Stable International Environmental Agreements: An Analytical Approach

Effrosyni Diamantoudi*
Department of Economics, Concordia University

Eftichios S. Sartzetakis
Department of Accounting and Finance, University of Macedonia

October 2001
(This version February 2003)

Abstract

In this paper we examine the formation of International Environmental Agreements (IEAs). We show that the welfare of the signatories does *not* increase monotonically with respect to the number of signatories. We provide an analytical solution of the leadership model. In particular, we find that the unique stable IEA consist of either two, three or four signatories if the number of countries is greater than 4. Furthermore, we show that the welfare of the signatories is almost at its lowest level when the IEA is *stable*. While in our model countries' choice variable is emissions, we extend our results to the case where the choice variable is abatement efforts.

*E-mail addresses: E. Diamantoudi: *ediamant@alcor.concordia.ca* and E. Sartzetakis: *esartz@uom.gr*. Corresponding author: Effrosyni Diamantoudi, Department of Economics, Concordia University, 1455 De Maisonneuve Blvd West, Montreal Quebec H3G 1M8, Canada. The authors would like to thank Licun Xue, Scott Barrett and Carlo Carraro and participants of the CREE 2001 conference for their suggestions. We are also indebted to two referees for their insightful comments

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 front, it is not surprising that there is a growing effort to analyze International Environmental Agreements (IEAs) at the theoretical front. A significant part of the literature on IEAs utilizes game theory to model the formation of a single coalition that reduces pollution. There are two main directions in which the IEA literature has been developed over the last fifteen years. The first argues that the formation of an IEA resembles the voluntary provision of a public good (with externalities) and formalizes countries' behavior as a cooperative game. It shows that an IEA ratified by all countries is stable (Chander and Tulkens (1995) and (1997)). The second direction uses the tools of non-cooperative game theory to model the formation of an IEA. The latter is the direction we follow in this paper.

The non-cooperative approach examines both the case where *all* countries (members or not of the IEA) make their decisions simultaneously (Carraro & Siniscalco (1993)), as well as the case where the countries that have ratified the IEA (signatories) act as a *leader*, whose decision precedes the decision of the countries that remain outside the IEA (Barrett (1994)). The simultaneous case has been resolved by De Cara & Rotillon (2001), Finus & Rundshagen (2001) and Rubio & Casino (2001) leading to the conclusion that, when cost and benefit functions are quadratic, the stable IEA will involve no more than 2 countries. In the leadership approach the most important contributions are based so far on simulations. While the size of the stable IEA remains unknown, simulations in Barrett (1994) suggest that a stable IEA may include a large number of countries, even the grand coalition. We believe that the leadership model, in which an individual country that decides unilaterally will wait to observe the decision of a coalition whose emission will influence significantly global pollution, is compelling enough to warrant further investigation. Thus, in this paper we adopt the leadership model which we solve analytically and our results complement the simulated ones in Barrett (1994).

In particular, each country's welfare (or payoff) is expressed as the difference between the benefits from the country's emissions and the damages from the aggregate emissions. In the leadership 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 and 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 (leader) maximizes its welfare in the first stage, it foresees and takes into account the non-signatories' (followers) behavior. Due to the lack of supra-national authorities that could enforce non-binding agreements, IEAs have to be self-enforcing in the sense that they are immune to deviation by the countries involved . An IEA is considered to be stable if none of its signatories has an incentive to withdraw (this aspect of stability is known as *Internal Stability*) and none of the non-signatories has an incentive to further participate in the agreement (this aspect of stability is known as *External Stability*)[1]. Such a coalitional stability notion was originally introduced by D'Aspremont et. al (1983) in the study of stable cartels in a price leadership model. However, our model and D' Aspremont et al.'s (1983) model differ significantly: (i) while in our model non-members behave strategically in theirs they behave as price takers, (ii) unlike the cartel formation case, in the IEA case members' welfare *does not* increase monotonically with respect to the size of the coalition. We study the problem of deriving the size of a stable IEA in a model very similar to Barrett (1994) with the main difference being the choice variable: in our model countries choose emission levels whereas in his they choose abatement efforts.

The main contribution of this paper is the complete analytical solution of the coalition formation model with quadratic benefit and damage functions. We find that a stable coalition consists of either 2, 3 or 4 members if the total number of countries is greater than 4. Furthermore, we show that the welfare level of the signatories is very close to its lowest value when the IEA is stable. Our results corroborate the outcome of the static models by anticipating very little participation in an IEA. In fact, the predicted size of a stable IEA is so small in both static and dynamic models

[1]The reader should take note that the notions of *Internal* and *External stability* introduced in this paper are completely different from those introduced by von Neumann and Morgenstern (1944) within the concept of the (abstract) stable set.

that contradicts empirical observations[2] and establishes the need for an alternative approach to modeling countries' behavior in international environmental negotiations. Along this vain, Hoel and Schneider (1997) propose a *simultaneous* model similar to that of Carraro and Siniscalco (1993) with one major difference: they introduce non-environmental costs incurred by the non-signatories[3]. In particular, in the primitive simultaneous model, when a country exits the coalition, there are two forces in effect: (i) the increase in its own emissions that results in higher benefits (e.g. due to cheaper production), and (ii) the increase in total emission levels that results in higher damages (e.g. due to an increase in global environmental pollution). If the increase in benefits exceeds the increase in damages the country has indeed an incentive to exit. Hoel and Schneider (1997) introduce an additional cost incurred by the exiting country representing non-environmental costs such as political ones. Naturally, if this additional cost is high enough it may reverse the original incentives, inducing thus the country to stay in the coalition. Due to this effect, Hoel and Schneider (1997) are able to support larger coalitions, including the grand coalition.

Our work parallels that of Konishi and Lin (1999) in terms of the coalition formation analysis employed. However, in Konishi and Lin (1999) the primitive model is cartel formation with Cournot fringe whereas in this paper it is IEA formation. While the two models share many common features, among which free-riding incentives by the coalition members, there are nevertheless, significant differences. As we show in this paper, an IEA can never contain more than 4 countries whereas a cartel, as Konishi and Lin (1999) show, may include a larger number of firms.

Our results, severely restricting the size of stable coalitions, complement Barrett's (1994) suggestion that stable IEAs could consist of any large number of countries by relating the size of a stable coalition to the domain of the choice variable. We convert our model's choice variable from emission levels to abatement efforts making, thus, our model directly comparable to his framework. In doing so, we formulate the

[2] For example, from the 194 members of the United Nations General Assembly, 184 have ratified the Montreal Protocol, 158 the Basel Convention, 164 the Convention on International Trade in Endangered Species and although the Kyoto Protocol has not come into effect yet, it has been ratified by 119 countries, 29 of which belong to the Annex I countries.

[3] Additionally the damage function is linear as opposed to quadratic but this makes their analysis simpler without disturbing the results. It is shown in De Cara & Rotillon (2001) and Finus and Rundshagen (2001) that this simpler version without non-environmental costs results also in very small coalitions.

link between the two approaches and show that our results survive such a conversion. Assuming that abatement cannot exceed the current flow of emissions we show that Barrett's (1994) model yields stable coalitions consisting of no more than 4 members. It is only when abatement can exceed the current flow of emissions, such as in the case of a stock pollutant whose stock could be technologically and economically viable to reduce, that Barrett's (1994) model could support stable IEAs consisting of more than 4 countries. 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 total benefits from country i's emissions, $B_i(e_i)$, and the damages $D_i(E)$ from the aggregate emissions, E, including country i's emissions. Since 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.[4]

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

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

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

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 equation (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, all countries generate the same level of emission at the equilibrium, denoted by e_{nc}. The solution of the reaction functions' system yields,

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

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

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

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 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 Carraro & Siniscalco (1993) and 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 S be $|S| = s$, the total emission 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 $e_{ns} = \frac{a-\gamma s e_s}{1+\gamma(n-s)}$ as a function of the signatory countries' emission e_s. The aggregate non-signatory emission level is $E_{ns} = \frac{(a-\gamma s e_s)(n-s)}{1+\gamma(n-s)}$.

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 maximization problem,

$$\max_{e_s} s\left[B(e_s) - D(se_s + (n-s)e_{ns}(e_s))\right] \ .$$

The first order condition yields the emission of the signatories,

$$e_s = a\left[1 - \frac{\gamma s n}{\Psi}\right] \ , \tag{4}$$

where $\Psi = X^2 + \gamma s^2$ and $X = 1 + \gamma(n-s)$. The aggregate emission level by the signatories is $E_s = sa\left[1 - \frac{\gamma s n}{\Psi}\right]$. Substituting the value of e_s into the reaction function of non-signatories yields,

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

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,

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

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 for interior solutions.

Proposition 1 *$e_s > 0$ if and only if $\gamma < \frac{4}{n(n-4)}$ and $n > 4$, $e_{ns} > 0$ 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 not very high (recall that $\gamma = c/b$). Although such a restriction may seem benign at first, it is of great importance since it is this 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 are established in Proposition 2.

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

1. $z^{\min} = \arg\min_{s\in\Re\cap[0,n]} \omega_s(s)$,

2. $\omega_s(s)$ *increases in* s *if* $s > z^{\min}$ *and it decreases in* s *if* $s < z^{\min}$,

3. $\omega_{ns}(s) \lesseqgtr \omega_s(s)$ *for all* $s \lesseqgtr z^{\min}$.

4. *If, moreover,* $z^{\min}$ *is an integer then the two indirect welfare levels are equal at* $s = z^{\min}$ *that is,* $\omega_{ns}(s^{\min}) = \omega_s(s^{\min})$.

We would like to point out that these indirect welfare functions do not exhibit the same properties with those in D' Aspremont et al. (1983). While in the latter paper the welfare functions are monotonically increasing, in our analysis 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. The difference stems from the fact that in the price leadership model the fringe behaves non-strategically, i.e., 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. Not surprisingly, the same observation is been made in Konishi and Lin (1999).

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,

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

respectively. The following proposition establishes all the possible sizes of the unique stable IEA.

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

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.

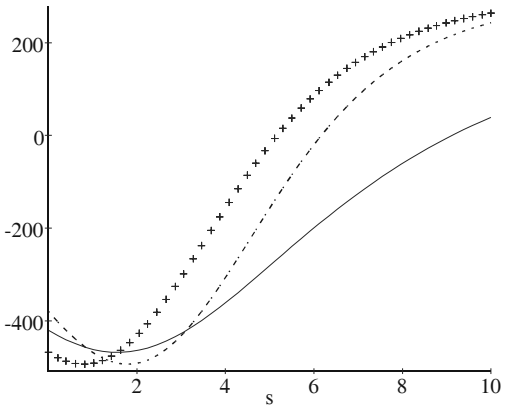

Figure 1

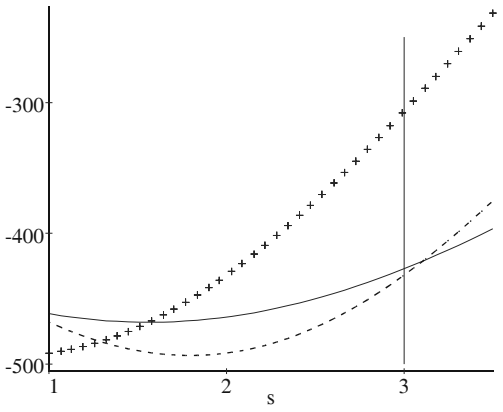

Figure 2

In both Figures 1 and 2 $\omega_s(s)$ is depicted by the solid line, $\omega_{ns}(s)$ by the crossed line and $\omega_{ns}(s-1)$ by the dashed line. All three indirect welfare functions are plotted against different coalition sizes 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, $s = 1, ..., 4$. 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.*

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 complements that of Barrett (1994) where the same type of quadratic benefits and costs functions are used. Although the main difference between Barrett's (1994) and our model is the choice variable, abatement effort and emissions respectively, the difference in the results is not due to the choice variable but rather to the restrictions imposed on the choice variable. As we show in this section the two models are equivalent and all results carry over, as long as abatement does not exceed the flow of emission. The difference arises in the case of a stock pollutant, whose stock could be technologically and economically viable to reduce.[5] In such case it could be possible to abate more than the current flow of emission, that is, it could be possible to have a negative net flow of emission. Then, Proposition 3 does not hold anymore eliminating, thus, the difference between our present work and Barrett's (1994) contribution. Lemma 4, however, applies to both models and the stable IEA remains the largest integer below $z^{\min}+1$, while the pessimistic observation, outlined by Remark 1, that a stable IEA is (almost) the least rewarding to its members is still in effect. That is, when the feasibility constraint is not binding we can have large coalitions whose members attain close to the lowest, in terms of coalition size, net welfare.

In the rest of this section we illustrate that the two models are directly comparable when abatement is defined as a reduction in the flow of emission. Barrett (1994) 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 in Barrett (1994, Proposition 1) 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

[5] For example, in the case of carbon dioxide serious consideration is given recently to the technological option of geological and oceanic storage, usually referred to as carbon sequestration. Although currently at the experimental stage, its pottential is explored both at national and international level. UNFCCC has invited IPCC to prepare a Special Report on the subject, and the results of the first workshop were presented in the 20th Session of IPCC, see IPCC-XX/Doc.19 (10.II.2003).

[6] We added hats in the symbols b, a and c to distinghuish them from the ones we have already used.

to the benefits derived from aggregate abatement. Although the model can be solved in a manner parallel to ours, the goal here is to derive the abatement model from the emission model, establishing the equivalence between them.[7]

If, as discussed earlier, abatement effort is defined as a reduction in the flow of emissions then abatement is meaningful only in the presence of emissions, and thus, the maximum level of abatement is constrained by the maximum uncontrolled flow of emissions. In other words, 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. According to the above definition of abatement the domain of Q, as captured by $B_i(Q)$, should be derived from the emissions model that independently determines the level of uncontrolled 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, country specific and aggregate abatements are then defined as $q_i = \overline{e} - e_i$, and $Q = \overline{E} - E = na - E$ respectively. Substituting these definitions into county i's welfare function defined in terms of abatement yields, $w_i = \frac{\hat{b}}{n}\left[\hat{a}\left(na - E\right) - \frac{1}{2}\left(na - E\right)^2\right] - \frac{\hat{c}}{2}(a - e_i)^2$ .This expression can take the following form which facilitates direct comparison with the welfare function specified in terms of emissions in equation (1).

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

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 $q_s = \frac{a\gamma sn}{\Psi}$.[8]

[7] We can provide the full solution to the interested reader on demand.

[8] Simple parameter transformation using the definitions in the begining of the paragraph yields

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 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' values are $n = 10$, $\hat{a} = 100$, $\hat{b} = 1$ and $\hat{c} = 0.25$, implying $\hat{\gamma} = \frac{\hat{c}}{\hat{b}} = 0.25$, 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 Barrett (1994, Table 1), 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.

5 Conclusions

The present paper studies the size of stable coalitions that ratify IEAs 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

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

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 is also invariant to the value of the model's parameters. 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. One explanation of the results is that when each country acts it does not foresee the disappointing outcome in which it may end up. Instead, it myopically concentrates on its own action ignoring the reactions of others. In a companion to this paper we study stable IEAs when countries behave in a more sophisticated manner and are forward looking.

There are, however, other venues one can explore. Asymmetry among countries has not yet been studied while in the real world it is widespread. For example, not all countries posses identical technologies, leading thus, to varying abating costs. Similarly, the (perceived) impact of environmental damages differs from country to country, hence the Damage function can vary as well. Such asymmetries can be incorporated in our model by indexing parameter γ by country. Then, a coalition will be characterized not only by its size but also by the identity of those in it. Spatial topology is another dimension that can be added to the basic model when regional pollution problems are studied. Emissions from a given country may affect only its neighbors instead of all the countries. A network will be more appropriate in modelling such a situation.

Lastly, in the present work it is assumed that there is only one IEA (hence one coalition). Although it is a natural assumption, it would be very interesting to examine whether it is also the outcome of a model that allows ex ante many coalitions to form. There are several works that model endogenous coalition formation, for example Block (1996) and Ray and Vohra (1999).

6 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. Block, F. (1996), "Sequential Formation of Coalitions in Games with Externalities and Fixed Payoff Division." *Games and Economic Behavior*, **38**, 201-230.

4. Carraro, C. and Siniscalco, D. (1993), "Strategies for the International Protection of the Environment." *Journal of Public Economics*, **52**, 309-328.

5. Chander, P. and Tulkens, H. (1995), "A Core-Theoretic Solution for the Design of Cooperative Agreements on Transfrontier Pollution, " *International Tax and Public Finance*, **2**, 279-93.

6. Chander, P. and Tulkens, H. (1997), "The Core of an Economy with Multilateral Environmental Externalities, " *International Journal of Game Theory,* **26**, 379-401.

7. De Cara S. and Rotillon G. (2001), "Multi Greenhouse Gas International Agreements." working paper.

8. Finus, M. and Rundshagen B. (2001), "Endogenous Coalition Formation Global Pollution Control." *working paper, FEEM,* Nota di Lavoro 43.2001.

9. Hoel, M. and Schneider, K. (1997), "Incentives to Participate in an International Environmental Agreement, " *Environmental and Resource Economics,* **9**, 153-170.

10. Konishi, H. and Lin P. (1999), "Stable Cartels with a Cournot Fringe in a Symmetric Oligopoly." *Keio Economics Studies*, **36**, 1-10.

11. Ray, D. and Vohra, R. (1999), "A Theory of Endogenous Coalition Structures," *Games and Economic Behavior*, **26**, 286-336.

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

13. von Neumann, J. and Morgenstern, O. (1944), "Theory of Games and Economic Behavior." Princeton University Press.

7 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. 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 $(\underline{s} - 2) = \frac{(n-2)(2\gamma+1)}{2\gamma+2}$ we have $A(\underline{s}) = \frac{4\gamma n - \gamma n^2 + 4}{4\gamma+4}$. 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.

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 $\Phi(s) = [1+\gamma(n-s)](1-\gamma s)+\gamma s^2$ and consider $\bar{s} = \arg\min_s \Phi(s) = \frac{\gamma n+2}{2\gamma+2}$. For $\Phi(s) > 0$ for all s it suffices that $\Phi(\bar{s}) > 0$. 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 $\Phi(\bar{s}) = \left[\frac{\gamma n(\gamma+2)+2}{2\gamma+2}\right]\left[\frac{2-\gamma^2 n}{2\gamma+2}\right] + \frac{\gamma(\gamma n+2)^2}{(2\gamma+2)^2}$. Notice that for $\Phi(\bar{s}) > 0$ 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 $\Phi(\bar{s}) = -\frac{1}{4}\frac{25\gamma^3 - 20\gamma - 4}{\gamma+1}$. For $\Phi(\bar{s}) > 0$ 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, $\frac{\partial \omega_s}{\partial s}|_{s=z^{\min}} = 0 \Leftrightarrow z^{\min} = \frac{1+\gamma n}{1+\gamma}$. Since $\frac{\partial^2 \omega_s}{\partial s^2} > 0$ for all γ and n the first order condition is sufficient. Moreover, observe that $\frac{\partial \omega_s}{\partial s} \lesseqgtr 0$ if $s \lesseqgtr X \Leftrightarrow s \lesseqgtr z^{\min}$.

3-4. 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. Unfortunately, allowing s to take non-integer values and then setting $\omega_s(z') = \omega_{ns}(z'-1)$ where $z' \in [0, n]$ does not provide an analytical solution for z' due to computational limitations and the model has remained 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 as we formally explain in the Stability section below.

We are able to bypass the difficulties of solving the complicated polynomial that results from $\omega_s(z') = \omega_{ns}(z'-1)$ by "guessing" some value $\bar{z}$, that satisfies both stability conditions, not necessarily with equality. Then we adjust it to the appropriate integer.

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. By its definition $z^{\min}$ is such that $\omega_s(z^{\min}) = \omega_{ns}(z^{\min})$. We now define $\bar{z}\colon = z^{\min} + 1$, and by Lemma 4 below we deduce that $\bar{z}$ satisfies the internal and external stability conditions: $\omega_s(\bar{z}) \geq \omega_{ns}(\bar{z}-1)$ and $\omega_s(\bar{z}+1) \geq \omega_{ns}(\bar{z})$ respectively.

Let z' be the smallest s such that $\omega_s(z') = \omega_{ns}(z'-1)$. It is straight forward to show[9] that $\omega_s(z^{\min}) > \omega_{ns}(s-1)$ for all $s < \bar{z}$. Then, from the internal and external stability of $\bar{z}$ we can conclude that $\bar{z} < z' < \bar{z}+1$ and hence $\omega_s(s) > \omega_{ns}(s-1)$ for all $s < z'$. Let $\lfloor x \rfloor$ 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 is $s^* = \lfloor z' \rfloor$. The internal stability of s^* ($\omega_s(s^*) \geq \omega_{ns}(s^*-1)$) is satisfied due to the fact that z' is, by definition, the first intersection between $\omega_s(s)$ and $\omega_{ns}(s-1)$, and since $\omega_s(\bar{z}) > \omega_{ns}(\bar{z}-1)$ and $\bar{z} < z'$ we have $\omega_s(s^*) \geq \omega_{ns}(s^*-1)$. Similarly, the external stability of s^*, i.e., $\omega_{ns}(s^*) \geq \omega_s(s^*+1)$, is satisfied since $\omega_{ns}(s-1) > \omega_s(s)$ for all $s > z'$ (as we illustrate below, under the Uniqueness section) and $s^* + 1 > z'$.

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

[9] The calculations are available upon request.

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^*.

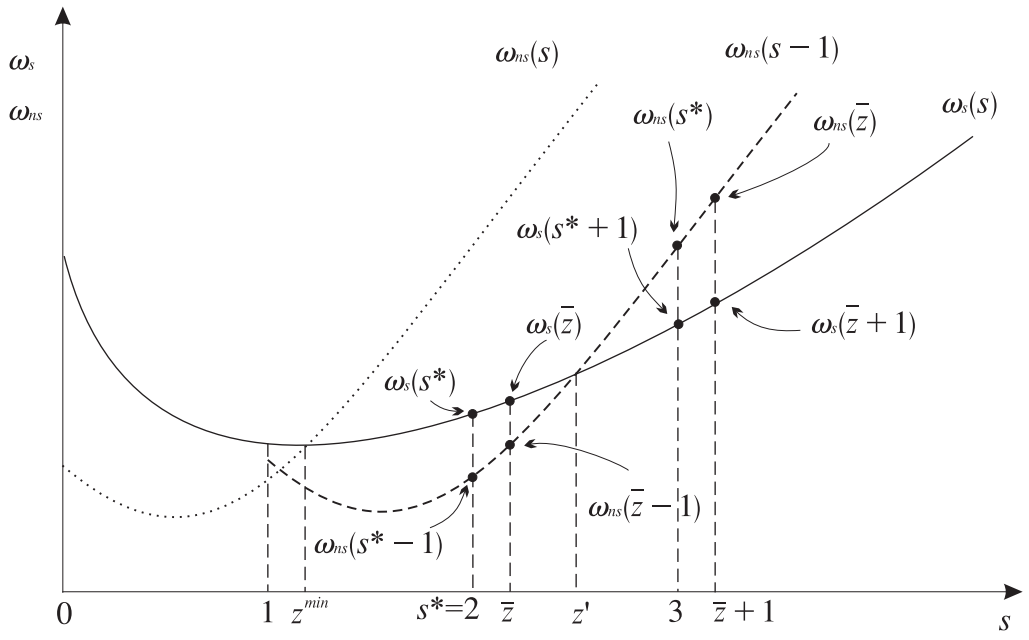

Figure 3

Uniqueness: We have already argued above that $\omega_s(s) > \omega_{ns}(s-1)$ for all $s < z'$. Thus, all coalitions $s < s^*$ are externally unstable since $\omega_s(s+1) > \omega_{ns}(s)$. In order to show that s^* is the only size of a stable IEA it suffices to show that all coalitions of size $s > z'$ are internally unstable, i.e., $\omega_{ns}(s-1) > \omega_s(s)$, for all $n > 4$.

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 $s>z'$suffices to show that $\Xi(s)=\Psi^2(s-1)-\Psi(s)\Psi(s-1)+\Psi(s)\gamma(s-1)^2-\Psi(s)\gamma X^2(s-1)>0$. Substituting all the relevant values, the expression can be further simplified to the following rather long polynomial:

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

We know that $\omega_{ns}(s-1)=\omega_s(s)$ at $s=z'$ for all $n>4$. We proceed by showing that $\Phi'(s)=\frac{d\Phi}{ds}>0$ for all $s\geq\bar{z}$ and for all $n>4$, where $\Phi(s)=\frac{1}{\gamma}\Xi(s)$. To do that we show that it is positive at its lowest value, i.e., $\Phi'(\tilde{s})>0$ where $\tilde{s}=\arg\min_{s\geq\bar{z}}\Phi'(s)$. We argue that $\tilde{s}=\bar{z}$ since $\frac{d\Phi'(s)}{ds}=\frac{d^2\Phi(s)}{ds^2}>0$. 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 $\gamma<1$. ■ ■