Department of Economics
University of Southampton
Southampton SO17 1BJ
UK

Discussion Papers in Economics and Econometrics

LEADERSHIP AND SELF-ENFORCING INTERNATIONAL ENVIRONMENTAL AGREEMENTS WITH NON-NEGATIVE EMISSIONS

Santiago J Rubio
and
Alistair Ulph

No. 0211

This paper is available on our website
http://www/soton.ac.uk/~econweb/dp/dp02.html

Leadership and Self-enforcing International Environmental Agreements with Non-Negative Emissions*

Santiago J. Rubio† and Alistair Ulph
University of Valencia, Spain
University of Southampton, UK

9 January 2003

*Research of Santiago J. Rubio was supported by the *Ministerio de Ciencia y Tecnología* (Spanish Ministry of Science and Technology) under grant BEC2000-1432. This research was initiated while the first author was a visiting fellow at the Department of Economics of the University of Southampton. The paper was presented at the 57^{th} European Meeting of the Econometric Society, Venice, August 25-28, 2002 and at XXVII Symposium of Economic Analysis, Salamanca, December 12-14, 2002. We want to thank Begoña Casino for her help with the figures.

†Corresponding author. Department of Economic Analysis, University of Valencia, Edificio Departamental Oriental, Avda. de los Naranjos s/n, 46022 Valencia, Spain. E-mail address: Santiago. Rubio@uv.es

Abstract

A model widely used by economists to study self-enforcing international environmental agreements (IEAs) is the model of stable IEAs introduced by Carraro and Siniscalco (1991,1993), Hoel (1992) and Barrett (1994). The key results emerging from the linear-quadratic emissions game version of this model are that, if the members of the IEA act in a Cournot fashion with respect to non-signatories, then a stable IEA consists of no more than 3 countries irrespective of the number of countries affected. See Carraro and Siniscalco (1991) and Hoel (1992). However, if the signatories act as Stackelberg leaders with respect to non-signatories, then, depending on parameters, a stable IEA can consist of any number of signatories, including the grand coalition, although, crucially, the number of signatories is inversely related to the potential gains from cooperation. See Barrett (1994).

However in the early paper by Barrett, these results were derived using numerical simulations and also ignored the fact that emissions must be non-negative. Recent attempts to use analytical approaches and to explicitly recognise the non-negativity constraints (e.g. Diamantoudi and Sartzetakis (2002)) have suggested that, even with Stackelberg leadership, the number of signatories of a stable IEA may be very small - no more than four. The way such papers have dealt with non-negativity constraints is to restrict parameter values to ensure interior solutions for emissions. Not surprisingly, this restricts the number of signatories.

We argue that a more appropriate approach is to directly impose the non-negativity constraint on emissions and recognise that for some parameter values this will entail corner solutions. When this is done we show, analytically, that the key results from the literature go through. With Cournot behaviour, a stable IEA will consist of at most 2 countries; we show that corner solutions cannot be stable. However with Stackelberg behaviour there can be stable IEAs which are corner solutions, and we derive parameter values for which the grand coalition is stable, and parameter values for which only a two-country IEA is stable. The number of signatories is directly related to marginal environmental damage.

Keywords: international externalities, self-enforcing international environmental agreements, Cournot equilibrium, Stackelberg equilibrium, non-negative emissions constraints

JEL classification:C72, D62, F02, Q20

1 Introduction

Over the last two decades, one of the factors driving an increased sense of interdependence between countries is the need to tackle global environmental problems such as climate change, ozone depletion, loss of biological diversity amongst others. Tackling such problems requires some form of agreement between countries, and the Framework Convention on Climate Change, the Montreal Protocol on Substances that Deplete the Ozone Layer, and the Convention on Biodiversity are important examples of such International Environmental Agreements (IEAs). However, the very different experience of these agreements illustrates the crucial importance of understanding how to design agreements which give countries incentives to both join and abide by such agreements. Economists have emphasised two important features: agreements must be profitable, that is there must be potential gains to all signatory countries; more importantly, in the absence of any international authority, agreements must be self-enforcing, i.e. there must be incentives for countries acting in their own self-interest to want to join or stay in an agreement.

One of the earliest definitions of a self-enforcing agreement was the concept of a stable IEA, which means that no individual signatory country has any incentive to leave the IEA, and no non-signatory country has an incentive to join, taking as given the membership decisions of all other countries.[1] Models based on this concept include Carraro and Siniscalco (1991,1993), Hoel (1992), Barrett (1994), Na and Shin (1998) amongst many others. If one uses a linear-quadratic emissions game version of this model, then the main findings are rather pessimistic. Carraro and Siniscalco (1991) and Hoel (1992) have shown that if signatory countries act in Cournot fashion with respect to non-signatories, then a stable IEA consists of 3 countries when marginal environmental damage is constant (i.e., when the countries' best-replay functions are orthogonal), and of 2 countries when marginal damage increases with emissions (i.e., when the best-replay functions has a negative slope), in both cases irrespective of the number of countries affected.[2] If

[1] There are a number of other concepts of what makes an agreement self-enforcing. Chander and Tulkens (1997) draw on cooperative game concepts (the gamma core). Other concepts, such as far-sightedness, have been developed, for instance, by Ecchia and Mariotti (1997,1998); see Finus (2001) and Wagner (2001) for excellent recent overviews.

[2] Carraro and Siniscalco (1991,1993) also show that the number of signatory countries can be increased by means of self-financed transfers. However, expanding coalitions re-

they act in a Stackelberg fashion, then, depending on parameter values, a stable IEA can have any number of signatories between two and the grand coalition of all countries. But the gain in global welfare from the stable IEA relative to the non-cooperative outcome is inversely related to the number of signatories. See Barrett (1994).[3] The rationale for the difference in outcomes between Cournot and Stackelberg models is that if one country was to leave the IEA, with Cournot behaviour, the non-signatories expand their emissions and the remaining signatory countries partially accommodate this by reducing their emissions. On the other hand with Stackelberg behaviour, if a signatory was to leave the IEA the remaining signatories would expand their emissions. Thus the incentives to leave an IEA are greater with Cournot behaviour than with Stackelberg.

Since this model has become something of a workhorse tool to study IEAs, it is important that its properties are well understood. There are two, related, weaknesses in the early paper developed by Barrett (1994). First, it relied on numerical simulations to derive the main findings. Second, it ignored the need to ensure that emissions would be non-negative (or, equivalently, in an abatement game, that abatement did not exceed the unabated level of emissions).[4] Recent papers have attempted to correct these weaknesses and to evaluate how they could affect the results. Finus (2001) presents an analyti-

quires some form of commitment. Petrakis and Xepapadeas (1996) extend this result to the case in which the countries are not identical using an emissions game with orthogonal best-replay functions as the one studied by Hoel (1992). Hoel and Schneider (1997) point out that the prospect of receiving a transfer tends to reduce the incentive a country might have to commit itself to cooperation so that if the disincentive is strong, total emissions will be higher with side payments. More recently, Barrett (2001) has shown that with strong asymmetry side payments become the vehicle for increasing participation in a cooperative agreement.

[3]These results apply to the case with increasing marginal damage. For the case of constant marginal damage, it is easy to show that the Cournot and Stackelberg equilibria coincide so that leadership does not increase the level of cooperation attained by a Cournot-IEA.

[4]These weaknesses do not appear in Carraro and Siniscalco's (1991) paper. These author develop an analytical solution for a symmetric Cournot equilibrium. In their model they assume an environmental damage function that is quadratic with respect to the local emissions but linear with respect to the imported emissions, i.e., environmental damage depends on the product between local emissions and total emissions that affect the country. As a result of this specification the solution of the game yields always interior solutions. In this paper, we focus on a global environmental problem so that we assume that environmental damage is a quadratic function of the aggregate emissions.

cal generalization of Barrett's results and shows that the higher the level of environmental damages, the greater the size of the stable IEA. However, his proof in Appendix X.1 assumes interior solutions, and it is easy to show that high environmental damages imply that, unconstrained, emissions will become negative. Diamantoudi and Sartzetakis (2002) and Rubio and Casino (2001) also use analytical approaches, but recognise the need to ensure that emissions are non-negative. They reach even more pessimistic conclusions - that even with Stackelberg behaviour the number of signatories of a stable IEA will be small - no greater than four.

However the way Diamantoudi and Sartzetakis deal with the non-negativity constraint is to compute an interior solution and then restrict parameter values to ensure that the resulting emissions are always strictly positive. It is not surprising that this restriction on parameters restricts the number of signatories in a stable IEA. Rubio and Casino go further and restrict parameters to ensure that payoffs are non-negative, which is difficult to justify. We argue that neither of these approaches is appropriate. In this paper we also use an analytical approach and deal with the non-negativity constraint by simply imposing it directly on the choice of emissions by both signatory and non-signatory countries and using Kuhn-Tucker conditions to derive the equilibrium of the game. Then for some parameter values, the emission game will result in corner solutions. We show that with Cournot behaviour a stable IEA consists of no more than two countries. We also show that a stable IEA cannot involve a corner solution, and derive a range of parameter values for which there will be no stable IEA, because of the non-negative emissions constraints. This will occur when the marginal damage is high. In that case, the free-riding incentive is so strong that no agreement is stable. Thus our analysis extends the results obtained by Carraro and Siniscalco (1991) for the case of a global environmental problem characterized by a quadratic damage function that depends on aggregate emissions.

With Stackelberg behaviour we find the opposite results. A stable IEA can involve a corner solution. Our findings show that when the marginal damage is enough high the unique stable IEA is the grand coalition and that the number of countries in a stable IEA is directly related to the level of marginal damage so that when marginal damage is enough low a stable IEA consists of at most 3 countries. The rationale for this kind of relationship is given by the fact that the interdependence among the countries occurs through the damage function. Thus when the marginal environmental damage cost is relatively high, the countries in the agreement choose emission levels which induce the non-signatories to select low emissions, making exit from the agreement unprofitable. With lower marginal environmental damage cost, these effects are weakened, so that some countries find it profitable to leave the agreement, i.e. the free-riding cannot be avoided by the leadership. Finally, we clarify the previous results in the literature which have been derived assuming interior solutions. According to our results restricting parameter values to guarantee interior solutions is a sufficient condition to get stable IEAs with a small number of signatories but it is not a necessary condition. In this paper we show that a stable IEA with a small number of countries can involve a corner solution.

In section 2 we present the basic model of an international emissions game, solve for the cooperative and non-cooperative equilibria and introduce the definition of a stable international environmental agreement. In section 3 we derive our stability results for Cournot behaviour by signatory countries. In section 4 we derive the stability results for Stackelberg behaviour. We compare the results of the Stackelberg model with and without non-negativity constraints, and show that imposing the non-negativity constraint makes a trivial difference to the size of IEA that might be expected to form. Section 5 concludes.

Thus we have shown in this paper that the results derived for the model of stable IEAs in paper by Barrett (1994), which used numerical calculations on the linear-quadratic version of the model and ignored the issue of non-negative emissions, all carry through when derived analytically in a model which takes seriously the need to ensure that emissions are non-negative. The reason why taking account of non-negative emissions does not change the main results of the literature is that, as we shall show, the definition of a stable agreement depends on the sign of the difference between payoffs to signatories and non-signatories as the number of signatories varies. Taking account of the need for emissions to be non-negative affects the value of these payoffs, but not the sign of differences in payoffs. Thus this paper not only derives analytically results for the Barrett (1994) model using an appropriate treatment of non-negative emission constraints, but disproves the claim that taking account of such constraints makes a significant difference to known results.

2 An International Emissions Game

2.1 The Basic Model

In this section we present the basic linear-quadratic model of an international emissions game.[5] There are N identical countries, $i = 1, ..., N$. We define $q_i \geq 0$ as the level of emissions generated by country i, $Q_i \equiv \sum_{k \neq i} q_k$ the total emissions generated by all countries other than i, and $Q = \sum_k q_k = Q_i + q_i$ as the total emissions generated by all N countries. Each country derives a gross benefit from its emissions (think of the economic benefits of burning fossil fuels) denoted: $B(q_i) \equiv \alpha q_i - (\beta/2)q_i^2$. Each country also suffers environmental damage which depends on the global level of emissions, and the damage cost function for each country is denoted: $C(Q) \equiv (\gamma/2)Q^2$. Then each country has a net benefit(payoff) function:

$$\pi(q_i, Q_i) \equiv \alpha q_i - \frac{\beta}{2}q_i^2 - \frac{\gamma}{2}(q_i + Q_i)^2.$$

We assume that $\alpha > 0, \beta > 0$ and $\gamma > 0$. It should be clear that w.l.o.g. we can normalise one of the parameters and we choose to normalise by setting $\gamma = 1$. To emphasise this normalisation we rewrite the net benefit function as:

$$\pi(q_i, Q_i) \equiv aq_i - \frac{b}{2}q_i^2 - \frac{1}{2}(q_i + Q_i)^2. \tag{1}$$

We shall think of b as β/γ - the ratio of the (absolute) slope of the marginal benefit curve and the slope of the marginal damage cost curve, so a low value of b is to be interpreted as a (relatively) high marginal damage cost.

2.2 Cooperative and Non-Cooperative Outcomes

When all countries cooperate, emissions for each country are chosen to maximize aggregate net payoffs. As is well known this requires that emissions for each country are chosen so that the marginal benefit it derives from an extra

[5] In this model, as in the early papers discussed in the introduction, we deal with a flow pollutant. For analysis using a stock pollutant see Rubio and Ulph (2002a, b). Note also that while we work with an emission game with no abatement, it is trivial to show that this is equivalent to a model of an abatement game with given level of unabated emissions.

unit of emissions equals the additional damage cost it imposes on all countries. It is straightforward to check that the cooperative levels of emissions and per country payoff are given by:

$$q^* \equiv \frac{a}{b+N^2}; \qquad \pi^* \equiv \frac{a^2}{2(b+N^2)}. \tag{2}$$

When countries act non-cooperatively, each country takes as given the emissions set by other countries, and chooses its own emissions to maximize its own net benefit. It will set its emissions so that the marginal benefit it derives from an extra unit of emissions equals the additional damage cost it imposes on itself. It is straightforward to check that the non-cooperative levels of emissions and per country payoff are given by:

$$\bar{q} \equiv \frac{a}{b+N}; \qquad \bar{\pi} \equiv \frac{a^2[b-N(N-2)]}{2(b+N)^2}. \tag{3}$$

As expected, $\bar{q} > q^*, \bar{\pi} < \pi^*$, so non-cooperative behaviour leads to higher emissions and lower payoffs than cooperative behaviour. Note also that emissions are strictly positive in both cases, so we have not worry about non-negativity constraints.

Finally we define the gains to full cooperation by:

$$G \equiv \frac{\pi^* - \bar{\pi}}{\bar{\pi}} = \frac{N^2(N-1)^2}{(b+N^2)[b-N(N-2)]}. \tag{4}$$

It is clear that the gains from cooperation are greater the smaller is b, i.e. the greater are marginal damage costs (relative to the (absolute) slope of the marginal benefit curve). This makes sense - the more damaging is global pollution the greater are the gains from cooperating to deal with it.

However, as noted in the introduction, it is not sufficient to show that countries will be better off if they cooperate than if they do not cooperate (the profitability issue). One also needs to ensure that entering an agreement is in the interest of the countries who do so- an agreement must be self-enforcing. In this paper we use the notion of *stability* of an IEA as our concept of an IEA being self-enforcing. We define this formally in the next subsection.

2.3 Stable International Environmental Agreements

We model an International Environmental Agreement as a two-stage game, in which in the first stage (the *Membership Game*) each country decides whether or not to join an IEA, and in the second stage (the *Emissions Game*) each country determines its emissions. We describe each game briefly, in reverse order.

The Emissions Game

Suppose that, as the outcome of the first-stage game, there are n signatory countries (a typical signatory being denoted by s) and $N - n$ non-signatory countries (a typical non-signatory being denoted by f, for fringe or free-rider). Non-signatory countries choose emissions in the same way that countries did when acting non-cooperatively - each country takes as given the emissions of all other countries and chooses its emissions to maximize its own net benefits. Using symmetry, this will define a *non-signatory reaction function* linking the emissions of a typical non-signatory country to the emissions of a typical signatory country. Signatory countries choose emissions in the same way countries did when they acted cooperatively - the emissions of each signatory are chosen to maximize the aggregate payoff of the n signatories.

Two issues need further clarification. First, there is the question of the timing, or really commitment, of the emissions of signatories relative to non-signatories. If neither signatories nor non-signatories can commit to their emissions, the we think of them choosing emissions simultaneously. Thus we have a *Cournot* model in which signatories take as given the emissions of non-signatories when choosing their emissions. This will define a *signatory reaction function* relating the emissions of a typical signatory country to the emissions of typical non-signatory. Equilibrium is the intersection of the two reaction functions. On the other hand, Barrett (1994) argues that membership of an IEA acts as a form of commitment device, which we model by thinking of signatories setting their emissions before non-signatories. In that case signatories can calculate what emissions non-signatories will choose (in terms of the non-signatory reaction function), and choose their own emissions to maximize aggregate net benefit. This is the *Stackelberg* model.

The second issue is that emissions by signatories and non-signatories have to be non-negative and we just impose these as constraints on the decision problems of signatories and non-signatories. So for some parameter values the emissions of either a signatory or non-signatory could be zero (a corner solution).

The outcome of the emission game then, is that, for any number of signatories n we can define the equilibrium payoffs to signatory and non-signatory countries: $\pi^s(n), \pi^f(n)$.

The Membership Game

We follow Hoel (1992), Carraro and Siniscalco (1993), Barrett (1994) and others in saying that an IEA is self-enforcing if it is stable, where the concept of stability is borrowed from the literature on cartel stability (d'Aspremont *et al* (1983)). For $2 \leq n \leq N$ we define $\Delta(n) = \pi^s(n) - \pi^f(n-1)$; then:

Definition 1 *An IEA with n signatories is* ***stable*** *if it satisfies the conditions: Internal Stability:* $\Delta(n) \geq 0$, *i.e.* $\pi^s(n) \geq \pi^f(n-1)$; *External Stability:* $\Delta(n+1) \leq 0$, *i.e.*$\pi^f(n) \geq \pi^s(n+1)$.

Internal stability simply means that any signatory country is at least as well off staying in the IEA as quitting, assuming that all other countries do not change their membership decisions. External stability similarly requires that any non-signatory is at least as well off remaining a non-signatory than joining the IEA, again assuming that all other countries do not change their membership decision.

We can also think of a stable IEA as a Nash equilibrium of the Membership Game. To see this, the strategies for each country in the Membership Game are to sign or not sign. A country takes as given the membership decisions of all other countries. Suppose these have resulted in a membership of $m,\ 0 \leq m \leq N-1$. Then the payoffs to a country are $\pi^s(m+1)$ if it signs and $\pi^f(m)$ if it does not. So it will join if $\pi^s(m+1) \geq \pi^f(m)$ and not join otherwise. For an IEA with n^* members to constitute a Nash equilibrium of the Membership Game, it must have paid each signatory to sign, so $\pi^s(n^*) \geq \pi^f(n^*-1)$. Similarly it must have paid each non-signatory not to join, so $\pi^f(n^*) \geq \pi^s(n^*+1)$. These are just the conditions for Internal and External Stability.

In this next section we analyse stable IEAs for the Cournot model with non-negative emissions, and in section 4 we analyse stable IEAs for the Stackelberg model with non-negative emissions.

3 Stable Cournot IEAs with Non-Negative Emissions

We now turn to the analysis of stable IEAs with Cournot behaviour and non-negative emissions. First we need to analyse the outcome of the emissions game.

3.1 Cournot Emissions Game with Non-Negative Emissions

Suppose there are n signatory countries and $N-n$ non-signatories. A non-signatory country k takes as given Q_{fk} and chooses q_{fk} to solve:

$$\max_{q_{fk}\geq 0}\ \pi_k^f = aq_{fk} - \frac{b}{2}q_{fk}^2 - \frac{1}{2}(q_{fk}+Q_{fk})^2.$$

The first order condition is:

$$\frac{\partial \pi_k^f}{\partial q_{fk}} = a - bq_{fk} - (q_{fk}+Q_{fk}) \leq 0,\ q_{fk}\geq 0,\ \ q_{fk}\frac{\partial \pi_k^f}{\partial q_f k} = 0. \tag{5}$$

(5) defines the *non-signatory reaction function* for a non-signatory country, $k = 1, ..., N-n$, allowing for the fact that emissions must be non-negative, so part of the reaction function has $q_{fk} = 0$.

Now signatories are assumed to coordinate in order to maximize their collective net benefits taking as given the emissions of non-signatories.

$$\max_{q_{s1},\dots,q_{sn}\geq 0}\ \Pi^s = \sum_{i=1}^{n}\pi_i^s = \sum_{i=1}^{n}[aq_{si} - \frac{b}{2}q_{si}^2 - \frac{1}{2}(q_{si}+Q_{si})^2].$$

The first-order condition is:

$$\frac{\partial \Pi^i}{\partial q_{si}} = a - bq_{si} - (q_{si}+Q_{si}) - \sum_{j\neq i}(q_{sj}+Q_{sj}) \leq 0,\ q_{si}\geq 0,\ \ q_{si}\frac{\partial \Pi^i}{\partial q_{si}} = 0,\ i = 1, ..., n. \tag{6}$$

(6) defines the *signatory reaction function* for a typical signatory country, allowing for the fact that emissions must be non-negative, so part of the reaction function has $q_{si} = 0$. Under the assumption of symmetry we have that $q_{f1} = ... = q_{f(N-n)} = q_f$, $q_{s1} = ... = q_{sn} = q_s$ and $Q = nq_s + (N-n)q_f$, so that (5) and (6) can be written as follows:

$$\frac{\partial \pi^f}{\partial q_f} = a - bq_f - [nq_s + (N-n)q_f] \leq 0,\ q_f \geq 0,\ \ q_f\frac{\partial \pi^f}{\partial q_f} = 0. \tag{7}$$

$$\frac{\partial \Pi^i}{\partial q_s} = a - bq_s - n[nq_s + (N-n)q_f] \leq 0,\ q_s \geq 0,\ \ q_s\frac{\partial \Pi^i}{\partial q_s} = 0. \tag{8}$$

Equilibrium in the emissions game involves solving (7) and (8) simultaneously, taking account of the non-negativity constraints. In principle there are three possibilities:

(i) **Interior solution** ($q_s > 0, q_f > 0$). From (7) and (8) we obtain:

$$q_s = \frac{a[b - (N-n)(n-1)]}{b[b + N + n^2 - n]}; \qquad q_f = \frac{a[b + n(n-1)]}{b[b + N + n^2 - n]},$$

so that $q_s > 0$ iff $b > b(n) = (N-n)(n-1)$.

(ii) **Signatory Corner Solution** ($q_s = 0, q_f > 0$).From (7) and (8) this requires:

$$q_f = \frac{a}{b + N - n}; \qquad a - n(N-n)q_f \leq 0,$$

iff $b \leq b(n)$.

(iii) **Non-Signatory Corner Solution** ($q_s > 0, q_f = 0$). From (7) and (8) this requires:

$$q_s = \frac{a}{b + n^2}; \qquad a - nq_s \leq 0$$

which implies that $b + n(n-1) \leq 0$ contradicting $b > 0$.

It is easy to show that a corner solution for both types of countries does not satisfy the first order conditions, so that the only two possible equilibria are the interior solution and the non-signatory corner solution, which are summarized in:

Lemma 1 *For any n, the equilibrium of the Cournot emissions game with non-negative emissions will depend on the value of the parameter b; the equilibrium outputs and net benefits for signatory and non-signatory countries are:*

(i) *Interior Solution* $\qquad (b > b(n))$

$$q_s = \frac{a\left[b - b(n)\right]}{b[b + N + n^2 - n]}; \quad q_f = \frac{a\left[b + n\left(n-1\right)\right]}{b[b + N + n^2 - n]},$$

$$\pi^s(n) = \frac{a^2}{2b}\left\{1 - \frac{N^2(b + n^2)}{[b + N + n^2 - n]^2}\right\}; \; \pi^f(n) = \frac{a^2}{2b}\left\{1 - \frac{N^2(1+b)}{[b + N + n^2 - n]^2}\right\}$$

(ii) *Corner Solution* $\qquad (b \leq b(n))$

$$q_s = 0; \quad q_f = \frac{a}{b+N-n},$$
$$\pi^s(n) = -\frac{a^2(N-n)^2}{2(b+N-n)^2}, \quad \pi_f(n) = \frac{a^2[b+1-(N-n-1)^2]}{2(b+N-n)^2}.$$

Note that for $n = 1$ and $n = N$, $b(n) = 0$, so, since $b > 0$, we always get an interior solution for these values of n. Indeed it is straightforward to see that when $n = 1$ or $n = 0$, $q_s = q_f = a/(b+N) = \bar{q}$, the emissions level in the non-cooperative equilibrium defined in Section 2.2, while if $n = N$ the emissions level for signatory countries is $q_s = a((b+N^2) = q^*$, the emissions level in the cooperative equilibrium defined in Section 2.2. In these two cases we have an interior solution for this reason we focus in the rest of this Section on $n = \{2, 3, ..., N-1\}$.

Now for a particular problem b is a fixed parameter, and what we are interested in is, for a given N, b for which values of n will we have an interior solution or a signatory corner solution. Note that $b(n) = (N-n)(n-1)$ is a quadratic concave function with $b(n) = b(N-n+1)$ and has a maximum $b(\hat{n}) = (N-1)^2/4$ for $\hat{n} = (N+1)/2$.[6] This maximum is an integer if N is an odd number. However, when N is an even number, the integers that maximize $b(n)$ are $\hat{n}_1 = N/2$ and $\hat{n}_2 = (N+2)/2$. Then there exists a unique integer, $\hat{n}$, that maximizes $b(n)$ if N is an odd number, and two integers, $\hat{n}_1$and $\hat{n}_2$, if N is an even number so that we can represent by $b(\hat{n})$ the maximum value of the function $b(n)$ given by an integer in domain $\{2,3,...,N-1\}$.

For $b \leq b(\hat{n})$, define $x_1 \leq x_2$ as positive roots of $b = b(n)$ and $\bar{n}_1 = I_1(x_1)$ as the smallest integer no less than x_1 and $\bar{n}_2 = I_2(x_2)$ as the biggest integer not greater than x_2.[7] The situation is shown in Figure 1 for the case when $x_1 = \bar{n}_1$ and $x_2 = \bar{n}_2$ and N is an odd number.

$\Rightarrow$ FIGURE 1 $\Leftarrow$

Then according to Lemma 1 we have:

[6]In order to study the behaviour of $b(n)$ we assume that n is a real number in interval $[1, N]$ and once we know the properties of $b(n)$ then we are able to characterize the values of $b(n)$ for n restricted to be an integer.

[7]If N is an odd number then $b(\hat{n}) = b(n)$ yields $\bar{n}_1 = x_1 = \bar{n}_2 = x_2 = (N+1)/2$. Instead if N is an even number then $b(\hat{n}) = b(n)$ yields $\bar{n}_1 = x_1 = N/2$ and $\bar{n}_2 = x_2 = (N+2)/2$.

Proposition 1 *For any b, N there exists a unique solution to the Cournot emissions game with non-negative emissions as follows: (i) If $b > b(\hat{n})$ we have that for all n, $2 \leq n \leq N-1$, the equilibrium will be an interior solution; (ii) If $b \leq b(\hat{n})$, there exist $\bar{n}_1, \bar{n}_2$ defined above such that the equilibrium will be an interior solution for n such that $2 \leq n < \bar{n}_1$, $\bar{n}_2 < n \leq N-1$ and a signatory corner solution for n when $\bar{n}_1 \leq n \leq \bar{n}_2$.*

Note that $b(2) = b(N-1) = N-2$ and so a corollary is:

Corollary 1 *If $0 < b \leq N-2$, $\bar{n}_1 = 2$ and $\bar{n}_2 = N-1$ so the only interior solutions are $n = 1$ and $n = N$.*

From Lemma 1 and Proposition 2, for any parameters b and N we can determine for all possible values of n what type of equilibrium there will be for the emissions game and the corresponding equilibrium payoffs to signatory and non-signatory countries $\pi^s(n)$, $\pi^f(n)$. We now turn to the equilibrium of the membership game.

3.2 Cournot Membership Game with Non-Negative Emissions

In section 2.3 we defined an equilibrium of the membership game as a stable IEA, where stability was defined using the properties of the function $\Delta(n) = \pi^s(n) - \pi^f(n-1)$. Clearly the value of $\Delta(n)$ depends on the nature of the emissions game equilibrium for n and $n-1$. We have the following result:

Proposition 2 *(i) For the Cournot membership game with non-negative emissions, an IEA with n signatories is stable only if for both n and $n-1$ the equilibrium of the emissions game is an interior solution; (ii) If $b \geq \underline{b}(N) = N-4+2(N^2-3N+3)^{1/2} > N-2$ the unique stable IEA of the Cournot membership game with non-negative emissions has membership 2; (iii) If $b < \underline{b}(N)$ there is no stable IEA of the Cournot membership game with non-negative emissions.*

The proof is provided in Appendix A.

Now from Corollary 1 and part (i) of Proposition 3, if $b \leq N-2$ the only values of n which yield interior solutions are $n = 1$ and $n = N$, so, because of the existence of corner solutions, a necessary condition for the existence of a stable IEA stable is $b > N - 2$. However from part (ii) of Proposition the critical minimum value of b for which there exists a unique stable IEA, with membership 2, is $\underline{b}(N) > N - 2$. So consideration of the possibility of corner solutions adds no additional constraint to the conditions under which an IEA of size 2 is stable.

The intuition behind Proposition 2 is essentially that given in the introduction. With Cournot behaviour, if one country leaves an IEA and expands its emissions, the remaining members of the IEA respond by cutting their emissions. This effectively "rewards" defection, and so no IEA with at least 3 members is internally stable. It is only when the IEA is driven down to membership of 2, and so any further defection by one country effectively dissolves the IEA, that there is a possibility of stabilizing an IEA, but only for a large enough value of b, i.e. a low enough (relative) value of marginal damage costs. Neither of these arguments is affected by the fact that for some values of n there may be a corner solution with signatory emissions constrained to zero. This does not affect the incentives to defect when $n = 3$, and it does not affect the critical size range of parameter values for b for which $n = 2$ is internally stable.

4 Stable Stackelberg IEAs with Non-Negative Emissions

In this section we analyse stable IEAs when signatory countries act collectively as a Stackelberg leader and emissions are restricted to be non-negative. We begin with the emissions game.

4.1 Stackelberg Emissions Game with Non-Negative Emissions

Suppose there are n signatories and $N - n$ non-signatories. Each non-signatory acts in the same way as in the Cournot model set out in 3.1 with reaction function given by (7). However, the countries in the agreement act as the leader of the game so that given that they are identical they coordinate for the same level of emissions in order to maximize their net benefits taking into account the reaction function of the followers:

$$\max_{q_s \geq 0,\ q_f \geq 0} n\pi_s = n[aq_s - \frac{b}{2}q_s^2 - \frac{1}{2}(nq_s + (N-n)q_f)^2]$$

$$s.t. \quad -a + bq_f + nq_s + (N-n)q_f \geq 0, \tag{9}$$

The Lagrangean for the problem is

$$\begin{aligned} L &= n[aq_s - \frac{b}{2}q_s^2 - \frac{1}{2}(nq_s + (N-n)q_f)^2] \\ &\quad + \lambda\left(-a + bq_f + nq_s + (N-n)q_f\right), \end{aligned}$$

and the KTCs are

$$\frac{\partial L}{\partial q_s} = n[a - bq_s - n(nq_s + (N-n)q_f) + \lambda] \leq 0, \tag{10}$$

$$q_s \geq 0, \quad q_s\frac{\partial L}{\partial q_s} = 0,$$

$$\frac{\partial L}{\partial q_f} = -n(N-n)(nq_s + (N-n)q_f) + \lambda(b + N - n) \leq 0, \tag{11}$$

$$q_f \geq 0, \quad q_f\frac{\partial L}{\partial q_f} = 0,$$

$$\frac{\partial L}{\partial \lambda} = -a + bq_f + nq_s + (N-n)q_f \geq 0, \tag{12}$$

$$\lambda \geq 0, \quad \lambda\frac{\partial L}{\partial \lambda} = 0.$$

Equilibrium in the emissions game involves solving (10)-(12) simultaneously, taking account of non-negativity constraints. In principle there are three possibilities:

(i) **Interior solution** ($q_s > 0, q_f > 0$). From (10)-(12) we obtain:

$$q_s = \frac{a\left[b^2 - (N-n)(n-2)b + (N-n)^2\right]}{b[(b+N-n)^2 + bn^2]}, \tag{13}$$

$$q_f = \frac{a\left[b^2 + (N+n^2-2n)b - (N-n)n\right]}{b[(b+N-n)^2 + bn^2]}, \tag{14}$$

so that $q_s > 0, q_f > 0$ iff

$$\begin{aligned} g(b,n) &= b^2 - (N-n)(n-2)b + (N-n)^2 > 0, \\ h(b,n) &= b^2 + (N+n^2-2n)b - (N-n)n > 0. \end{aligned}$$

(ii) **Signatory Corner Solution** ($q_s = 0, q_f > 0$). From (10)-(12) this requires:

$$q_f = \frac{a}{1 + c(N-n)}, \quad g(b,n) \leq 0.$$

(iii) **Non-Signatory Corner Solution** ($q_s > 0, q_f = 0$). From (10)-(12) this requires:

$$q_s = \frac{a}{n}, \quad h(b,n) \leq 0.$$

It is easy to show that for the Stackelberg equilibrium a solution $q_s = q_f = 0$ does not satisfy the KTCs.

Summarizing:

Lemma 2 *For any n, the equilibrium of the Stackelberg emissions game with non-negative emissions will depend on the value of the parameter b; the equilibrium outputs and net benefits for signatory and non-signatory countries are:*

(i) Interior Solution $\qquad (g(b,n), h(b,n) > 0)$

$$q_s = \frac{ag(b,n)}{b[(b+N-n)^2 + bn^2]}, \quad q_f = \frac{ah(b,n)}{b[(b+N-n)^2 + bn^2]}.$$

with net benefits given by

$$\begin{aligned} \pi^s(n) &= \frac{a^2}{2b}\left\{1 - \frac{N^2 b}{(b+N-n)^2 + bn^2}\right\}, \\ \pi^f(n) &= \frac{a^2}{2b}\left\{1 - \frac{(b+1)N^2(b+N-n)^2}{[(b+N-n)^2 + bn^2]^2}\right\}. \end{aligned}$$

(ii) Signatory Corner Solution $\qquad (g(b,n) \leq 0,\ h(b,n) > 0)$

$$q_s = 0, \quad q_f = \frac{a}{b+N-n}$$

with net benefits given by

$$\pi^s(n) = -\frac{a^2(N-n)^2}{2(b+N-n)^2}, \quad \pi^f(n) = \frac{a^2[b - (N-n)(N-n-2)]}{2(b+N-n)^2}.$$

(iii) Non-Signatory Corner Solution $\qquad (g(b,n) > 0,\ h(b,n) \leq 0)$

$$q_s = \frac{a}{n}, \quad q_f = 0$$

with net benefits given by

$$\pi^s(n) = -\frac{a^2(b + n(n-2))}{2n^2}, \quad \pi^f(n) = -\frac{a^2}{2}.$$

Now note that for the interior solution the full-cooperative level of emissions is given by Eq. (13) for $n = N$ and that the full-noncooperative Cournot level of emissions is given by Eq. (14) for $n = 0$. In these two cases we have an interior solution, for this reason we focus in the rest of this Section on $n = \{1, 2, ..., N-1\}$.

We now want to determine more precisely for which parameter values the three different solutions occur. This clearly depends on the signs of $g(b, n)$ and $h(b, n)$. The next proposition fixes n and considers for which values of b we get each of the three solutions, and shows that there are no values of b and n for which both $g(b, n) \leq 0$ and $h(b, n) \leq 0$. For given n, $h(b, n) = 0$ presents a unique positive root that we denote by $b_1(n)$, $g(h, n)$ is always positive for $b > 0$ if $n = \{1, 2, 3\}$, if $n = 4$ $g(b, 4)$ is zero for $b = N - 4$ and positive otherwise, and if $n > 4$ $g(b, n) = 0$ presents two positive roots that we denote by $b_2(n)$, $b_3(n)$ such that $b_1(n) < b_2(n) < b_3(n)$. Then we have:

Proposition 3 *For any n, there exists a unique solution to the Stackelberg emissions game with non-negative emissions as follows: (i) for $n = \{1, 2, 3\}$, there exists $b_1(n)$ defined above such that for $b \leq b_1(n)$ the equilibrium is the non-signatory corner solution while for $b > b_1(n)$ the equilibrium is the interior solution; (ii) For $n = 4$, we have that $b_1(4) < b = N - 4$ so that: (a) for $b \leq b_1(4)$ the equilibrium is the non-signatory corner solution, (b) for $b_1(4) < b < N - 4$ the solution is the interior solution, (c) for $b = N - 4$ the equilibrium is the signatory corner solution, (d) for $b > N-4$ the equilibrium is the interior solution; (iii) For $4 < n < N$, we have that $b_1(n) < b_2(n) < b_3(n)$ so that : (a) for $b \leq b_1(n)$ the equilibrium is the non-signatories corner solution, (b) for $b_1(n) < b < b_2(n)$ the equilibrium is the interior solution, (c) for $b_2(n) \leq b \leq b_3(n)$ the equilibrium is the signatories corner solution, and (d) for $b > b_3(n)$ the solution is the interior solution.*

The proof is in Appendix B. As with the Cournot model, we are more interested in the question, for a given b what is the solution to emissions game for different values of n.

First we focus on *non-signatories*. The non-signatories' emissions depend on whether b is greater or less than $b_1(n)$ so that we need to know how $b_1(n)$

changes with respect to n in order to establish whether the equilibrium will be an interior solution or a corner solution for non-signatories. For $b_1(n)$ we have that $b_1(0) = b_1(N) = 0$ and it is easy to show that the function presents a unique extreme in the interior of the interval $(1, N-1)$ which is a maximum.[8] Then there exists a unique integer, $\hat{n}$, in that interval that maximizes $b_1(n)$ so that $b_1(\hat{n})$ is the maximum value of the function given by an integer in the domain $\{1, 2, ..., N-1\}$.[9]

For $b \leq b_1(\hat{n})$, define $x_1 \leq x_2$ as positive roots of $b = b_1(n)$ and $\bar{n}_1 = I_1(x_1)$ as the smallest integer no less than x_1 and $\bar{n}_2 = I_2(x_2)$ as the biggest integer not greater than x_2.[10] In this case we obtain the following results:

Lemma 3 *(i) If $b > b_1(\hat{n})$ we have that for all n, $1 \leq n \leq N-1$, the equilibrium will be an interior solution; (ii) If $b \leq \hat{b}_1(\hat{n})$, there exist $\bar{n}_1, \bar{n}_2$ defined above such that the equilibrium will be an interior solution for n such that $1 \leq n < \bar{n}_1, \bar{n}_2 < n \leq N-1$ and a corner solution for n when $\bar{n}_1 \leq n \leq \bar{n}_2$.*

The proof is provided in Appendix C.

Moreover, as $b_1(N-1)$ is the minimum value of $b_1(n)$ for $n = \{1, 2, ..., N-1\}$ we can establish that:

Corollary 2 *If $0 < b \leq b_1(N-1)$, $\bar{n}_1 = 1$ and $\bar{n}_2 = N-1$ so the only interior solutions are $n = 0$ and $n = N$. In other words, the equilibrium is a corner solution for non-signatories for all n, $1 \leq n \leq N-1$.*

The proof is provided in Appendix D.

For *signatories* emissions depend on whether b belongs to the interval $[b_2(n), b_3(n)]$. $b_2(n)$ is a strictly convex, decreasing function defined in the

[8] To study the behaviour of $b_1(n)$ in that interval we do as we did for the Cournot equilibrium we assume that n is a real number and once we know the properties of $b_1(n)$ then we are able to characterize the values of $b_1(n)$ with respect to n but now with n restricted to be an integer number. The same approach is followed to study $b_2(n)$ and $b_3(n)$.

[9] In the Cournot model there existed the possibility of two maxima given the symmetry of the function $b(n)$, however as $b_1(n)$ does not present this property there is no reason to expect that the maximum of the function be reached for two integers.

[10] If $\hat{n}$ in the domain $\{1, 2, ..., N-1\}$ maximizes $b_1(n)$ in the real interval $[1, N-1]$ then $b_1(\hat{n}) = b_1(n)$ yields $x_1 = x_2 = \hat{n}$. However, if the value of n that maximizes $b_1(n)$ is not an integer number then $b_1(\hat{n}) = b_1(n)$ yields $x_1 < x_2$ and one of these two values is equal to $\hat{n}$.

interval $[4, N]$ with $b_2(4) = N-4$ and $b_2(N) = 0$. Then for $n = \{4, 5, ..., N-1\}$ the maximum value of the function is $N - 4$ and the minimum value is $b_2(N-1)$.

For $b_2(N-1) \leq b \leq N-4$, define x_3 as the unique positive root of $b = b_2(n)$ and $\bar{n}_3 = I_3(x_3)$ as the smallest integer no less than x_3.

On the other hand, $b_3(n)$ is a strictly concave function with a maximum value equal to $N(N-4)/4$ in interval $(4, N)$ with $b_3(4) = N-4$ and $b_3(N) = 0$. Then there exists a unique integer, n^*, that maximizes $b_3(n)$. We denote by $b_3(n^*)$ the maximum value of this function given by an integer in the domain $\{4, 5, ..., N-1\}$.

For $N - 4 < b \leq b_3(n^*)$, defines $x_3 \leq x_4$ as the positive roots of $b = b_3(n)$ and $\bar{n}_3 = I_3(x_3)$ as the smallest integer no less than x_3 and $\bar{n}_4 = I_4(x_4)$ as the biggest integer not greater than x_4. For $b_2(N-1) \leq b \leq N-4$, define x_4 as the unique positive root of $b = b_3(n)$ and $\bar{n}_4 = I_4(x_4)$ as above. Then we have:

Lemma 4 *(i) If $b > b_3(n^*)$ we have that for all n, $1 \leq n \leq N-1$, the equilibrium will be an interior solution;(ii) If $b_2(N-1) \leq b \leq b_3(n^*)$, there exist $\bar{n}_3, \bar{n}_4$ defined above depending on whether b is greater or less than $N-4$ such that the equilibrium will be an interior solution for n when $1 \leq n < \bar{n}_3$, $\bar{n}_4 < n \leq N-1$ and a signatory corner solution for n when $\bar{n}_3 \leq n \leq \bar{n}_4$;(iii) If $0 < b < b_2(N-1)$, we have that for all n, $1 \leq n \leq N-1$, the equilibrium will be an interior solution.*

The proof is provided in Appendix E and we illustrate it in Figure 2.

⇒ FIGURE 2 ⇐

Now we can establish the type of solution for each kind of country depending on the number of countries in the agreement using Lemmas 3 and 4.

Proposition 4 *(i) If $b > b_3(n^*)$ we have that for all $n, 1 \leq n \leq N-1$, the equilibrium will be an interior solution for signatories and non-signatories; (ii) If $b_1(\hat{n}) < b \leq b_3(n^*)$, the equilibrium will be an interior solution for non-signatories for all n, however for signatories there exist $\bar{n}_3, \bar{n}_4$ defined above such that the equilibrium will be an interior solution for n when $1 \leq n < \bar{n}_3$, $\bar{n}_4 < n \leq N-1$ and a corner solution when $\bar{n}_3 \leq n \leq \bar{n}_4$;(iii) If $b_2(N-1) \leq b \leq b_1(\hat{n})$, there exist $\bar{n}_1, \bar{n}_2, \bar{n}_3$ and $\bar{n}_4$ defined above such* *that the equilibrium will be an interior solution for non-signatories and for n when $1 \leq n < \bar{n}_1, \bar{n}_2 < n \leq N-1$ and a non-signatory corner solution for n when $\bar{n}_1 \leq n \leq \bar{n}_2$, moreover the equilibrium will be an interior solution for signatories and for n when $1 \leq n < \bar{n}_3$, $\bar{n}_4 < n \leq N-1$ and a signatory corner solution for n when $\bar{n}_3 \leq n \leq \bar{n}_4$;(iv) If $b_1(N-1) < b < b_2(N-1)$, the equilibrium will be an interior solution for signatories for all n and a corner solution for non-signatories also for all n except for $n = N-1$;(v) If* $0 < b \leq b_1(N-1)$, *the equilibrium will be an interior solution for signatories for all n and a corner solution for non-signatories also for all n.*

The proof is in Appendix F and we illustrate it in Figure 3.[11]

$\Rightarrow$ FIGURE 3 $\Leftarrow$

As can be seen, Proposition 4 just expresses the results of Proposition 3 in terms of n as a function of b, rather than b as a function of n. The previous results apply for $N > 5$ although with minimal changes they are also valid for $N = 5$.

Thus for any parameters N and b, Proposition 4 indicates for, any number of signatories n, what type of solution there is to the Stackelberg emissions game, with corner solutions to take account of non-negative emissions constraints, and Lemma 2 indicates the corresponding outputs and equilibrium payoff functions for signatories and non-signatories. This is what we need to know to conduct the stability analysis.

4.2 Membership Game for Stackelberg Model with Non-Negative Emissions

In this Section we show that the scope of the international cooperation for controlling an environmental problem depends critically on the level of the marginal environmental damage. We begin analyzing the stability for (relatively) high marginal damages, i.e., for low values of b.

Proposition 5 *If $b \leq b_1(N-1)$ the unique stable IEA of the Stackelberg model with non-negative emissions is the grand coalition.*

[11]Notice that as $b_2(n)$ is a decreasing, strictly convex function and $b_2(N) = b_1(N) = 0$, $b_1(n)$ must be also a strictly convex function for big enough values of n.

The proof is in Appendix G. But the intuition is simply that, as we know from Proposition 4, for this range of parameter values, for $n = N$ the equilibrium of the emissions game is the cooperative equilibrium, in which all signatories get a positive payoff, while for $n < N$ the equilibrium of the emissions game is the non-signatories corner solution, in which, from Lemma 2, there is a fixed total of emissions independent of the number of signatories. Thus, for all $n < N$ non-signatories will get a negative payoff which is less than the payoff to signatories (since signatories get the benefit of producing emissions), so it does not pay a signatory to leave the grand coalition, and it always pays non-signatories to join.

Next we show that the grand coalition cannot be a stable agreement for lower values of damage costs.

Proposition 6 *If $b \in [b_2(N-2), b_2(4) = N-4]$, there exists an upper bound given by $\bar{n}_3$ for the number of countries that belong to a self-enforcing IEA. This upper bound decreases when b increases.*

The proof is in Appendix H. This result establishes that the scope of cooperation is very sensitive to changes in the level of marginal environmental damage. So that we have to expect that a reduction in the marginal damage leads to a reduction in the level of cooperation reached by a self-enforcing IEA. The explanation for this kind of relationship is given by the fact that the interdependence among the countries occurs through the damage function. Thus, when the marginal environmental damage is relatively high (a low b), the leadership of the countries in the agreement is strong and the signatories choose emission levels which induce non-signatories to select low values of emission, making exit from the agreement unprofitable. These effects are weakened as environmental damage costs get smaller.

Finally, we focus on the scope of cooperation when $b > N - 4$.

Proposition 7 *There exists a critical value $\bar{b}$ higher than $N-4$ such that if $b \in [N-4, \bar{b}]$, then an IEA of three countries is self-enforcing. If $b > \bar{b}$, only two countries can sign a self-enforcing IEA.*

See Appendix I for the proof and the definition of the critical value $\bar{b}$. Comparing $N(N-4)/4$ and $\bar{b}$ we obtain that $N(N-4)/4 < \bar{b}$ for N in interval

(5,26) and that the relationship is the contrary for $N \geq 26$.[12] Consequently we can conclude that:

Corollary 3 *If $b > N(N-4)/4$, i.e., when the Stackelberg equilibrium is an interior solution for signatories and non-signatories, the number of countries in a self-enforcing IEA is three if N is lower than 26 and two if N is greater than or equal to 26.*

These conclusions clarify the previous results in the literature which have been derived assuming that there are interior solutions. According to our results restricting parameter values to guarantee interior solutions is a sufficient condition to get stable IEAs with a small number of signatories but it is not a necessary condition. We have obtained that it is enough with $b > N - 4$ to have a maximum of three countries in an IEA. But between $N - 4$ and $N(N-4)/4$ the Stackelberg equilibrium is a corner solution for different values of n depending on the value of b. This means that what is necessary and sufficient to get a small degree of cooperation is a high value of b and not interior solutions for signatories and non-signatories.

Thus we have shown that even if we take seriously non-negative emission constraints, the Stackelberg model can have stable IEAs as large as the grand coalition and as low as a bilateral agreement depending on the value of the marginal environmental damage.

Finally, although we have shown that, allowing for non-negative emission constraints, it is still possible to get the grand coalition as a stable IEA, it could still be the case that imposing non-negative emissions has a significant effect on the size of a stable IEA in the sense that for any particular set of parameter values the size of IEA is significantly smaller than would be calculated if one simply ignored the constraints. To test this we have taken values of $a = 1000$, values of $N = 10, 20, , 150$, and 1500 values of b. For each set of parameter values we calculated the size of the stable IEA imposing non-negative emission constraints and without imposing such constraints. Three points emerged: (i) first, we confirmed, that, for all N, by varying b the maximum size of stable IEA obtained was the grand coalition, whether or not the non-negative emission constraints were imposed; (ii) for any set of parameter values, the size of the stable IEA with the non-negative emissions

[12] Remember that $N(N-4)/4$ is the maximum value for $b_3(n)$. This means that $b > N(N-4)/4$ is a sufficient condition to have an interior solution for signatories and non-signatories.

constraints imposed was never greater than the size of stable IEA when the constraints were ignored; (iii), but crucially, the differences in size of stable IEA were small. We illustrate this in Table 1 by showing for a range of values of N the average size of stable IEA (averaged over different values of b) with and without the constraints. As can be seen the difference in average size by imposing the constraints is tiny. One has to go to the third decimal place to detect a difference in average size.

Table 1
Average Size of Stable IEA
With and Without Constraints

N	With Constraints	Without Constraints
10	6.268	6.272
30	16.677	16.680
50	27.148	27.151
70	37.629	37.631
90	48.110	48.111
110	58.589	58.590
130	69.068	69.069
150	79.545	79.546

So, as with the Cournot model, introducing non-negative emissions constraints does not significantly change the results that exist in the literature.

5 Conclusions

In this paper we have provided analytical proofs of the main results of the linear-quadratic version of the widely used model of stable IEAs introduced by Carraro and Siniscalco (1991) and Barrett (1994). Morover, we have shown, analytically, that these results are robust to the introduction of constraints that emissions must be non-negative. While such constraints significantly complicate the analysis, they leave the main findings of the original literature almost completely unaffected. Since it is clearly right that such non-negative constraints should be taken into account, it is important to know the original results are robust and this is one of the contributions of our paper. Another contribution is that our results disprove the claim by Diamantoudi and Sartzetakis (2002) that imposing non-negative emissions constraints significantly reduces the size of the stable IEAs that can be found with non-negative emissions. This paper argues that these claims are wrong and arise from the inappropriate way non-negative emissions constraints were taken into account.

Of course there are many other respects in which the original models of stable IEAs need to be extended - richer concepts of self-enforcing agreements, asymmetric countries, concepts of fairness, dealing with stock pollutants, allowing for uncertainty and learning - and the authors of the original papers and many others have made important contributions to addressing these extensions (see again Finus (2001) for an excellent survey). We too have addressed some of these extensions (Rubio and Ulph (2002 a,b), Ulph (2002 a, b)). However, since the basic model continues to attract interest, it is important to make sure that its properties rest on thorough analysis, and this paper contributes to that purpose.

A Proof of Proposition 2

(i) Stability depends on the value of the function $\Delta(n)$, which as noted in the text, depends on whether there are interior or corner solutions for n and $n-1$. We show that unless n and $n-1$ are both interior solutions, then $\Delta(n) < 0$, and so does not satisfy the condition for internal stability.

(a) Both Corner Solutions

From Lemma 1 we have that for this case

$$\Delta(n) = -\frac{a^2}{2}\left[\frac{(N-n)^2}{(b+N-n)^2} + \frac{b+1}{(b+N-n+1)^2} - \frac{(N-n)^2}{(b+N-n+1)^2}\right] < 0,$$

since the first and third term of parenthesis have the same numerator but the denominator of the third is greater.

(b) n corner solution, $n-1$ interior solution (i.e. $n = \bar{n}_1$)

For this case n is defined in the interval $[2, (N+1)/2]$. Now we are going to compare the increase in benefits with the increase in costs when one country leaves the agreement. This increase is equal to the benefits of non-signatories since for a corner solution the emissions of a signatory are zero.

$$B^f(n-1) = aq_f(n-1) - \frac{b}{2}q_f(n-1)^2 = \frac{a^2\left\{[b+A(n)]^2 + 2N[b+A(n)]\right\}}{2b[b+N+A(n)]^2},$$

where

$$A(n) = n^2 - 3n + 2 \geq 0 \; for \; all \; n \geq 1$$

On the other hand, the increase in costs is given by

$$C^f(n-1) - C^s(n) = \frac{a^2}{2}\frac{N^2(b+N-n)^2 - (N-n)^2[b+N+A(n)]^2}{[b+N+A(n)]^2(b+N-n)^2},$$

so that the increase in net benefits is given by

$$-\Delta(n) = \frac{a^2}{2b}\frac{b^4 + B(n)b^3 + C(n)b^2 + D(n)b + E(n)}{[b+N+A(n)]^2(b+N-n)^2},$$

where

$$B(n) = 2N + n(n-2) + 2A(n) > 0 \; for \; n \geq 2$$

$$\begin{aligned} C(n) &= (N-n)(5N-(2N+1)n) + A(n)[2(N-n)(2+N-n) \\ &\quad + A(n) + 2N] \end{aligned}$$

$$D(n) = 2N(N-n)^2 + A(n)(N-n)\left[2(2N-n) + 2N(N-n) + A(n)(N-n+2)\right] > 0 \; for \; n \in [2, (N+1)/2]$$

$$E(n) = A(n)(N-n)^2[A(n) + 2N] \geq 0 \; for \; n \geq 2.$$

Next, we check the sign of $C(n)$. Expanding this term and substituting $A(n)$ we obtain

$$\begin{aligned} C(n) &= 3n^4 - 4(4+N)n^3 + 2(N^2+10N+15)n^2 \\ &\quad -(8N^2+32N+20)n + 9N^2 + 12N + 4. \end{aligned}$$

It can be easily shown that $C''(n)$ is a strictly convex function with a minimum at $n = (4+N)/3$. For this minimum we have that the second derivative is positive for $N \geq 3$.[13] This implies that $C'(n)$ is increasing for all n in interval [2,$(N+1)/2$] and that moreover the first derivative is positive in this interval since $C'(2)$ is positive. From this result we can conclude that $C(n)$ is increasing and moreover that is positive since $C(2)$ is positive. Thus, the final conclusion is that $C(n)$ is positive in interval $[2,(N+1)/2]$ and consequently the internal stability condition is not satisfy when n is a corner solution and $n-1$ is an interior solution.

(c) <u>n interior solution, $n-1$ corner solution</u> (i.e. $n = \bar{n}_2$)

For this case n is defined in the interval $((N+1)/2, N]$. As in the previous case we compare the increase in benefits with the increase in costs when one country leaves the agreement. The benefits for signatories are

$$B^s(n) = aq_s(n) - \frac{b}{2}q_s(n)^2 = \frac{a^2\left[(b+N+n^2-n)^2 - N^2n^2\right]}{2b(b+N+n^2-n)^2},$$

and for the non-signatory countries

$$B^f(n-1) = aq_f(n-1) - \frac{b}{2}q_f(n-1)^2 = \frac{a^2\left[b+2(N-n+1)\right]}{2(b+N-n+1)^2},$$

so that the increase in benefits is

[13] We assume in this analysis that the minimum number of countries is 3.

$$B^f(n{-}1){-}B^s(n) = \frac{a^2}{2}\frac{N^2n^2(b+N-n+1)^2-(N-n+1)^2(b+N+n^2-n)^2}{b(b+N-n+1)^2(b+N+n^2-n)^2}.$$

On the other hand, the increase in costs is given by

$$C^f(n{-}1){-}C^s(n) = \frac{a^2}{2}\frac{(N-n+1)^2(b+N+n^2-n)^2-N^2(b+N-n+1)^2}{(b+N-n+1)^2(b+N+n^2-n)^2},$$

so that the increase in net benefits is given by

$$-\Delta(n) = \frac{a^2}{2b}\frac{F(n)b^3+G(n)b^2+H(n)b+I(n)}{(b+N-n+1)^2(b+N+n^2-n)^2},$$

where

$$\begin{aligned}
F(n) &= N^2-(N-n+1)^2, \\
G(n) &= N^2n^2+(N-n+1)^2+2N^2(N-n+1)-2(N-n+1)(N+n^2-n), \\
H(n) &= N^2(N-n+1)^2-(N-n+1)^2(N+n^2-n)^2+2N^2n^2(N-n+1) \\
&\quad -2(N-n+1)^2(N+n^2-n), \\
I(n) &= (N-n+1)^2(N^2n^2-(N+n^2-n)^2).
\end{aligned}$$

Next, we check the sign of coefficients $F(n)$, $G(n), H(n)$ and $I(n)$. $F(n)$ is positive for all $n \geq 2$ since then $N^2 > (N-n+1)^2$.

Expanding $G(n)$ we obtain

$$G(n) = -2n^4+n^3(4N+6)-n^2(7+N^2+10N)+n(4+10N+4N^2)-1-3N^2-4N.$$

It can be easily shown that $G''(n)$ is a strictly concave function with a maximum for $n = (3+2N)/4$. For this maximum we have that $G''((3+2N)/4)$ is positive for $N \geq 3$. Moreover we also have that $G''((N+1)/2)$ is positive and $G''(N)$ is positive for $N = \{3,4,5,6\}$, zero for $N = 7$ and negative for $N > 7$. Let us assume for the moment that $N > 7$, then we can conclude that

$$n_1 < \frac{N+1}{2} < \frac{3+2N}{4} < n_2 < N,$$

where $n_1 < n_2$ are the two positive roots of equation $G''(n) = 0$. Thus, we can establish that the first derivative is increasing from $(N+1)/2$ until n_2 where the first derivative presents a maximum and afterwards decreasing. Moreover as $G'((N+1)/2)$ is positive and $G'(N)$ is also positive we can establish that the first derivative is positive in the interval $((N+1)/2, N]$. This means that $G(n)$ is increasing in such interval, then as $G((N+1)/2)$ is positive we can conclude that $G(n)$ is positive in the interval $((N+1)/2, N]$. If $N = 7$ then $n_2 = N$ but again the first derivative is positive in the interval $((N+1)/2, N]$. Finally, if $N = \{3, 4, 5, 6\}$ we have that $N < n_2$ and the sign of the first derivative does not change. The result is that $G(n)$ is positive in the interval $((N+1)/2, N]$.

Expanding $H(N)$ we obtain

$$\begin{aligned} H(n) &= -n^6 + n^5(4+2N) - n^4(8+N^2+8N) + n^3(10+4N^2+16N) \\ &\quad -n^2(9N^2+7+18+n(10N^2+2+10N+2N^3) \\ &\quad -2N - 2N^3 - 4N^2 \end{aligned}$$

In this case $H(n)$ is positive around $(N+1)/2$ if $N \in [3, 8]$ and presents positive values near of N for $N \geq 3$. However, for $N > 8$ we cannot establish the sign of this coefficient.

Finally, we expand $I(N)$ obtaining

$$I(n) = (N-n+1)^2(-n^4 + 2n^3 + n^2(N^2 - 2N - 1) + 2Nn - N^2),$$

so that the sign of this coefficient depends on the sign of the second term on the right-hand side.

It is easy to show that the second derivative of this term is a strictly concave function with a positive maximum for $n = 1/2$.As the second derivative is negative for $(N+1)/2$ and N we can conclude that the first derivative is decreasing in the interval $((N+1)/2, N]$. Calculating the sign of the first derivative for these two values we find that the first derivative is positive for $(N+1)/2$ and negative for N which implies that the first derivative is zero for some value n_1 between $(N+1)/2$ and N and consequently we have a maximum for the term for n_1.Then as the term is positive for $(N+1)/2$ and zero for N we can conclude that $I(n)$ is positive or zero in the interval $((N+1)/2, N]$.

The result is that the sign of the internal stability condition depends on the sign of the following function of b :

$$f(b) = F(n)b^3 + G(n)b^2 + H(n)b + I(n),$$

where $F(n)$, $G(n)$ are positive and $I(n)$ is positive or zero for $n \in ((N+1)/2, N]$ and $H(n)$ can be positive or negative. When $H(n)$ is positive we have that the internal stability condition is not satisfy when n is an interior solution and $n-1$ is a corner solution. When $H(n)$ is negative we have that the second derivative is a strictly convex function with a minimum for a negative value of b. This implies that the first derivative is increasing for b positive and as $H(0)$ is negative the first derivative intersects the horizontal axis for a unique positive value $\hat{b}$. This value is a minimum for $f(b)$ so that the function is increasing on the right of this value.

On the other hand if n is an interior solution and $n-1$ is a corner solution, b must belong to the following interval

$$(N-n)(n-1) < b \leq (N-n+1)(n-2) \tag{15}$$

Remember that $(N-n)(n-1)$ is lower than $(N-n+1)(n-2)$ when n belongs to the interval $((N+1)/2, N]$. Then if $(N-n)(n-1)$ is bigger than $\hat{b}$ and $f((N-n)(n-1))$ is positive we can conclude that $f(b)$ is positive for (15) since then $f(b)$ is increasing. This implies that $\Delta(n)$ is negative and that the internal stability condition is not satisfy. Next, we show that $\hat{b} < (N-n)(n-1)$

Let's assume that

$$\hat{b} = \frac{-G(n) + (G(n)^2 - 3F(n)H(n))^{1/2}}{3F(n)} \geq (N-n)(n-1),$$

where $\hat{b}$ is the positive root of equation $f'(b) = 0$.

Reordering terms we obtain

$$(G(n)^2 - 3F(n)H(n))^{1/2} \geq 3F(n)(N-n)(n-1) + G(n) > 0,$$

now squaring and reordering terms we get

$$0 \geq 3F(n)(3F(n)(N-n)^2(n-1)^2 + 2(N-n)(n-1)G(n) + H(n)),$$

which implies that

$$0 \geq 3F(n)(N-n)^2(n-1)^2 + 2(N-n)(n-1)G(n) + H(n).$$

Substitution of $F(n)$, $G(n)$ and $H(n)$ yields

$$
\begin{aligned}
0 \geq\ & 2Nn^5 - (6N + 6N^2)n^4 + (4N + 14N^2 + 4N^3)n^3 \\
& -n^2(9N^2 - 2N + 8N^3) + n(6N^3 - 2N) - 2N^3 + N^2. \qquad (16)
\end{aligned}
$$

It is easy to show that the third derivative of the right-hand side of this inequality is a strictly convex function with a minimum for $N = (3N+3)/5$. For this minimum the third derivative is negative. Moreover, we find that the third derivative is also negative for $(N+1)/2$ and N, so that we can conclude that the second derivative is decreasing in the interval $((N+1)/2, N]$. Then as the second derivative is negative for $(N+1)/2$ we have that the second derivative is negative and consequently the first derivative is decreasing. Now, calculating the first derivative for $(N+1)/2$ and N we find that it intersects the horizontal axis for a value between $(N+1)/2$ and N since the first derivative is positive for $(N+1)/2$ and negative for N. This value defines a maximum for the right-hand side of inequality (16). Then we can conclude that right-hand side of this inequality is positive since it takes positive values for $(N+1)/2$ and N. The value for $(N+1)/2$ is on the left of the maximum and the value for N is on the right of the maximum. This conclusion implies a contradiction and we have to admit that

$$\hat{b} < (N-n)(n-1),$$

or in words, that for the values of b in the interval (15) function $f(b)$ is increasing.

Thus, the final step of this part of the proof is to show that $f((N-n)(n-1))$ is positive. Substitution of $b = (N-n)(n-1)$ in $f(b)$ yields

$$
\begin{aligned}
f(n) =\ & n^6(2+N^2) - n^5(8 + 8N + 2N^2 + 2N^3) \\
& +n^4(12 + 28N + 14N^2 + 4N^3 + N^4) \\
& -n^3(8 + 36N + 37N^2 + 13N^3 + 2N^4) \\
& +n^2(2 + 20N + 37N^2 + 23N^3 + 5N^4) \\
& -n(4N + 15N^2 + 15N^3 + 6N^4) + 2N^2 + 3N^3 + 2N^4 \qquad (17)
\end{aligned}
$$

It is easy to show that the fourth derivative is a strictly convex function with a minimum for

$$\frac{4 + 4N + N^2 + N^3}{3(2 + N^2)}$$

and that

$$\frac{4 + 4N + N^2 + N^3}{3(2 + N^2)} < \frac{N + 1}{2} < N.$$

Moreover, for $(N + 1)/2$ we have that the fourth derivative is negative for $N \geq 7$ and positive for $N = \{3, 4, 5, 6\}$ whereas for N the four derivative is positive. Then for $N \geq 7$, the third derivative presents a minimum for a value between $(N+1)/2$ and N. Calculating the value of the third derivative for $(N + 1)/2$ and for N we can conclude that it intersects the horizontal axis for a value again between $(N + 1)/2$ and N since the third derivative is negative for $(N + 1)/2$ and positive for N. This intersection point defines a minimum for the second derivative so that this derivative is decreasing on the left of this minimum and increasing on the right. The result is that the second derivative also presents an intersection point with the horizontal axis between $(N+1)/2$ and N since the second derivative is negative for $(N+1)/2$ and positive for N. This intersection point defines a minimum for the first derivative between $(N + 1)/2$ and N since the value of the first derivative is positive for $(N+1)/2$ and negative for N. So that we have another intersection point for the first derivative between $(N + 1)/2$ and N which defines in this case a maximum for the function (17). Then as the value of this function for $(N + 1)/2$ is positive and zero for N, we can conclude that the function is positive in interval $((N + 1)/2, N)$ and zero for N since the function is increasing on the left of the maximum and decreasing on the right. Finally, for $N = \{3, 4, 5, 6\}$ we have that the fourth derivative is positive in interval $((N + 1)/2, N]$ so that the third derivative is increasing but as it is negative for $(N + 1)/2$ and positive for N it also presents an intersection point which defines a minimum for the second derivative. The argument from this point is the same than before, as is the conclusion.

Then the final result is that $f(b)$ for $b = (N - n)(n - 1)$ is positive when n is in interval $((N + 1)/2, N)$ and zero when $n = N$, so that it will be positive for the values of b in interval (15) since $f(b)$ is increasing. Thus, we can conclude that the internal stability condition is not satisfied when n is an interior solution and $n - 1$ is a corner solution.

(ii) We now consider the case where for both n and $n-1$ there are interior solutions. From Corollary 1, a necessary condition for this to hold is that $b > N-2$.

When n and $n-1$ are interior solutions we know from Lemma 1 that $\Delta(n)$ is

$$\Delta(n) = -\frac{a^2N^2(n-1)}{2b}\frac{(n-3)b^2 + J(n)b + K(n)}{\left[b+N+n(n-1)\right]^2\left[b+N+(n-1)(n-2)\right]^2} \quad (18)$$

where

$$\begin{aligned} J(n) &= 2[n^3 - 4n^2 + (N+3)n - N - 2] > 0 \; for \; n \geq 3, \\ K(n) &= (n+1)N^2 + 2n(n^2-2n-1)N + n^2(n-1)^2(n-3) > 0 \; for \; n \geq 3. \end{aligned}$$

Thus for $n \in [3, N]$ $\Delta(n)$ is negative, and so any such n is internally unstable.

Under what conditions might $n = 2$ be stable? For $n = 2$ we have that $\Delta(2)$ is written as follows

$$\Delta(2) = \frac{a^2N^2}{2b}\frac{b^2 - 2(N-4)b - (3N^2-4N-4)}{(b+N+2)^2(b+N)^2},$$

so that the sign of the internal stability condition depends on the sign of $b^2 - 2(N-4)b - (3N^2-4N-4)$ from which it follows that

$$\Delta(2) \geq 0 \Leftrightarrow \; b \geq N - 4 + 2(N^2 - 3N + 3)^{1/2} > N - 2.$$

Finally, we have to check the external stability condition, i.e., the sign of $\Delta(3)$. If for $n = 3$ it applies an interior solution we know from (18) that $\Delta(3) < 0$ so that $\pi^s(3) < \pi^f(2)$ and the external stability is satisfied. If for $n = 3$ it applies a corner solution we know from (i)-(b) that $\Delta(3)$ is also negative. Thus, we can conclude that for the Cournot equilibrium only a bilateral agreement can be stable.

B Proof of Proposition 3

In this Appendix we show that there does not exist any value of b such that $g(b,n),\; h\,(b,n) \leq 0$ for a given value of n. For given n $h(b,n)$ is strictly convex with respect to b with a minimum for a negative value of b and an intersection point with the vertical axis also negative. This implies that $h(b,n)=0$ has only a positive real solution given by

$$b_1(n)=\frac{1}{2}\left\{-(N+n^2-2n)+(n^4-4n^3+2Nn^2+N^2)^{1/2}\right\}, \tag{19}$$

so that $h(b,n)$ will be negative if $b\in(0,b_1(n))$ and positive for $b>b_1(n)$.

On the other hand, function $g(b,n)$ is strictly convex with respect to b and presents a minimum for $b=(N-n)(n-2)/2$. For this minimum the value of the function is $n(N-n)^2(4-n)/4$ which implies that $g(b,n)>0$ for $n=\{1,2,3\}$ and $b>0$. Then we can conclude that there does not exist any value of b such that $g(b,n),h(b,n)\leq 0$ for these values of n since $g(b,n)$ is always positive. For $n=4$, $g(b,4)=0$ for $b=N-4$ and positive for $b\neq N-4$. Moreover $b=N-4$ is bigger than $b_1(4)$ if $N>4$, so that if $b\leq b_1(4)$ we have that the Stackelberg equilibrium yields $q_s(4)>0$ and $q_f(4)=0$, if $b_1(4)<b<N-4$ then yields $q_s(4),\ q_f(4)>0$, if $b=N-4$ then yields $q_s(4)=0$ and $q_f(4)>0$, and finally if $b>N-4$ then yields $q_s(4),\ q_f(4)>0$. Thus, it does not occur that signatories and non-signatories select zero emissions at the same time. If $N=4$, the interesting cases are $n=\{1,2,3\}$ and we know that for these cases $g(b,n)$ is positive since the minimum value of the function is positive. For $n>4$, $g(b,n)=0$ has two positive real solutions

$$b_2(n) = \frac{1}{2}\left\{(N-n)\left[n-2-(n^2-4n)^{1/2}\right]\right\}, \tag{20}$$

$$b_3(n) = \frac{1}{2}\left\{(N-n)\left[n-2+(n^2-4n)^{1/2}\right]\right\} \tag{21}$$

and $g(b,n)$ will be negative if $b\in(b_2(n),b_3(n))$.

Finally, we show that $b_1(n)$ is lower than $b_2(n)$. Let's suppose now that $b_1(n)\geq b_2(n)$ which yields

$$\begin{aligned} & -(N+n^2-2n)+(n^4-4n^3+2Nn^2+N^2)^{1/2} \\ \geq\ & (N-n)(n-2)-(N-n)(n^2-4n)^{1/2}>0, \end{aligned} \tag{22}$$

simplifying terms we have that

$$(n^4-4n^3+2Nn^2+N^2)^{1/2}\geq N(n-1)-(N-n)(n^2-4n)^{1/2}>0.$$

Then squaring and simplifying terms again we get

$$n(n^2 - (3+N)n + 3N) \geq -(n-1)(N-n)(n^2-4n)^{1/2},$$

where the left-hand side of the inequality is negative for $n \in (4, N)$, then multiplying by -1 we obtain

$$0 < -n(n^2 - (3+N)n + 3N) \leq (n-1)(N-n)(n^2-4n)^{1/2}.$$

Finally, squaring again and simplifying terms we get a contradiction

$$4n(n^2 - 2Nn + N^2) \leq 0,$$

since $n^2 - 2Nn + N^2$ is positive for $n < N$. Consequently, we can establish that $b_1(n) < b_2(n)$ for all $n > 4$.

Given this relationship we can order the critical values of $b : b_1(n) < b_2(n) < b_3(n)$, so that if $b > b_3(n)$ we have that the Stackelberg equilibrium yields $q_s > 0, q_f > 0$, if $b \in [b_2(n), b_3(n)]$ then yields $q_s = 0, q_f > 0$, if $b \in (b_1(n), b_2(n))$ then yields again $q_s > 0, q_f > 0$, and finally if $b \leq b_1(n)$ then yields $q_s > 0, q_f = 0$. Thus, summarizing we can conclude that for a given value of n there does not exist any value of b such that $g(b,n),\ h(b,n) \leq 0$.

C Proof of Lemma 3

The results in Lemma 3 are shown from the properties of $b_1(n)$. Thus, what we show first is that $b_1(0) = b_1(N) = 0$ and that $b_1(n)$ presents a unique extreme in the interval $(1, N-1)$ which is a maximum. As we have written in footnote 3 we assume that n is a real number and once we know the properties of $b_1(n)$ then we focus on the values of $b_1(n)$ for $n = \{1, 2, ..., N-1\}$.

By substitution it is easy to check that $b_1(0) = b_1(N) = 0$. To show that the unique extreme of $b_1(n)$ is a maximum we use the inverse function of $b_1(n) = 0$. In order to obtain this function we rewrite $h(b,n)$ as

$$h(b,n) = (b+1)n^2 - (2b+N)n + (b+N)b,$$

and then from $h(b,n) = 0$ we get:

$$n = \frac{2b + N \pm (N^2 - 4b^3 - 4b^2)^{1/2}}{2(b+1)}. \tag{23}$$

So that for $N^2 - 4b^3 - 4b^2 \geq 0$ we can define $n^+(b_1)$ and $n^-(b_1)$ and their first derivatives:

$$\begin{aligned}
\frac{dn^+}{db_1} &= -\frac{2b^3 + 6b^2 + 4b + N^2 + (N-2)(N^2 - 4b^3 - 4b^2)^{1/2}}{2(b+1)^2(N^2 - 4b^3 - 4b^2)^{1/2}}, \\
\frac{dn^-}{db_1} &= \frac{2b^3 + 6b^2 + 4b + N^2 - (N-2)(N^2 - 4b^3 - 4b^2)^{1/2}}{2(b+1)^2(N^2 - 4b^3 - 4b^2)^{1/2}}.
\end{aligned}$$

On the other hand, it is easy to show that $N^2 - 4b^3 - 4b^2 = 0$ has a unique positive solution that we represent by $\tilde{b}$. For this value we have that

$$\lim_{b \to \tilde{b}^-} \frac{dn^+}{db_1} = -\infty; \quad \lim_{b \to \tilde{b}^-} \frac{dn^-}{db_1} = +\infty,$$

which implies that $db_1/dn = 0$ when $n = \tilde{n}$ that is the value defined by (23) for $\tilde{b}$:

$$\tilde{n} = \frac{2\tilde{b} + N}{2(\tilde{b} + 1)}.$$

Then given the sign of the limits we can establish that $\tilde{n}$ is a maximum for $b_1(n)$ in the real interval $(1, N-1)$ so that $b_1(n)$ is increasing for $n < \tilde{n}$ and decreasing for $n > \tilde{n}$. In order to show that $\tilde{n}$ belongs to that interval the only thing that we need to do is to calculate the differences $\tilde{n} - 1$ and $N - 1 - \tilde{n}$.

$\tilde{n}$ may or may not be an integer. If $\tilde{n}$ is an integer $\tilde{n} = \hat{n}$ and $b_1(\hat{n}) = b_1(n)$ has a unique solution $x_1 = x_2 = \hat{n}$ where x_1 and x_2 stand by the solutions to the equation. If $\tilde{n}$ is not an integer then there will exist an integer $\hat{n}$ such that $b_1(\hat{n})$ yields the maximum value of $b_1(n)$ for $n = \{1, 2, ..., N-1\}$. In that case equation $b_1(\hat{n}) = b_1(n)$ has two solutions $0 < x_1 < x_2$ and one of them will be equal to $\hat{n}$ by definition. In both cases according to Lemma 2 the emissions of non-signatories are zero only when $n = \hat{n}$ since for $n \neq \hat{n}$, $b_1(\hat{n}) > b_1(n)$. Remember that $b_1(n)$ characterizes the pairs (b, n) that satisfy $h(b, n) = 0$ and that for $b > b_1(n)$ Proposition 3 establishes that the equilibrium is the interior solution for non-signatories, i.e. $h(b, n) > 0$. If $b < b_1(\hat{n})$, $b = b_1(n)$ has two solutions and $I_i(x_i) = \bar{n}_i$, $i = 1, 2$ define two integers such that $\bar{n}_1 < \bar{n}_2$. Then given the properties of $b_1(n)$ we have that $b \leq b_1(n)$ for those values of n in the interval $[\bar{n}_1, \bar{n}_2]$ so that according to Proposition 3 the equilibrium is a corner solution. For the rest of values of n, $b > b_1(n)$ and again according to Proposition 3 the equilibrium for non-signatories is an interior solution. Finally, it is obvious that the distance $x_2 - x_1$ increases when b decreases which implies that $\bar{n}_2 - \bar{n}_1$ also increases although not monotonically.

D Proof of Corollary 2

This result is immediate from Lemma 3 always that $b_1(N-1)$ be the minimum value of $b_1(n)$ for $n = \{1, 2, ..., N-1\}$. As $b_1(n)$ is first increasing for $n < \tilde{n}$ and afterwards decreasing. This will occur if $b_1(N-1)$ is lower than $b_1(1)$.

First we calculate these two values:

$$\begin{aligned} b_1(1) &= \frac{1}{2}\left\{-(N-1) + (N^2+2N-3)^{1/2}\right\}, \\ b_1(N-1) &= \frac{1}{2}\left\{-(N^2-3N+3) + (N^4-6N^3+15N^2-14N+5)^{1/2}\right\}. \end{aligned}$$

Let´s suppose now that $b_1(1) \leq b_1(N-1)$ which yields

$$-(N-1)+(N^2+2N-3)^{1/2} \leq -(N^2-3N+3)+(N^4-6N^3+15N^2-14N+5)^{1/2},$$

simplifying terms we have that

$$0 < N^2 - 4N + 4 + (N^2+2N-3)^{1/2} \leq (N^4-6N^3+15N^2-14N+5)^{1/2}.$$

Then squaring and simplifying terms again we get

$$0 < 2(N^2-4N+4)(N^2+2N-3)^{1/2} \leq 2N^3 - 10N^2 + 16N - 8.$$

Finally, squaring again and simplifying terms we get a contradiction

$$0 \leq -16N^5 + 144N^4 - 512N^3 + 896N^2 - 768N + 256,$$

since the right-hand side of the inequality is negative for $N \geq 3$. Then we can conclude that $b_1(1) > b_1(N-1)$.

E Proof of Lemma 4

As with Lemma 3, the proof of Lemma 4 derives from the properties of the functions $b_2(n)$ and $b_3(n)$. By substitution we get that $b_2(4) = N - 4$ and $b_2(N) = 0$. Remember that signatories' emissions are always positive for $n = \{1, 2, 3\}$ and $b > 0$.

On the other hand, if we take the first derivative of $b_2(n)$ (see (20) in the proof of Proposition 3) we obtain

$$\frac{db_2}{dn} = \frac{1}{2}\left\{N - 2n + 2 + \frac{2n^2 - (6+N)n + 2N}{(n^2 - 4n)^{1/2}}\right\},$$

that presents the following limits:

$$\begin{aligned} \lim_{n\to 4} \frac{db^2}{dn} &= -\infty, \\ \lim_{n\to N} \frac{db^2}{dn} &= \frac{1}{2}\left\{-(N-2) + \frac{N(N-4)}{(N^2-4N)^{1/2}}\right\} < 0 \text{ for } N > 4. \end{aligned}$$

Moreover, its second derivative is:

$$\frac{d^2b}{dn^2} = \frac{1}{2}\left(-2 + \frac{2n^3 - 12n^2 + 12n + 4N}{(n^2 - 4n)^{1/2}}\right).$$

Let ´s suppose that this second derivative is negative or zero. This implies that

$$0 < 2n^3 - 12n^2 + 12n + 4N \leq 2(n^2 - 4n)^{1/2},$$

squaring and reordering terms we obtain the following inequality

$$4n^6 - 48n^5 + 192n^4 + (16N - 288)n^3 - (96N - 140)n^2 + (96N + 16)n + 16N^2 \leq 0.$$

It is easy but tedious to show that the left-hand side of this inequality if positive for $n \geq 4$ yielding a contradiction. So that we can conclude that $d^2b_2/dn^2 > 0$ which allows us to establish that db_2/dn is increasing and, consequently, that $b_2(n)$ is a decreasing, strictly convex function in interval $[4, N]$. Thus, for $b_2(N-1) \leq b \leq N-4$, $b = b_2(n)$ has a unique solution that we call x_3.

Next, we study the properties of function $b_3(n)$ (see (21) in the proof of Proposition 3). By substitution we get that $b_3(4) = N - 4$ and $b_3(N) = 0$.

Moreover, it is easy to show that $b_3(n)$ is a strictly concave function with a maximum in interval $(4, N)$ equal to $N^2/2(N-2)$. So that for $4 < n < N^2/2(N-2)$ $b_2(n)$ increases and for $N^2/2(N-2) < n < N$ decreases. This implies that $b_3(N^2/2(N-2)) = N(N-4)/4 > b_2(4) = N-4$.

$N^2/2(N-2)$ can be or cannot be an integer. If $N^2/2(N-2)$ is an integer $N^2/2(N-2) = n^*$ and $b_3(n^*) = b_3(n)$ has a unique solution $x_3 = x_4 = n^*$ where x_3 and x_4 stand by the solution to the equation. In order to have a maximum n^* different from 4 we assume that $N > 5$. Notice that for $N = 5$, $b_2(n)$ and $b_3(n)$ are defined in interval $[4, 5]$ and, consequently, $n^* = 4$.[14] If $N^2/2(N-2)$ is not an integer then there will exist an integer n^* such that $b_1(n^*)$ yields the maximum value of $b_3(n)$ for $n = \{4, 5, ..., N-1\}$. In that case equation $b_3(n^*) = b_3(n)$ has two solutions $0 < x_3 < x_4$ and one of them will be equal to n^* by definition. In both cases according to Lemma 2 the emissions of signatories are zero only when $n = n^*$ since for $n \neq n^*, b_3(n^*) > b_3(n)$. Notice that $b_3(n)$ characterizes the pairs (b, n) that satisfy $g(b, n) = 0$ and that for $b > b_3(n)$ Proposition 3 establishes that the equilibrium is the interior solution, i.e. $g(b, n) > 0$. If $N-4 < b < b_3(n^*)$, $b = b_3(n)$ has two solutions and $I_i(x_i) = \bar{n}_i, i = 3, 4$ define two integers such that $\bar{n}_3 < \bar{n}_4$. Then given the properties of $b_3(n)$ we have that $b \leq b_3(n)$ for those values of n in interval $[\bar{n}_3, \bar{n}_4]$ so that according to Proposition 3 the equilibrium for signatories is a corner solution. For $b_2(N-1) \leq b \leq N-4$, we have that $b = b_3(n)$ has a unique positive solution x_4 that along with x_3 obtained from $b = b_2(n)$ define applying $I_i(x_i)$ an interval $[\bar{n}_3, \bar{n}_4]$ for which the equilibrium is also a corner solution. If $n \notin [\bar{n}_3, \bar{n}_4]$, $b > b_3(n)$ when $N-4 < b < b_3(n^*)$ or $b > b_2(n)$ and $b > b_3(n)$ when $b_2(N-1) \leq b \leq N-4$ and again according to Proposition 3 the equilibrium for signatories is an interior solution.

It is obvious that the distance $x_4 - x_3$ increases when b decreases always that $b > N-4$ which implies that $\bar{n}_4 - \bar{n}_3$ also increases although not monotonically. However, if $b < N-4$ the relationship is the contrary. Finally, as the minimum value of $b_2(n)$ for $n = \{4, 5, ..., N-1\}$ is given by $b_2(N-1)$ we have that if $b < b_2(N-1)$, $b < b_2(n)$ for all n, $4 \leq n \leq N-1$ and the equilibrium is an interior solution for $n = \{1, 2, ..., N-1\}$.

[14]In the rest of the paper we focus on $N > 5$ that is the interesting case. Anyway, for $N = 5$ the results are the same adjusting the sign of some inequalities.

F Proof of Proposition 4

Points (i)-(iii) and (v) follow from Lemmas 3 and 4 and Corollary 2 always that $b_1(\hat{n}) < b_3(n^*)$. Point (iv) occurs only if $b_1(1) > b_2(N-1)$ and $b_1(N-2) > b_2(N-1)$ too, since $b_1(N-1) < b_2(N-1)$. First we show that $b_1(\hat{n}) < b_3(n^*)$ for $N > 5$. From the proofs of Proposition 3 and Lemma 4 we know that $b_1(4) < b_2(4) = b_3(4) = N - 4 < b_3(n^*)$.

On the other hand, $b_1(n)$ can be increasing or decreasing at $n = 4$ depending on the number of countries, N. Let´s suppose that

$$\frac{db_1(4)}{dn} = -3 + \frac{16 + 4N}{(32N + N^2)^{1/2}} \geq 0.$$

This implies that

$$16 + 4N \geq 3(32N + N^2)^{1/2},$$

that squaring and simplifying yields

$$256 - 160N + 7N^2 \geq 0.$$

It is easy to show that the left-hand side of the inequality is negative for $N \in (5, 21]$. So that we can conclude that $db_1(4)/dn < 0$ which means that $\hat{n} \leq 4$. The previous result also allows us to establish that for $N > 21$, the first derivative of $b_1(n)$ at $n = 4$ is positive which means that $\hat{n} \geq 4$.

Next we suppose that $N \in (5, 21]$. In that case, $\hat{n}$ is an integer in domain $\{1, 2, 3, 4\}$. If $\hat{n} = 4$, we already know that $b_1(4) < N - 4 < b_3(n^*)$ and it is established that $b_1(\hat{n}) < b_3(n^*)$. If $\hat{n}$ were an integer different from 4, it is easy to show that $b_1(n) < N-4$ for $n = \{1, 2, 3\}$ so that we can also conclude that $b_1(\hat{n}) < b_3(n^*)$ for this values of n.

Next, we suppose that $N > 21$. In that case, $\hat{n} \geq 4$. For $\hat{n} = 4$ the same argument than the one we have just used is applied. For $\hat{n} > 4$, we know from Proposition 3 that $b_1(\hat{n}) < b_2(\hat{n})$ and from the proof of Lemma 4 that $b_2(\hat{n}) < N - 4 < b_3(n^*)$ so that we find that $b_1(\hat{n}) < b_3(n^*)$ as we wanted to establish.

Finally, we show that $b_1(1) > b_2(N-1)$ and $b_1(N-2) > b_2(N-1)$. Let´s suppose first that $b_1(1) \leq b_2(N-1)$. This implies that

$$0 < \left((N-1)^2 + 4(N-1)\right)^{1/2} \leq 2N - 4 - \left((N-1)^2 - 4(N-1)\right)^{1/2}.$$

Squaring yields

$$2(2N-4)\left((N-1)^2 - 4(N-1)\right)^{1/2} \leq (2N-4)^2 - 8(N-1),$$

squaring again, simplifying and reordering terms we obtain the following inequality

$$32N^3 - 240N^2 + 448N - 256 \leq 0.$$

This inequality yields a contradiction for $N > 5$ so that we can conclude that $b_2(N-1) < b_1(1)$.

Next, we suppose that $b_1(N-2) \leq b_2(N-1)$. This implies that

$$0 < (N^4 - 10N^3 + 41N^2 - 72N + 48)^{1/2} \leq -4N + N^2 + 5 - (N^2 - 6N + 5)^{1/2}.$$

Squaring yields

$$2(-4N + N^2 + 5)(N^2 - 6N + 5)^{1/2} \leq 2N^3 - 14N^2 + 26N - 18,$$

squaring again, simplifying and reordering terms we obtain the following inequality

$$16N^4 - 144N^3 + 400N^2 - 464N + 176 \leq 0$$

This inequality yields a contradiction fro $N > 5$ so that we can conclude that $b_2(N-1) < b_1(N-2)$.

G Proof of Proposition 5

It is straightforward to show that the grand coalition is self-enforcing since π^*, the net benefits for the grand coalition, is positive, see (2), and $\pi^f(N-1)$ is negative according to Lemma 2. This means that the internal stability condition is satisfied. It is also easy to show that the followers are always interested in joining the agreement. Using the net benefits expressions that appear in Lemma 2, the external stability condition is given by the difference

$$\triangle(n+1) = \frac{a^2[2(n+1) - b]}{2(n+1)^2}.$$

On the other hand, it is easy to show that $b_1(N-1) < 1$. Let's suppose that $b_1(N-1) \geq 1$. This implies that[15]

$$b_1(N-1) = \frac{1}{2}(-(N^2 - 3N + 3) + (N^4 - 6N^3 + 15N^2 - 14N + 5)^{1/2}) \geq 1,$$

[15]See Appendix B for the expression of $b_1(n)$.

which yields

$$(N^4 - 6N^3 + 15N^2 - 14N + 5)^{1/2} \geq N^2 - 3N + 5 > 0.$$

Squaring and ordering terms we obtain the following contradiction:

$$0 \geq 4N^2 - 16N + 20.$$

Thus, for $b \leq b_1(N-1) < 1$ the numerator of $\triangle(n+1)$ is positive for all $n \geq 1$ and a non-signatory is always interested in entering the agreement.

H Proof of Proposition 6

For $b_2(N-2) \leq b \leq N-4$, we have that $\bar{n}_3 = I_3(x_3) < \bar{n}_4 = I_4(x_4) = N-1$ so that we have a signatory corner solution for $n = \{\bar{n}_3, ..., N-1\}$. This is a consequence of the fact that $b_3(N-1) > b_2(4) = N-4$.[16] Now given a value of b we select n such that $\bar{n}_3 < n \leq N-1$ and we check if the internal stability condition can be satisfied. In this case we have that both n and $n-1$ are signatory corner solutions so that according to Lemma 2 the internal stability condition is given by

$$\begin{aligned} \triangle(n) &= -\frac{a^2(N-n)^2}{2(b+N-n)^2} - \frac{a^2[b-(N-n+1)(N-n-1)]}{2(b+N-n+1)^2} \\ &= -\frac{a^2[b^3 + L(n)b^2 + M(n)b + P(n)]}{2(b+N-n)^2(b+N-n+1)^2}, \end{aligned}$$

where

$$L(n) = 2(N-n) + 1 > 0, \quad M(n) = 3(N-n)^2 + 2(N-n) > 0$$

$$P(n) = 2(N-n)^3 + 2(N-n)^2 > 0$$

So, $\triangle(n)$ is negative, in fact, for all $b > 0$. Consequently if there exists a self-enforcing IEA the number of countries in the agreement will be equal or less than $\bar{n}_3$. Finally, we know that $b_2(n)$ is a strictly convex, decreasing function defined in interval $[4,N]$, then as $\bar{n}_3$ is defined as the smallest integer no less than x_3 being x_3 the unique positive root of equation $b = b_2(n)$, we can conclude that $\bar{n}_3$ decreases when b increases.

[16] This is very easy to show so that we omit it.

I Proof of Proposition 7

According to Proposition 6 if $b = N - 4$ the maximum number of countries in a self-enforcing IEA is four then the question to answer now is whether an agreement of four countries can be self-enforcing for $b > N-4$. For $b > N-4$ and $n = 4$, according to Proposition 3, the equilibrium is the interior solution for signatories and non-signatories. Then using Lemma 2 we get

$$\triangle(4) = -\frac{a^2N^2k(b)}{2[(b+N-3)^2+9b]^2[(b+N-4)^2+16b]},$$

where

$$\begin{aligned} k(b) &= 3b^4 + (6N+48)b^3 + (2N^2 - 40N + 101)b^2 \\ &\quad -(2N^3 - 7N^2 - 10N + 39)b - (N^4 - 14N^3 + 73N^2 - 168N + 144). \end{aligned}$$

It can be easily shown that $k''(b)$ is a strictly convex function with a minimum at $b = (8+N)/2$. For this minimum we have that the second derivative is positive for all N. This implies that $k'(b)$ is increasing for all $b > 0$ and consequently is also increasing for $b > N-4$. Then as $k'(N-4) = 252N^3 - 1212N^2 - 3408N + 16716$ is positive for $N > 5$ we can conclude that the first derivative is positive for $b > N-4$, which implies that $k(b)$ is increasing for $b > N-4$. Finally, as $f(N-4) = 8N^4 - 99N^3 + 362N^2 - 207N - 676$ is positive for $N > 5$ we have that $k(b)$ is positive for $b > N-4$ so that $\triangle(4)$ is negative and the internal stability condition is not satisfied.

Next, we check whether an agreement with three countries can be stable. For $n = 3$ we have that

$$\triangle(3) = -\frac{a^2N^2l(b)}{2[(b+N-2)^2+4b]^2[(b+N-3)^2+9b]},$$

where

$$\begin{aligned} l(b) &= 8b^3 - (N^2 + 10N - 23)b^2 - (2N^3 - 8N^2 + 10N - 4)b \\ &\quad -(N^4 - 10N^3 + 37N^2 - 60N + 36). \end{aligned}$$

It is easy to show that this is a strictly convex function that first decreases until reaching a minimum value that is negative and afterwards increases. This means that the equation $l(b) = 0$ has a unique, positive solution that we denote by $\bar{b}$ such that if $b < \bar{b}$, $l(b)$ is negative and if $b > \bar{b}$, $l(b)$ is positive.

Then as $l(N-4) = -N^4 + 32N^3 - 88N^2 + 144N - 196$ is negative we can conclude that $N - 4 < \bar{b}$ so that for b in interval $[N-4, \bar{b}]$, $l(b) \leq 0$ and consequently $\triangle(3)$ is positive or zero and the internal stability condition holds. Moreover as $\triangle(4)$ will be negative and this implies that $\pi^f(3) > \pi^s(4)$, the external stability condition is also satisfy and the agreement is self-enforcing.

Finally for $b > \bar{b}$, $l(b)$ is positive and the internal stability condition for $n = 3$ is not satisfy. In this case only an agreement of two countries is self-enforcing. For $n = 2$ we have that

$$\triangle(2) = \frac{a^2 N^2 m(b)}{2[(b+N-1)^2 + b]^2[(b+N-2)^2 + 4b]},$$

where

$$\begin{aligned} m(b) &= b^4 + 2Nb^3 + (2N^2-1)b^2 + (2N^3 - 7N^2 + 10N - 5)b \\ &\quad + (N-1)^2(N-2)^2. \end{aligned}$$

$\triangle(2)$ is positive for $b > 0$ and $N > 5$ which implies that $\pi^s(2) > \pi^f(1)$ and that, consequently, the internal stability condition for an agreement of two countries holds. The external stability condition is also satisfied since for $b > \bar{b}$, $\triangle(3)$ is negative.

References

[1] Barrett, S. (1994), "Self-enforcing international environmental agreements", *Oxford Economic Papers* **46**: 878-94.

[2] Barrett, S. (2001), "International cooperation for sale", *European Economic Review* **45**:1835-50.

[3] Carraro, C. and Siniscalco, D. (1991), "Strategies for the international protection of the environment", CEPR Discussion Paper 568.

[4] Carraro, C. and Siniscalco, D. (1993), "Strategies for the international protection of the environment", *Journal of Public Economics* **52**: 309-28.

[5] Chander, P. and Tulkens, H. (1997), "The core of an economy with multilateral environmental externalities", *International Journal of Game Theory* **26**: 379-401.

[6] D'Aspremont, C.A., Jacquemin, J., Gadszeweiz, J. and Weymark, J. A. (1983), "On the stability of collusive price leadership", *Canadian Journal of Economics* **16**:17-25.

[7] Diamantoudi, E. and Sartzetakis, E. S. (2002), "Stable international environmental agreements: An analytical approach", Nota Di Lavoro 7.2002, FEEM, Venice.

[8] Ecchia, G. and Mariotti, M. (1997), "The stability of international environmental coalitions with farsighted countries: Some theoretical observations" in C. Carraro (ed.), *International environmental negotiations*, Cheltenham, UK: Edward Elgar.

[9] Ecchia, G. and Mariotti, M. (1998), "Coalition formation in international environmental agreements and the role of institutions", *European Economic Review* **42**:573-82.

[10] Finus, M. (2001), *Game theory and international environmental cooperation*, Cheltenham, UK: Edward Elgar.

[11] Hoel, M. (1992), "International environmental conventions: the case of uniform reductions of emissions", *Environmental and Resource Economics* **2**: 141-59.

[12] Hoel, M. and Schneider, K. (1997), "Incentives to participate in an international environmental agreement", *Environmental and Resource Economics* **9**:153-70.

[13] Na, S.L. and Shin, H.S. (1998), "International environmental agreements under uncertainty", *Oxford Economic Papers* **50**: 173-85.

[14] Petrakis, E. and Xepapadeas, A. (1996), "Environmental consciousness and moral hazard in international agreements to protect the environment", *Journal of Public Economics* 60:95-110.

[15] Rubio, S. and Casino, B. (2001), "International cooperation in pollution control", IVIE Working Paper WP-AD 2001-21.

[16] Rubio, S. and Ulph, A. (2002a), "A simple dynamic model of international environmental agreements with a stock pollutant", Discussion Paper in Economics and Econometrics 0209, Department of Economics, University of Southampton.

[17] Rubio, S. and Ulph, A. (2002b), "An infinite-horizon model of dynamic membership of international environmental agreements", mimeo, Department of Economics, University of Southampton.

[18] Ulph, A. (2002a), "International environmental agreements with a stock pollutant, uncertainty and learning", mimeo, Department of Economics, University of Southampton.

[19] Ulph, A. (2002b), "Environmental agreements with uncertainty and learning - the case of stock-dependent unit costs", mimeo, Department of Economics, University of Southampton.

[20] Wagner, U.J. (2001), "The design of stable international environmental agreements: Economic theory and political economy", *Journal of Economic Surveys* 15:377-411.

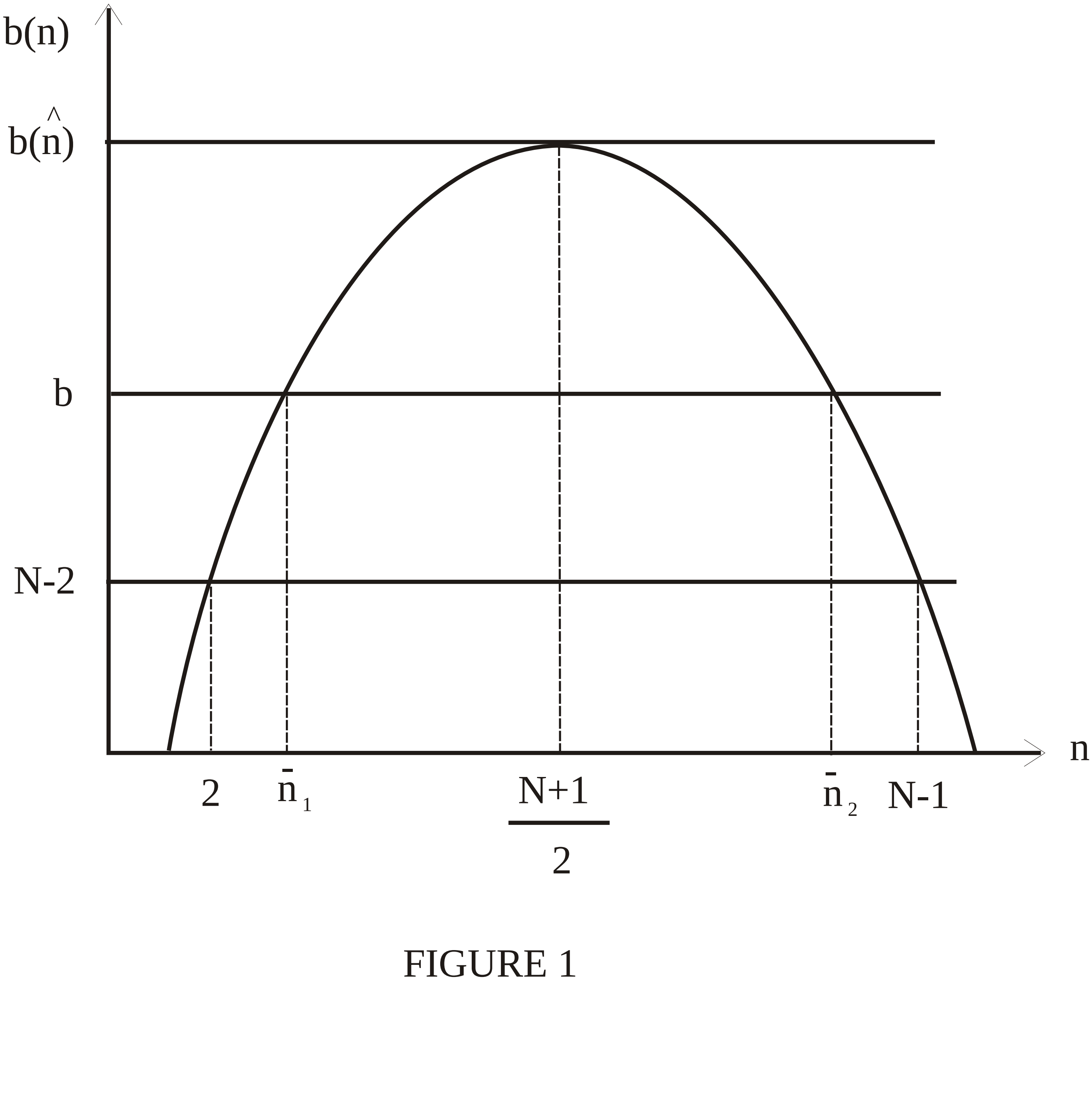

FIGURE 1

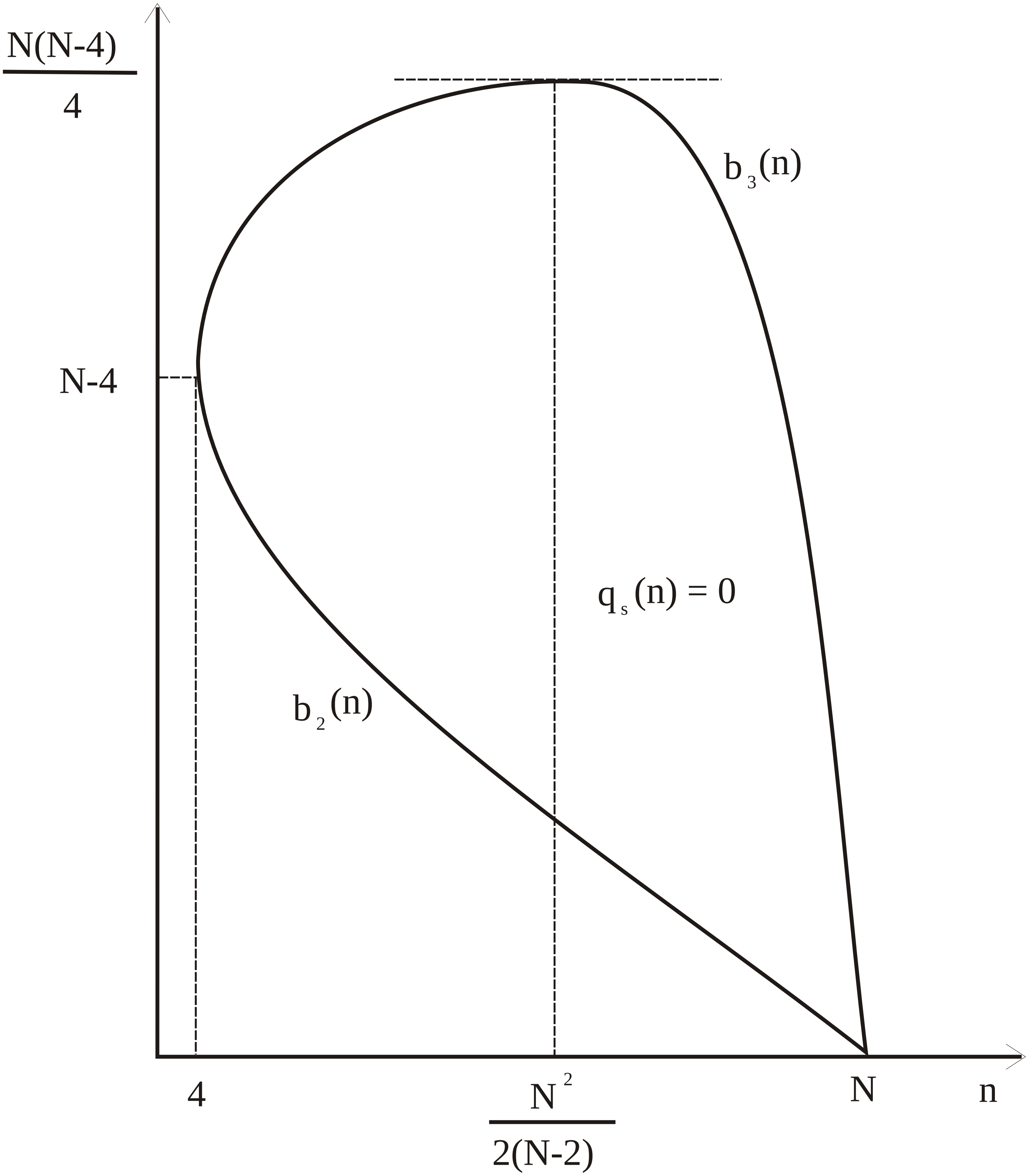

FIGURE 2

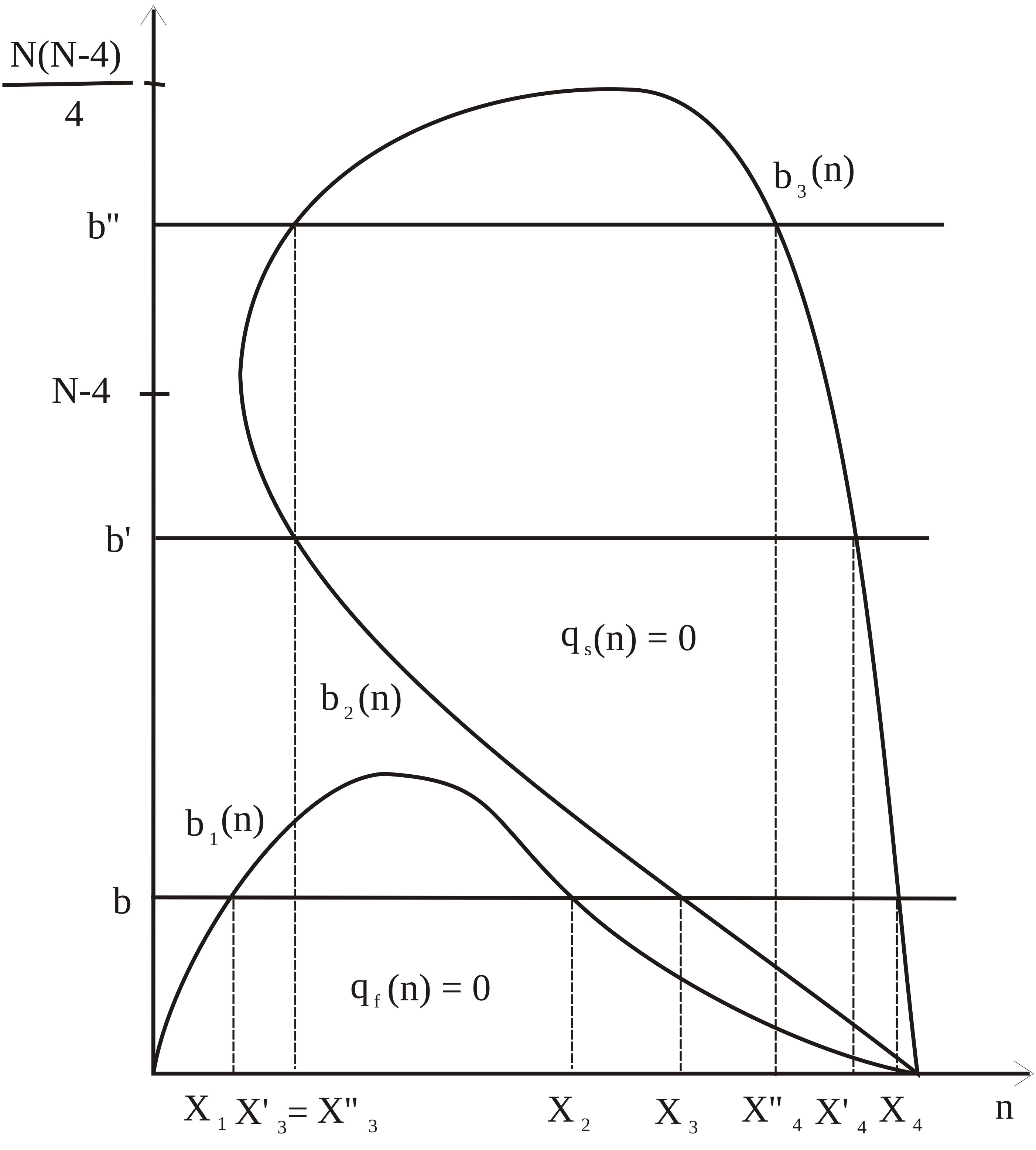

FIGURE 3