Stable International Environmental Agreements: An Analytical Approach

Effrosyni Diamantoudi
Department of Economics, University of Aarhus

Eftichios S. Sartzetakis*
Department of Accounting and Finance, University of Macedonia &
Department of Economics, University College of the Cariboo

September 2001

Abstract

In this paper we examine the formation of International Environmental Agreements (IEAs) modeled as a two-stage non-cooperative game. We provide a rigorous analytical treatment of the main model used in the literature and we offer, for the first time, a formal solution, while we clarify some misconceptions that exist in the literature. We find that the unique stable IEA consist of either two, three or four signatories if the number of countries is greater than or equal to 5. Furthermore, we show that the welfare of the signatories of a *stable* IEA is very close to its lowest level relative to the welfare of signatories of other non-stable IEAs. While in our model countries' choice variable is emissions, we extend our results to the case where the choice variable is abatement effort.

Keywords: International Environmental Agreements, Coalition Formation.
JEL classification: D62, Q28.

*E-mail addresses: E. Diamantoudi: *faye@econ.au.dk* and E. Sartzetakis: *esartz@uom.gr*. Corresponding author: Eftichios S. Sartzetakis, Department of Accounting and Finance, University of Macedonia, 156 Egnatia Str., Thessaloniki 54006, Greece.

1 Introduction

Some of the most important environmental problems urgently calling for solution are problems related to transboundary pollution. Environmental problems such as ozone depletion, climate change and marine pollution have been the focus of intense negotiations at the international level over the past two decades. Given the high priority environmental problems receive at the policy level, it is not surprising that there is a growing effort to analyze International Environmental Agreements (IEAs) at the theoretical level[1]. A significant part of the literature on IEAs studies the formation of a coalition that reduces pollution in the presence of free riding incentives by its members. Although other directions[2] have been explored, in the present work we adopt the approach that models the formation of IEAs as a two stage, non-cooperative game. Notably, no analytical solution of such a model exists in the literature, and thus, the most important contributions are based on simulations. In this paper we provide a rigorous analytical treatment of the model while we offer a formal solution of it and clarify some misconceptions that exist in the literature.

A critical characteristic of IEAs is the luck of a supra-national authority that could dictate and enforce environmental policies on sovereign states. Thus, IEAs have to be self-enforcing in the sense that they are immune to deviation by the countries involved. In the literature it is assumed that, in the first stage, countries signing the IEA form a coalition and behave cooperatively by maximizing the coalition's aggregate welfare. In the second stage, the countries that do not participate in the agreement observe the results of the agreement and behave non-cooperatively by maximizing their individual welfare. Naturally when the coalition maximizes its welfare, in the first stage, it explicitly foresees and takes into account the non-signatories' behavior that is about to follow. An IEA is considered to be stable if no

[1] For excellent reviews of this literature see Finus (2000), Ioannidis, Papandreou and Sartzetakis (2000) and Folmer, Hanley and Mibfeldt (1998).

[2] The literature has explored other directions as well. Chandler and Tulkens (1992) and (1997) have analysed IEAs as cooperative games, while a number of papers, among which Barrett (1994) and Finus and Rundshagen (1998) have employed repeated games.

one of its signatories has an incentive to withdraw (this aspect of stability is known as *Internal Stability*) while no other country outside of the IEA has an incentive to participate in the agreement (this aspect of stability is known as *External Stability*). Such a coalition formation analysis was originally undertaken by D'Aspremont et. al (1983) to model collusive behavior in price leadership and was first introduced to the study of IEAs by Barrett (1994).[3]

We study the problem of deriving the size of a stable IEA in a model very similar to Barrett (1994) with the only difference being the choice variable, namely, in our model countries choose their emission level whereas in Barrett's (1994) they choose their abatement effort. Moreover, we adopt specific functional forms (quadratic benefit and damage functions), corresponding to those in Barrett (1994). As mentioned earlier the stability notion applied to the model was first introduced by D' Aspremont et al. (1983). We show that although the environmental literature often borrows results from the price leadership model, the two models are not the same. The most important difference is that in the environmental agreements case the members' welfare does not monotonically increase with respect to the size of the coalition. In fact, we show that there exist situations (with sufficiently small coalitions), where a country is better off as a member of the coalition than outside of the coalition and as the coalition grows its members' welfare drops. This difference stems from the fact that in the price leadership model the fringe behaves non-strategically, that is, its members behave as price-takers, not conceptualizing the impact of their actions on the market price. Whereas, in the IEAs case the non-signatories behave strategically by explicitly taking into account the negative effect their individual emissions have on their welfare via global pollution.

However, the main contribution of this paper is the complete analytical solution of the coalition formation model with the afore mentioned functional forms. We find that a stable coalition consist of either 2, 3 or 4 members if the total number of countries is greater than or equal to 5. Furthermore, we show that the welfare level of the signatories of a stable IEA is very close to

[3]See also the works of Carraro and Siniscalco (1993) and (1998).

its lowest value in comparison with the welfare level of signatories of other, non-stable IEAs.

Our results severely restricting the size of stable coalitions, contradict Barrett's (1994) suggestion that stable IEAs could consist of any large number of countries.[4] We resolve this seeming inconsistency by converting our model's choice variable from emission levels to abatement efforts, thus making the model directly comparable to Barrett's (1994) framework. In doing so, we formulate the link (and hence the equivalence) between the two approaches and show that our results survive such a conversion. The proofs of all the results presented in the paper are delineated in the appendix.

2 The model

We assume that there exist n identical countries, $N = \{1, ..., n\}$. Production and consumption in each country i generates emissions $e_i \geq 0$ of a global pollutant as an output. The term *global* pollutant indicates that we assume pollution to be a public bad, and that individual emission impose negative externalities on all other countries. Similarly, in Section 4 where the model is specified in terms of abatement effort, individual abatement effort is assumed to be a public good. The social welfare of country i, w_i, is expressed as the net between the benefits from country i's emissions, $B_i(e_i)$, and the damages $D_i(E)$ from the aggregate emissions, E. Since the countries are assumed to be identical we henceforth drop the subscripts from the functions. As each country i's emission level increases, its benefits $B(e_i)$ increase as well. We consider the following quadratic benefit function for each country $i \in N$, $B(e_i) = b\left[ae_i - \frac{1}{2}e_i^2\right]$, where a and b are positive parameters. Country i's damages from pollution depend on aggregate pollution, E, where $E = \sum_{i \in N} e_i$. We assume a quadratic damage function for each country $i \in N$, of the following form $D(E) = \frac{1}{2}c(E)^2$, where c is a positive parameter.[5]

[4] Although Barrett (1994) clearly supports the existence of large coalitions, he also observes that large coalitions offer very small increases in global net benefits.

[5] An alternative form of the damage function is also used in the literature, see for example Barrett (1994) and Finus (2000). According to this variation, each country's damages are a share of aggregate emissions, that is, $D(E) = \frac{1}{2n}c(E)^2$. The difference

With these specifications, each country i's welfare function becomes:

$$w = b\left[ae_i - \frac{1}{2}e_i^2\right] - \frac{c}{2}\left(\sum_{i \in N} e_i\right)^2 \quad . \tag{1}$$

The (pure) non-cooperative case: In the non-cooperative case each country chooses its emission level taking the other countries' emissions as given. That is, country i behaves in a typical Cournot fashion maximizing (1). The first order condition of the above maximization problem yields country i's emission reaction function, $e_i = \frac{ba - c\sum_{j \neq i} e_j}{b+c}$.

Since we have assumed complete symmetry, $e_i = e_{nc}$ for every $i \in N$, the above reaction function yields the equilibrium emission level per country,

$$e_{nc} = \frac{a}{1 + \gamma n} \quad , \tag{2}$$

where $\gamma = \frac{c}{b}$. Consequently, the aggregate emission level under the (purely) non-cooperative case is, $E_{nc} = ne_{nc} = \frac{na}{1+\gamma n}$.

Full cooperation: Under full cooperation, the grand coalition maximizes the joint welfare. The first order condition yields the aggregate emission level, $E_c = \frac{an}{\gamma n^2 + 1}$. Since each country contributes $\frac{1}{n}$ of the total emissions, the per country emission level, e_c, is

$$e_c = \frac{E_c}{n} = \frac{a}{\gamma n^2 + 1} \quad . \tag{3}$$

It is easily verifiable that each country emits less and is better off in the case of full cooperation than under non-cooperation, that is, $e_c < e_n$ and $w_c > w_n$.

However, in this one stage, purely simultaneous framework each country has an incentive to cheat on the agreement and free-ride on the emission reduction achieved by the countries complying with the agreement. In what follows we examine the two stage framework where the incentive to free ride

between the two forms is a difference in parameter specification and it does not affect the results. The full analysis using this alternative functional form is available to the interested reader upon request. We believe that the specification we use is more appropriate in describing global pollution problems such as ozone depletion and global warming.

on the coalition's cooperating efforts may be offset by the adjustment of the coalition's emissions upon a member's deviation. The equilibrium number of countries participating in an IEA, is derived by applying the notions of internal and external stability of a coalition as was originally developed by D'Aspremont et. al (1983) and extended to IEAs by Barrett (1994).

Coalition Formation: Assume that a set $S \subset N$ of countries sign an agreement and $N \backslash S$ do not. Let the size of coalition be $|S| = s$, the total emissions generated by the coalition be E_s, while each member of the coalition emits e_s such that $E_s = se_s$. In a similar manner, each non-signatory country emits e_{ns}, yielding a total emission level $E_{ns} = (n - s)e_{ns}$.

The non-signatories behave non-cooperatively after having observed the choice of signatories. Their maximization problem results to a best response function of the form presented earlier. However, now only $n - s$ countries stay outside of the emission reduction agreement emitting e_{ns}, while the rest s countries emit in total E_s, that is $\sum_{i \in N} e_i = (n - s)e_{ns} + se_s$. Substituting this into the reaction function yields each non-signatory country's emissions $e_{ns} = \frac{a - \gamma E_s}{1 + \gamma(n-s)}$ as a function of the signatory countries' aggregate emission E_s. The aggregate non-signatory emission level is $E_{ns} = \frac{(a-\gamma E_s)(n-s)}{1+\gamma(n-s)}$.

Signatories choose their emission level by maximizing their collective welfare while taking into account the behavior of non-signatories. That is, signatories choose E_s by solving the following constrained maximization problem,

$$\max_{E_s} \sum_{i \in S} w_s$$
$$\text{subject to} \quad E_{ns} = \frac{(a - \gamma E_s)(n - s)}{1 + \gamma (n - s)}$$

where w_s is the welfare function of each signatory. The first order condition yields the aggregate emission level of the signatories, $E_s = sa\left[1 - \frac{\gamma sn}{\Psi}\right]$, where $\Psi = X^2 + \gamma s^2$ and $X = 1 + \gamma (n - s)$. The individual country's emission level is,

$$e_s = \frac{E_s}{s} = a\left[1 - \frac{\gamma sn}{\Psi}\right] \quad . \tag{4}$$

Substituting the value of E_s into the reaction function of non-signatories yields,

$$e_{ns} = e_s + \frac{a\gamma n(s - X)}{\Psi} \quad . \tag{5}$$

The total emission level by non signatories is $E_{ns} = (n - s)\left[e_s + \frac{a\gamma n(s-X)}{\Psi}\right]$.

The full-cooperative and the pure non-cooperative solutions can be derived as special cases of the above solution. That is, when $s = n$, the problem reduces to the full cooperative solution and $e_s = e_c$, while when $s = 0$, it reduces to the pure non-cooperative solution, and, $e_{ns} = e_{nc}$.

The aggregate emission level $E = E_{ns} + E_s$ is,

$$E = \frac{naX}{\Psi} \quad . \tag{6}$$

Unlike the previous two cases where $e_{nc} > 0$ and $e_c > 0$ always hold, in the coalition formation case we have to restrict the parameters of the model in order to guarantee that our solutions are interior, that is, we need to restrict the parameters so that $e_s > 0$ and $e_{ns} > 0$. The following Proposition establishes the necessary conditions that yield interior solutions.

Proposition 1 *$e_s > 0$ and $e_{ns} > 0$ if and only if $\gamma < \frac{4}{n(n-4)}$ and $n > 4$.*

The intuitive explanation behind these conditions is that for emissions to be positive it must be that the relative impact of damages to benefits is very low (recall that $\gamma = c/b$). Although such a restriction may seem benign at first, it is of great importance since it is this particular condition that restricts the size of the stable coalition to 2, 3 or 4 countries as we formally show in Section 3.

Despite its importance, this condition has been overlooked so far, simply because the model is most commonly defined in terms of abatement efforts rather than in terms of emissions (the prominent example is the work of Barrett (1994)). In Section 4 we convert our model's choice variable to abatement effort and, while establishing the direct link between the two models, we extend the constraint to the converted model as well, validating, thus, the immunity of our results to the selection of choice variable.

The last step in fully formulating our model is the determination of the welfare level of signatories and non-signatories for any given s. This is done by simply substituting the emission levels e_s, e_{ns} and E with their equilibrium values from equations 4, 5 and 6 respectively into the corresponding welfare functions. We denote the indirect welfare function of the signatories by ω_s while that of the non-signatories by ω_{ns}, which take the following form:

$$\omega_s = ba^2 \left[\frac{1}{2} - \frac{n^2\gamma}{2\Psi} \right], \text{ and } \quad \omega_{ns} = ba^2 \left[\frac{1}{2} - \frac{n^2\gamma X^2(1+\gamma)}{2\Psi^2} \right]. \tag{7}$$

The properties of these indirect welfare functions have been neglected in the literature upon the assumption that they are the same with those of the profit functions in the price leadership model developed in D' Aspremont et al. (1983). As we have already argued in the Introduction the two models are not the same and such an assumption is baseless. In fact, in Proposition 2 we illustrate that the properties differ.

Proposition 2 *Consider the indirect welfare functions of signatory and non-signatory countries, $\omega_s(s)$ and $\omega_{ns}(s)$ respectively and let $z^{\min} = \frac{1+\gamma n}{1+\gamma}$.*

1. *Then, $z^{\min} = \arg\min_{s\in\Re\cap[0,n]} \omega_s(s)$.*
2. *$\omega_s(s)$ increases in s if $s > z^{\min}$ and it decreases in s if $s < z^{\min}$,*
3. *the welfare level of non-signatories is less than that of signatories, $\omega_{ns}(s) < \omega_s(s)$ for all $s < z^{\min}$ while,*
4. *the welfare level of non-signatories is more than that of signatories, $\omega_{ns}(s) > \omega_s(s)$ for all $s > z^{\min}$.*
5. *If, moreover, $z^{\min}$ is an integer then the two are equal at $s = z^{\min}$ $\omega_{ns}(z^{\min}) = \omega_s(z^{\min})$.*

Despite the fact that non-monotonic indirect welfare functions, derived from simulations appear in Barrett (1994), the assumption of monotonically increasing indirect welfare functions is made in Carraro and Siniscalco (1997). Moreover, as Proposition 2 shows there exist sufficiently small coalition sizes ($s < z^{\min}$) where a country is better off as a member of the coalition than outside the coalition.

3 The size of stable IEAs

We now proceed with the determination of the size of the stable IEA, denoted by s^*, using the internal and external stability conditions. Recall that the internal stability condition ensures that if a country were to defect unilaterally, its gains from free riding would be outweighed by the adjustment (due to its defection) of the emission levels of the remaining members of the IEA. The external stability condition ensures that no other non-signatory country finds it beneficial to unilaterally join the IEA. Formally, the internal and external stability conditions are,

$$\omega_s(s^*) \geq \omega_{ns}(s^* - 1) \quad \text{and} \quad \omega_s(s^* + 1) \leq \omega_{ns}(s^*) \quad ,$$

respectively.

Unfortunately, allowing s to take non-integer values and then setting $\omega_s(s)$ and $\omega_{ns}(s-1)$ equal, does not provide an analytical solution for z' such that $\omega_s(z') = \omega_{ns}(z'-1)$, and the model has remained, to the best of our knowledge, unsolved. Fortunately, it is not z' that we are interested in per se. Instead, it is the largest integer $s^* \leq z'$ that we are looking for.

We were able to bypass the difficulties of solving the complicated polynomial by "guessing" some value $\bar{z}$, that satisfies the stability conditions, not necessarily with equality, and then adjust it to the appropriate integer. In particular, we show that $\bar{z} = z^{\min} + 1$, satisfies the stability conditions. Next, using the interior solution constraints from Proposition 1 we identify the range of $z^{\min}$ and hence the range of $\bar{z}$. Lastly, since s^* is an integer, we locate the closest integer(s) to $\bar{z}$ that can satisfy the internal and external stability conditions. We conclude that the stable coalition consists of either two, three, or four countries, depending on the parameters of the model.

Proposition 3 *For $n > 4$ there exists a unique stable IEA whose size s^* such that $s^* = \{2, 3, 4\}$.*

We illustrate the results presented in Proposition 3 by considering a numerical example that leads to $s^* = 3$. We assume $n = 10$, $a = 10$, $b = 6$

and $c = 0.39999$, which result in $\gamma = 0.066665$. Observe that $\gamma < \frac{4}{n(n-4)} \Leftrightarrow 0.066665 < 0.066667$ satisfying the interior solution constraint.

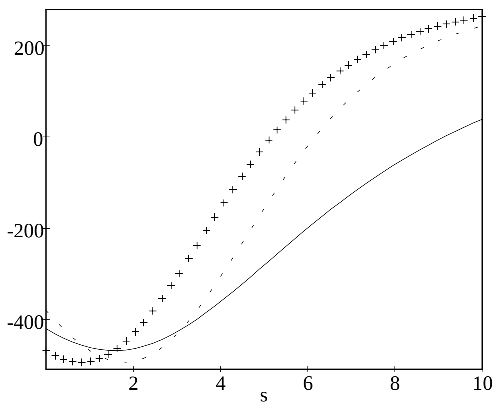

Figure 1

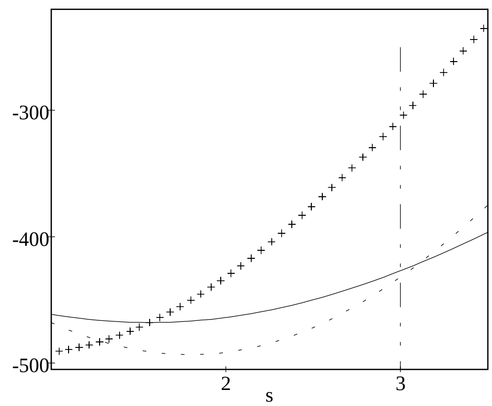

Figure 2

In both Figures 1 and 2 $\omega_s(s)$ is denoted by the solid line, $\omega_{ns}(s)$ is denoted by the crossed line and $\omega_{ns}(s-1)$ is denoted by the dashed line. All three indirect welfare functions are plotted against different coalition sizes s. Observe that $\omega_{ns}(s-1)$ is a horizontal shift of $\omega_{ns}(s)$. While Figure 1 plots the functions for all possible values of $s = 0, ..., 10$, Figure 2 focuses on the values of interest, that is, $s = 0, ..., 4$. Observe that coalition $s^* = 3$ is internally stable, i.e., $\omega_s(s^*) > \omega_{ns}(s^* - 1)$ since the dashed curve is below the solid curve. Moreover, $s^* = 3$ is externally stable, i.e., $\omega_s(s^* + 1) < \omega_{ns}(s^*)$ since $s^* + 1$ is after the intersection of the dashed and the solid curves. Therefore, the coalition of size $s^* = 3$ is stable.

Remark 1 *An important observation stemming from the above analysis is that the size of the stable coalition is slightly larger than that for which the welfare of the signatories is at its minimum. This implies that the welfare of the countries that choose to be members of the IEA, is very close to its lowest possible value.*

Closer to our results, Rubio and Casino (2001) have suggested that a coalition consisting of two countries is the only stable coalition, but, their result is derived by constraining the indirect welfare levels to be positive. Such a constraint is unjustified since welfare functions are invariant to positive monotonic transforms and hence their cardinal values are insignificant.

4 Emissions vs Abatement

As we mentioned in the previous Section, our result in Proposition 3 regarding the size of the stable coalition seems to contradict that of Barrett (1994) where the same type of quadratic benefits and costs functions are used. The only difference between the two models is that in Barrett (1994) the choice variable is abatement effort instead of emission.

In particular, Barrett assumes that countries derive benefits from aggregate abatement Q, with country i's benefits given by $B_i(Q) = \frac{\hat{b}}{n}(\hat{a}Q - \frac{1}{2}Q^2)$. Each country's costs depend on its own abatement, that is, $C_i(q_i) = \frac{\hat{c}}{2}q_i^2$, where $\hat{b}$, $\hat{a}$ and $\hat{c}$ are parameters and n denotes the number of countries.[6] Within this framework, it is asserted (Proposition 1, on page 886) that stable IEAs can be signed by a large number of countries for low values of $\hat{\gamma} = \frac{\hat{c}}{\hat{b}}$, that is, when the importance of own abatement costs is small relative to the benefits derived from aggregate abatement. Although Barrett's findings are based on simulations, the model can be solved in a manner parallel to ours.[7]

Given that the functional forms used in this paper and in Barrett's work are the same while the results differ significantly, one is tempted to conclude that there is no direct correspondence between the two models.[8] However, by its definition abatement effort is a reduction in emissions. In other words, abatement is meaningful only in the presence of emissions, and thus, the level of abatement is constrained by the maximum uncontrolled level of emissions, that is, the abatement model is derived from the emission model.

Denote by $\overline{E}$ the uncontrolled, aggregate emissions level, that is, the level of emissions associated with zero abatement, and by E the controlled emissions level we derived in the previous Section. Observe that the domain of Q stemming from $B_i(Q)$ is some exogenous unconstrained parameter $\hat{a}$. If $\hat{a}$ is to reflect a meaningful upper bound on abatement it should be derived from the emissions model that independently determines the level of uncontrolled

[6] Barrett uses the sympols b, a and c respectively to denote the parameters but we have already used these symbols.

[7] We do not present the analytical solution here, since the process is similar to that presented in the previous Section. We can provide the full solution to the interested reader on demand.

[8] See for example Rubio and Casino (2001), p.5 and footnote 6.

emissions. That is, each country's uncontrolled level of emissions is derived directly from its benefit function $B_i(e_i)$ and it is $\overline{e} = a$, and thus, $\overline{E} = na$. By extension, aggregate and country specific abatement levels are defined as $Q = \overline{E} - E = na - E$, and $q_i = \overline{e} - e_i = a - e_i$ respectively. Substituting these definitions into county i's welfare function defined in terms of abatement yields,

$$w_i = \frac{\hat{b}}{n}\left[\hat{a}\left(na - E\right) - \frac{1}{2}\left(na - E\right)^2\right] - \frac{\hat{c}}{2}(a - e_i)^2 \quad .$$

This expression can take the following form which facilitates direct comparison with the welfare function specified in terms of emissions in equation (1).

$$w_i = \hat{c}\left[ae_i - \frac{1}{2}e_i^2\right] - \frac{\hat{b}}{2n}E^2 + \frac{\hat{b}}{n}(na - \hat{a})E + \left[\hat{b}\hat{a}a - \frac{\hat{b}na^2}{2} - \frac{\hat{c}a^2}{2}\right] \quad . \qquad (8)$$

By setting $\hat{c} = b$, $\hat{b} = nc$ and $\hat{a} = na$, equation (8) reduces to $w_i = b\left[ae_i - \frac{1}{2}e_i^2\right] - \frac{c}{2}E^2 + \frac{cna^2}{2}\left(n - \frac{1}{\gamma n}\right)$, where γ has been defined in Section 2 as $\gamma = \frac{c}{b}$. Note that the last term is just a constant that only scales welfare levels and does not affect the solution of the problem. Therefore, the same solution is derived whether we specify welfare in terms of emissions, that is, $w_i = b\left[ae_i - \frac{1}{2}e_i^2\right] - \frac{c}{2}E^2$, or in terms of abatement, that is, $w_i = \frac{\hat{b}}{n}\left[\hat{a}Q - \frac{1}{2}Q^2\right] - \frac{\hat{c}}{2}q_i^2$, as long as $\hat{c} = b$, $\hat{b} = nc$, $\hat{a} = na$, and $\hat{\gamma} = \frac{\hat{c}}{\hat{b}} = \frac{1}{\gamma n}$. For example, one can derive the abatement level of signatory countries using equation (4) in Section 2 ($e_s = a - \frac{a\gamma sn}{\Psi}$), simply by recalling the definition of abatement, that is, $e_s = \overline{e} - q_s$ which implies that $q_s = \frac{a\gamma sn}{\Psi}$.[9]

Using the above equivalence between the two models we can now support the derived abatement model specification with the necessary constraints from the primary emission model. Recall that Proposition 2 provides the necessary conditions to ensure that the choice variables are positive, that is, $e_s \geq 0$ and $e_{ns} \geq 0$. These constraints though, imply the following conditions for the corresponding abatement levels, $q_s \leq a$, and $q_{ns} \leq a$. Note that the latter constraints are equivalent with the ones stemming from the benefit

[9] Simple parameter transformation using the definitions in the begining of the paragraph yields $q_s = \frac{\hat{a}\alpha\hat{\gamma}}{(\hat{\gamma}+1-\alpha)^2+\alpha^2 n\hat{\gamma}}$, which if multiplied by $n\alpha$ yields the total abatement level of signatory countries, given in equation (6), p. 882, Barrett (1994).

function $B_i(Q)$, that is $Q \leq \hat{a}$ which implies $q \leq \frac{\hat{a}}{n} = \frac{an}{n} = a$. Since the parameters $\hat{a}, \hat{b}$ and $\hat{c}$ are directly derived from the emission model, they carry over the constraints imposed on a, b and c, namely, $\gamma < \frac{4}{n(n-4)} \Longleftrightarrow \frac{c}{b} < \frac{4}{n(n-4)}$. Replacing c and b yields $\frac{\hat{b}/n}{\hat{c}} < \frac{4}{n(n-4)}$ which is equivalent to $\hat{\gamma} = \frac{\hat{c}}{\hat{b}} > \frac{n-4}{4}$.

If these conditions are taken into account, it is immediate that the admissible sizes of a stable coalition reduce to 2, 3, and 4 as was the case in Section 3. To illustrate the equivalence between the two models consider the first example constructed in Barrett (1994). The parameters are $n = 10$, $\hat{a} = 100$, $\hat{b} = 1$ and $\hat{c} = 0.25$, which implies $\hat{\gamma} = \frac{\hat{c}}{\hat{b}} = 0.25$, and the stable coalition allegedly consists of four countries. However, the chosen values of $\hat{b}$ and $\hat{c}$ clearly violate the maximum abatement constraint established earlier, requiring that $\hat{\gamma} > 1.5$. The violation of the maximum abatement constraint is evident from the data presented in Table 1, p. 883, Barrett (1994), since the abatement of signatory countries exceeds the corresponding uncontrolled level of emissions $\overline{e} = \frac{\hat{a}}{n} = 10$. That is, each signatory abates more than it can ever emit. In this case, restricting $\hat{\gamma} > 1.5$ yields stable coalitions consisting of either two or three countries depending on the value of $\hat{\gamma}$. In general, restricting the value of $\hat{\gamma}$ to the admissible range, we find that the stable coalition consists of either two, three or four countries, depending on how close the value of $\hat{\gamma}$ is to its lower bound. Not surprisingly, this result is in accordance with the results in Proposition 3.

In this Section we have established that the results presented in Sections 2 and 3 are independent of the selection of the choice variable. Whether the model is defined in terms of emissions or abatement, or whether it is assumed that each country enjoys a share or the total of benefits from aggregate abatement, has no impact on the size of stable coalitions. The seeming divergence is due to the inability of deriving the domain of abatement independently of the emissions model. In fact, Barrett (1994) recognizes the necessity to impose a maximum constraint on the value of abatement (see footnote 4, p. 880), but observing that $q \leq \hat{a}$ is always true, he proceeds with the assumption that q need not be constraint. However, as we have shown, the upper bound on abatement can only be provided from the primary emission model.

5 Conclusions

The present paper studies the size of stable coalitions of countries that ratify agreements concerning transboundary environmental problems. A coalition is considered stable when no signatories wish to withdraw while no more countries wish to participate. Within this framework we show that, contrary to the general perception in the literature, the welfare levels of both the signatories and the non-signatories do *not* monotonically increase in the size of the coalition. Furthermore, in the case of small coalitions, signatories are better off than non-signatories, while as the coalition grows sufficiently the opposite is true.

We find that the size of the stable coalition is not only very small, but it also does not change when the parameters of the model change. Moreover, it is very close to the worst, in terms of the members' welfare, coalition size.

All these problematic features of a stable coalition suggest that there exists a caveat in the model. An explanation of the results is that when each country acts it does not foresee the disappointing outcome in which it will end up. Instead, it myopically concentrates on its own action ignoring the actions of others. In a companion to this paper we study stability of IEAs when countries behave in a more sophisticated manner and are forward looking.

6 Appendix

Although in our model s is a non-negative integer smaller than n, for the ease of exposition and calculations in the proofs we assume that s is a real number taking values from $[0, n]$. When necessary, at the end of some proofs we convert s back to being an integer.

Proof of Proposition 1. *(e_s>0)* From equation (4) we know that $e_s = a\left[1 - \frac{\gamma s n}{\Psi}\right]$. Hence $e_s > 0 \Leftrightarrow [1+\gamma(n-s)]^2 - \gamma s(n-s) > 0$. Let $A(s) = [1+\gamma(n-s)]^2 - \gamma s(n-s) = 1 + \gamma(n-s)[\gamma(n-s) - (s-2)]$ and consider $\underline{s} = \arg\min_s A(s) = \frac{2\gamma n+2+n}{2\gamma+2}$. For $A(s) > 0$ for all s it suffices that $A(\underline{s}) > 0$. Observe that since $(n-\underline{s}) = \frac{n-2}{2\gamma+2}$ and $(\underline{s}-2) = \frac{(n-2)(2\gamma+1)}{2\gamma+1}$ we have $A(\underline{s}) = 4\gamma n - \gamma n^2 + 4$. Then $A(\underline{s}) > 0 \Leftrightarrow 4\gamma n - \gamma n^2 + 4 > 0 \Leftrightarrow \gamma < \frac{4}{n(n-4)}$ and the latter is true from our hypothesis.

(e_{ns}>0) From equation (5) we know that $e_{ns} = e_s + \frac{a\gamma n(s-X)}{\Psi} = a\left[1 - \frac{\gamma s n}{\Psi}\right] + \frac{a\gamma n(s-X)}{\Psi}$. For $e_{ns} > 0$ it suffices that $[1+\gamma(n-s)](1-\gamma s) + \gamma s^2 > 0$. Let $B(s) = [1+\gamma(n-s)](1-\gamma s) + \gamma s^2$ and consider $\bar{s} = \arg\min_s B(s) = \frac{\gamma n+2}{2\gamma+2}$. For $B(s) > 0$ for all s it suffices that $B(\bar{s}) > 0$. Observe that since $1+\gamma(n-\bar{s}) = \frac{\gamma n(\gamma+2)+2}{2\gamma+2}$ and $(1-\gamma\bar{s}) = \frac{2-\gamma^2 n}{2\gamma+2}$ we have $B(\bar{s}) = \left[\frac{\gamma n(\gamma+2)+2}{2\gamma+2}\right]\left[\frac{2-\gamma^2 n}{2\gamma+2}\right] + \frac{\gamma(\gamma n+2)^2}{(2\gamma+2)^2}$. Notice that for $B(\bar{s}) > 0$ it suffices that $\frac{2-\gamma^2 n}{2\gamma+2} > 0 \Leftrightarrow \gamma < \sqrt{\frac{2}{n}}$. But we already know from our hypothesis that $\gamma < \frac{4}{n(n-4)}$ and since $\frac{4}{n(n-4)} < \sqrt{\frac{2}{n}}$ for all $n \geq 6$ it is indeed the case that $\gamma < \sqrt{\frac{2}{n}}$ if $n \geq 6$. Moreover, when $n = 5$ we have $B(\bar{s}) = -\frac{1}{4}\frac{25\gamma^3 - 20\gamma - 4}{\gamma+1}$. For $B(\bar{s}) > 0$ it suffices that $25\gamma^3 - 20\gamma - 4 < 0$ which is true since $\gamma < \frac{4}{5}$. ■

Proof of Proposition 2.

1-2 Observe that $\frac{\partial \omega_s}{\partial s} = \frac{ba^2\gamma^2 n^2}{\Psi^2}(s-X)$. Thus, $\frac{\partial \omega_s}{\partial s}|_{s=z^{\min}} = 0 \Leftrightarrow z^{\min} = \frac{1+\gamma n}{1+\gamma}$ ($\frac{\partial^2 \omega_s}{\partial s^2} > 0$ for all γ and n satisfying the interior solution constrains). Moreover, observe that $\frac{\partial \omega_s}{\partial s} \lesseqgtr 0$ if $s \lesseqgtr X \Leftrightarrow s \lesseqgtr z^{\min}$.

3-5. Combining the expressions in (7), the welfare of non-signatory countries can be expressed as a function of signatories' welfare as follows: $\omega_{ns} = \omega_s + \frac{ba^2\gamma^2 n^2}{2\Psi^2}(X+s)(s-X)$. Then it is obvious that $\omega_{ns} \lesseqgtr \omega_s$, for $s \lesseqgtr X \Leftrightarrow s \lesseqgtr z^{\min}$. If, moreover, $z^{\min}$ is an integer, then when $s =$

$z^{\min} \Leftrightarrow s = X$ and $\omega_{ns}(z^{\min}) = \omega_s(z^{\min})$.

■

Proof of Proposition 3.

Stability: To illustrate our analysis we use Figure 3 below. The curve $\omega_s(s)$ denotes the welfare of the signatories for a size of coalition s, while curves $\omega_{ns}(s)$ and $\omega_{ns}(s-1)$ denote the welfare of the non-signatories when the size of coalition is s and $s-1$ respectively. Observe that $z^{\min}$ is such that $\omega_s(z^{\min}) = \omega_{ns}(z^{\min})$, while $\bar{z} = z^{\min} + 1$ and notice that $\bar{z}$ satisfies the internal $\omega_s(\bar{z}) > \omega_{ns}(\bar{z}-1)$ and the external $\omega_s(\bar{z}+1) > \omega_{ns}(\bar{z})$ stability conditions in accordance with Lemma 4 below.

Let z' be such that $\omega_s(z') = \omega_{ns}(z'-1)$. From the internal and external stability of $\bar{z}$ we know that $\bar{z} < z' < \bar{z}+1$. Let $I[x]$ denote the largest integer that is less than or equal to (if x is an integer itself) x. Then, the size of the stable coalition s^* is $s^* = I[z']$.

Recall that $z^{\min} = \frac{\gamma n+1}{\gamma+1}$, rearranging the expression yields $\gamma = \frac{z^{\min}-1}{n-z^{\min}}$. We know that $0 < \gamma < \frac{4}{n(n-4)}$, thus, $0 < \frac{z^{\min}-1}{n-z^{\min}} < \frac{4}{n(n-4)}$. From $0 < \frac{z^{\min}-1}{n-z^{\min}}$ we get that $z^{\min} > 1$. From $\frac{z^{\min}-1}{n-z^{\min}} < \frac{4}{n(n-4)}$ we get that $z^{\min} < \frac{n^2}{n^2-4n+4} < 2$ if $n > 6$. Therefore, $1 < z^{\min} < 2$, and by extension $2 < \bar{z} < 3$, and $3 < \bar{z}+1 < 4$, hence $2 < z' < 4$.

Since we know that $2 < z' < 4$ we can conclude that if $z' < 3$ then $s^* = 2$ (this is the case depicted in Figure 3), whereas if $3 \leq z'$ then $s^* = 3$.

Moreover, $1 < z^{\min} < 3$ if $4 < n \leq 6$, hence $2 < \bar{z} < 4$ and $3 < \bar{z}+1 < 5$, and thus $2 < z' < 5$. Then, the size of the stable coalition s^* can take the values

$$\begin{aligned} s^* &= 2 \text{ if } z' < 3 \\ s^* &= 3 \text{ if } z' < 4 \\ s^* &= 4 \text{ if } z' \geq 4 \end{aligned}$$

if $4 < n < 6$. In the special case where $n = 6$ the possibility of $s^* = 4$ is ruled out below when we show the uniqueness of s^*.

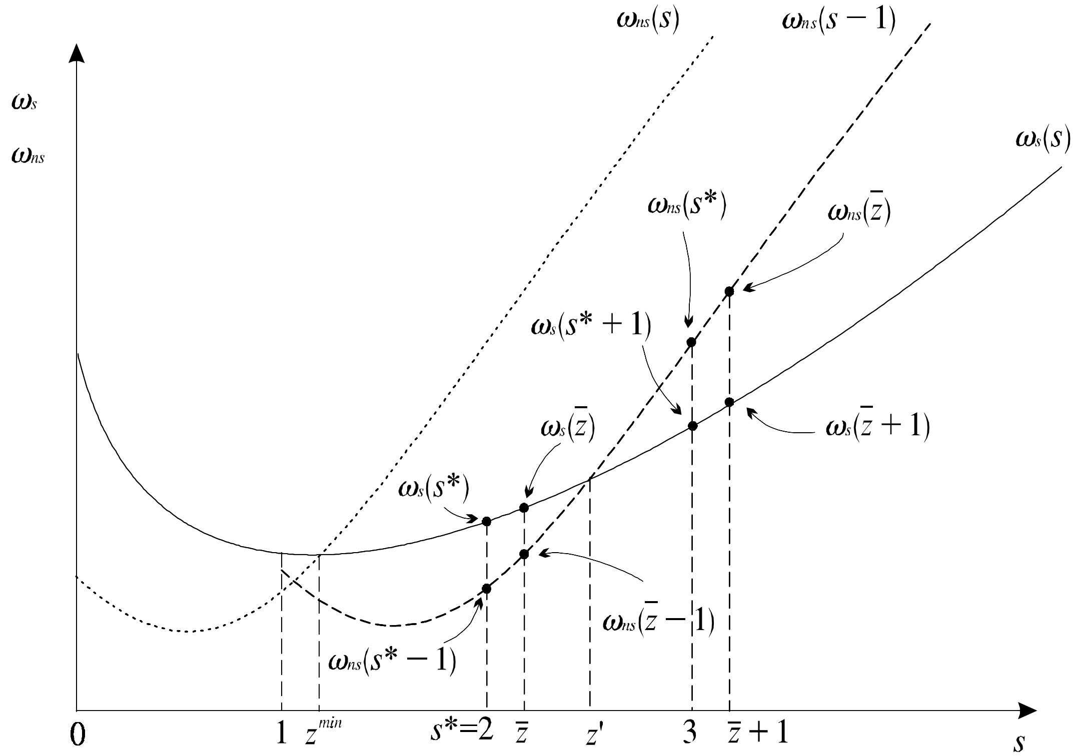

Figure 3

Uniqueness: In order to show that $s^* = 2$, $s^* = 3$ and $s^* = 4$ are the only sizes of stable IEAs -recall that they are mutually exclusive- it suffices to show that all coalitions of size $s \geq 4$ are internally unstable, i.e., $\omega_{ns}(s-1) > \omega_s(s)$ for all $n \geq 6$ since $s = 0$ and $s = 1$ are externally unstable.

Using the expressions in (7) we derive that

$$\omega_{ns}(s-1) - \omega_s(s) =$$

$$\frac{ba^2n^2\gamma}{2}\left[\frac{\Psi^2(s-1) - \Psi(s)\Psi(s-1) + \Psi(s)\gamma(s-1)^2 - \Psi(s)\gamma X^2(s-1)}{\Psi(s)\Psi^2(s-1)}\right].$$

To show that $\omega_{ns}(s-1) - \omega_s(s) > 0$ for all $s \geq 4$ suffices to show that $B(s) = \Psi^2(s-1) - \Psi(s)\Psi(s-1) + \Psi(s)\gamma(s-1)^2 - \Psi(s)\gamma X^2(s-1) > 0$. Substituting all the relevant values, the expression can be further simplified to the following rather long polynomial:

$$\begin{gathered}B=-8\gamma ns+3-4s-12\gamma^2sn-2\gamma^2ns^3+2\gamma^3ns+2\gamma ns^2+\gamma^3\\+5\gamma^2+8\gamma^2n-12\gamma^2s+9\gamma^2s^2+15\gamma s^2+8\gamma-18\gamma s+6\gamma n-2\gamma^3s\\-2\gamma^3s^2+2\gamma^3n-\gamma^3n^2+2\gamma^2n^2-6\gamma^4ns^2+4\gamma^3s^3-6\gamma s^3-2\gamma^3n^3\\-\gamma^4s^2+2\gamma^4s^3-\gamma^4s^4+\gamma^2s^4-4\gamma^2s^3-\gamma^4n^2-2\gamma^4n^3-\gamma^4n^4\\-8\gamma^3ns^2-\gamma^3s^4+4\gamma^4n^3s+\gamma s^4+6\gamma^4n^2s+2\gamma^3s^3n-\gamma^3s^2n^2+2\gamma^4ns\\-4\gamma^2n^2s+8\gamma^2ns^2-6\gamma^4n^2s^2+\gamma^2n^2s^2+4\gamma^4ns^3+6\gamma^3n^2s+s^2\end{gathered}$$

We start by showing that $\omega_{ns}(s-1)-\omega_s(s)>0$ at $s=4$ for all $n\geq 6$ and then we proceed by showing that $B'=\frac{dB}{ds}>0$ for all $s\geq 4$ and for all $n\geq 6$. To do that we show that it is positive at its lowest value, i.e., $B'(\tilde{s})>0$ where $\tilde{s}=\arg\min\ _sB'(s)$. We argue that $\tilde{s}=4$ since $\frac{dB'}{ds}=\frac{d^2B}{ds^2}>0$. The calculations are omitted due to their length and are available upon request.

Lemma 4 *Consider $\bar{z}$ such that $\bar{z}=z^{\min}+1$, then $\bar{z}$ satisfies the internal and external stability conditions.*

Proof.

Internal stability: From Proposition 1 we know that $\omega_s(z^{\min})=\omega_{ns}(z^{\min})$ and that $\omega_s(s)$ increases in s if $s>z^{\min}$. Then, $\omega_s(z^{\min}+1)>\omega_s(z^{\min})$, thus, $\omega_s(z^{\min}+1)>\omega_{ns}(z^{\min})$ which is equivalent to the internal stability condition $\omega_s(\bar{z})>\omega_{ns}(\bar{z}-1)$.

External stability: External stability is shown by substituting $\bar{z}=\frac{\gamma n+1}{\gamma+1}+1$ into the external stability condition $\omega_{ns}(\bar{z})>\omega_s(\bar{z}+1)$. The inequality reduces to $\gamma\frac{2\gamma^2n^3+\left(-3\gamma^2+4\gamma-\gamma^3\right)n^2+\left(8\gamma^3+2\gamma+14\gamma^2+2\right)n+6-\gamma^2-4\gamma^4-11\gamma^3+14\gamma}{(\gamma+1)^3}\geq 0$. It suffices to show that the following inequality holds:

$$\left[\begin{array}{c}2\gamma^2n^3+\left(4\gamma-\gamma^3-3\gamma^2\right)n^2\\+\left(2+14\gamma^2+8\gamma^3+2\gamma\right)n\\+6+14\gamma-\gamma^2-4\gamma^4-11\gamma^3\end{array}\right]\geq 0.$$

Observe that $4\gamma-\gamma^3-3\gamma^2\geq 0$ for $\gamma\leq 1$, while $6+14\gamma-\gamma^2-4\gamma^4-11\gamma^3\geq 0$ for $\gamma<1.0937$. Therefore, the external stability condition is satisfied since $\gamma<\frac{4}{n(n-4)}$ and $n>4$ imply that $\gamma<1$. ■ ■

7 References

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

2. Barrett, S. (1994), "Self-enforcing international environmental agreements." *Oxford Economic Papers*, **46**, 878-894.

3. Carraro, C. and Siniscalco, D. (1993), "Strategies for the international protection of the environment." *Journal of Public Economics*, **52**, 309-328.

4. Carraro, C. and Siniscalco, D. (1998), "International environmental agreements: incentives and political economy." *European Economic Review*, **42**, 561-572.

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. Finus, M. (2000), "Game theory and International environmental cooperation: A survey with an application to the Kyoto-Protocol." *working paper, FEEM*, Nota di Lavoro 86.2000.

7. Finus, M. and Rundshagen B. (1998), "Toward a positive theory of coalition formation and endogenous instrumental choice in global pollution control." *Public Choice*, **96**, 145-186.

8. Folmer, H., Hanley, N. and Mibfeldt, F. (1998), "Game-theoretic modeling of environmental and resource problems: an introduction." In Nick Hanley and Henk Folmer (editors): *Game Theory and the Environment*, chapter 1, pages 1-29. E. Elgar, Cheltenham, UK.

9. Ioannidis, A., Papandreou, A. and Sartzetakis E. (2000), "International environmental agreements: A literature review", *working*

paper, GREEN, Universite Laval, Cahier de recherche du GREEN 00-08.

10. Rubio, J. S. and Casino, B. (2001), "International Cooperation in Pollution Control" *mimeo.*