{ij}^t(\theta_n) E_i(\sigma^{t+\theta_n}) \end{aligned} \qquad \text{(A6)}$$

And by Lemma 4 (as $n \to \infty$) $E_i(\sigma^t) = U_i(X_j)$ independently of i's strategy. Therefore ψ_j is countervailing for i. But since any strategy η_i yields the same expectation $E_i(\sigma^t) = U_i(X_j)$ to i on any path σ^t, the countervailing ψ_i is trivially a best reply after any history and the pair Ψ forms a SPE. □

We now extend the above results to the incomplete information case. For each type $l \in \mathcal{I}$ consider the continuation probability $H_l^t(s) = 1 - F_l^t(s) - \sum_{k\in\mathcal{J}} b_k(t) F_k^t(s)$. Formula (A3) in Lemma 2 extends easily to

$$E_l(\sigma^t) = \sum_{k\in\mathcal{J}} b_k(t) \int_{[0,\theta)} e^{-r_l s} dG_{lk}^t(s) + e^{-r_l\theta} H_l^t(\theta) E_l(\sigma^{t+\theta}) \qquad \text{(A7)}$$

As they evolve according to Bayes' Law some beliefs may fall to zero. If and when that happens the development will be called an event. By (17) this can only happen if the survival probability of the corresponding type also falls to zero. With this understanding of events, Lemma 4 clearly extends to $\lim_{n\to\infty} e^{-r_l\theta_n} H_l^t(\theta_n) = 0$.

The extension of the concept of countervailing strategy revolves around the notion of "active" types, one for each side, who actively countervail each other while all non-active types find it best to wait until circumstances change.

Definition 1**:** The strategy profile[40] $\Psi_{\mathcal{J}}$ with active types $i \in \mathcal{I}$ and $j \in \mathcal{J}$ is defined at any time t, with $\lambda_i = r_i \frac{U_i(\Xi_{\mathcal{J}}) - U_i(\text{NE})}{U_i(\Xi_{\mathcal{I}}) - U_i(\Xi_{\mathcal{J}})}$, $\tau = \frac{-ln(1-b_j(t))}{\lambda_i}$, $b_j(t) \neq 0$, and all $s \in [0,\tau)$ by

- Offer constantly $\Xi_{\mathcal{J}}$ and accept instantly $\Xi_{\mathcal{J}}$;
- Play constantly $y_{\mathcal{J}}^{\text{NE}}$ of the common Nash equilibrium NE;
- Wait for an additional s with probability

$\phi_j^t(s) = \frac{e^{-\lambda_i s} + b_j(t) - 1}{b_j(t)}$ for active type j;

$\phi_k^t(s) \equiv 1$ for non-active $k \in \mathcal{J} - \{j\}$ such that $b_k(t) \neq 0$;

$\phi_k^t(s) \equiv 0$ for k such that $b_k(t) = 0$.[41]

Lemma 5**:** Assume that for all $s \in [0,\tau)$ the players choose constant $X_{\mathcal{I}}^t(s) \equiv X_{\mathcal{I}}$, $X_{\mathcal{J}}^t(s) \equiv \Xi_{\mathcal{J}}$, and $Y^t(s) \equiv Y = (y_{\mathcal{I}}, y_{\mathcal{J}}^{\text{NE}})$. If the types in $\mathcal{J}$ play according to the profile $\Psi_{\mathcal{J}}$ then any survival strategy φ_l^t for any type $l \in \mathcal{I}$ within time period $[0,\tau)$ yields

$$\sum_{k\in\mathcal{J}} b_k(t) \int_{[0,\tau)} e^{-r_l s} dG_{lk}^t(s) = (1 - e^{-r_l\tau} H_l^t(\tau)) U_l(X_{\mathcal{J}}) + \rho_{li}(X_{\mathcal{I}}, Y) \int_{[0,\tau)} e^{-(r_l+\lambda_i)s} \varphi_l^t(s) ds$$

with $\rho_{li}(X_{\mathcal{I}}, Y) = \lambda_i (U_l(X_{\mathcal{I}}) - U_l(\Xi_{\mathcal{J}})) - r_l (U_l(\Xi_{\mathcal{J}}) - U_l(Y))$.

***Proof*:** Since

$$dG^t_{lk}(s) = -U_l(\Xi_{\mathcal{J}})\phi^t_k(s)d\varphi^t_l(s) - U_l(X_{\mathcal{I}})\varphi^t_l(s)d\phi^t_k(s) + r_l U_l(Y)\varphi^t_l(s)\phi^t_k(s)ds$$

replacing according to $\Psi_{\mathcal{J}}$ yields

$$\sum_{k\in\mathcal{J}} b_k(t)dG^t_{lk}(s) = b_j(t)dG^t_{lj}(s) + \sum_{k\neq j} b_k(t)dG^t_{lk}(s) = e^{-\lambda_i s}\left(-U_l(\Xi_{\mathcal{J}})d\varphi^t_l(s) + (\lambda_i U_l(X_{\mathcal{I}}) + r_l U_l(Y))\varphi^t_l(s)ds\right) \quad \text{(A8)}$$

But, integrating by parts $\int_{[0,\tau)} e^{-r_l s}e^{-\lambda_i s}d\varphi^t_l(s)$ and replacing in the integration of (A8) yields the result. ☐

***Theorem 2*:** At time t let the active type in $\mathcal{I}$ be $i = \text{argmax}\{r_l \frac{U_l(\Xi_{\mathcal{J}})-U_l(\text{NE})}{U_l(\Xi_{\mathcal{I}})-U_l(\Xi_{\mathcal{J}})} \mid l \in \mathcal{I}, b_l(t) \neq 0\}$ and symmetrically for the active $j \in \mathcal{J}$, *mutatis mutandis*. Then the strategy profiles $\Psi_{\mathcal{I}}$ and $\Psi_{\mathcal{J}}$ defined above, together with beliefs according to (17), form a PBE.

***Proof*:** Assume that all types in $\mathcal{J}$ play according to $\Psi_{\mathcal{J}}$. Consider an arbitrary path σ^t and an arbitrary survival strategy $\varphi^t_l(s)$ for $s \in [0,\tau)$. Clearly, $\phi^t_{\mathcal{J}}(s) = e^{-\lambda_i s}$ and thus $H^t_l(s) = e^{-\lambda_i s}\varphi^t_l(s)$. By Lemma 5 and formula (A7)

$$E_l(\sigma^t) = U_l(\Xi_{\mathcal{J}}) + e^{-(r_l+\lambda_i)\tau}\varphi^t_l(\tau)(E_l(\sigma^{t+\tau}) - U_l(\Xi_{\mathcal{J}})) + \rho_{li}(X_{\mathcal{I}}, Y)\int_{[0,\tau)} e^{-(r_l+\lambda_i)s}\varphi^t_l(s)ds \quad \text{(A9)}$$

By its definition, and our incentive compatibility assumption, ρ_{li} is maximum for $X_{\mathcal{I}} = \Xi_{\mathcal{I}}$ and $Y = \text{NE}$. If $b_l(t) = 0$ then $\rho_{li}(\Xi_{\mathcal{I}}, \text{NE}) \leq 0$ for all present and future choices of i.[42] By (A9) for any such l and any future τ

$$E_l(\sigma^t) \leq U_l(\Xi_{\mathcal{J}}) + e^{-r_l\tau}(E_l(\sigma^{t+\tau}) - U_l(\Xi_{\mathcal{J}})) \leq ... \leq U_l(\Xi_{\mathcal{J}})$$

But since any l can always accept $\Xi_{\mathcal{J}}$ at any time $E_l(\sigma^t) \geq U_l(\Xi_{\mathcal{J}})$.[43] So, $E_l(\sigma^t) = U_l(\Xi_{\mathcal{J}})$ if $b_l(t) = 0$, and $\varphi^t_l(s) \equiv 0$ is optimal. Moreover, if $b_l(t) \neq 0$ then $\rho_{li}(\Xi_{\mathcal{I}}, \text{NE}) \geq 0$ by definition of i and $E_l(\sigma^t)$ is clearly maximum in (A9) when $\varphi^t_l(s) \equiv 1$.

Finally, since the maximum for $l = i$ occurs when $\rho_{ii}(\Xi_{\mathcal{I}}, \text{NE}) = 0$, (A9) yields

$$E_i(\sigma^t) = U_i(\Xi_{\mathcal{J}}) + e^{-(r_i+\lambda_i)\tau}\varphi^t_i(\tau)(E_i(\sigma^{t+\tau}) - U_i(\Xi_{\mathcal{J}})) \quad \text{(A10)}$$

But, by Bayes Law in (17), $b_i(t+\tau) > 0$ if and only if $\varphi_i^t(\tau) > 0$. So, i remains active for as long as it takes to reach $\varphi_i^t(\tau) = 0$.[44] Whether or not there exists such a τ, it follows that $E_i(\sigma^t) = U_i(\Xi_{\mathcal{J}})$ regardless of the choice φ_i^t. Therefore $\varphi_i^t(s) \equiv \phi_i^t(s)$ is as good as any other and is optimal. $\Psi_{\mathcal{I}}$ is therefore a best reply to $\Psi_{\mathcal{J}}$. □

One should note that target j only lasts for $\tau = \frac{-ln(1-b_j(t))}{\lambda_i}$ from time t on, at which point $\phi_j^t(\tau) = 0$ and therefore $b_j(t+\tau) = 0$. Also, since λ_i is chosen as the maximum of $r_l \frac{U_l(\Xi_{\mathcal{J}}) - U_l(\text{NE})}{U_l(\Xi_{\mathcal{I}}) - U_l(\Xi_{\mathcal{J}})}$ for $b_l(t) \neq 0$, as information is gained the coefficient λ_i^t *decreases* and the probability of acceptance decreases accordingly. So, incomplete information can actually *speed up* acceptance rather than slow it down as argued elsewhere.

References

Baldwin, David, 2000, "The Sanctions Debate and the Logic of Choice," *International Security*, Vol. 24, No. 3, Winter 1999-2000, pp. 80-107.

Baldwin, David, 1985, *Economic Statecraft*, Princeton University Press, Princeton, New Jersey.

Bolks, Sean and Dina Al-Sowayel, 2000, "How Long Do Economic Sanctions Last? Examining the Sanctions Process Through Duration," *Political Research Quarterly*, Vol. 53, No. 2, June, pp.241-265.

Cleves, Mario, William Gould and Roberto Gutierrez, 2004, *An Introduction to Survival Analysis Using Stata*, Stata Press, College Station, Texas.

Daoudi, M.S. and M.S. Dajani, 1983, *Economic Sanctions: Ideals and Experience*, Routledge and Kegan Paul.

Dashti-Gibson, Jaleh, Patricia Davis and Benjamin Radcliff, 1997, "On the Determinants of the Success of Economic Sanctions: An Empirical Analysis," *American Journal of Political Science*, Volume 41, No. 2, April, pp. 608-618.

Dorussen, Han and Jongryn Mo, 2001, "Ending Economic Sanctions: Audience Costs and Rent Seeking as Commitment Strategies," *The Journal of Conflict Resolution*, Vol. 45, No. 4, August, pp.395-426.

Doxey, Margaret, P. 1996, *International Sanctions in Contemporary Perspective*, 2nd Edition, St Martin's Press Inc, New York.

Dresher, Melvin, 1981, *The Mathematics of Games of Strategy: Theory and Applications*, London, Dover Publications.

Drezner, Daniel, 2003, "The Hidden hand of Economic Coercion," *International Organization*, Vol. 57, Summer, pp. 643-659.

Drezner, Daniel, 2000, "The Complex Causation of Sanctions Outcomes," in *Sanctions as Economic Statecraft*, Steve Chan and Cooper Drury, Editors, pp. 212-249, International Political Economy Series, St Martin's Press, New York.

Drezner, Daniel, 1999, *The Sanctions Paradox. Economic Statecraft and International Relations*, Cambridge University Press, Cambridge, England.

Drezner, Daniel, 1998, "Conflict Expectations and the Paradox of Economic Coercion," *International Studies Quarterly*, Vol. 42, pp. 709-731.

Eaton, Jonathan and Maxim Engers, 1999, "Sanctions, Some Simple Analytics," *The American Economic Review*, Vol. 89, No. 2, May pp. 409-414.

Eaton, Jonathan and Maxim Engers, 1992, "Sanctions," *The Journal of Political Economy*, Vol. 100, No. 5, October, pp. 899-928.

Greene, William, 1993, *Econometric Analysis*, 2nd Edition, Prentice Hall, New Jersey.

Helms, Jesse, 1999, "What Sanctions Epidemic? US Business' Curious Crusade," *Foreign Affairs*, Vol. 78 No. 1 January/February, pp. 2-8.

Hufbauer, Gary, Jeffrey Schott, and Kimberly Elliott, 1990a, *Economic Sanctions Reconsidered: History and Current Policy*, 2nd Edition, Washington DC: Institute for International Economics.

Hufbauer, G., J. Schott, and K. Elliott, 1990b, *Economic Sanctions Reconsidered: Supplemental Case Studies*, 2nd Edition, Washington DC: Institute for International Economics.

Kaempfer, William and Anton Lowenberg, 1988, "The Theory of Economic Sanctions: A Public Choice Approach," *The American Economic Review*, Vol. 78 No. 4, September, pp. 786-793.

Martin, Lisa, 1992, *Coercive Cooperation: Explaining Multilateral Economic Sanctions*," Princeton University Press, Princeton.

McGillivray, Fiona, and Allan Stam, 2004, "Political Institutions, Coercive Diplomacy, and the Duration of Economic Sanctions," *Journal of Conflict Resolution*, Vol. 48, No. 2, April, pp.154-172.

Miyagawa, Makio, 1992, *Do Economic Sanctions Work*, New York, St. Martin's Press.

Morgan, T. Clifton and Valerie Schwebach, 1997, "Fools Suffer Gladly: The Use of Economic Sanctions in International Crises," *International Studies Quarterly*, Vol. 41, pp. 27-50.

Pape, Robert A., 1997, "Why Economic Sanctions Do Not Work," *International Security*, Vol. 22, No. 2, Fall, pp. 90-136.

Schelling, Thomas, 1956, "An Essay on Bargaining," *The American Economic Review*, Volume 46, No. 3, June.

Smith, Alastair, 1996, "The Success and Use of Economic Sanctions," *International Interactions*, Vol. 21, January, pp.229-245.

Smith, Lones and Ennio Stacchetti, 2003, "Aspirational Bargaining," Unpublished Working Paper.

Tsebelis, George, 1990, "Are Sanctions Effective? A Game Theoretic Analysis," *The Journal of Conflict Resolution*, Vol. 34, No. 1, March, pp.3-28.

Wagner, R. Harrisson, 1988, "Economic Interdependence, Bargaining Power, and Political Influence," *International Organization*, Vol. 42, No 3, Summer, pp. 461-483.

Footnotes

[1]Thomas Schelling, 1956, p.282.

[2]Wagner (1988), Morgan and Schwebach (1997) are notable exceptions. Dorussen and Mo (2001) model the *end* of sanctions as the objective of interstate bargaining.

[3]Kaempfer and Lowenberg (1988) also develop a theoretical model of sanctioning behavior but theirs is a public choice approach that highlights the incentives of interest groups within the sanctioning and target polities.

[4]McGillivray and Stam (2004) posit that leadership change in sender or target states will impact sanctions duration if these states are non-democratic. The hazard model they estimate supports the claim. The duration of sanctions is also examined by Dorussen and Mo (2001) and Bolks and Al-Sowayel (2000). Bolks and Al-Sowayel (2000) point out that the duration of sanctions episodes should depends on the countermeasures deployed by the target and find, indeed, that factors that facilitate the development of countermeasures are associated to increased sanctions duration. Dorussen and Mo (2001) develop a model in which states bargain over the ending of sanctions. Their empirical analysis of sanctions duration emphasizes the role played by domestic politics and rent seeking in target and sender states.

[5]All theoretical models of sanctions do indeed assume that sanctions are costly to both sender and target.

[6]This means subgame perfect in the complete information case and perfect Bayesian in the incomplete information case. Authors such as Eaton and Engers (1992) or Smith (1996), who have modeled sanctioning behavior generally, conclude that if states are fully informed, extended sanctioning episodes should not be observed. Eaton and Engers do exhibit a case in which sanctions are imposed "fleetingly...(as a result) of indivisibilities in actions as well as in sanctions," (Eaton and Engers, 1992, p.919). But theoretical explanations of extended sanctioning rely on imperfect information. Smith, for example, shows that if sanctioning behavior is modeled in war of attrition terms with one-sided imperfect information, "the model's predictions vary markedly with different distributional assumptions" about the

target's type, (Smith,1996, p. 233). But Smith (1996) examines Bayesian Nash Equilibria (not perfect Bayesian equilibria) and does not allow any bargaining to take place. By contrast, the equilibria we will examine are perfect Bayesian, and arise in a context where the target can make a counteroffer to the sanctioner.

[7]Only very few authors have attempted to model bargaining in *unrestricted* continuous time although it is the only way to rid bargaining models of "temporal monopoly" or the assumption that each player has exclusive use of a period of time in which to make or accept an offer. Yet allowing the bargainors to intervene at any point in time brings bargaining theory closer to the real-world bargaining environment, endowing the theory with "strong behavioral foundations." This has been argued by Smith and Stachetti (2003) who, however, assume that any offer must be immediately accepted or rejected. So, there can be no standing offers. In our approach, there is no time monopoly either but we make no such restriction on bargainor behavior.

[8]For instance, the *expectation* of i accepting reads, replacing X by D and integrating (2) with density $-\phi_i(s)d\phi_j(s)$:

$$\int_0^\infty ((1-e^{-r_i s})U_i(\mathrm{S})+e^{-r_i s}(U_i(\mathrm{D})))\mu_i e^{-(\mu_i+\mu_j)s}ds$$
$$=\frac{\mu_i U_i(\mathrm{S})}{\mu_i+\mu_j}+\frac{\mu_i(U_i(\mathrm{D})-U_i(\mathrm{S}))}{r_i+\mu_i+\mu_j}$$

and similarly for the expectation of j's acceptance.

[9]Countervailing strategies do not require constant offers or sanction point. A more general framework will be described in our theoretical section.

[10]Our discussion focusses on the counteroffer O after it has been made. Of course, that counteroffer could be expected beforehand but this would only slightly modify the formulae as explained in Note 20.

[11]Bolks, Sean and Dina Al-Sowayel propose that sanctions duration will depend on the target's ability to implement countermeasures that dull the pain of the sanctions. In our

model, if the cost to the target is dulled by countermeasures it will decrease the probability that the sanctioner will give in. As such, it will impact sanctions duration.

[12]Indeed recall, by contrast, that in the case where one party defects in the classic prisoner's Dilemma, the other's defection in response is a clear rational choice since it makes him better off than if he suffered the sucker's payoff by continuing to cooperate.

[13]We assume that jumps in $Y(t)$ do not occur infinitely frequently. Such exotic behavior is possible in the time continuum but it is unrealistic in this context and would require unnecessarily sophisticated mathematics for its handling.

[14]In the most general case, the timing decision can be described by a distribution function which may or may not imply some randomization. In that case, we will assume that each player privately monitors the random device used in his strategy and that the devices are independent.

[15]Recall that all points on the Pareto frontier are such that in any move from such a point at least one side must be losing utility.

[16]Note that y_i and y_j need not have the same dimension. In our example y_i is two dimensional while y_j is one dimensional.

[17]Alternatives to countervailing strategies are strategies for which one party or the other is non-acceptant over intervals of time. This characterization goes beyond the scope of this paper and requires more complicated mathematical developments.

[18]As mentioned before, a countervailing strategy is not limited to constant offers and sanction points. For instance, if j is expected to switch from X_j to some Z_j at some future date $(t+\theta)$, where θ is a random variable of density $\nu_j e^{-\nu_j s}$, countervailing holds with

$$\lambda_i = \frac{r_i\big(U_i(X_j) - U_i(y_i^t(s), y_j^{\mathrm{S}})\big) - \nu_j(U_i(Z_j) - U_i(X_j))}{U_i(X_i^t(s)) - U_i(X_j)}$$

in (13) *before* the switch occurs. This new formula does not affect the basic relationships that are estimated in our empirical section. Also note that $y_i^t(s)$ in formula (13) is *not* necessarily i's sanction state strategy y_i^{S} and that $X_i^t(s)$ is not necessarily X_i.

[19]Recall that pooling strategies are non revealing about the type while separating strategies reveal the player's type

[20]Namely $\lambda_j = \max_{k \in \mathcal{J}} \{\lambda_k = r_k \frac{U_k(\Xi_{\mathcal{J}}) - U_k(\mathrm{NE})}{U_k(\Xi_{\mathcal{I}}) - U_k(\Xi_{\mathcal{J}})}; b_k(t) > 0\}$.

[21]All types k with nil beliefs $b_k(t) = 0$ are assumed in PBE to prefer to accept instantly with $\phi_k^t(s) \equiv 0$. But they play no role since they are not believed to be involved any more.

[22]We refer here to Harsanyi's work on the purification of mixed equilibria. Although the motives are different, the insight remains: uncertainty is always present even if seemingly negligible.

[23]We thank Daniel Drezner for generously sharing his data with us.

[24]Drezner also disagregates some of the cases to distinguish between what are essentially separate sanctioning episodes between a given pair of states (for example US vs Israel, 1956-1982). We adopted this disagregation.

[25]Hufbauer, Schott and Elliott (1990a,b) code sender costs on a scale of 1 to 4 where 1=net gain to sender, 2=little effect on sender, 3=modest welfare loss on sender, 4=major loss to sender. In all these cases, sender costs were either coded as 1=net gain to sender by Hufbauer, Schott and Elliott, or the cost to the sender as computed by Drezner was negative. In some cases, negative costs were calculated by Drezner although Hufbauer, Schott and Elliot coded sender costs as a 2. Our 53 cases include all cases coded 1 by Hufbauer *et al*, and those cases in which Drezner computed sender costs as being negative although Hufbauer *et al* coded sender costs as 2.

[26]It turns out that, in all these cases, sender costs were insignificant according to both Drezner's calculations and Hufbauer, Schott and Elliott's coding.

[27]The Hufbauer, Schott and Elliott data was updated to 1996 by Drezner. We coded those episodes that were ongoing as of 1996 as censored for estimation purposes.

[28]In practice compromise offers are made by the target in response to the imposition of sanctions by the sender. The compromise outcome eventually reached may be the result

of an interchange between sender and target, possibly blurring the issue of who accepts whose terms. However, our reading of the cases suggests that those costly sanctioning episodes that end in compromise do so because the sender accepts a compromise that is initiated by the target.

[29]All explanatory variables used except CSENDER and OUTCOME are the measures provided by Hufbauer, Schott and Elliott. ALLY codes prior relationship as follows:1=antagonistic relationship,2=neutral relationship,3=cordial relationship. ASSIST coded as 1 when another country extends significant economic or military assistance to the target. CTARGET is the cost of sanctions to the target as a percentage of GDP. COOPERATION is measured on a scale of 1 to 4 where 1=no cooperation, 2=minor cooperation, 3=modest cooperation, 4=significant cooperation. CSENDER is the cost to the sender as a percentage of GDP as estimated by Drezner (1998,1999). OUTCOME is measured as follows: Hufbauer *et al* provide a policy outcome score on a scale of 1 to 4, with 1=failed outcome, 2=unclear but possible positive outcome, 3=positive outcome, 4=successful outcome. We interpret a score of 1 as maintenance of the *status quo*, scores of 2 and 3 as compromise outcomes accepted by the sender, and a score of 4 as target acquiescence to the sender's demand. However, we agree with Drezner that these policy scores need to be adjusted for issue salience. Drezner (1998,1999) rates the importance of the issue at hand on a scale of 1 to 2. We multiplied policy scores of 2, 3 or 4 by the issue salience score to generate a coding of outcome that corresponds to the outcomes of our model.

[30]In the costly sanctions cases, for all explanatory variables included in our hazard estimations whose t-statistic fell below $|1.96|$, the hypothesis that the corresponding coefficient was 0 was rejected using a log likelihood ratio test.

[31]These are the cases in which the sender gains or incurs no costs and includes two cases in which the target incurs no cost, and the target acquiesces to the sender's demand.

[32]We also estimated a hazard function for the costly sanctions cases adding the following set of traditional explanatory variables from the Hufbauer *et al* data set : ALLY which codes the historical relationship between target and sender, DTARG=1 if the target is a democracy, HEALTH=which codes the stability of the target government, COOPERATION=degree of international cooperation with the sender and POWER=ln($\frac{\text{Sender GDP}}{\text{Target GDP}}$). However, using a log likelihood ratio test, we could not reject the hypothesis that the addition of these variables did not contribute significantly to the explanation of duration. Importantly addition of these variables did not compromise the signs or significance of CTARGET, OUTCOME and ALLY. We also estimated a wider hazard model for the 14 costly sanctions cases where the target acquiesces to the sender's demand. There again the enlarged model was rejected on the basis of a log likelihood test and the signs on the coefficients we are interested in did not change. However, the small number of observations for these cases made the addition of extra explanatory variables of dubious value.

[33] Right censored cases as of 1996 are: US-Iran, US-Syria and US-Iraq over support of terrorism. The US-Cuba sanctions case was eliminated because it was a clear outlier. The costs imposed by the sanctions on Cuba are many times higher, as a percentage of Cuban GDP, than the largest target cost in the data base. When the US-Cuba sanctions data is included, coefficient signs are reversed and significance levels drop. A data point that drives coefficient signs is a typical outlier and can legitimately be removed.

[34]Hufbauer Schotta and Elliott rate the policy result of the sanctioning episode on a scale of 1 to 4 with 1=failed outcome, 2=unclear but possible positive outcome, 3=positive outcome, 4=successful outcome.

[35]A distribution function is usually assumed right-continuous whereas F_i^t is here left-continuous because of the word "before" in the definition of A_i^t. This has no consequences on the standard theory of Lebesgue-Stieltjes integrals but it simplifies several of our arguments.

[36]If p_i^t and p_j^t denote the *discrete* probabilities that i and j issue acceptances at precisely time t then $p_i^t p_j^t$ is the probability of failure according to our assumption. The discontinuity $F_i^t(0^+) - F_i^t(0) = p_i^t(1 - p_j^t)$ is thus the probability that i *successfully* accepts, and $H_{ij}^t(0^+) - H_{ij}^t(0) = (1 - p_i^t)(1 - p_j^t) + p_i^t p_j^t$.

[37]Formally, we assume that X_j^t can have only finitely many discontinuities in any finite time interval. (A1) is then always well defined as a Lebesgue-Stieltjes integral.

[38]This holds despite a possible right-hand-side discontinuity of φ_i^t at $t = 0$.

[39]There may be finitely many such θ_n and even none at all. In that case, we simply add an arbitrary sequence of θ_n increasing to ∞ in this argument.

[40]We need to define one strategy ψ_k for each type $k \in \mathcal{J}$.

[41]Implicitly, $\phi_k^t(s) \equiv 0$ for $s > 0$ if $b_k(t) = 0$ but this is inconsequential for the expectations of types in $\mathcal{I}$.

[42]Because the successive λ_i form a *decreasing* sequence.

[43]Since $\phi_{\mathcal{J}}^t$ is continuous the probability of simultaneous acceptance attempts is nil.

[44]However, the active type $j \in \mathcal{J}$ changes when $\phi_j^t(\tau) = 0$.