International Environmental Agreements - The Impact of Heterogeneity among Countries on Stability

Effrosyni Diamantoudi*
Concordia University

Eftichios Sartzetakis†
University of Macedonia

Stefania Strantza‡
Concordia University

February 2018 (preliminary version)

Abstract

The present paper examines the stability of self-enforcing International Environmental Agreements (IEAs) among heterogeneous countries in a two-stage emission game. In the first stage each country decides whether or not to join the agreement, while in the second stage the quantity of emissions is chosen simultaneously by all countries. We use quadratic benefit and environmental damage functions and assume two types of countries differing in their sensitivity to the global pollution. We solve the model analytically and show that introducing heterogeneity does not enhance the size of the coalition. On the contrary, under heterogeneity, when stable coalitions exist their size is very small, and, if the asymmetry is strong enough, they include only one type of countries. Moreover, heterogeneity can reduce the scope of cooperation relative to the homogeneous case. We show, using an example, that introducing asymmetry into a stable, under symmetry, agreement can disturb stability.

Keywords: Environmental Agreements, *JEL*: D6, Q5, C7

*Effrosyni Diamantoudi, Department of Economics, Concordia University, 1455 De Maisonneuve Blvd West, Montreal Quebec H3G 1M8, Canada (effrosyni.diamantoudi@concordia.ca).

†Eftichios Sartzetakis, Department of Economics, University of Macedonia, 156 Egnatia Str., Thessaloniki 54006, Greece (esartz@uom.edu.gr).

‡Stefania Strantza, Department of Economics, Concordia University, 1455 De Maisonneuve Blvd West, Montreal Quebec H3G 1M8, Canada (stefania.strantza@concordia.ca).

1 Introduction

The most important environmental problem, that of climate change, has an international dimension and thus, it can only be addressed effectively through international cooperation. However, the absence of a supranational authority that could enforce environmental actions on sovereign states necessitates the development of self enforcing agreements. The agreements should provide countries with incentives to join and remain as members. A country will join an agreement if its net benefits as a signatory exceed its net benefits as a non-signatory. D'Aspremont et al. (1983) introduced a concept of coalitional stability whereby no member has an incentive to leave (internal stability) and no non-member has an incentive to join (external stability), assuming that the rest of the agents do not change their membership decision. The notion is essentially a Nash equilibrium where the strategy choice is to join a coalition or not.

The main body of the literature models the formation of IEAs as a two-stage non-cooperative game: in the first stage countries decide whether to join the coalition, while in the second they choose their emission level depending on their membership status. In the second stage, it is assumed that either all countries (signatories or not) choose emissions simultaneously or that the coalition acts as a leader and the non-signatories follow[1]. The subgame perfect Nash equilibrium of the resulting two-stage game is usually derived by applying the notions of the aforementioned internal and external stability conditions.

Although it is clear that all countries benefit from cooperation, each country has strong incentives to free ride on the coalition's efforts. Free-riding incentives increase as the costs of reducing emissions increase. The literature shows that the size of a stable coalition is small, regardless of the total number of countries. Assuming quadratic cost and benefit functions and simultaneous choice of emissions, it has been shown that stable coalitions consist of no more than two countries (De Cara and Rotillon (2001), Finus and Rundshagen (2001), and Rubio and Casino (2001) among others). If the coalition is assumed to be a leader, a stable coalition could have more than two members, but still a maximum of four countries (Barrett (1994), Diamantoudi and Sartzetakis (2006))[2].

[1] The two approches yield similar results.

[2] Barrett 1994 suggests that a stable coalition may achieve a high degree of cooperation, including the grand coalition, but only when an accumulation of stock pollutant is assumed and

One of the most restrictive assumptions of the literature so far is the homogeneity of countries' costs and benefits. It is widely accepted that both damages suffered from a global pollutant and benefits derived from emitting the pollutant (related to production and consumption) differ significantly among countries. The present paper addresses this issue by introducing two types of countries that differ in their sensitivity towards a global pollutant. We find that the introduction of heterogeneity does not yield larger stable coalitions. In particular, we show that the internal stability condition holds only for coalitions with maximum two members from each type of countries. Furthermore, the external stability condition holds only for coalitions consisting of one type of countries, if the asymmetry is strong. Only for very small asymmetry, a mixed coalition consisting of one country from each type is stable. Finally, we prove that coalitions that are stable under asymmetry they may become unstable when asymmetry is introduced. Therefore, the assumption of homogeneity is not the determining factor driving the pessimistic result of a very small coalition.

Despite its apparent importance, only a few papers have addressed the issue of heterogeneity within a theoretical framework, albeit in a limited way. Assuming two types of countries, Barrett (1997) finds no substantial difference in the size of the stable coalition relative to the homogeneous case. On the contrary, McGinty (2007), allowing for transfer payments through a permit system, finds that heterogeneity can increase the coalition size. Chou and Sylla (2008) consider two types of countries (denoted developed and developing) and provide a theoretical framework to explain why it is more likely that some developed countries form a small stable coalition first and then engage in monetary transfers to form the grand coalition. Osmani and Tol (2010) assume also two types of countries but allow the formation of two separate coalitions. They demonstrate that in the case of high environmental damages, forming two coalitions yields higher welfare and better environmental quality relative to a unique coalition. Biancardi and Villani (2010) introduce asymmetry in environmental awareness and find that the coalitions' stability depends on the level of the asymmetry and that the grand coalition can be obtained only by transfers. Fuentes-Albero and Rubio (2010) assume that countries differ either in abatement costs or environmental damages (which are

therefore per period abatement can exceed per period emissions. In contrast, Diamantoudi and Sartzetakis (2006) demostrate that when no stock pollutant is present and emissions must be positive (interior solution) the stable coalition cannot have more than four members.

assumed to be linear) and find that heterogeneity has no important effect without transfers, but if transfers are allowed the level of cooperation increases with the degree of heterogeneity. Finally, Pavlova and Zeeuw (2013) assuming differences in both emission-related benefits and environmental damages (which are assumed to be linear), find that large stable coalitions are possible without transfers if the asymmetries are sufficiently large, however, the gains of cooperation are very low and that transfers could improve gains from cooperation. As the above review indicates, results of the theoretical literature are mixed. Most of the literature based on simulations finds that under some circumstances, heterogeneity may improve coalitions' effectiveness. Some papers support the idea that the introduction of heterogeneity yields larger stable coalitions, with and without transfers, while some others find that transfers are necessary to induce larger stable coalitions.

The present paper derives analytical results and proves that introducing heterogeneity in environmental damages does not increase the size of the coalition. On the contrary, if heterogeneity is strong enough, a smaller stable coalition results relative to the homogeneous case. The main difference between our model and those developed by Fuentes-Albero and Rubio (2010) and Pavlova and Zeeuw (2013) is the functional form of the environmental damages, since they use a linear damage function. With a quadratic environmental damage function the analysis becomes more complex but more interesting as well, since we can capture the interaction effects between heterogeneous countries due to the aggregate global emission level. Our results demonstrate that, introducing asymmetry into a stable under symmetry agreement can disturb stability. Moreover, when stable coalitions exist their size is small and, when the asymmetry is strong enough, they can not include both types of countries. Our analysis also confirms that the symmetric approach is a special case of the asymmetric approach. When we simplify the asymmetric analysis, assuming that there exist only one type of countries, the results from our model can be paralleled with those in Rubio and Casino (2001).

The rest of the paper is structured as follows. Section 2 describes the model and presents the coalition formation. Section 3 solves for the countries' choice of emissions. Section 4, analyses the existence and stability of an IEA when countries are asymmetric in environmental damages, it provides the appropriate constraints to establish stability and presents a counterexample where a stable coalition is not possible. Section 5 concludes the paper.

2 The Model

We consider two types of countries, $j \in \{A, B\}$. We assume that for each type j there exist N^j countries, $N^j = \{1, 2, 3, \ldots, n^j\}$, each of which generate emissions $e_i^j > 0$[3] as a result of their economic activity. The set of all the countries is defined by N, where $N = N^A \cup N^B$. Each country i of type j derives benefits from the economic activity, expressed as function of its emissions, $B_i^j(e_i^j)$ and assumed to be strictly concave, $B_i^j(0) = 0$, $B_i^{j\prime} \geq 0$ and $B_i^{j\prime\prime} < 0$. It also suffers damages from the aggregate emissions of the global pollutant, $D_i^j(E)$, which are assumed to be strictly convex, $D_i^j(0) = 0$, $D_i^{j\prime} \geq 0$ and $D_i^{j\prime\prime} > 0$. In particular we use the following functional forms: $B_i^j(e_i^j) = b^j(a^j e_i^j - \frac{1}{2}(e_i^j)^2)$ and $D_i^j(E) = \frac{1}{2}c^j E^2$, where a^j, b^j and c^j are type specific, positive parameters, and $E = \sum_{j,i=1}^{n^j} e_i^j$ is the aggregate emission level, with $j = \{A, B\}$, that is, $E = \sum_{i=1}^{n^A} e_i^A + \sum_{i=1}^{n^B} e_i^B$.

The social welfare of each country i of type j, W_i^j, is defined as total benefits from its own emissions minus environmental damages from aggregate emissions,

$$W_i^j = B_i^j\left(e_i^j\right) - D_i^j(E).$$

Substituting the specific functional forms, country i's of type j social welfare is,

$$W_i^j = b^j\left(a^j e_i^j - \frac{1}{2}\left(e_i^j\right)^2\right) - \frac{1}{2}c^j\left(\sum_{j,i=1}^{n^j} e_i^j\right)^2, \tag{1}$$

where $j = \{A, B\}$ and $i \in N^j = \{1, 2, 3, \ldots, n^j\}$.

2.1 Coalition formation

We model the process of the heterogeneous countries' decision as a non-cooperative two stage game and we examine the existence and stability of a self-enforcing coalition aiming at controlling emissions. In the first stage, each country i of type j decides whether or not to join the coalition, while in the second stage, chooses its emission level. We assume that only a single coalition can be formed and we determine the equilibrium number of countries participating in the coalition by applying the notions of internal and external stability of a coalition as was

[3] The type of the country is denoted by superscript j and each country is denoted by subscript i.

originally developed by D'Aspremont et. al (1983) and extended to international environmental agreements (IEAs) by Carraro and Siniscalco (1993) and Barrett (1994). We also assume that when a country contemplates joining or defecting from the coalition, assumes that no other country will change its decision regarding participation in the coalition. Furthermore, we assume that members of the coalition act cooperatively, maximizing their joint welfare, while non-members act in a non-cooperative way, maximizing own welfare. Finally, we assume that in the second stage all countries decide their emission level simultaneously.

In particular, for each type $j \in \{A, B\}$ a set of countries $S^j \subset N^j$ sign an agreement to reduce the emissions of the global pollutant and $N^j \backslash S^j$ do not. Each signatory of type j emits e_s^j, such that $E_{s^j} = s^j e_s^j$, where $s^j = |S^j|$and thus the coalition's total emissions are $E_s = E_{s^A} + E_{s^B}$. Similarly, each non-signatory of type j emits e_{ns}^j, such that $E_{ns^j} = (n^j - s^j)e_{ns}^j$, yielding aggregate emissions of non-signatories, $E_{ns} = E_{ns^A} + E_{ns^B}$. Therefore, global emissions are, $E = E_s + E_{ns} = s^A e_s^A + s^B e_s^B + (n^A - s^A)e_{ns}^A + (n^B - s^B)e_{ns}^B$.

3 Choice of emission

Signatories maximize the coalition's welfare given by $W_s = \sum_j s^j W_s^j$, with $j = \{A, B\}$, that is, $W_s = s^A W_s^A + s^B W_s^B$. Therefore, signatories choose e_s^j by solving the following maximization problem,

$$\max_{e_s^j} s^A[B_s^A(e_s^A) - D_s^A(E)] + s^B[B_s^B(e_s^B) - D_s^B(E)], \tag{2}$$

where, $E = s^A e_s^A + s^B e_s^B + (n^A - s^A)e_{ns}^A + (n^B - s^B)e_{ns}^B$. The first order conditions of the signatories' maximization problem (2) yields the equilibrium emission levels,

$$e_s^A = a^A - \frac{\gamma^A(a^A n^A + a^B n^B)(s^A + c^{-1}s^B)}{\Psi}, \tag{3}$$

$$e_s^B = a^B - \frac{\gamma^B(a^A n^A + a^B n^B)(cs^A + s^B)}{\Psi}, \tag{4}$$

where $\gamma^j = \frac{c^j}{b^j}$ indicates the relationship between environmental damages and benefits due to emissions for all countries i in each type $j = \{A, B\}$. Moreover, we define the parameters $c = \frac{c^A}{c^B}$ and $b = \frac{b^A}{b^B}$ as the ratio of the slopes of the marginal environmental damage and marginal benefits respectively, of type $j = A$ over type

$j = B$ countries. Finally, we determine $X = 1 + \gamma^A(n^A - s^A) + \gamma^B(n^B - s^B)$ and $\Psi = X + \gamma^A(s^A)^2 + \gamma^B(s^B)^2 + \gamma^A(b + c^{-1})s^A s^B$. Note that the both X and Ψ are positive since $s^A \leq n^A$ and $s^B \leq n^B$ and $\gamma^j > 0$.

The aggregate emission level by the signatories is, $E_s = s^A e_s^A + s^B e_s^B$. Substituting the equilibrium emission levels we derive,

$$E_s = \frac{(a^A n^A + a^B n^B)X}{\Psi} - a^A(n^A - s^A) - a^B(n^B - s^B). \tag{5}$$

Non-signatories maximize their own welfare given by $W_{ns,}^j$ with $j = \{A, B\}$, by choosing e_{ns}^j.

$$\max_{e_{ns}^j} B_{ns}^j(e_{ns}^j) - D_{ns}^j(E), \tag{6}$$

where $j = \{A, B\}$, and $E = s^A e_s^A + s^B e_s^B + (n^A - s^A)e_{ns}^A + (n^B - s^B)e_{ns}^B$. The first order conditions of the non-signatories' maximization problem (6) yields the equilibrium emission levels,

$$e_{ns}^A = a^A - \frac{\gamma^A(a^A n^A + a^B n^B)}{\Psi}, \tag{7}$$

$$e_{ns}^B = a^B - \frac{\gamma^B(a^A n^A + a^B n^B)}{\Psi}. \tag{8}$$

The aggregate emission level by the non-signatories is, $E_{ns} = (n^A - s^A)e_{ns}^A + (n^B - s^B)e_{ns}^B$. Substituting the equilibrium emission levels we derive,

$$E_{ns} = a^A(n^A - s^A) + a^B(n^B - s^B) - \frac{(a^A n^A + a^B n^B)(X - 1)}{\Psi}. \tag{9}$$

From (5) and (9), the aggregate global emission level is,

$$E = \frac{(a^A n^A + a^B n^B)}{\Psi}. \tag{10}$$

Substituting the equilibrium values of the choice variables from (3), (4), (7) and (8) into equation (1), we derive the indirect welfare function of signatories (W_s^A and W_s^B) and non-signatories (W_{ns}^A and W_{ns}^B) for both types of countries.

The welfare functions of signatories, W_s^A and W_s^B are,

$$W_s^A = \frac{1}{2}b^A \left[(a^A)^2 - \frac{\gamma^A(a^A n^A + a^B n^B)^2(1 + \gamma^A(s^A + c^{-1}s^B)^2)}{\Psi^2}\right], \tag{11}$$

$$W_s^B = \frac{1}{2}b^B \left[(a^B)^2 - \frac{\gamma^B(a^A n^A + a^B n^B)^2(1 + \gamma^B(cs^A + s^B)^2)}{\Psi^2}\right]. \tag{12}$$

The welfare functions of non-signatories, W_{ns}^A and W_{ns}^B are,

$$W_{ns}^A = \frac{1}{2}\left[b^A(a^A)^2 - \frac{\gamma^A(a^An^A + a^Bn^B)^2(1+\gamma^A)}{\Psi^2}\right], \tag{13}$$

$$W_{ns}^B = \frac{1}{2}b^B\left[(a^B)^2 - \frac{\gamma^B(a^An^A + a^Bn^B)^2(1+\gamma^B)}{\Psi^2}\right]. \tag{14}$$

3.1 The case of homogeneity

Before we proceed we compare our result to the homogeneous case. Without loss of generality, we assume that $n^A = n^B = \frac{n}{2}$ and $s^A = s^B = \frac{s}{2}$. Moreover, we simplify the parameters as follows[4]: $a^A = a^B = a$, $b^A = b^B = b^I$, and $c^A = c^B = c^I$. Therefore, in the symmetric case, $c = b = 1$ since $c = \frac{c^A}{c^B}$ and $b = \frac{b^A}{b^B}$. In addition, we define $\gamma = \frac{c^I}{b^I}$, which indicates the relationship between environmental damages and benefits due to emissions for all countries $i \in N = \{1, 2, 3, \ldots, n\}$. Emissions of signatories are e_s, and of non-signatories e_{ns}. The welfare of signatories and non-signatories are W_s and W_{ns}, respectively.

The signatories' equilibrium emission level, e_s, is,

$$e_s = a\left(1 - \frac{\gamma sn}{\Psi}\right), \tag{15}$$

where, $X = 1 + \gamma(n-s)$ and $\Psi = X + \gamma s^2$. The aggregate emission level of signatories is, $E_s = se_s$. Substituting the equilibrium emission level yields,

$$E_s = sa\left(1 - \frac{\gamma sn}{\Psi}\right). \tag{16}$$

The non-signatories' equilibrium emission level, e_{ns}, is,

$$e_{ns} = a\left(1 - \frac{\gamma n}{\Psi}\right), \tag{17}$$

The aggregate emission level by the non-signatories is, $E_{ns} = (n-s)e_{ns}$. Substituting the equilibrium emission level yields,

$$E_{ns} = (n-s)a\left(1 - \frac{\gamma n}{\Psi}\right). \tag{18}$$

From (16) and (18), the aggregate global emission level is,

$$E = \frac{an}{\Psi}. \tag{19}$$

[4]The superscript I is used to define that countries are indentical.

Substituting the equilibrium values of the choice variables from (15) and (17) into equation (1), assuming that there is only one type of countries, we derive the indirect welfare functions of both signatories (W_s) and non-signatories (W_{ns}).

$$W_s = \frac{1}{2} b^I a^2 \left[1 - \frac{n^2 \gamma (1 + \gamma s^2)}{\Psi^2} \right], \tag{20}$$

$$W_{ns} = \frac{1}{2} b^I a^2 \left[1 - \frac{n^2 \gamma (1 + \gamma)}{\Psi^2} \right]. \tag{21}$$

By collapsing the results of the previous Section to homogenous countries, we get the same results derived in Rubio and Casino (2001), noting that we use different notations and a slightly different benefit function.[5] Consequently, as expected, the symmetric country is a special case of the asymmetric assumption.

3.2 Generalizing for k types of countries

We assume that there are k asymmetric types of countries. Let $K = \{1, ..., k\}$ denote the set of types and the letters $m, j \in K$ denote types. For each type $j \in K$ there exists a set of N^j countries, $N^j = \{1, 2, 3, \ldots, n^j\}$, each of which emits $e_i^j > 0$ [6]. Let the total set of countries be $N = \bigcup_{j \in K} N^j$ and the total number of countries be $n = \sum_{j \in K} n^j$.

For each type $j \in K$ a set of countries $S^j \subset N^j$ sign an agreement to reduce the emissions of the global pollutant and $N^j \backslash S^j$ do not. Let $s^j = |S^j|$ for all $j \in K$.Each signatory emits e_s^j, such that $E_{s^j} = s^j e_s^j$, and thus the coalition's total emissions are $E_s = \sum_{j \in K} s^j e_s^j$. Similarly, each non-signatory of type j emits e_{ns}^j, such that $E_{ns^j} = (n^j - s^j) e_{ns}^j$, yielding aggregate emissions of non-signatories, $\sum_{j \in K} (n^j - s^j) e_{ns}^j$. Therefore, the aggregate global emission level is, $E = E_s + E_{ns}$, hence $E = \sum_{j \in K} s^j e_s^j + \sum_{j \in K} (n^j - s^j) e_{ns}^j$, for $j \in K$.

For each type $j \in K$ let the parameter γ^j be the ratio between environmental damages and benefits due to emissions for all countries N^j. Thus,

$$\gamma^j = \frac{c^j}{b^j}. \tag{22}$$

[5] Rubio and Casino (2001) assume that the quadratic benefit function for each country takes the form: $Bi(qi) = aqi - \frac{b}{2} qi^2$ with $a > 0$ and $b > 0$.

[6] The subscript i is used to denote a particular country of some type j.

Moreover, we define the expression Ψ as,

$$\Psi = 1 + \sum_{j\in K} \gamma^j (n^j - s^j) + \sum_{j\in K} \gamma^j (s^j)^2 + \sum_{j\in K} \frac{s^j}{b^j} (\sum_{j\in K} c^j s^j - c^j s^j). \tag{23}$$

The expression Ψ is always positive since $n^j \geq s^j$ and is not type specific. The value of the parameter depends only on the total number of the asymmetric types.

The equilibrium emission level for each type $m \in K$ signatory country is,

$$e_s^m = a^m - \frac{1}{b^m} \frac{\left(\sum_{j\in K} a^j n^k\right)\left(\sum_{j\in K} c^j s^j\right)}{\Psi}, \tag{24}$$

Substituting the equilibrium emission levels yields,

$$E_s = \sum_{j\in K} s^j a^j - \frac{\left(\sum_{j\in K} a^j n^j\right)}{\Psi} \sum_{j\in K} s^j \frac{1}{b^j} \left(\sum_{j\in K} c^j s^j\right). \tag{25}$$

The equilibrium emission level for some non-signatory country of type $m \in K$ is:

$$e_{ns}^m = a^m - \gamma^m \frac{\left(\sum_{j\in K} a^j n^j\right)}{\Psi}, \tag{26}$$

The aggregate emission level by non-signatories is:

$$E_{ns} = \sum_{j\in K} (n^j - s^j) a^j - \frac{\left(\sum_{j\in K} a^j n^j\right)}{\Psi} \sum_{j\in K} (n^j - s^j) \gamma^j. \tag{27}$$

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

$$E = \frac{\sum_{j\in K} a^j n^j}{\Psi}. \tag{28}$$

The indirect welfare function for some signatory of type $m \in K$ is:

$$W_s^m = \frac{1}{2} b^m \left[(a^m)^2 - \gamma^m \frac{\left(\sum_{j\in K} a^j n^j\right)^2}{\Psi^2} \left(1 + \frac{1}{c^m b^m} \left(\sum_{j\in K} c^j s^j\right)^2\right)\right]. \tag{29}$$

The indirect welfare function for some non-signatory country of type $m \in K$ is:

$$W_{ns}^{j} = \frac{1}{2} b^{j} \left[(a^{j})^{2} - \gamma^{j} \frac{\left(\sum_{m=1}^{k} a^{m} n^{m} \right)^{2}}{\Psi^{2}} \left(1 + \gamma^{j} \right) \right]. \quad (30)$$

4 Coalition's membership

To determine the existence and stability of a coalition, we use the notions of the internal and external stability as was originally developed by D'Aspremont et. al (1983) and extended to IEAs by Carraro and Siniscalco (1993) and Barrett (1994). The internal stability implies that no coalition member has an incentive to leave the coalition, while the external stability implies that no country outside the coalition has an incentive to join the coalition. In our case, these conditions are specified for the two types of countries, i.e. $j = \{A, B\}$, since stability requirements should be satisfied by both types. We denote the size of a stable coalition by $\left(s^{A*}, s^{B*}\right)$, where s^{j*} is the number of type j countries signing the agreement.

Formally, for type A and B countries respectively, the internal and external stability conditions take the following form:
internal stability condition,

$$W_{s}^{A}(s^{A*}, s^{B*}) \geq W_{ns}^{A}(s^{A*} - 1, s^{B*}) \text{ and } W_{s}^{B}(s^{A*}, s^{B*}) \geq W_{ns}^{B}(s^{A*}, s^{B*} - 1), \quad (31)$$

external stability condition,

$$W_{s}^{A}(s^{A*} + 1, s^{B*}) \leq W_{ns}^{A}(s^{A*}, s^{B*}) \text{ and } W_{s}^{B}(s^{A*}, s^{B*} + 1) \leq W_{ns}^{B}(s^{A*}, s^{B*}). \quad (32)$$

In this context, a coalition is characterized stable only if all four (two internal and two external) conditions are satisfied at the equilibrium. If even one of these conditions is violated, then the coalition is not stable implying that at least one country of either type $j = A$ or $j = B$ wants to either enter or withdraw from the coalition.

Substituting the values of the indirect welfare functions from (11), (12), (13) and (14), the internal and external stability conditions are derived. The internal stability conditions (I^{j}) for the two types of countries are the following:
Type A countries,

$$I^A = \frac{\gamma^A b^A \Theta}{2}\left[\frac{1+\gamma^A}{(\Psi - 2\gamma^A(s^A-1) - \gamma^A(b+c^{-1})s^B)^2} - \frac{1+\gamma^A\left(s^A + c^{-1}s^B\right)^2}{\Psi^2}\right] \geq 0, \tag{33}$$

where, $\Theta = \left(a^A n^A + a^B n^B\right)^2$.
Type B countries,

$$I^B = \frac{\gamma^B b^B \Theta}{2}\left[\frac{1+\gamma^B}{(\Psi - 2\gamma^B(s^B-1) - \gamma^B(b^{-1}+c)s^A)^2} - \frac{1+\gamma^B\left(cs^A + s^B\right)^2}{\Psi^2}\right] \geq 0. \tag{34}$$

The external stability conditions (E^j) for the two types of countries are the following:
Type A countries,

$$E^A = \frac{\gamma^A b^A \Theta}{2}\left[\frac{1+\gamma^A\left(1+s^A+c^{-1}s^B\right)^2}{\left(\Psi + 2\gamma^A s^A + \gamma^A\left(b+c^{-1}\right)s^B\right)^2} - \frac{1+\gamma^A}{\Psi^2}\right] \geq 0, \tag{35}$$

Type B countries,

$$E^B = \frac{\gamma^B b^B \Theta}{2}\left[\frac{1+\gamma^B\left(1+cs^A+s^B\right)^2}{\left(\Psi + 2\gamma^B s^B + \gamma^A\left(b^{-1}+c\right)s^A\right)^2} - \frac{1+\gamma^B}{\Psi^2}\right] \geq 0. \tag{36}$$

4.1 Existence and stability of a coalition assuming heterogeneity in environmental damages

In order to derive analytical results we restrict the asymmetry between the two types of countries in the environmental damage function, that is, we assume $c^A \neq c^B$ while $a^A = a^B = a$ and $b^A = b^B = b^I$[7]. For simplicity and without any loss of generality we assume $n^A = n^B = n$. Furthermore, without any loss of generality, we assume that $c > 1$, implying that $c^A > c^B$ and since $b = \frac{b^A}{b^B} = 1$, $\gamma^A > \gamma^B$. Therefore, in this context, type $j = A$ countries have a steeper marginal environmental damage function compared to type $j = B$ countries, hence the former are more sensitive to environmental pollution.

[7]The superscript I in parameter b, i.e. b^I, is used to define that countries are indentical with respect to benefits.

Under these assumptions, the internal stability conditions (I^j) given by the equations (33) and (34) can be simplified as follows:
Type A countries,

$$I^A(s^A, s^B) = \frac{\gamma^A b^I (an)^2}{2}\left[\frac{1+\gamma^A}{(\Psi - 2\gamma^A(s^A-1) - \Gamma s^B)^2} - \frac{1+\gamma^A\left(s^A + c^{-1}s^B\right)^2}{\Psi^2}\right], \tag{37}$$

where, $\Gamma = (\gamma^A + \gamma^B)$, $c = \frac{\gamma^A}{\gamma^B}$ (since $b = 1$) and $\Psi = X + \gamma^A(s^A)^2 + \gamma^B(s^B)^2 + \Gamma s^A s^B$ with $X = 1 + \gamma^A(n - s^A) + \gamma^B(n - s^B)$.
Type B countries,

$$I^B(s^A, s^B) = \frac{\gamma^B b^I (an)^2}{2}\left[\frac{1+\gamma^B}{(\Psi - 2\gamma^B(s^B-1) - \Gamma s^A)^2} - \frac{1+\gamma^B\left(cs^A + s^B\right)^2}{\Psi^2}\right]. \tag{38}$$

Similarly, the external stability conditions (E^j) given by the equations (35) and (36) can be simplified as follows:
type A countries,

$$E^A(s^A, s^B) = \frac{\gamma^A b^I (an)^2}{2}\left[\frac{1+\gamma^A\left(1 + s^A + c^{-1}s^B\right)^2}{(\Psi + 2\gamma^A s^A + \Gamma s^B)^2} - \frac{1+\gamma^A}{\Psi^2}\right], \tag{39}$$

type B countries,

$$E^B(s^A, s^B) = \frac{\gamma^B b^I (an)^2}{2}\left[\frac{1+\gamma^B\left(1 + cs^A + s^B\right)^2}{(\Psi + 2\gamma^B s^B + \Gamma s^A)^2} - \frac{1+\gamma^B}{\Psi^2}\right]. \tag{40}$$

We start by checking the internal stability conditions. Lemma 1, states that the internal stability conditions hold only for coalitions with fewer than three members from each type.

Lemma 1 *For all $s^j \geq 3$, the internal stability condition $I^j(s^A, s^B)$ is not satisfied, i.e. $I^j(s^A, s^B) < 0$, for all $j = \{A, B\}$.*

Proof. See appendix A. ■

Therefore, if a stable coalition exists, it can consist of maximum four members since $s^j < 3$, for all $j = \{A, B\}$. Table 1, presents all possible combinations and the appropriate constraints under which a stable agreement is feasible.

Table 1. Stable agreements

		s^B		
		0	1	2
s^A	0		$(0,1)$	$(\mathbf{0},\mathbf{2})$ constraint (43)
	1	$(1,0)$	$(\mathbf{1},\mathbf{1})$ constraint (41)	$(1,2)$
	2	$(\mathbf{2},\mathbf{0})$ constraint (44)	$(2,1)$	$(2,2)$

For all $n > 3$, $\gamma^B > 0$ and $\gamma^A > \gamma^B$, only the combinations along the main diagonal in the above table can support stable agreements. In particular, we have the following three cases:

Case 1:

$(s^A = 1, s^B = 1)$ can support a stable agreement only when $I^j(1,1) \geq 0$ and $E^j(1,1) \geq 0$, for $j = \{A, B\}$. The appropriate constraints that ensure stability are,

$$\gamma^A \leq \frac{1}{2\left(n-2+\sqrt{4n^2-6n+3}\right)},$$

$$\sqrt{\begin{array}{c}(1+\gamma^B)(2+\gamma^B(1+2n))\left(\gamma^B\right)^2+2(2+\left(1-(n-2)n)\gamma^B\right)\gamma^A(\gamma^B)^2\\ +\left[\left(1+\gamma^B-n\left(4+(5n-2)\gamma^B\right)\right)\gamma^B-1\right]\left(\gamma^A\right)^2\\ -2\left(1+2\gamma^B n\right)n\left(\gamma^A\right)^3-n^2\left(\gamma^A\right)^4\end{array}} \leq \gamma^B \leq \gamma^A,$$

$$\gamma^B \geq \frac{2}{5(n-1)+4\sqrt{4n^2-6n+3}}. \tag{41}$$

In the specific case where $\gamma^A = \gamma^B$ the model represents the symmetric case. Under symmetry, a coalition consisting of two countries is the unique self-enforcing IEA as long as,

$$\gamma^A = \gamma^B (= \gamma) \leq \frac{1}{n-4+2\sqrt{n^2-3n+3}}. \tag{42}$$

The derived restriction (42) is identical to the one presented in the literature with symmetric countries (De Cara and Rotillon 2001; Rubio and Casino 2001) under which an agreement of size two is stable.

Case 2:

$(s^A = 0, s^B = 2)$ can support a stable agreement only when $I^B(0,2) \geq 0$ and $E^j(0,2) \geq 0$ for $j = \{A, B\}$. The appropriate constraints that ensure stability are,

$$\gamma^A < \frac{2\sqrt{(1+\gamma^B)(1+4\gamma^B)} - (1+(3n-2))\gamma^B}{3n},$$
$$\frac{2}{5(n-1)+4\sqrt{4n^2-6n+3}} \leq \gamma^B < \frac{1}{2\left(n-2+\sqrt{4n^2-6n+3}\right)}. \tag{43}$$

Case 3:

$(s^A = 2, s^B = 0)$ can support a stable agreement only when $I^A(2,0) \geq 0$ and $E^j(2,0) \geq 0$ for $j = \{A, B\}$. The appropriate constraints that ensure stability are,

$$\gamma^A < \frac{2\left(2+\sqrt{3+n\left[n\left(1-\gamma^B\right)\left(1-4\gamma^B\right)-3\left(1-2\gamma^B\right)\right]}-\left(1+(3n-2)\gamma^B\right)n\right)}{(n-2)(2+3n)},$$
$$\gamma^B < \frac{1}{2\left(n-2+\sqrt{4n^2-6n+3}\right)}. \tag{44}$$

The following Proposition summarizes the results of the above analysis.

Proposition 2 *i) The mixed coalition $(s^A = 1, s^B = 1)$ is stable only under minimal asymmetry, that is, when countries are almost identical (c^A very close to c^B).*
ii) When asymmetry increases, the coalition consists only of one type of countries, When the coalition $(s^A = 0, s^B = 2)$ is stable, the coalition $(s^A = 2, s^B = 0)$ is stable as well.
iii) When the mixed coalition is stable, the other two coalitions, $(s^A = 0, s^B = 2)$ and $(s^A = 2, s^B = 0)$, are stable as well.

Proof. See appendix B. ■

4.1.1 Emission levels

According to the above analysis, a stable agreement can exist under the following cases: Case 1: $(s^A = 1, s^B = 1)$, Case 2: $(s^A = 0, s^B = 2)$, and Case 3: $(s^A = 2, s^B = 0)$. We can now compare the aggregate emission levels among these three possible cases.

Case 1: $(s^A = 1, s^B = 1)$

The aggregate global emission level is,

$$E = \frac{2an}{1 + \Gamma(n+1)}. \tag{45}$$

Case 2: $(s^A = 0, s^B = 2)$
The aggregate global emission level is,

$$E = \frac{2an}{1 + \Gamma n + 2\gamma^B}. \tag{46}$$

Case 3: $(s^A = 2, s^B = 0)$
The aggregate global emission level is,

$$E = \frac{2an}{1 + \Gamma n + 2\gamma^A}. \tag{47}$$

Since $\gamma^A > \gamma^B$, we can easily verify that global emissions are lower in Case 3 and larger in Case 2. Hence, with a high level of asymmetry such that only the coalition $(s^A = 2, s^B = 0)$ satisfies stability, we can achieve the lower level of global emissions.

Lemma 3 *The constraints presented in Section 4.1 guarantee that emissions from both signatories and non-signatories are always positive*

Proof. See appendix C. ■

4.1.2 Cases of instability under heterogeneity

The literature (De Cara and Rotillon 2001; Finus and Rundshagen 2001; Rubio and Casino 2001) has shown that when countries are symmetric, a coalition consisting of two countries is the unique self-enforcing agreement. Nonetheless, when we allow countries to be heterogeneous, the analysis shows that asymmetry can have an inverse effect on stability. Our results, presented in Proposition 4, indicate that heterogeneity has grave implications on the scope of cooperation relative to the homogeneous case. Specifically, we demonstrate that introducing asymmetry into an, under symmetry, stable coalition could disturb stability.

Proposition 4 *Assuming heterogeneous countries, a stable agreement where* $s^{j*} > 1$ *for some* $j = \{A, B\}$ *may not exist, unlike the case of homogeneous countries.*

Proof. To prove the above proposition, we provide a numerical counter-example where a stable coalition does not exist when we relax the homogeneity assumption. We set the following values of the parameters: $a = 10$, b^{I}[8] $= 6$ and $n = 5$ implying that $N = 10$, while $c^A = 0.55$ and $c^B = 0.25$. Using these values, we derive, $\gamma^A = \frac{c^A}{b^I} = 0.091\bar{6}$ and $\gamma^B = \frac{c^B}{b^I} = 0.041\bar{6}$.

Consider first the case that all countries are symmetric and of type B. The condition for the coalition $(s^A = 1, s^B = 1)$ to be stable is given in (42). For the numerical example we examine $n = 5$, the stability condition requires that $\gamma \leq 0.0433125$, which is satisfied given that $b^I = 6$, $n = 10$, and $c^I = 0.25$. Therefore, in the case of ten type B countries, a coalition of two countries is stable (in accordance to the literature) and the aggregate emission level is $E = \frac{2an}{1+2\gamma^B(n+1)} = 66.\bar{6}$.

We now examine stability in the case of two types of countries. Table 2 presents the stability conditions that fail in each of the possible coalitions. As already noted, only the combinations along the main diagonal of the three by three table can support stable agreements as long as the appropriate constraints hold (Section 4.1). However, in all three possible coalitions, $(s^A = 1, s^B = 1)$, $(s^A = 0, s^B = 2)$ and $(s^A = 2, s^B = 0)$, the internal stability condition $I^j(s^A, s^B)$, for $j = \{A, B\}$, is violated. Consequently, stability can be achieved only under the trivial coalition $(s^A = 1, s^B = 0)$ indicating that there is no stable agreement where $s^{j*} > 1$ for some $j = \{A, B\}$.

Table 2. No stable agreement

		s^B		
		0	1	2
s^A	0		$-$	$(0,2)$ $I^B(0,2) < 0$
	1	$(\mathbf{1}, \mathbf{0})$	$(1,1)$ $I^j(1,1) < 0$	$-$
	2	$(2,0)$ $I^A(2,0) < 0$	$-$	$-$

We first check the stability conditions for coalition $(s^A = 2, s^B = 0)$, i.e. constraints (44). The second of these constraints, given that $n = 5$, yields the following range of values for the parameter γ^B, $0.0369167 \leq \gamma^B < 0.0433125$. This

[8]Following the same notation as in Section 3.1, b^I is used to define that countries are indentical with respect to benefits.

condition is satisfied since for the values of the parameters of the present example, $\gamma^B = 0.041\bar{6} \in [0.0369, 0.0433)$.

The first of the conditions in (44), given that $n = 5$ and $\gamma^B = 0.041\bar{6}$, requires that $\gamma^A < 0.0463334$. This condition is not satisfied, since for the values of the parameters in our example, $\gamma^A = 0.091\bar{6} > 0.0463$. Therefore, the first of the constraints in (44) is violated, implying that the coalition $(s^A = 2, s^B = 0)$ is unstable. Note that, both constraints (41) and (43) are more restrictive for the parameter γ^A relative to (44) (Proof, see Appendix B). As a consequence, none of the other two coalitions $(s^A = 1, s^B = 1)$ and $(s^A = 0, s^B = 2)$ can be stable as well. Thus, stability is achieved only under the trivial coalition $(s^A = 1, s^B = 0)$ and the aggregate global emission level is $E = \frac{2an}{1+(\gamma^A+\gamma^B)n} = 60$. ■

Therefore, in the case of symmetric type B countries, a stable agreement of size two is possible. On the contrary, if half of the countries are more sensitive to pollution (higher value of c) relative to the other halve of type B countries, a stable agreement is not always possible. The latter result holds when asymmetry is very strong, that is, the parameters c^A and c^B differ significantly.

Note that aggregate emissions in the case of ten symmetric type B countries, two of which form a coalition to reduce their emissions, are $E = 66.\bar{6}$. In the case of five type A and five type B countries, case that does not allow the formation of any stable coalition, aggregate emissions are $E = 60$. Although this is expected since type A countries, being more sensitive to pollution, emit less than countries B it is worth noting that the existence of stable coalitions is not necessary related to the aggregate emission level.

Figure 1 illustrates the effect of heterogeneity on the stability in the case of the above numerical counterexample. We set $s^B = 0$ and investigate at which s^A the internal stability condition of type A countries is satisfied. In particular, we plot the indirect welfare functions of type A countries against different coalition size s^A when $s^B = 0$. The welfare for the signatories, i.e. $W_s^A(s^A, s^B)$, is depicted by the solid line and the welfare for the non-signatories, i.e. $W_{ns}^A(s^A, s^B)$, is depicted by the dotted line. Moreover, the welfare $W_{ns}^A(s^A - 1, s^B)$ is depicted by the dashed line and represents the welfare for the non-signatories sifted by one (we use that line to represent graphically the internal stability condition).

As indicated in the figure, when $s^B = 0$ the internal stability condition of type A countries, condition (31), is satisfied only at $s^A = 1$. In particular, at this point

Figure 1: Indirect welfare functions of type A countries when $s^B = 0$

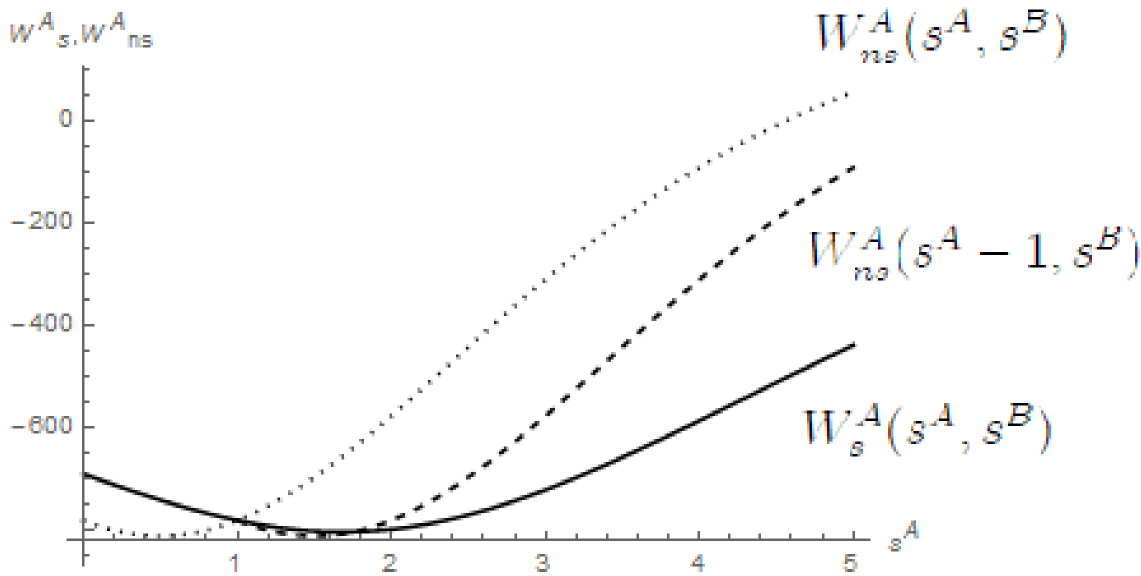

Figure 2:

$(s^A = 1, s^B = 0)$ the internal stability condition is satisfied with equality, i.e. $W_s^A(1,0) = W_{ns}^A(0,0)$. Obviously, at the point $(s^A = 2, s^B = 0)$ the condition is violated since $W_{ns}^A(1,0) > W_s^A(2,0)$. Hence, the only stable coalition is the trivial coalition $(s^A = 1, s^B = 0)$ confirming once more that a stable agreement where $s^{j*} > 1$ for some $j = \{A, B\}$ does not exist.

5 Conclusions

The present paper examines the existence and stability of international environmental agreements in a two stage non-cooperative game assuming heterogeneous countries. In particular, we introduce two types of countries differing in their sensitivity to a global pollutant. A coalition is considered stable when none of the coalition's members wish to withdraw and no country outside the coalition wishes to join. We use quadratic functions and further assume that in the second stage all countries make their decisions simultaneously.

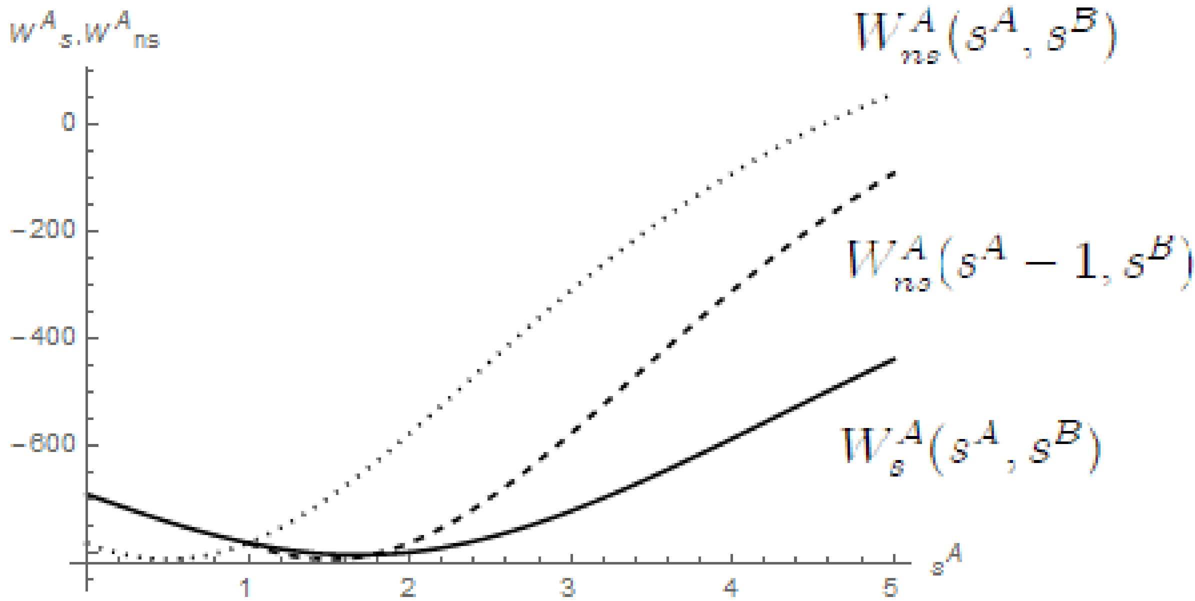

Figure 3

Our results show that, relaxing the, widely used in the literature, assumption of symmetric countries, the size of stable coalitions attempting to mitigate environmental problems remains small. The largest possible coalition that can be achieved includes only two countries and the membership of the coalition is mainly driven by the degree of the asymmetry. In particular, the mixed coalition that includes one country of each type, i.e. $(s^A = 1, s^B = 1)$, is possible only when asymmetry is very small. This case is close to the symmetric case, where according to the literature a coalition of two countries is the unique self-enforcing agreement. When heterogeneity is strong enough, a possible coalition consists of two countries again but they belong only to one type of countries, either type A or type B countries, depending on the level of asymmetry. Under mild heterogeneity, a coalition can contain either two type B countries, i.e. $(s^A = 0, s^B = 2)$, or two type A countries, i.e. $(s^A = 2, s^B = 0)$. However, when the level of heterogeneity is even stronger, the stable coalition can consist only of two type A countries, i.e. $(s^A = 2, s^B = 0)$, and this coalition supports the lower level of aggregate global emissions.

The most important outcome of our analysis is that, heterogeneity can have grave implications on the scope of cooperation in comparison with the homogeneous case. We show that, introducing asymmetry into a stable under symmetry agreement may disturb stability. We provide a counterexample where a coalition does not exist when countries exhibit a strong level of asymmetry in environmental damages. Consequently, heterogeneity leads to an asymmetric distribution of the gains and losses from participating in an agreement, and thus creates stronger free-riding incentives. Our work can be extended by investigating strategies, such as side payments, that can offset these incentives and expand the size of an agreement when countries are heterogeneous.

6 References

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

Barrett, S. (1997). "Heterogenous international environmental agreements". *In: Carraro, C. (Ed.), International Environmental Negotiations*, Edward Elgar, Cheltenham, UK, (Chapter 2).

Biancardi, M. and Villani, G. (2010) "International environmental agreements with asymmetric countries". *Computational Economics*, **36**, 69–92.

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

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

Chou, P.B. and Sylla, C. (2008). "The formation of an international environmental agreement as a two-stage exclusive cartel formation game with transferable utilities". *International Environmental Agreements*, **8**, 317–341.

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

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

Diamantoudi, E. and Sartzetakis, E. (2006). "Stable International Environmental Agreements: An Analytical Approach". *Journal of Public Economic Theory,* ***8***, 247–263.

FINUS, M. AND RUNDSHAGEN, B. (2001). "Endogenous coalition formation global pollution control". *Working paper*, FEEM, Nota di Lavoro 43.2001.

FUENTES-ALBERTO, C. AND RUBIO, S.J. (2010). "Can international environmental cooperation be bought?". *European Journal of Operational Research*, **202**, 255–264.

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

MCGINTY, M. (2007). "International environmental agreements among asymmetric nations". *Oxford Economic Papers*, **59**, 45–62.

OSMANI, D. AND TOL, R.T.J. (2010). "The case of two self-enforcing international agreements for environmental protection with asymmetric countries". *Computational Economics*, **36**, 93–119.

OXFAM INTERNATIONAL (2015)."Extreme Carbon Inequality". *Series: Oxfam Media Briefings*.

PAVLONA, Y. AND DE ZEEUW, A. (2012). "Asymmetries in International Environmental Agreements". *Environment and Development Economics*, **18**, 51–68.

RUBIO, J. S. AND CASINO, B. (2001). "International Cooperation in Pollution Control". *Mimeo*.

RUBIO, J. S. AND ULPH, A. (2006). "Self-Enforcing International Environmental Agreements Revisited". *Oxford Economic Papers*, **58**, 233–263.

SARTZETAKIS, E. AND STRANTZA S. (2013). "International Environmental Agreements: An Emission Choice Model with Abatement Technology". *Discussion paper*, No. 5/2013.

7 Appendices

In what follows we present the proofs of Lemmas and Propositions in the document.

7.1 Appendix A

The internal stability condition of type $j = A$ countries is satisfied only when,

$$I^A(s^A, s^B) = \frac{1}{2}\gamma^A b^I (an)^2$$
$$\left[\frac{(1+\gamma^A)\Psi^2 - (1+\gamma^A(s^A + c^{-1}s^B)^2)(\Psi - 2\gamma^A(s^A-1) - \Gamma s^B)^2}{\Psi^2(\Psi - 2\gamma^A(s^A-1) - \Gamma s^B)^2}\right] \geq 0. \quad (48)$$

The sign of this condition depends on the sign of the expression in the numerator. Hence, for all $\gamma^B > 0$ and $\gamma^A > \gamma^B$ it should be,

$$(1+\gamma^A)\Psi^2 - (1+\gamma^A(s^A + c^{-1}s^B)^2)(\Psi - 2\gamma^A(s^A-1) - \Gamma s^B)^2 \geq 0, \quad (49)$$

Recalling that, $c = \frac{\gamma^A}{\gamma^B}$, $\Gamma = \gamma^A + \gamma^B$, $\Psi = X + \gamma^A(s^A)^2 + \gamma^B(s^B)^2 + \Gamma s^A s^B$ and $X = 1 + \gamma^A(n - s^A) + \gamma^B(n - s^B)$ and rearranging terms we obtain,

$$(1+\gamma^A)(1 + n\Gamma + \gamma^A s^A(s^A-1) + \gamma^B s^B(s^B-1) + \Gamma s^A s^B)^2 -$$
$$(1+\gamma^A(s^A + c^{-1}s^B)^2)(1 + n\Gamma + \gamma^A s^A(s^A-3) + (\gamma^B s^B - \gamma^A)(s^B-2) + \Gamma s^A s^B)^2. \quad (50)$$

The term $(1 + \gamma^A(s^A + c^{-1}s^B)^2)$ is always greater (or at least equal) than the term $(1 + \gamma^A)$ as long as $s^A \geq 1$. Thus, the expression can be positive only if $s^A < 3$ given s^B. For all $s^A \geq 3$ (ceteris paribus) the second term: $(1 + \gamma^A(s^A + c^{-1}s^B)^2)(1 + n\Gamma + \gamma^A s^A(s^A-3) + (\gamma^B s^B - \gamma^A)(s^B-2) + \Gamma s^A s^B)^2$, is greater than the first term: $(1+\gamma^A)(1 + n\Gamma + \gamma^A s^A(s^A-1) + \gamma^B s^B(s^B-1) + \Gamma s^A s^B)^2$, and the internal stability condition is negative.

The internal stability condition of type $j = B$ countries is satisfied only when,

$$I^B(s^A, s^B) = \frac{1}{2}\gamma^A b^I (an)^2$$
$$\left[\frac{(1+\gamma^B)\Psi^2 - (1+\gamma^B(cs^A + s^B)^2)(\Psi - 2\gamma^B(s^B-1) - \Gamma s^A)^2}{\Psi^2(\Psi - 2\gamma^A(s^A-1) - \Gamma s^B)^2}\right] \geq 0. \quad (51)$$

The sign of this condition depends on the sign of the expression in the numerator. Hence, for all $\gamma^B > 0$ and $\gamma^A > \gamma^B$ it should be,

$$(1+\gamma^B)\Psi^2 - (1+\gamma^B(cs^A + s^B)^2)(\Psi - 2\gamma^B(s^B-1) - \Gamma s^A)^2 \geq 0, \quad (52)$$

Rearranging terms we obtain,

$$(1+\gamma^B)(1+n\Gamma+\gamma^A s^A(s^A-1)+\gamma^B s^B(s^B-1)+\Gamma s^A s^B)^2-$$
$$(1+\gamma^B(cs^A+s^B)^2)(1+n\Gamma+(\gamma^A s^A-\gamma^B)(s^A-2)+\gamma^B s^B(s^B-3)+\Gamma s^A s^B)^2. \tag{53}$$

Similarly, the term $(1+\gamma^B(cs^A+s^B)^2)$ is always greater (or at least equal) than the term $(1+\gamma^B)$ as long as $s^B \geq 1$. Thus, the expression can be positive only if $s^B < 3$ given s^A. For all $s^B \geq 3$ (ceteris paribus) the second term: $(1+\gamma^B(cs^A+s^B)^2)(1+n\Gamma+(\gamma^A s^A-\gamma^B)(s^A-2)+\gamma^B s^B(s^B-3)+\Gamma s^A s^B)^2$ is greater than the first term: $(1+\gamma^B)(1+n\Gamma+\gamma^A s^A(s^A-1)+\gamma^B s^B(s^B-1)+\Gamma s^A s^B)^2$, and the internal stability condition is negative.

Therefore, we can conclude that the two internal stability conditions are satisfied simultaneously, i.e. $I^j(s^A, s^B) \geq 0$ for all $j = \{A, B\}$, only if $s^j < 3$.

7.2 Appendix B

In Table 3, we present more analytically all the cases under which we can achieve stable agreements. According to Lemma 1, the two internal stability conditions are satisfied, i.e. $I^j(s^A, s^B) \geq 0$, for all $n > 3$, $\gamma^B > 0$ and $\gamma^A > \gamma^B$ only if $s^j < 3$ for all $j = \{A, B\}$. Consequently, we have to examine only the cases where $s^j \leq 2$ for all $j = \{A, B\}$. Moreover, an agreement is considered stable only if the four stability conditions presented by the equations: (37), (38), (39) and (40) are fulfilled simultaneously at the equilibrium point. The table below includes all the possible coalitions. For each stable coalition, we state the appropriate constraints that ensure stability, while for each non-stable coalition we mention the condition that is violated.

Table 3. Possible coalitions

		s^B		
		0	1	2
s^A	0		$(0,1)$ Trivial coalition	$(\mathbf{0},\mathbf{2})$ constraints (58) and (64)
	1	$(1,0)$ Trivial coalition	$(\mathbf{1},\mathbf{1})$ constraints (60) and (64)	$(1,2)$ $I^B(1,2)<0$ eq. (56)
	2	$(\mathbf{2},\mathbf{0})$ constraint (63)	$(2,1)$ $I^B(2,1)<0$ eq. (55)	$(2,2)$ $I^B(2,2)<0$ eq. (57)

– Trivial coalition

The combinations above the main diagonal of Table 3, i.e. $(s^A = 1, s^B = 0)$ and $(s^A = 0, s^B = 1)$, consist a trivial coalition.

At $(s^A = 1, s^B = 0)$ the internal stability condition of type A countries is satisfied, i.e. $I^A(1,0) = 0$. Moreover, the external stability conditions of both types are satisfied, i.e. $E^j(1,0) \geq 0$, for all $n > 3$, $\gamma^B > 0$, $\gamma^A > \gamma^B$, and $j \in \{A, B\}$.

At $(s^A = 0, s^B = 1)$ the internal stability condition of type B countries is satisfied, i.e. $I^B(0,1) = 0$. The external stability conditions of type B countries is satisfied, i.e. $E^B(0,1) \geq 0$ for all $n > 3$, $\gamma^B > 0$ and $\gamma^A > \gamma^B$. However, the external stability conditions of type A countries is satisfied if,

$$\gamma^B > \frac{1}{2(n-2+\sqrt{4n^2-6n+3})}. \tag{54}$$

– Violation of internal stability

The combinations below the main diagonal of Table 3, i.e. $(s^A = 2, s^B = 1)$, $(s^A = 1, s^B = 2)$ and $(s^A = 2, s^B = 2)$, fail to satisfy the internal stability condition of type B countries. In particular, for all $n > 3$, $\gamma^B > 0$ and $\gamma^A > \gamma^B$ we have:

At $(s^A = 2, s^B = 1)$ the internal stability condition of type B countries is violated, i.e. $I^B(2,1) < 0$, since,

$$I^B(2,1) = -\frac{4\Gamma}{\gamma^B}[(n+2)^2(\gamma^A)^3 + 2(n+2)(1+\gamma^B n)(\gamma^A)^2 + \\ (1+(n-3+((n-1)n-3)\gamma^B)\gamma^B))\gamma^A - (1+\gamma^B)(1+\gamma^B(1+n))\gamma^B] < 0, \tag{55}$$

At ($s^A = 1, s^B = 2$) the internal stability condition of type B countries is violated, i.e. $I^B(1,2) < 0$, since,

$$I^B(1,2) = -\frac{(\Gamma + 2\gamma^B)}{\gamma^B}[(1+(1+n)\gamma^A)^2\gamma^A +$$
$$((1+2n)\gamma^A + 3(1+n)^2(\gamma^A)^2 - 1)\gamma^B + ((4+3n)n\gamma^A - 5)(\gamma^B)^2 + (n^2-4)(\gamma^B)^3] < 0, \tag{56}$$

At ($s^A = 2, s^B = 2$) the internal stability condition of type B countries is violated, i.e. $I^B(2,2) < 0$, since,

$$I^B(2,2) = -\frac{1}{\gamma^B}[4(\gamma^A + (n+4)(\gamma^A)^2)^2 + 4(1+(n+3)(5+4(n+4)\gamma^A)\gamma^A)\gamma^A\gamma^B +$$
$$(2(5n-2)a + (n+2)(86+23n)(\gamma^A)^2 - 5)(\gamma^B)^2 +$$
$$2((7(n+4)n+12)\gamma^A - 14 - n)(\gamma^B)^3 + (n-2)(10+3n)(\gamma^B)^4] < 0. \tag{57}$$

– Possible stable agreements

The only cases under which we can achieve a stable agreement consist of the combinations lying along the main diagonal of Table 3, i.e. $(0,2)$, $(1,1)$ and $(2,0)$.

When $s^A = 0$, ($s^A = 0, s^B = 2$) can support a stable agreement only if the following conditions are satisfied simultaneously:

Internal stability condition of type B countries is satisfied, i.e. $I^B(0,2) \geq 0$. The condition holds if,

$$\gamma^A < \frac{2\sqrt{(1+\gamma^B)(1+4\gamma^B)} - ((3n-2)\gamma^B + 1)}{3n},$$
$$\gamma^B < \frac{1}{2(n-2+\sqrt{4n^2-6n+3})}. \tag{58}$$

External stability condition of type A countries is satisfied, i.e. $E^A(0,2) \geq 0$. The condition holds if,

$$\mathit{Root}[\gamma^B + (n-3+((n-1)n-3)\gamma^A)\gamma^A\gamma^B + 2(n+2)(1+\gamma^A n)(\gamma^B)^2 +$$
$$(n+2)^2(\gamma^B)^3 - (1+\gamma^A)(1+\gamma^A(1+n))\gamma^A\&, 3] \leq \gamma^B < \gamma^A. \tag{59}$$

External stability condition of type B countries is satisfied, i.e. $E^B(0,2) \geq 0$. This condition holds for all $n > 3$, $\gamma^B > 0$ and $\gamma^A > \gamma^B$.

When $s^A = 1$, $(s^A = 1, s^B = 1)$ can support a stable agreement only if the following conditions are satisfied simultaneously:

Internal stability condition of type j countries is satisfied, i.e. $I^j(1,1) \geq 0$ for all $j = \{A, B\}$. The condition holds if,

$$\gamma^A \leq \frac{1}{2(n-2+\sqrt{4n^2-6n+3})},$$
$$\sqrt{\frac{(1+\gamma^B)(2+\gamma^B(1+2n))(\gamma^B)^2 + 2(2+(1-(n-2)n)\gamma^B)\gamma^A(\gamma^B)^2 +}{((1+\gamma^B-n(4+(5n-2)\gamma^B))\gamma^B-1)(\gamma^A)^2 - 2(1+2\gamma^B n)n(\gamma^A)^3 - n^2(\gamma^A)^4}} \leq \gamma^B \leq \gamma^A. \tag{60}$$

External stability condition of type A countries is satisfied, i.e. $E^A(1,1) \geq 0$. The condition holds if,

$$\begin{aligned} Root[(\gamma^B(1+2n) - 1 + 3(1+n)^2(\gamma^B)^2)\gamma^A + ((4+3n)\gamma^B n - 5)(\gamma^A)^2 + \\ (n^2-4)(\gamma^A)^3 + (1+\gamma^B(1+n))^2\gamma^B \&, 3] \leq \gamma^B \leq \gamma^A. \end{aligned} \tag{61}$$

External stability condition of type B countries is satisfied, i.e. $E^B(1,1) \geq 0$. This condition holds for all $n > 3$, $\gamma^B > 0$ and $\gamma^A > \gamma^B$.

We have to mention that, when countries are identical an agreement consisting of two countries is stable as long as,

$$\gamma^A = \gamma^B (= \gamma) \leq \frac{1}{n-4+2\sqrt{n^2-3n+3}}. \tag{62}$$

The constraint (62) is derived by replacing n with $\frac{n}{2}$ in the expression $\frac{1}{2(n-2+\sqrt{4n^2-6n+3})}$, since in the symmetric case $n^A = n^B = \frac{n}{2}$ while in the asymmetric case we assume that $n^A = n^B = n$.

When $s^A = 2$, the point $(s^A = 2, s^B = 0)$ can support a stable agreement only if the following conditions are satisfied simultaneously:

Internal stability condition of type A countries is satisfied, i.e. $I^A(2,0) \geq 0$. The condition holds if,

$$\gamma^A < \frac{2(2+\sqrt{3+(n(1-\gamma^B)(1-4\gamma^B)-3(1-2\gamma^B))n}-(1+(3n-2)\gamma^B)n}{(n-2)(2+3n)},$$

$$\gamma^B < \frac{1}{2(n-2+\sqrt{4n^2-6n+3})}. \tag{63}$$

External stability condition of type j countries is satisfied, i.e. $E^j(2,0) \geq 0$. This condition holds for all $n > 3$, $\gamma^B > 0$, $\gamma^A > \gamma^B$, and $j \in \{A, B\}$.

The constraints (59) and (61) regarding the external stability conditions of type A countries, i.e. $E^A(0,2) \geq 0$ and $E^A(1,1) \geq 0$, are non-binding as long as,

$$\frac{2}{5(n-1)+4\sqrt{4n^2-6n+3}} \leq \gamma^B < \frac{1}{2(n-2+\sqrt{4n^2-6n+3})}. \tag{64}$$

According to the results, for all $n > 3$, $\gamma^B < \frac{1}{2(n-2+\sqrt{4n^2-6n+3})}$ and $\gamma^A > \gamma^B$, the constraint (58) regarding parameter γ^A is always stricter than the constraint (63). Thus,

$$\frac{2(2+\sqrt{3+(n(1-\gamma^B)(1-4\gamma^B)-3(1-2\gamma^B))n}-(1+(3n-2)\gamma^B)n}{(n-2)(2+3n)} >$$
$$\frac{2\sqrt{(1+\gamma^B)(1+4\gamma^B)}-((3n-2)\gamma^B+1)}{3n}$$
$$\text{when } \gamma^B < \frac{1}{2(n-2+\sqrt{4n^2-6n+3})} \text{ for all } n > 3. \tag{65}$$

Consequently, the stability of the coalition $(s^A = 0, s^B = 2)$ ensures the stability of the coalition $(s^A = 2, s^B = 0)$ as well.

Moreover, for all $n > 3$, $\gamma^B < \frac{1}{2(n-2+\sqrt{4n^2-6n+3})}$ and $\gamma^A > \gamma^B$, the constraint (60) regarding parameter γ^A is always stricter than the constraint (58). Thus,

$$\frac{2\sqrt{(1+\gamma^B)(1+4\gamma^B)}-((3n-2)\gamma^B+1)}{3n} >$$
$$\frac{1}{2(n-2+\sqrt{4n^2-6n+3})}$$
$$\text{when } \gamma^B < \frac{1}{2(n-2+\sqrt{4n^2-6n+3})} \text{ for all } n > 3. \tag{66}$$

Therefore, the stability of the coalition ($s^A = 1, s^B = 1$) ensures the stability of the coalition ($s^A = 0, s^B = 2$) as well, and as a consequence the stability of the coalition ($s^A = 2, s^B = 0$).

Consequently, we can conclude that when the coalition ($s^A = 1, s^B = 1$) is stable the other two coalitions, ($s^A = 0, s^B = 2$) and ($s^A = 2, s^B = 0$), are stable as well and when the coalition ($s^A = 0, s^B = 2$) is stable, the coalition ($s^A = 2, s^B = 0$) is also stable. Therefore, if the coalition ($s^A = 2, s^B = 0$) fails to satisfy stability requirements, none of the other two coalitions, i.e. ($s^A = 0, s^B = 2$) and ($s^A = 1, s^B = 1$), can be stable.

7.3 Appendix C

The emissions of signatories are given by the following equations:

$$e_s^A = a^A - \frac{\gamma^A(a^A n^A + a^B n^B)(s^A + c^{-1}s^B)}{\Psi}, \tag{67}$$

$$e_s^B = a^B - \frac{\gamma^B(a^A n^A + a^B n^B)(cs^A + s^B)}{\Psi}. \tag{68}$$

The emissions of non-signatories are given by the following equations:

$$e_{ns}^A = a^A - \frac{\gamma^A(a^A n^A + a^B n^B)}{\Psi}, \tag{69}$$

$$e_{ns}^B = a^B - \frac{\gamma^B(a^A n^A + a^B n^B)}{\Psi}. \tag{70}$$

Where $X = 1 + \gamma^A(n^A - s^A) + \gamma^B(n^B - s^B)$ and $\Psi = X + \gamma^A(s^A)^2 + \gamma^B(s^B)^2 + \gamma^A(b + c^{-1})s^A s^B$.

When countries differ only in environmental damages, the above equations can be simplified as follows:

$$e_s^A = e_s^B = a - \frac{2an(\gamma^A s^A + \gamma^B s^B)}{1 + \gamma^A(n - s^A) + \gamma^B(n - s^B) + \gamma^A(s^A)^2 + \gamma^B(s^B)^2 + \Gamma s^A s^B}, \tag{71}$$

$$e_{ns}^A = a - \frac{2an\gamma^A}{1 + \gamma^A(n - s^A) + \gamma^B(n - s^B) + \gamma^A(s^A)^2 + \gamma^B(s^B)^2 + \Gamma s^A s^B}, \tag{72}$$

$$e^B_{ns} = a - \frac{2an\gamma^B}{1 + \gamma^A(n - s^A) + \gamma^B(n - s^B) + \gamma^A(s^A)^2 + \gamma^B(s^B)^2 + \Gamma s^A s^B}. \quad (73)$$

Emissions from both signatories and non-signatories are positive for all $n > 3$, $\gamma^B > 0$ and $\gamma^A > \gamma^B$, as long as the following conditions are satisfied:

When $(s^A = 0, s^B = 0)$, $(s^A = 1, s^B = 0)$ or $(s^A = 0, s^B = 1)$,

$$\gamma^A < \frac{1 + \gamma^B n}{n}. \quad (74)$$

In all other cases,

$$\gamma^B < \frac{1}{(2n - s^A - s^B)(s^A + s^B - 1)},$$
$$\gamma^A \leq \frac{1 + \gamma^B(n - s^B) + \gamma^B s^B(s^B + s^A - 2n)}{(1 + 2n - s^A - s^B)s^A - n}. \quad (75)$$

Consequently, for the possible stable coalitions $(s^A = 1, s^B = 1)$, $(s^A = 2, s^B = 0)$ and $(s^A = 0, s^B = 2)$, the constraints in (75) simplified as follows,

$$\gamma^B < \frac{1}{2(n - 1)},$$
$$\gamma^A \leq \frac{1}{n - 1} - \gamma^B. \quad (76)$$

We can verify that $2(n-2+\sqrt{4n^2 - 6n + 3}) > 2(n-1)$ for all $n > 3$. Hence, the constraint regarding parameter γ^B, i.e. $\gamma^B < \frac{1}{2(n-2+\sqrt{4n^2-6n+3})}$, is always stricter than the constraint $\gamma^B < \frac{1}{2(n-1)}$. Thus,

$$\frac{1}{2(n - 2 + \sqrt{4n^2 - 6n + 3})} < \frac{1}{2(n - 1)} \text{ for all } n > 3.$$

Therefore, emissions from both signatories and non-signatories are always positive under any possible stable coalition.