ORIGINAL PAPER

Check for
updates

An adaptation-mitigation game: does adaptation promote participation in international environmental agreements?

Miguel Borrero[1] · Santiago J. Rubio[1]

Abstract

This paper studies how the investment in adaptation can influence the participation in an international environmental agreement (IEA) when countries decide in adaptation before they choose emissions. Three types of agreements are studied, a mitigation agreement for which countries coordinate their decisions only on emissions; an adaptation agreement for which there is only coordination when countries decide their levels of adaptation and a complete agreement when there is coordination in both emissions and adaptation levels. In every case, we assume that the degree of effectiveness of adaptation is bounded from above, in order words, adaptation can alleviate the environmental problem, but it cannot solve it by itself leading the vulnerability of the country to almost zero. Our first results show that in our symmetric model where signatories select the same level of adaptation there are not signatory-signatory international externalities and the complete agreement coincides with the mitigation agreement, and moreover it does not matter when adaptation is chosen with respect to emissions. The main contribution of this paper is to show that the grand coalition could be stable for all types of agreement, but only for *extremely* high degrees of effectiveness of adaptation. If this condition is not satisfied, the model predicts low levels of membership. The standard result of three countries is found for the mitigation/complete agreement. For the adaptation agreement participation can be higher than three, but not higher than six countries. In any case, we can conclude that under reasonable values for the degree of effectiveness of adaptation, in our model adaptation does not promote participation in an IEA.

Keywords International environmental agreements · Adaptation-mitigation game · Vulnerability · Effectiveness of adaptation · Complete agreement · Adaptation agreement

Extended author information available on the last page of the article

1 Introduction

Countries can choose between *mitigation* and *adaptation* to face transboundary pollution problems as global warming. The former reduces the amount of emissions and the latter reduces environmental damages without affecting the level of pollution. An important difference between these two types of policies is that mitigation has public/international good characteristics, while adaptation has private/national good characteristics. The previous distinction between adaptation and mitigation states at least two important issues to address. One is the optimal policy-mix the countries should implement. The other is whether adaptation plays against or in favor of international cooperation. The recent literature indicates that adaptation can promote cooperation. Bayramoglu et al. (2018) solve a mitigation-adaptation game and find that the participation in an emission agreement can be high when emissions are strategic complements and both signatories and non-signatories choose their mitigation and adaptation levels simultaneously.[1] On the other hand, Breton and Sbragia (2019) solve an adaptation-mitigation game and find that the participation in an environmental agreement can be high provided that countries cooperate when they decide on their levels of adaptation. The authors analyze two types of agreements with cooperation in adaptation. A complete agreement where signatory countries agree to coordinate both their adaptation and mitigation policies, and an adaptation agreement where signatory countries coordinate only their adaptation policies, while each country decides on emissions individually. In both cases, they consider situations where investments in adaptation requires a prior commitment. Using numerical simulations, they find that the agreement that best performs in terms of participation is the adaptation agreement.[2] This is a very interesting result because the literature on technology agreements is not so optimistic about participation. For instance, Rubio (2017) concludes that for linear damages and quadratic investment costs, the grand coalition could be stable if marginal damages are large enough to justify the development of a "breakthrough" technology and technology spillovers are not very important. Otherwise, participation is low.

For this reason, we think that this is an issue that deserves more attention. In this paper, we analyze the impact that adaptation has on participation when countries decide first on their levels of adaptation as in Breton and Sbragia (2019) paper.[3] Examples of such measures include building infrastructures for water management (dykes, dams, canal systems), change in land use and housing planning. Thus, adaptation can be interpreted as an investment to avoid or reduce damages coming from future emissions, and in this case the adaptation stage must occur before the emissions stage.

---

[1] Marrouch and Ray Chaudhuri (2016) extend the stability analysis of these authors to consider also the Stackelberg scenario where signatories simultaneously choose first their mitigation and adaptation levels and then non-signatories do the same as followers.

[2] Masoudi and Zaccour (2017) also find that an adaptation agreement, where countries decide on investment in adaptation before they select their emissions, can lead to a high level of participation, but they focus on a type of investment in adaptation that presents *imperfect international/public good* characteristics.

[3] This approach has been followed by others authors as Zehaie (2009), Masoudi and Zaccour (2017), Breton and Sbragia (2017) and Harstad et al. (2019) for analyzing different issues.

We propose an extension of the standard model with only mitigation, ex-ante symmetric countries and linear damages for which the maximum level of participation is three countries. In such model, the vulnerability of a country is given by its marginal damages. In the extension we propose in this paper, investment in adaptation reduces marginal damages decreasing in this way the vulnerability of the country. However, what is *new* in our analysis is that we assume that the investment in adaptation can reduce the marginal damages, but not below a positive lower bound.[4] This means that we are assuming that adaptation can alleviate the environmental problem, but cannot solve it taking the marginal damages very close to zero.[5] In other words, we suppose that the degree of the effectiveness of adaptation is bounded from above to eliminate from the model what we could call an *"almost" corner solution*. To evaluate the impact of this assumption in the formation of an international environmental agreement (IEA), we solve an adaptation-mitigation game in three stages considering three types of agreements: a complete agreement, a mitigation agreement and adaptation agreement as in Breton and Sbragia (2019) . In the first stage, countries decide unilaterally on participation. In the second stage, they select the levels of adaptation. For a complete agreement and an adaptation agreement, signatories coordinate their actions at this stage. Finally, countries choose their emissions. At this stage, countries coordinate their actions if they have signed a complete agreement or if they have decided to participate in a mitigation agreement. The game is solved by backward induction.

A first result to highlight is that there are no differences between the complete agreement and the mitigation agreement. As damages are linear, we obtain that in the third stage signatories' emissions and total emissions increase linearly with signatories' total investment in adaptation. Hence, signatory investment in adaptation will produce indirect effects on other signatories' net benefits modifying its national/private good nature due to the timing of the game. However, if all signatories select the same level of adaptation, as expected when all the countries are ex-ante symmetric, and there is cooperation at the emission stage then the signatory-signatory international externalities in the second stage of the game cancel out, and consequently cooperation at this stage of the game is irrelevant.[6] This means that in practice we have only two agreements that yield different outcomes: the adaptation agreement and the mitigation agreement, and, moreover, in the second case it does not matter whether adaptation is chosen before or after emissions.

From the analysis of these two kinds of agreements, we can conclude that the properties of the adaptation subgame played in the second stage coincide with the properties of the model without adaptation. The unique difference is that in the model without adaptation, emissions are strategic substitutes, but with adaptation we find that the levels of adaptation are strategic complements. However, it seems that complementarity does not have a significative influence on the scope on cooperation. In both cases, the grand coalition can be stable but only for a very high degree of adaptation effectiveness. This is our *main contribution* to this literature. We define the way to link the effectiveness of adaptation with

---

[4] Lazkano et al. (2016) have used this kind of model to analyze the consequences that differences in adaptation costs have on the incentives to participate in an IEA. They present conditions under which adaptation can strengthen or weaken free rider incentives. However, they do not impose any lower bound on vulnerability.

[5] This is a standard assumption in the literature of technology innovation. See for instance (Montero 2002).

[6] But this is not the only consequence of the fact that for a symmetric equilibrium there are not signatory-signatory international externalities. It is also straightforward that in this case it does not matter if adaptation is selected before countries choose emissions or if it is selected after or at the same time they choose emissions.

the level of participation in an IEA and we find that only for extremely high values of the degree of effectiveness of adaptation the grand coalition is stable and otherwise the levels of participation are low. For instance, for the mitigation agreement the grand coalition is stable for one hundred countries if signatories are able through investing in adaptation to reduce the marginal damages in a 99.96%. In the case of an adaptation agreement, the figure is very similar. Thus, if we consider that these figures are not reasonable, the results are the standard ones. For the mitigation agreement, only an agreement consisting of three countries can be stable as occurs in the standard model without adaptation. Thus, we can argue that the extension to an adaptation-mitigation game we propose in this paper is a generalization of the pure mitigation game with linear damages. For the adaptation agreement, participation can be higher but not larger than six countries.

The stability analysis we present in the next sections enrolls in a large strand of the literature on the game-theoretic analysis of international environmental agreements (IEAs) which can be traced back to the seminal papers by Carraro and Siniscalco (1993) and Barrett (1994).[7] Surprisingly, in spite of the huge number of paper published on this topic, only a few papers have analyzed formally the effects of adaptation on the participation in an IEA. Aside from the ones already quoted earlier, the list of papers includes Barrett (2020),Bayramoglu et al. (2017), Marrouch & Ray Chaudhuri (2011) and Benchekroun et al. (2017) . Barrett (2020) examines the stability of a mitigation agreement using a linear model in which adaptation and mitigation are both binary actions and the countries select the level of adaptation individually after they have chosen the level of mitigation. His main result establishes that the feasibility of adaptation may increase the participation level in the mitigation agreement, however signatories benefit only marginally from the IEA relative to the noncooperative outcome. Marrouch and Ray Chaudhuri (2011) present a model with linear damages where the non-signatories' emissions are strategic complements of signatories' emissions and the countries decide simultaneously on their levels of emissions and adaptation. In their model the signatories act as the leader of the coalition formation game. Using a numerical example, they show that the more effective the adaptive measure in terms of reducing the marginal damages from emissions, the larger the stable size of the IEA. Benchekroun et al. (2017) show for a model with a quadratic damage function and identical countries where countries' emissions are strategic substitutes and both types of countries, signatories and non-signatories decide simultaneously on the levels of adaptation and emissions that a more efficient adaptation technology diminishes the incentives of individual countries to free-ride on a global agreement over emissions. However, they do not clarify whether the grand coalition could be stable.[8]Battaglini and Harstad (2016); Bayramoglu et al. (2018) claim that if adaptation does that emissions are complements in the second stage of the game when countries select their level of emissions, adaptation will always lead to larger stable agreements with lower aggregate emissions and higher global welfare. In all these models, investment in adaptation is considered a private/national good and countries select the level of adaptation at the same time they select their levels of emissions or after this decision has been taken. They also assume that emissions can

---

[7] A nice collection of the most influential papers in the field has been published by Finus and Caparrós (2015). A very complete review of the literature on IEAs can be found in Marrouch and Ray Chaudhuri (2016).

[8] Li and Rus (2019) extend this model for heterogeneous countries showing that technological progress in adaptation can foster an IEA. They use a numerical example with parameters estimated from climate change data.

be strategic substitutes or strategic complements.[9] In our model, we focus on investment in adaptation involving long-term planning. For this kind of investments, countries must act in anticipation of mitigation policies. One of the most similar papers which share this last aspect with our paper is Breton and Sbragia (2019). However, it is difficult to compare results with the ones developed by these authors, given that their damage function is different depending quadratically on vulnerability and they do not link participation levels with adaptation effectiveness. Moreover, their numerical results are more optimistic than ours except for the mitigation agreement where they coincide.

To end this review of the literature, we would like to add that besides the investment in adaptation other papers have studied the impact of investment in green technologies that reduces the abatement costs on the stability of IEAs. Among other papers, we could mention those published by Barrett (2006), Hoel and de Zeeuw (2010), Harstad (2012, 2016), Hong and Karp (2012), El-Sayed and Rubio (2014), Helm and Schmidt (2015), Battaglini and Harstad (2016), Goeschl and Perino (2017), Rubio (2017) and Harstad et al. (2019). One of the issues examined by this literature is to know whether a technology agreement could be a good alternative to an emission agreement.[10]

The paper is organized in four sections. In the next section, Section 2, we present the model and in Section 3 we analyze the scope of cooperation in a mitigation agreement. Section 4 analyzes the case of an adaptation agreement, in Section 5 an agreement comparison is presented, Section 6 closes the paper with the conclusions and the presentation of different issues for future research.

2 The model

We consider a model with N countries where each country emits a global pollutant as a result of its consumption and production activities. We let e_i stand for the emission level of country i where $i = 1, ..., N$, and $E = \sum_{i=1}^{N} e_i$ are total emissions. While total emissions damage all countries, each country can reduce the negative effects of pollution by mitigation and/or investing in adaptation. Let a_i represent the adaptation level of country i. A key difference in our paper between emissions and adaptation lies in the international public good nature of pollution and the national private good nature of adaptation.[11] While each country's emissions are a national decision, pollution is a global public bad that creates free-riding incentives on emission abatement. Instead, adaptation is a national decision with country-specific benefits and costs.

---

[9] Masoudi and Zaccour (2018) analyze the stability of a complete agreement on investment in adaptation and emissions where countries decide simultaneously on the levels of these two variables. However, they assume as in Masoudi and Zaccour (2017) that investment in adaptation is an imperfect global public good.

[10] We would like to quote also the paper by Caparrós (2018). This author shows that short-term agreements following an incomplete long-term agreement, as the Paris Agreement, cannot achieve the first best solution but it can improve upon the situation without a long-term agreement. In his model, countries invest to reduce the abatement costs after the long-term agreement is signed but before the state of nature that determines the benefit of total abatement is realized.

[11] One might argue that adaptation could also have an international dimension. We abstract from this possibility because our aim in this paper is to study how country incentives to participate in an IEA change when national adaptation is available. See (Masoudi and Zaccour 2017, 2018), for the analysis of international cooperation when adaptation presents an imperfect international public good characteristic.

Each country's net benefits consist of benefits from pollution activities minus emission damages and adaptation costs. Global pollution damages all countries; however, each country has the option to offset damages through adaptation. Country $i's$ benefits from emissions are

$$B(e_i) = \alpha e_i - \frac{\gamma}{2}e_i^2, \quad \alpha, \gamma > 0,$$

and the damage function is[12]

$$D(a_i, E) = (d - a_i)E, \quad d > a_i > 0.$$

As usual we assume that environmental damages cannot be completely eliminated through adaptation. The cost of reducing the marginal damages is increasing and is given by $C(a_i) = ca_i^2/2, \ c > 0.$[13] Thus, the net benefit for country i are

$$W_i(a_i, e_i, E_{-i}) = \alpha e_i - \frac{\gamma}{2}e_i^2 - (d - a_i)(e_i + E_{-i}) - \frac{c}{2}a_i^2, \tag{1}$$

where $E_{-i} = \sum_{j \neq i} e_j$.

3 The complete agreement and the mitigation agreement

The formation of an IEA is modeled as a three-stage game. Each stage will be now described briefly in reverse order as the subgame-perfect equilibrium of the game is computed by backward induction.

Given the participation in the agreement and the investment in adaptation of all countries, in the third stage, the emission subgame, signatory countries choose their emissions so as to maximize the agreement net benefits taking as given non-signatories' emissions. Non-signatories choose the level of emissions acting non-cooperatively and taking the emissions of all other countries as given in order to maximize their national net benefits. Signatories and non-signatories choose emissions levels simultaneously. In the second stage, the adaptation subgame, we are going to consider both the possibility of cooperation (the complete agreement) and the possibility of no cooperation (the mitigation agreement). In the first case, countries act as in the third stage, but now they decide on investment in adaptation. In the second case, there is no cooperation when the countries decide on adaptation and the agreement only obligates the countries to coordinate their decisions on emissions. Finally, it is assumed that in the first stage countries play a *simultaneous open membership game with a single binding agreement*. In a single agreement formation game, the strategies for each country are to sign or not to sign and the agreement is formed by all players who have chosen to sign. Under open membership, any country is free to join the agreement. Lastly, we assume that the signing of the agreement is binding on signatories. The game finishes when the emission subgame is over.

---

[12] This specification of the damage function is based on the one proposed by d'Aspremont and Jacquemin (1988) to study the effects of R&D on the cooperation in a duopolistic market. Since then it has been intensively used in the IO literature. The authors represent the R&D variable as a reduction in the marginal cost of production.

[13] Notice that the marginal cost is increasing indicating that the resources invested to reduce damages present decreasing returns.

3.1 The third stage: an equilibrium in dominant strategies

As for the two kinds of agreements non-signatories countries do not cooperate in the third stage, optimal emissions can be calculated by maximizing (1) given that participation is decided in the first stage and adaptation in the second stage.

The first-order condition (FOC) for an *interior solution* are

$$\alpha - \gamma e_i^f = d - a_i^f, \quad i = 1, ..., N - n, \tag{2}$$

where f stands for a non-signatory countries and n represents the number of signatories so that $N - n$ is the number of non-signatories. This condition establishes that the marginal benefits of emissions must be equal to the national marginal damages. Thus, non-signatories only take into account the effect that emissions have on its national damages.

Then, emission are given by

$$e_i^f = \frac{\alpha - d}{\gamma} + \frac{a_i^f}{\gamma}. \tag{3}$$

Notice that an increase in adaptation leads to higher emissions.

On the other hand, signatories choose the level of emissions to maximize the agreement net benefits taking as given the non-signatories' emissions

$$\max_{\{e_1^s,...,e_n^s\}} W_A = \sum_{j=1}^{n} \left\{ \alpha e_j^s - \frac{\gamma}{2}(e_j^s)^2 - (d - a_j^s)(e_j^s + E_{-j}) - \frac{c}{2}(a_j^s)^2 \right\},$$

where s stands for a signatory country. The FOCs for this problem are

$$\alpha - \gamma e_j^s = \sum_{k=1}^{n}(d - a_k^s) = nd - A^s, \quad j = 1, ..., n, \tag{4}$$

where $A^s = \sum_{k=1}^{n} a_k^s$.

As in condition (2), the LHS is the marginal benefit of emissions. However, the signatories take into account the increase in damages for the rest of signatories caused by the increase in its own emissions.

Thus emissions for signatories are given by

$$e_j^s = \frac{\alpha - nd}{\gamma} + \frac{A^s}{\gamma}. \tag{5}$$

Emissions increase with adaptation, but in this case signatories' emissions depend on the total adaptation of signatory countries. Moreover, it is clear that all signatories will choose the same level of emissions.

As environmental damages are linear, the countries' reaction functions for emissions are orthogonal for both signatories and non-signatories, and the optimal emissions are given by an equilibrium in dominant strategies.

Using (3) and (5) we obtain the following expression for total emissions

$$E = \sum_{i=1}^{N-n} e_i^f + \sum_{j=1}^{n} e_j^s = \frac{1}{\gamma}(N\alpha - (N - n + n^2)d + A^f + nA^s), \tag{6}$$

where $A^f = \sum_{i=1}^{N-n} a_i^f$.

Next, using (1), net benefits can be written as follows for non-signatories

$$W_i^f = \frac{\alpha}{\gamma}(\alpha - d + a_i^f) - \frac{1}{2\gamma}(\alpha - d + a_i^f)^2 - (d - a_i^f)E - \frac{c}{2}(a_i^f)^2, \quad i = 1, ..., N - n, \quad (7)$$

and as follows for signatories

$$W_j^s = \frac{\alpha}{\gamma}(\alpha - nd + A^s) - \frac{1}{2\gamma}(\alpha - nd + A^s)^2 - (d - a_j^s)E - \frac{c}{2}(a_j^s)^2, \quad j = 1, ..., n, \quad (8)$$

where total emissions are given by (6).

Observe that although the investment in adaptation is a national good, if countries decide on adaptation before they select the level of emissions, the investment in adaptation generates indirectly *international externalities* through the effect that adaptation has on total emissions. The following derivatives capture this effect

$$\frac{\partial W_i^f}{\partial a_j^s} = -(d - a_i^f)\frac{\partial E}{\partial a_j^s}, \quad \frac{\partial W_j^s}{\partial a_i^f} = -(d - a_j^s)\frac{\partial E}{\partial a_i^f}, \quad \frac{\partial W_i^f}{\partial a_l^f} = -(d - a_i^f)\frac{\partial E}{\partial a_l^f} < 0,$$

$$i, l = 1, ..., N - n, \quad i \neq l, \quad j = 1, ..., n,$$

that taking into account that $\partial E/\partial a_j^s = n/\gamma$ and $\partial E/\partial a_i^f = \partial E/\partial a_l^f = 1/\gamma$ according to the expression for total emissions given by (6) can be rewritten as follows

$$\frac{\partial W_i^f}{\partial a_j^s} = -(d - a_i^f)\frac{n}{\gamma} < 0, \quad \frac{\partial W_j^s}{\partial a_i^f} = -(d - a_j^s)\frac{1}{\gamma} < 0, \quad \frac{\partial W_i^f}{\partial a_l^f} = -(d - a_i^f)\frac{1}{\gamma} < 0. \quad (9)$$

Thus, an increase in non-signatories' adaptation has a negative effect on net benefits both for signatory countries and non-signatory countries and the same occurs for an increase in adaptation of a signatory country on non-signatories' net benefits. However, the sign of signatory-signatory international externalities depends on the symmetry of the solution of the coalition formation game. If we calculate the partial derivative of a signatory's net benefits given by (8) with respect to the level of adaptation of another signatory, we obtain the following expression

$$\frac{\partial W_j^s}{\partial a_k^s} = \frac{\alpha}{\gamma} - \frac{1}{\gamma}(\alpha - nd + A^s) - (d - a_j^s)\frac{\partial E}{\partial a_k^s},$$

where $\partial E/\partial a_k^s = n/\gamma$. Simplifying terms we obtain the following expression

$$\frac{\partial W_j^s}{\partial a_k^s} = \frac{1}{\gamma}(na_j^s - A^s), \quad (10)$$

so that if all signatories choose the same level of adaptation $A^s = a^s n$ we have that $\partial W_j^s/\partial a_k^s = 0$. Thus, if all the signatories select the same level of adaptation, the effect that adaptation of one signatory has on the net benefits of another signatory is zero. On the one hand, an increase in the adaptation of a signatory increases the country emissions of the rest of signatories increasing their benefits and, on the other hand, it also augments their total emissions by a quantity equal to n/γ resulting in an increase in damages. The result is that these variations cancel if signatories choose the emissions that maximize the

agreement's net benefits in the third stage, and consequently there are not signatory-signatory international externalities. This means that for signatories, adaptation, even if it is selected before countries decide on emissions, does not originate any externality. Thus, we can establish the following result[14]

Lemma 1 *If all signatories select the same level of adaptation there are not signatory-signatory international externalities in the second stage of the game.*

This result has important consequences in terms of the effects of cooperation in the second stage of the game and also on the effects that the timing of adaptation has on the levels of adaptation chosen by the countries.

3.2 The second stage (I): the complete agreement

In this subsection, we solve stage two assuming that in the first stage n countries, with $n \geq 1$, have signed the agreement.[15] Each non-signatory country chooses its level of adaptation as to maximize (7) taking as given the other countries' adaptation levels.

The FOCs for non-signatories are

$$\frac{\partial W_i^f}{\partial a_i^f} = \frac{\alpha}{\gamma} - \frac{1}{\gamma}(\alpha - d + a_i^f) + E - (d - a_i^f)\frac{\partial E}{\partial a_i^f} - ca_i^f = 0,$$

that taking into account that $\partial E / \partial a_i^f = 1/\gamma$ this expression simplifies to give

$$E = ca_i^f, \quad i = 1, ..., N - n. \tag{11}$$

The LHS of this conditions stands for the marginal benefit of adaptation given by the reduction in damages because the decrease in the marginal damages caused by adaptation that in our model is given by total emissions, and the RHS stands for the marginal costs of adaptation. Moreover, from this condition we can conclude that all non-signatories will choose the same level of adaptation and hence the same emissions too. Taking into account (6), the condition (11) implicitly defines the non-signatory reaction function. Applying the implicit function theorem we obtain that

$$\frac{\partial a_i^f}{\partial a_l^f} = -\frac{1}{1 - \gamma c}, \ i, l = 1, ..., N - n, \ i \neq l, \quad \frac{\partial a_i^f}{\partial a_j^s} = -\frac{n}{1 - \gamma c}, \ j = 1, ..., n.$$

The second-order condition (SOC) for the maximization of net benefits requires that $\gamma c > 1$ and consequently the adaptation of a non-signatory is a *strategic complement* of the rest of countries' adaptation.

On the other hand, signatories choose the level of adaptation to maximize the agreement net benefits taking as given the non-signatories' adaptation

---

[14] In the next two subsections we show that both for the complete agreement and the mitigation agreement all signatories choose the same level of adaptation and consequently this lemma applies.

[15] If $n = 1$, no agreement is signed and the outcome of the game is the fully non-cooperative equilibrium. If $n = N$, the agreement is the grand coalition and the efficient solution is implemented by the agreement.

$$\max_{\{a_1^s,\dots,a_n^s\}} W_A = \sum_{j=1}^{n} \left\{ \frac{\alpha}{\gamma}(\alpha - nd + A^s) - \frac{1}{2\gamma}(\alpha - nd + A^s)^2 - (d - a_j^s)E - \frac{c}{2}(a_j^s)^2 \right\}.$$

The FOCs for this problem can be written as follows

$$\frac{\partial W_A}{\partial a_k^s} = \frac{n\alpha}{\gamma} - \frac{n}{\gamma}(\alpha - nd + A^s) + E - \frac{\partial E}{\partial a_k^s} \sum_{j=1}^{n}(d - a_j^s) - ca_k = 0, \ \ k = 1,\dots,n.$$

As $\partial E/\partial a_j^s = n/\gamma$ the previous condition yields

$$\frac{n\alpha}{\gamma} - \frac{n}{\gamma}(\alpha - nd + A^s) + E - \frac{n}{\gamma}(nd - A^s) - ca_k = 0, \ \ k = 1,\dots,n,$$

that finally simplifies to give

$$E = ca_k^s, \ \ k = 1,\dots,n. \tag{12}$$

This condition establishes that all signatories select the same level of adaptation and then according to Lemma 1 there are not signatory-signatory international externalities in this stage of the game.

This condition implicitly defines the reaction function of the representative signatory. Applying the implicit function theorem again, we obtain that

$$\frac{\partial a^s}{\partial a_i^f} = -\frac{1}{n^2 - \gamma c}, \ i = 1,\dots,N - n.$$

For the maximization of the agreement net benefits, the SOC is $\gamma c > n^2$ what establishes that the signatory's adaptation is a *strategic complement* of the non-signatories' adaptation. As $n \in [1, N]$, we assume that $\gamma c > N^2$ that guarantees that SOC are satisfied for both signatories and non-signatories regardless of the level of participation in the agreement for $N > 2$. This condition acts as a concavity requirement for each signatory: in the sense that it establishes a lower bound on the values of the concavity coefficients within the emission and adaptation net benefit functions such that these are strictly concave for any possible value of participation $n \in [1, N]$.

Conditions (11) and (12) establish that both signatories and non-signatories choose the same level of adaptation.

3.3 The second stage (II): the mitigation agreement

The previous analysis shows that as signatories select the same level of adaptation according to Lemma 1 there are not signatory-signatory international externalities. In this case, it is obvious that

Proposition 1 *The mitigation agreement coincides with the complete agreement.*

In other words, the international cooperation in the provision of a national good that does not generate indirectly any international externality does not alter the national provision of the good. It is straightforward to show this. The third stage of the game is the same for both the complete agreement and the mitigation agreement. However, for the mitigation agreement there is no cooperation between signatories in the second stage. In this case, non-signatories

and signatories select the level of adaptation that maximizes their net benefits. Thus, for non-signatories we would obtain the same condition (11) that we obtained for the complete agreement, whereas for signatories, the FOC for the maximization of their net benefits now is

$$\frac{\partial W_j^s}{\partial a_j^s} = \frac{\alpha}{\gamma} - \frac{1}{\gamma}(\alpha - nd + A^s) + E - (d - a_j^s)\frac{\partial E}{\partial a_j^s} - ca_j^s = 0, \quad j = 1, ..., n,$$

that taking into account that $\partial E / \partial a_j^s = n/\gamma$, can be written as follows

$$\frac{1}{\gamma}(na_j^s - A^s) + E - ca_j^s = 0, \quad j = 1, ..., n,$$

that implies that all signatories choose the same level of adaptation. Then, the first term cancels out and the same condition that characterizes the equilibrium for the complete agreement $E = ca^s$ is obtained. Notice that the first term of the left-hand side of the previous equation is equal to $\partial W_j^s / \partial a_k^s$ according to (10), but as all signatories select the same level of adaptation the term vanishes and Lemma 1 applies.

But this is not the only consequence of the fact that for a symmetric equilibrium there are not signatory-signatory international externalities. It is also straightforward to show that

Proposition 2 *The outcome of the second stage of the mitigation agreement is the same regardless of whether adaptation is selected before the countries choose emissions or if it is selected after or at the same time they choose emissions.*

In other words, the equilibrium of the mitigation agreement formation game is the same regardless of the adaptation timing. Moreover, as we claimed above, the mitigation agreement yields the same outcome as the complete agreement.

Suppose now that adaptation is chosen after the countries select emissions. In this case, adaptation will not originate any externality because of the timing of the game as occurs when countries choose adaptation before emissions. For this reason, the outcome of the game will be the same if countries choose adaptation and emissions at the same time. If we assume that adaptation and emissions are chosen simultaneously, the agreement formation game has two stage. In the first stage, countries decide on participation and in the second stage they select the level of adaptation and emissions simultaneously, the signatories to maximize the agreement's net benefits and the non-signatories to maximize their national net benefits. Thus, non-signatories will choose these two variables to maximize (1). In this case, the FOCs for an interior solution are

$$\alpha - \gamma e_i^f = d - d_i^f,$$
$$E = ca_i^f, \quad i = 1, ..., N - n.$$

On the other hand, for signatories the FOCs for an interior solution are

$$\alpha - \gamma e_j^s = \sum_{k=1}^{n}(d - a_k^s) = nd - A^s,$$

where $A^s = \sum_{k=1}^{n} a_k^s$ and

$$E = ca_j^s, \quad j = 1, ..., n.$$

This set of conditions coincide with conditions (2), (4), (11) and (12) that characterize the complete/mitigation agreement when adaptation is selected before countries choose emissions. Thus Lemma 1 has important consequences for the number of equilibria that our coalition formation game admits. On the one hand, we see that there is no difference between a complete agreement and a mitigation agreement, and, on the other hand, it does matter whether countries decide the level of adaptation before emission or after emissions or at the same time. In the next section, we will show that an adaptation agreement yields a different outcome to the one corresponding to the complete/mitigation agreement. Cooperation only on the second stage modifies the decisions countries take on adaptation, but now we proceed with the analysis of the second stage of the complete/mitigation agreement.[16]

Conditions (11) and (12) establish that both signatories and non-signatories choose the same level of adaptation that is given by

$$a = \frac{N\alpha - (n^2 - n + N)d}{\gamma c - (n^2 - n + N)},$$ (13)

and multiplying by c would obtain total emissions. Substituting this expression in (3) and (5) allows us to calculate emissions

$$e^f = \frac{\gamma c(\alpha - d) - n(n-1)\alpha}{\gamma(\gamma c - (n^2 - n + N))}, \quad e^s = \frac{\gamma c(\alpha - nd) + \alpha(N - n)(n-1)}{\gamma(\gamma c - (n^2 - n + N))}.$$ (14)

Observe that if $\gamma c > N^2$ the denominator of these expression is positive for all $n \in [1, N]$. On the other hand, as $n^2 - n + N$ increases with n, $\alpha/N > d$ will give a positive numerator for a for all $n \in [2, N]$. If adaptation is positive this condition also guarantees that emissions are positive for both signatories and non-signatories according conditions (3) and (5). Moreover, using these conditions we obtain that

$$e^f - e^s = \frac{1}{\gamma}(n-1)(d-a),$$

and we can conclude that if marginal damages are positive the non-signatories' emissions are larger than the signatories' emissions for all levels of cooperation. Using (13) marginal damages can be written as follows

$$d - a = \frac{\gamma cd - N\alpha}{\gamma c - (n^2 - n + N)}.$$ (15)

Given this expression, $c > \alpha N/d\gamma$ guarantees that there is no over-adaptation. However, when c is close to this lower bound we will have what we could call an *"almost" corner solution* with marginal damages close to zero. The investment in adaptation is boosted by low adaptation costs leading the marginal damages close to zero. We think that this is a very optimistic assumption about what we can expect from adaptation. To avoid this kind of solutions we are going to introduce a lower bound on marginal damages larger than zero.

---

[16] As the complete agreement coincides with the mitigation agreement, there are no reasons to form an agreement that obligates countries to cooperate both on emissions and adaptation. Thus, from now on we will assume that our model only admits two types of agreements, the mitigation agreement and the adaptation agreement.

We will assume that marginal damages with adaptation cannot be lower than a fraction $\beta \in (0, 1)$ of the marginal damages without adaptation. In this case, we require that

$$(1 - \beta)d - a = \frac{(1 - \beta)d\gamma c - N\alpha + \beta d(n^2 - n + N)}{\gamma c - (n^2 - n + N)} \geq 0,$$

that imposes a lower bound on d

$$d \geq \frac{N\alpha}{(1 - \beta)\gamma c + \beta(n^2 - n + N)}.$$

This lower bound on d is simply a minimum distance requirement between d and a, which means that as mentioned above, we impose a minimum vulnerability given that we assume marginal damages with adaptation cannot be lower than a fraction $\beta \in (0, 1)$ of the marginal damages without adaptation.

The RHS of this inequality is decreasing with respect to n. Thus, it takes its highest value for $n = 1$

$$d \geq \frac{N\alpha}{(1 - \beta)\gamma c + \beta N}.$$

But this lower bound must be compatible with the upper bound for d, α/N, defined above that requires that

$$\gamma c > \frac{N^2 - \beta N}{1 - \beta} > N^2 \text{ for } \beta \in (0, 1).$$

We can summarize all these conditions in the following assumption[17]

Assumption 1 We assume that $N^2 < (N^2 - \beta N)/(1 - \beta) < \gamma c$ and $d \in [\alpha N/(\gamma c(1 - \beta) + \beta N), \alpha/N)$ for $\beta \in (0, 1)$.

Thus, this assumption guarantees that the non-negativity constraints are satisfied, that marginal damages are higher than a positive lower bound and that the SOC are also satisfied. These parameter restrictions simply enforce the technological requirements assumed in our model. For example, no over adaptation parameter constraint ensures that countries will not consider selecting a level of adaptation above d, which would convert environmental damages into environmental benefits of pollution. This is just a consistency requirement which translates real-world characteristics to an a priori unrestricted initial model. Therefore the feasible set on parameter values derived here simply ensures that through their net benefit functions, countries will be aware of and will act consistently with the real-world assumptions we impose into the model.

Notice that $1 - \beta$ defines de *degree of effectiveness* of adaptation since multiplying by one hundred we would obtain the percentage reduction in marginal damages because of the investment in adaptation. In the next subsection, we will study how the level of participation in an IEA depends on this parameter. As in this model, the marginal damages represent the *vulnerability* of the country to total emissions, we could interpret β as a measure of

---

[17] Notice thtat $(N^2 - \beta N)/(1 - \beta)$ is an increasing strictly convex function of β.

the country's vulnerability, so that the higher the degree of effectiveness of adaptation, the lower the vulnerability.

Next, we compare net benefits. The non-signatories pollute more than signatories and invest the same in adaptation than signatories, consequently their net benefits are higher than the net benefits signatories get. On the other hand, it is easy to check that adaptation for both non-signatories and signatories decreases as the number of signatories increases. Thus, cooperation decreases both adaptation and emissions because emissions depend positively on adaptation. The same occurs with total emissions. Now, if we look at net benefits, we know that benefits and adaptation costs are going to decrease with an increase in participation. However, it is not so clear what occurs with damages. On the one hand, marginal damages increase because of the reduction in adaptation. On the other hand, total emissions decrease with cooperation. Next, we evaluate how damages change with the participation. Damages are given by the following expression

$$D(n) = (d - a)E = \frac{c(\gamma cd - N\alpha)(N\alpha - (n^2 + N - n)d)}{(\gamma c - (n^2 + N - n))^2},$$

where $\gamma cd - N\alpha$ is positive according to Assumption 1. The first derivative with respect to n is

$$\frac{\partial D(n)}{\partial n} = c(d\gamma c - N\alpha)(2n - 1)\frac{2N\alpha - (\gamma c + n^2 + N - n)d}{(\gamma c - (n^2 + N - n))^3}. \qquad (16)$$

Thus, the sign of this first derivative depends on the sign of the numerator. As the numerator decreases with n, we can define two threshold values for d

$$d_1 = d(n = N) = \frac{2N\alpha}{\gamma c + N^2} < d_2 = d(n = 1) = \frac{2N\alpha}{\gamma c + N},$$

such that if $d < d_1$ damages are increasing for all $n \in [1, N]$ provided that $d_1 > \alpha N/(\gamma c(1 - \beta) + \beta N)$, and if $d > d_2$ damages are decreasing for all $n \in [1, N]$ provided that $d_2 < \alpha/N$. For d in the interval (d_1, d_2) there will exist a critical value n^* defined by $\partial D/\partial n = 0$, so that for $n < n^*$ damages are increasing and for $n > n^*$ damages decrease. Next, we investigate when damages are increasing comparing d_1 with the bounds for d defined in Assumption 1.

$$\frac{\alpha}{N} - d_1 = \frac{\alpha(\gamma c - N^2)}{N(\gamma c + N^2)} > 0 \text{ for } \gamma c > \frac{N^2 - \beta N}{1 - \beta},$$

$$d_1 - \frac{\alpha N}{\gamma c(1 - \beta) + \beta N} = \frac{N\alpha((1 - 2\beta)\gamma c - (N - 2\beta)N)}{(\gamma c + N^2)(\gamma c(1 - \beta) + \beta N)}.$$

The numerator of this expression is negative for all $\gamma c > 0$ if $\beta \geq 1/2$. However, if $\beta < 1/2$ then there exists a threshold value $(\gamma c)'$ equal to $(N - 2\beta)N/(1 - 2\beta) > (N^2 - \beta N)/(1 - \beta)$ such that

$$\text{if } \gamma c \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} (\gamma c)' \text{ then } d_1 \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} \frac{\alpha N}{\gamma c(1 - \beta) + \beta N},$$

and we can conclude that

Proposition 3 *If $\beta < 1/2$ and $\gamma c > (\gamma c)' = (N - 2\beta)N/(1 - 2\beta)$ then $d_1 \in (\alpha N/(\gamma c(1 - \beta) + \beta N), \alpha/N)$ and damages are increasing with the participation for all $n \in [1, N]$ when $d \in [\alpha N/(\gamma c(1 - \beta) + \beta N), d_1)$.*

Thus, damages can increase with participation if the country's vulnerability is low. In this case, the increase in marginal damages because the reduction in adaptation when the participation steps up is strong enough as to compensate the reduction in damages because the reduction in total emissions yielding that an increase in participation leads to an increase in damages.

Next, we compare d_2 with the bounds for d defined in Assumption 1

$$d_2 - \frac{\alpha N}{\gamma c(1 - \beta) + \beta N} = \frac{N\alpha(\gamma c - N)(1 - 2\beta)}{(\gamma c + N)(\gamma c(1 - \beta) + \beta N)}.$$

This difference is negative for all $\gamma c > N$ if $\beta > 1/2$ which implies that d_2 is also lower than α/N. For $\beta < 1/2$ we need to compare d_2 with α/N.

$$\frac{\alpha}{N} - d_2 = \frac{\alpha(-2N^2 + N + \gamma c)}{N(\gamma c + N)}.$$

This difference is zero for the threshold value $(\gamma c)'' = 2N^2 - N > (N^2 - \beta N)/(1 - \beta)$ so that

$$\text{if } \gamma c \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} (\gamma c)'' \text{ then } \frac{\alpha}{N} \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} d_2,$$

and we can conclude that

Proposition 4 *If $\beta \geq 1/2$ then damages are decreasing with the participation for all $n \in [1, N]$ and all $d \in [\alpha N/(\gamma c(1 - \beta) + \beta N), \alpha/N)$. If $\beta < 1/2$ and $\gamma c > (\gamma c)'' = N(2N - 1)$ then $d_2 \in (\alpha N/(\gamma c(1 - \beta) + \beta N), \alpha/N)$ and damages are decreasing with the participation for all $n \in [1, N]$ when $d \in (d_2, \alpha/N)$.*

Thus, damages decrease if the country's vulnerability is high for all values of d, although they could also decrease if the vulnerability is low, but then damages and the product γc must be high. In this case, although the marginal damages increase with the participation, the reduction in total emissions is strong enough as to cause a reduction in total damages. In the rest of cases that are not included in the previous propositions, there will exist a critical value $n^* \in [2, N]$, so that for $n < n^*$ damages are increasing and for $n > n^*$ damages decrease.

However, regardless of whether damages increase or decrease with cooperation, cooperation has a positive effect on net benefit for both non-signatories and signatories. Signatories internalize the negative externality caused by pollution and as a result of this, the net benefits increase monotonically with membership.

Taking the first derivative of net benefits with respect to n for signatories yields

$$\frac{\partial W^s}{\partial n} = \alpha \frac{\partial e^s}{\partial n} - \gamma e_s \frac{\partial e^s}{\partial n} + \frac{\partial a^s}{\partial n} E - (d - a^s) \frac{\partial E}{\partial n} - ca^s \frac{\partial a^s}{\partial n}, \tag{17}$$

that considering that

$$\frac{\partial E}{\partial n} = e^s + n\frac{\partial e^s}{\partial n} - e^f + (N - n)\frac{\partial e^f}{\partial n}, \tag{18}$$

can be rewritten as follows:

$$\frac{\partial W^s}{\partial n} = \left(\alpha - \gamma e_s - n(d - a^s)\right)\frac{\partial e^s}{\partial n} + (E - ca^s)\frac{\partial a^s}{\partial n} - (d - a^s)\left(e^s - e^f + (N - n)\frac{\partial e^f}{\partial n}\right),$$

where the first term of the RHS is zero according to FOC (4) and $E = ca^s$ according to (12) resulting in

$$\frac{\partial W^s}{\partial n} = (d - a^s)\left(e^f - e^s - (N - n)\frac{\partial e^f}{\partial n}\right) > 0, \tag{19}$$

since $e^f > e^s$ and non-signatories' emissions decrease with the cooperation.

For non-signatories, we obtain the following expression

$$\frac{\partial W^f}{\partial n} = (\alpha - \gamma e_f - (d - a^f))\frac{\partial e^f}{\partial n} + (E - ca^f)\frac{\partial a^f}{\partial n} - (d - a^f)(e^s + n\frac{\partial e^s}{\partial n} - e^f + (N - n - 1)\frac{\partial e^f}{\partial n}),$$

where again the first term is zero by the FOCs of the third stage and $E = ca^f$ according to (11). Thus, we obtain the following expression

$$\frac{\partial W^f}{\partial n} = (d - a^f)(e^f - e^s - n\frac{\partial e^s}{\partial n} - (N - n - 1)\frac{\partial e^f}{\partial n}) > 0. \tag{20}$$

Thus, we find that there are *positive spillovers* for non-signatories stemming from cooperation, i.e., cooperation increases the non-signatories' net benefits. Moreover, it is easy to show that the difference in net benefits also increases with the participation. Lastly, we show that the game presents the property of *full cohesiveness.*[18] This property states that total net benefits increase when the coalition is enlarged gradually and obtains its maximum for the grand coalition. This property justifies the search for large stable agreements. For this reason, it deserves some discussion. If we look at the expression of total net benefits $W = nW^s(n) + (N - n)W^f(n)$ we have that the increase in participation is driving by three variations as the following derivative shows

$$\frac{\partial W}{\partial n} = W^s(n) - W^f(n) + n\frac{\partial W^s(n)}{\partial n} + (N - n)\frac{\partial W^f(n)}{\partial n},$$

where the first term is negative and the other two terms positive as we have just showed, thus the effect of an increase in the number of signatories could be negative or positive depending of the magnitude of each term in the expression. Our analysis shows that the addition of the two last terms is greater than the difference in net benefits that represents the first term for all levels of cooperation. This term is negative because we have showed that signatories obtain a lower net benefit of non-signatories whatever is the number of signatories. It is important to highlight the role that the positive spillovers of cooperation has on this result since it reinforces the positive effect that an increase in participation has on signatories' net benefits resulting finally in an increase in the aggregate net benefits for all levels of cooperation. Thus, we can conclude that this result is not an artifact of the

---

[18] These two properties are showed respectively in Appendix A.1 and Appendix A.2.

adaptation-emissions game we analyze in the paper, but a feature that characterizes the PANE of the second stage of the game. To end this subsection, we would like to point out that all these features of the PANE of the second stage already appear in the model without adaptation. In other words, the introduction of adaptation does not modify the features of the PANE of the second stage of the model without adaptation except in one point, whereas emissions are strategic substitutes in the model without adaptation, investment in adaptation are strategic complements in our model with adaptation.

3.4 The first stage: the Nash equilibrium of the membership game

In this subsection, we investigate which is the level of participation that can be achieved with a mitigation agreement. First, we present the definition of coalition stability from d'Aspremont et al. (1983), which has been extensively used in the literature on international environmental agreements.

Definition 1 An agreement consisting of n signatories is stable if $W_k^s(n) \geq W_k^f(n-1)$ for $k = 1, ..., n$ and $W_j^f(n) \geq W_j^s(n+1)$ for $j = 1, ..., N - n$.

The first inequality, which is also known as the *internal stability condition*, simply means that any signatory country is at least as well-off staying in the agreement as withdrawing from it, assuming that all other countries do not change their membership status. The second inequality, which is also known as the *external stability condition*, similarly requires any non-signatory to be at least as well-off remaining as a non-signatory that joining the agreement, assuming once again, that all other countries do not change their membership status. In order to develop the stability analysis we define the stability function $S(n) = W^s(n) - W^f(n-1)$. Notice that if $S(n)$ is positive and $S(n+1)$ is negative an agreement consisting of n countries is stable and the stability analysis can be reduced to find out whether $S(n) = 0$ has a solution that satisfies the stability conditions.[19]

For our model, the stability function $S(n)$ reads as follows

$$S(n) = \frac{(n-1)(\gamma cd - \alpha N)^2 F(n)}{2\gamma(\gamma c - ((n-1)^2 + N - n + 1))^2(\gamma c - (n^2 + N - n))^2},\tag{21}$$

where the denominator is positive and $F(n)$ is a polynomial of fifth degree

$$F(n) = -n^5 + 5n^4 + f_3 n^3 - f_2 n^2 - f_1 n + f_0,\tag{22}$$

with

$$f_3 = 2\gamma c - 2N - 7 > 0,$$
$$f_2 = 8\gamma c - 4N - 3 > 0,$$
$$f_1 = (\gamma c)^2 - 2(N+3)\gamma c + (N-2)N > 0,7$$
$$f_0 = 3(\gamma c)^2 - 2(N+2)\gamma c - N^2 > 0,$$

---

[19] Notice that this implies that $S'(n^*)$ must be negative where n^* is the solution for $S(n) = 0$. If n^* is not a natural number, the stable agreement is given by the first natural number on the left of n^* provided that for the first natural number on the right of n^*, $S(n)$ is negative.

for $N \geq 3$ provided that $\gamma c > N^2$.Analyzing this polynomial, we can conclude that[20]

Proposition 5 *For interior solutions and $N \geq 7$, if the degree of effectiveness of adaptation $1 - \beta$ is larger than or equal to $(N^2 - 3N)/(N^2 - 3N + 4)$ the grand coalition is stable. However, if it is lower than this threshold value the only stable agreement consists of three countries regardless of the severity of environmental damages.*

Proof See Appendix A.3 □

This result establishes that incorporating the investment in adaptation to an agreement on emissions, the grand coalition could be stable. However, the limit of the lower bound for the effectiveness of adaptation that defines the interval for this variable for which the grand coalition is stable converges to one very quickly with the number of countries involved in the international environmental problem. For instance, for $N = 10$, the lower bound for the effectiveness of adaptation is $1 - \beta = 0.9459$ that means that the grand coalition through the investment in adaptation is able to reduce marginal damages in a 94.59%, leading the vulnerability of the country to 0.0541. For $N = 100$, we have that the lower bound is $1 - \beta = 0.9996$ that reduces the vulnerability to 0.0004. We believe that this is a very optimistic assumption about we can expect from the investment in adaptation. In fact, it implies that the environmental problem could be solved by investing in adaptation. Thus, the answer to the question in the title of this paper is that we cannot expect that adaptation promotes the participation in an IEA under reasonable assumptions about the degree of effectiveness of adaptation.

To conclude the analysis of the complete/mitigation agreement we would like to add some words about the scope of this result and the relationship between participation and the gains of cooperation. Our result establishes that unless adaptation technology reduces a country's vulnerability virtually to 0, which is completely unrealistic given the nature of such technology, we have that for all combinations of parameter values that satisfy Axiom 1, only an agreement consisting of three countries is stable. There is no more room for cooperation. Thus, as it occurs for the standard model with linear damages when there is no adaptation, the scope of cooperation is extremely limited. On this note, it therefore does not matter whether the potential gains from full cooperation are low or high. Suppose that the gains of full cooperation are high when the vulnerability is sufficiently low such that the grand coalition could be stable, but this is an uninterested case because we do not expect that adaptation leads the vulnerability of a country to these extremely low values. Suppose instead that the gains of full cooperation are high for larger values of the vulnerability.[21] Then the problem is that a reduced proportion of these gains can be achieved through cooperation because only three countries will sign the agreement. Thus, this result is very pessimistic because it establishes that even in the case that cooperation brings to the countries important increases in their net benefits, the free-rider incentives reduce significantly the scope of cooperation.

---

[20] The lower bound on N that appears in the proposition guarantees that the external stability of an agreement consisting of three countries is satisfied. On the other hand, Assumption 1 guarantees that the internal stability condition is also satisfied. See the details in Appendix A.3.

[21] It can be shown that the gains of full cooperation increase in absolute and relative terms with the vulnerability. The proof is available from the authors upon request.

4 The adaptation agreement

In this section, we focus on an agreement on investment in adaptation. As we have concluded in the previous section that there are not signatory-signatory international spillovers, we could think that also for this kind of agreements cooperation has no influence in the decisions on adaptation taken in the second stage. But this is not the case as we will show in the next subsection. The lack of cooperation in the third stage when countries decide on emissions changes the effects that the variation of adaptation in a signatory country has on the other signatory countries' net benefits. Now, there are negative externalities and if the countries internalize these externalities, they select levels of adaptation different from those they would select without cooperation. The formation of an adaptation agreement is also modeled as a three-stage game as in the case of a complete agreement with the difference that there is no cooperation in the third stage.

4.1 The third stage: an equilibrium in dominant strategies

Without cooperation in the third stage, the FOCs for an *interior solution* are given by (2)

$$\alpha - \gamma e_i = d - a_i, \ i = 1, ..., N. \tag{23}$$

As there is no cooperation in this stage, all countries only take into account the effect that emissions have on its national damages.

Thus, emission is given by

$$e_i = \frac{1}{\gamma}(\alpha - d + a_i). \tag{24}$$

As environmental damages are linear, the countries' reaction functions are orthogonal and the optimal emissions are given by an equilibrium in dominant strategies. Notice that as in the complete agreement an increase in adaptation leads to higher emissions.

Adding for all countries allows us to calculate total emissions

$$E = \frac{N(\alpha - d) + A}{\gamma}, \tag{25}$$

where $A = \sum_{i=1}^{N} a_i$ is total investment in adaptation.

Next, using (1), net benefits can be written as follows

$$W_i = \frac{\alpha}{\gamma}(\alpha - d + a_i) - \frac{1}{2\gamma}(\alpha - d + a_i)^2 - (d - a_i)E - \frac{c}{2}(a_i)^2, \tag{26}$$

where total emissions are given by (25).

Observe that although the investment in adaptation is a national good, if countries decide on adaptation before they select the level of emissions, the investment in adaptation generates *negative international externalities* as the following derivative shows

$$\frac{\partial W_i}{\partial a_j} = -\frac{1}{\gamma}(d - a_i) < 0, \ i,j = 1, ..., N, \ i \neq j, \tag{27}$$

and this occurs regardless the countries cooperate or not cooperate in the second stage of the game.

4.2 The second stage: the PANE of the adaptation game

In this subsection, we solve stage two assuming that in the first stage n countries have signed the agreement. Thus, in this stage we have to distinguish between non-signatory countries and signatory countries. Notice that as total emissions are positively related to total adaptation, this variable can be seen as a global public bad like total emissions.

As (26) is equal to (7) there are no differences with the FOCs obtained for non-signatories in the case of a mitigation agreement and we have that $E = ca_i^f$. All non-signatories countries select the same level of adaptation.

On the other hand, signatories choose the level of adaptation to maximize the agreement net benefits taking as given the non-signatories' adaptation.

$$\max_{\{a_1^s,...,a_n^s\}} W_A = \sum_{j=1}^{n} \left\{ \frac{\alpha}{\gamma}(\alpha - d + a_j^s) - \frac{1}{2\gamma}(\alpha - d + a_j^s)^2 - (d - a_j^s)E - \frac{c}{2}a_j^2 \right\}.$$

The FOCs for this problem are

$$\frac{\partial W_A}{\partial a_k^s} = \frac{\alpha}{\gamma} - \frac{1}{\gamma}(\alpha - d + a_k^s) + E - \frac{\partial E}{\partial a_k^s}\sum_{j=1}^{n}(d - a_j^s) - ca_k^s = 0, \ \ k = 1,...,n,$$

that taking into account that $\partial E/\partial a_k^s = 1/\gamma$ yield

$$\frac{\alpha}{\gamma} - \frac{1}{\gamma}(\alpha - d + a_j^s) + E - \frac{1}{\gamma}(nd - A) - ca_j = 0, \ k = 1,...,n.$$

that establishes that all signatories select the same level of adaptation. Then as in the symmetric case $A = na^s$, finally we obtain the following condition for signatories

$$E = \frac{(n-1)}{\gamma}(d - a^s) + ca^s, \tag{28}$$

where the LHS is, as in condition (11), the marginal benefit of adaptation. However, the signatories take into account the increase in damages for the rest of signatories caused by the increase in adaptation. Remember that adaptation increases national emissions. This condition implicitly defines the reaction function of the representative signatory. Applying the implicit function theorem again, we obtain that

$$\frac{\partial a^s}{\partial a_i^f} = \frac{1}{\gamma c + 1 - 2n}, \ i = 1,...,N - n.$$

For signatories, the SOC requires that $\gamma c > 2n - 1$ what establishes that the signatory's adaptation is a *strategic complement* of the non-signatories' adaptation. As $n \in [1, N]$, we assume that $\gamma c > 2N - 1$ that guarantees that SOC for both signatories and non-signatories are satisfied regardless of the level of participation in the agreement for $N > 2$.

Thus, using (11) and (28), we may obtain the level of adaptation of the Partial Agreement Nash Equilibrium (PANE) of the second stage

$$a^s = \frac{\gamma c \alpha N + ((N-n)(n-1) - \gamma c(N+n-1))d}{(\gamma c)^2 - (N+n-1)\gamma c + (n-1)(N-n)}, \tag{29}$$

$$a^f = \frac{\alpha(\gamma c - n + 1)N + ((N-n)(n-1) - \gamma c N)d}{(\gamma c)^2 - (N+n-1)\gamma c + (n-1)(N-n)}. \tag{30}$$

It is easy to show that denominator of both expressions is positive for all $n \in [1, N]$ if $\gamma c > 2N - 1$, and that the numerator of (29) is also positive provided that $\alpha N/(2N-1) > d$. [22] Thus, these two constraints on parameter values guarantee that a^s is positive for all $n \in [1, N]$ and also that e^s is positive since $\alpha N/(2N-1) > d$ implies that $\alpha > d$.

Next, we verify if the marginal damages are positive

$$d - a^s = \frac{\gamma c(\gamma cd - \alpha N)}{(\gamma c)^2 - (N+n-1)\gamma c + (n-1)(N-n)}. \tag{31}$$

Given this expression, $d > \alpha N/(\gamma c)$ guarantees that there is no over-adaptation. This condition also guarantees that a^f is larger than a^s, that e^f is positive and obviously larger than e^s, and that marginal damages for non-signatories are also positive.

$$d - a^f = \frac{(\gamma c - n + 1)(\gamma cd - \alpha N)}{(\gamma c)^2 - (N+n-1)\gamma c + (n-1)(N-n)}. \tag{32}$$

Now, using (29) and (30) we can calculate the emissions for each type of country

$$e^s = \frac{1}{\gamma} \frac{(\gamma c - N)\alpha + (\alpha - d)(\gamma c)^2 - (\gamma c - 1 - N)\alpha n - \alpha n^2}{(\gamma c)^2 - (N+n-1)\gamma c + (n-1)(N-n)}, \tag{33}$$

$$e^f = \frac{1}{\gamma} \frac{\gamma c(\alpha - d)(1 + \gamma c) + (\alpha - \gamma c(\alpha - d))n - \alpha n^2}{(\gamma c)^2 - (N+n-1)\gamma c + (n-1)(N-n)}, \tag{34}$$

and adding for all countries we obtain total emissions

$$E = ne^s + (N-n)e^f = \frac{c(N(1+\gamma c)(\alpha - d) + (d - N(\alpha - d))n - dn^2)}{(\gamma c)^2 - (N+n-1)\gamma c + (n-1)(N-n)}. \tag{35}$$

As in the previous section, we will assume that marginal damages with adaptation cannot be lower than a fraction $\beta \in (0, 1)$ of the marginal damages without adaptation. If we impose this constraint on the non-signatories, it will also be satisfied for signatories since $a^f > a^s$.

$$(1 - \beta)d - a^f = \frac{(\gamma c - n + 1)(\gamma cd - \alpha N) - \beta d((\gamma c)^2 - (N+n-1)\gamma c + (n-1)(N-n))}{(\gamma c)^2 - (N+n-1)\gamma c + (n-1)(N-n)} \geq 0,$$

that imposes a lower bound on d

$$d \geq \frac{(\gamma c - n + 1)\alpha N}{(1-\beta)(\gamma c)^2 + (\beta N - (1-\beta)(n-1))\gamma c - \beta(n-1)(N-n)}.$$

---

[22] Notice that the numerator is decreasing in n.

The RHS of the inequality is decreasing with n. Thus, it takes the highest value for $n = 1$

$$d \geq \frac{\alpha N}{\gamma c(1 - \beta) + \beta N},$$

which is the same constraint we obtain for the complete agreement.[23] But this lower bound must be lower than the upper bound for d, $\alpha N/(2N - 1)$ we have defined above and ensures no over adaptation, that requires that

$$\gamma c > \frac{(2 - \beta)N - 1}{(1 - \beta)} > 2N - 1 \text{ for } \beta \in (0, 1).$$

We can summarize all these constraints on parameters values in the following assumption

Assumption 2 We assume that $2N - 1 < ((2 - \beta)N - 1)/(1 - \beta) < \gamma c$ and $d \in [\alpha N/((1 - \beta)\gamma c + \beta N), \alpha N/(2N - 1))$ for $\beta \in (0, 1)$.

Thus, this assumption guarantees that the non-negativity constraints are satisfied, that marginal damages are higher than a positive lower bound and that the SOC are also satisfied.

Next, we compare net benefits. The non-signatories invest more in adaptation and pollute more than signatories. Thus, the non-signatories will have larger benefits and lower damages than signatories, but higher adaptation costs. In order to compare net benefits, we need to calculate net benefits for both non-signatories and signatories. Using the previous expressions for emissions, marginal damages, total emissions and the adaptation level, we obtain the following expression for the non-signatories net benefits

$$W^f = \frac{w_4^f n^4 + w_3^f n^3 + w_2^f n^2 + w_1^f n + w_0^f}{2((\gamma c)^2 - (N + n - 1)\gamma c + (n - 1)(N - n))^2}, \tag{36}$$

where

$$w_4^f = \alpha^2 - \gamma c d^2,$$
$$w_3^f = 2(\gamma c d^2 - \alpha^2)(N - \gamma c + 1),$$
$$w_2^f = 2(\gamma c)^3 d^2 + (\gamma c)^2(3d^2 - 2Nd\alpha - \alpha^2) - \gamma c((4N + N^2 + 1)d^2 - 2Nd\alpha - (N^2 - 4)\alpha^2) + \alpha^2(4N + 1),$$

$$w_1^f = -2(\gamma c + 1)((\gamma c)^2(\alpha^2 - 2Nd(\alpha - d)) - \gamma c(N(N + 1)d^2 - 2Nd\alpha - (N^2 - 1 - 2N)\alpha^2) + N\alpha^2),$$
$$w_0^f = \gamma c(\gamma c + 1)^2(\alpha - d)((\alpha - (2N - 1)d)\gamma c + N(N - 2)\alpha + N^2 d).$$

and the following expression for signatories countries

$$W^s = \frac{w_4^s n^4 + w_3^s n^3 + w_2^s n^2 + w_1^s n + w_0^s}{2((\gamma c)^2 - (N + n - 1)\gamma c + (n - 1)(N - n))^2}, \tag{37}$$

where

---

[23] Notice that as expected this is a stronger constraint than $d > \alpha N/\gamma c$.

$$w_4^s = \alpha^2 - \gamma c d^2,$$
$$w_3^s = 2(\gamma c d^2 - \alpha^2)(N - \gamma c + 1),$$
$$w_2^s = (\gamma c d^2 - \alpha^2)((\gamma c)^2 + 4\gamma c - N^2 - 4N - 1),$$
$$w_1^s = -2((\gamma c)^3(2Nd^2 - 2Nd\alpha + \alpha^2) - (\gamma c)^2((N^2 - 1 - N)d^2 - N(N+2)\alpha^2)$$
$$-\gamma c(N(N+1)d^2 - (N^2 - 1 - N)\alpha^2) + N\alpha^2(N+1))$$
$$w_0^s = (\gamma c + 1)(-(\gamma c)^3(\alpha - d)((2N-1)d - \alpha) + (\gamma c)^2(+2Nd^2 - N^2 d^2 + (N^2 - 2N + 1)\alpha^2)$$
$$-N\gamma c(+Nd^2 - (N-2)\alpha^2) + N^2\alpha^2).$$

Next, we calculate the difference in net benefits using (36) and (37)

$$W^f - W^s = \frac{(n-1)(1 + \gamma c + n(\gamma c - 1))(\gamma c d - \alpha N)^2}{2((\gamma c)^2 - (N + n - 1)\gamma c + (n-1)(N-n))^2}, \tag{38}$$

that is positive since we have assumed that $\gamma c > 2N - 1$. As occurs for the case of a complete agreement, non-signatories have larger net benefits than signatories for all levels of cooperation.

On the other hand, it is easy to check that emissions for both non-signatories and signatories decreases as the number of signatories increases. Thus, cooperation decreases both emission and adaptation. The same occurs with total emissions as the following expressions show

$$\frac{\partial e^s}{\partial n} = -\frac{1}{\gamma} \frac{\gamma c(\gamma c d - \alpha N)(\gamma c + 2n - N - 1)}{((\gamma c)^2 - (N + n - 1)\gamma c + (n-1)(N-n))^2} < 0,$$

$$\frac{\partial e^s}{\partial n} = -\frac{1}{\gamma} \frac{(\gamma c d - \alpha N)((2n-1)\gamma c - (n-1)^2)}{((\gamma c)^2 - (N + n - 1)\gamma c + (n-1)(N-n))^2} < 0,$$

$$\frac{\partial E}{\partial n} = -\frac{c(\gamma c d - \alpha N)((2n-1)\gamma c - (n-1)^2)}{((\gamma c)^2 - (N + n - 1)\gamma c + (n-1)(N-n))^2} < 0.$$

All these derivatives are negative if conditions of Assumption 2 are satisfied. Now, if we look at net benefits, it is clear that benefits and adaptation costs decrease with participation, but it is not clear what occurs with damages. As in the mitigation agreement, damages can increase or decrease with participation both for signatories and non-signatories, but again regardless of damages increase or decrease, cooperation has a positive effect on net benefit for both non-signatories and signatories.[24] Signatories internalize the negative externality caused by pollution and as a result of this the net benefits increase monotonically with membership.

Taking the first derivative of the net benefits with respect to n for signatories yields

$$\frac{\partial W^s}{\partial n} = \alpha \frac{\partial e^s}{\partial n} - \gamma e_s \frac{\partial e^s}{\partial n} + \frac{\partial a^s}{\partial n} E - (d - a^s) \frac{\partial E}{\partial n} - ca^s \frac{\partial a^s}{\partial n},$$

that taking into account that the effect of participation in total emissions is given by (18) can be reorganized as follows:

---

[24] We omit the details of this claim in order to shorten the extension of the paper.

$$\frac{\partial W^s}{\partial n} = (\alpha - \gamma e^s - (d - a^s))\frac{\partial e^s}{\partial n} + (E - ca^s)\frac{\partial a^s}{\partial n} - (d - a^s)(e^s + (n-1)\frac{\partial e^s}{\partial n} - e^f + (N-n)\frac{\partial e^f}{\partial n}),$$

where the first term of the RHS is zero according to FOC (23) and

$$E - ca^s = \frac{c(\gamma cd - N\alpha)(n-1)}{(\gamma c)^2 - (N + n - 1)\gamma c + (n-1)(N-n)} = (d - a^s)(n-1),$$

that gives

$$\frac{\partial W^s}{\partial n} = (d - a^s)\left((n-1)(\frac{\partial a^s}{\partial n} - \frac{\partial e^s}{\partial n}) + e^f - e^s - (N-n)\frac{\partial e^f}{\partial n}\right),$$

where $\partial a^s / \partial n = \partial e^s / \partial n$ according to (24). Thus, we obtain the following expression for the derivative of signatories' net benefits with respect to n

$$\frac{\partial W^s}{\partial n} = (d - a^s)\left(e^f - e^s - (N-n)\frac{\partial e^f}{\partial n}\right) > 0, \tag{39}$$

since $e^f > e^s$ and non-signatories emissions decrease with the number of signatories.

Proceeding in the same way, the derivative of non-signatories net benefits can be written as follows:

$$\frac{\partial W^f}{\partial n} = (E - ca^f)\frac{\partial a^f}{\partial n} - (d - a^f)(e^s + n\frac{\partial e^s}{\partial n} - e^f + (N - n - 1)\frac{\partial e^f}{\partial n}),$$

where from second stage FOCs $E = ca^f$ resulting in

$$\frac{\partial W^f}{\partial n} = (d - a^f)(e^f - e^s + n\frac{\partial e^s}{\partial n} + (N - n - 1)\frac{\partial e^f}{\partial n}) > 0, \tag{40}$$

since $e^f > e^s$ and emissions for both signatories and non-signatories decrease with participation. Thus, we find that the there are *positive spillovers* for non-signatories stemming from cooperation, i.e., cooperation increases the non-signatories' net benefits as occurs in the case of the complete agreement. Moreover, it is easy to show that the difference in net benefits given by (38) also increases with the participation. Lastly, we claim that the game presents the property of *full cohesiveness*.[25] To end this subsection we would like to point out that all these features of the PANE of the second stage are the same we have found for the case of the mitigation agreement.

4.3 The first stage: the Nash equilibrium of the membership game

In this subsection, we investigate which is the level of participation an adaptation agreement can achieve. For this type of agreement, the stability function $S(n)$ reads as follows:

$$S(n) = -\frac{(n-1)(\gamma cd - \alpha N)^2 F(n)}{((\gamma c)^2 - (n + N - 2)\gamma c + (N + 1 - n)(n - 2))^2((\gamma c)^2 + (N + n - 1)\gamma c - (n-1)(N-n))^2}, \tag{41}$$

---

[25] We omit the proof of these two properties because it follows the same steps we have used to show them for the case of a complete agreement.

where the denominator is positive and $F(n)$ is a polynomial of fifth degree

$$F(n) = f_5 n^5 + f_4 n^4 + f_3 n^3 + f_2 n^2 + f_1 n + f_0, \qquad (42)$$

with

$$f_5 = \gamma c - 1 > 0,$$

$$f_4 = 2(\gamma c)^2 - (5 + 2N)\gamma c + 5 + 2N > 0,$$

$$f_3 = -(\gamma c)^3 - 11(\gamma c)^2 + (12 + N^2 + 6N)\gamma c - 10N - N^2 - 8 < 0,$$

$$f_2 = -2(\gamma c)^4 + (4N - 1)(\gamma c)^3 - (2N^2 - 6N - 19)(\gamma c)^2 - (N^2 + 6N + 16)\gamma c + 16N + 5N^2 + 4 < 0$$

$$f_1 = (\gamma c)^5 - (2N - 11)(\gamma c)^4 + (N^2 - 12N + 4)(\gamma c)^3 + (N^2 - 6N - 16)(\gamma c)^2 - (6N^2 - 8N + 8)\gamma c - 8N - 8N^2 > 0,$$

$$f_0 = -3(\gamma c)^5 + (2N - 7)(\gamma c)^4 + (N^2 + 2N + 4)(\gamma c)^3 + (5N^2 - 8N + 8)(\gamma c)^2 + (8N^2 - 8N)\gamma c + 4N^2 < 0,$$

for $N \geq 3$ provided that $\gamma c > 2N - 1$.[26] Analyzing this polynomial, we can conclude that

Proposition 6 *For interior solutions and $N \geq 11$, if the degree of effectiveness of adaptation $1 - \beta$ is larger than or equal to $(N^2 - 4N + 3)/(2 + (N - 2)\sqrt{(N^2 - 4N + 7)})$ the grand coalition can be stable. However, if it is lower than this threshold value there exists only one stable adaptation agreement with a minimum of participation of three countries and a maximum of six countries regardless of the severity of environmental damages.*

Proof See Appendix A.5. □

As occurs for the mitigation agreement, the grand coalition could be stable too for an adaptation agreement, but we also have in this case that the limit of effectiveness of adaptation that defines the interval for which the grand coalition is stable converges to one very quickly with the total number of countries. For instance, for $N = 20$, $1 - \beta = 0.9863$ what implies that the grand coalition is able to reduce the marginal damages in a 98.63% through the investment in adaptation yielding a vulnerability for the country equal to 0.0137. For $N = 100$, we obtain that $1 - \beta = 0.9995$ that gives a vulnerability of 0.0005. The levels of effectiveness of adaptation that stabilize the grand coalition for an adaptation agreement are very similar to those we obtain for a mitigation agreement. Thus, there are not big differences between the two types of agreements except that if the grand coalition is not stable, an adaptation agreement could be formed with the double of countries that a mitigation agreement allows, six instead of three.[27] But, in practical terms this is not a great difference because in any case the participation in an IEA is very low.

Finally, we would like to highlight that the model suggests that the participation decreases as the effectiveness of adaptation decreases. For instance, if we evaluate $F(n)$ for $n = 6$, we obtain the following polynomial in γc

$$F(\gamma c; n = 6) = 3(\gamma c)^5 - (\gamma c)^4(10N + 13) + (\gamma c)^3(7N^2 + 74N - 224)$$
$$- (\gamma c)^2(-172N + 61N^2 - 812) + 8\gamma c(-184N + 19N^2 + 408) - 80(N - 6)^2,$$

---

[26] We study the sign of these coefficients in Appendix A.4.

[27] Interestingly, Barrett (2020) also obtain for a technological agreement that the maximum participation consists of six countries. In their model, the investment reduces the abatement costs and the marginal damages are also linear.

so that the polynomial equation could have until five positive real roots. But we know that

$$F(\gamma c = 2N - 1; n = 6) = -8N^5 + 136N^4 - 986N^3 + 2096N^2 + 4590N - 5124,$$

is negative for $N \geq 6$. With five roots, the function will have three inflection points given by the solution to

$$F''(\gamma c; n = 6) = 60(\gamma c)^3 - 12(\gamma c)^2(10N + 13) + 6\gamma c(7N^2 + 74N - 224) - 2(61N^2 - 172N - 812),$$

where the second derivative is taken respect to γc. If we evaluate this derivative for $\gamma c = 2N - 1$, we obtain that $F''(\gamma c = 2N - 1; n = 6)$ is positive for $N \geq 5$. This means that $\gamma c = 2N - 1$ could be between the first inflection point and the second inflection point or on the right of the third inflection point. To advance in the analysis, we need to calculate the third derivative

$$F'''(\gamma c; n = 6) = 180(\gamma c)^2 - 24\gamma c(10N + 13) + 6(7N^2 + 74N - 224),$$

that says us that the second derivative has two extremes, first a maximum and then a minimum. As $F'''(\gamma c = 2N - 1; n = 6)$ is positive for $N \geq 4$, $\gamma c = 2N - 1$ could be on the left of the maximum or on the right of the minimum, but it is easy to show that the slope of $F'''(\gamma c = 2N - 1; n = 6)$ is positive that implies that $\gamma c = 2N - 1$ is greater than the minimum of $F''(\gamma c; n = 6)$, but moreover $F''(\gamma c = 2N - 1; n = 6)$ is positive which means that is greater than the third inflection point and we also know that $F(\gamma c = 2N - 1; n = 6) < 0$. With all this information, we can conclude that $\gamma c = 2N - 1$ is between the fourth root and the fifth root of $F(\gamma c; n = 6) = 0$, so that in the interval between $\gamma c = 2N - 1$ and the fifth root, $F(\gamma c; n = 6) < 0$ and consequently $S(6) > 0$ and $n = 6$ is a stable agreement.[28] However, if γc is higher than the fifth root, $F(\gamma c; n = 6) > 0$ and $n = 6$ becomes an unstable agreement. But, as $((2 - \beta)N - 1)/(1 - \beta)$ is an increasing strictly convex function of β, that is equal to $2N - 1$ for $\beta = 0$ it is clear that $\beta > 0$ for the fifth root of $F(\gamma c; n = 6) = 0$. Thus, as β is inversely related with the degree of effectiveness of adaptation, we can conclude that if the effectiveness of adaptation is below the level defined by the fifth root of $F(n) = 0$, an agreement consisting of 6 countries cannot be stable. In other words, there exists a threshold value for the degree of effectiveness of adaptation for $n = 6$ below which this agreement cannot be stable.[29]

This argument is illustrated in Fig.1. In this graph, we plot the implicit function defined by $F(n, \gamma c, N) = 0$ for $N = \{10, 50, 100\}$.

The figure shows that the participation is decreasing with adaptation costs for the different values of N. Using the argument we have just presented we can also say that the membership is directly related with the degree of effectiveness of adaptation or in other words inversely related to the vulnerability of the country to the environmental problem. In the graph, $F(n)$ is positive above the curves and negative below the curves. For instance, for $N = 50$ and $n = 6$, the distance between the point defined by $\gamma c = 2N - 1 = 99$ and the value γc determined by the curve for $n = 6$ defines the interval of values for γc that makes stable the agreement. This distance increases as n decreases, but in the three cases we find

---

[28] Notice that in the proof of Proposition 6, we show that $S(7) < 0$.

[29] We would like to highlight that the same kind of argument leading to the same conclusion can be developed if $F(\gamma c; 6) = 0$ has three roots or only one, and also for $n = \{4, 5\}$.

the curves are decreasing with respect to γc. The conclusion is obvious, if the effectiveness of adaptation is very low, the only stable agreement consists of three countries.

5 Agreement comparison

Lets start the comparative analysis at the country level. First note that for the complete/mitigation agreement both signatories and non-signatories invest equally in adaptation. We will start with the comparison of signatories' adaptation in the adaptation agreement (recall $a^f > a^s$ in this agreement type) with the adaptation in the mitigation agreement that is identical for signatories and non-signatories[30]

$$a_A^s - a_M = \frac{(n-1)(\gamma c(n-1) + N - n)(\gamma cd - \alpha N)}{(\gamma c - (n^2 - n + N))((\gamma c)^2 - (N + n - 1)\gamma c + (n-1)(N-n))}.$$

In order to be able to compare both agreements we have to restrict to the intersection of feasible sets. Knowing that $\gamma cd > \alpha N$ to avoid over-adaptation for both types of agreement, see expressions (15) and (31), we can trivially conclude that the numerator is positive. Meanwhile, in the denominator the right term is the denominator of a_A^s and the left term the denominator of a_M, and we know that the right term is positive if $\gamma c > 2N - 1$ and that the left term is positive if $\gamma c > N^2$. Then as $N^2 > 2N - 1$ for all $N \geq 2$, if we apply the stronger constraint on γc we have that both terms are positive and consequently the denominator of the difference is also positive and we can conclude that $a_A^s > a_M$ and hence that $a_A^f > a_M$. So we can conclude that both signatories and non-signatories in the adaptation agreement adapt more that both signatories and non-signatories in the mitigation agreement. Consequently, it is straightforward to conclude as expected that $A_A > A_M$, i.e., the aggregate level of adaptation when countries firm an adaptation agreement, A_A, is higher than the aggregate level of adaptation for all n when they form a mitigation agreement, A_M.

Next, we compare emissions. If we take into account that countries do not cooperate in the third stage when they sign an adaptation agreement, we have that emissions for signatories in the adaptation agreement are given by the expression: $e_{A_f}^s = (\alpha - d + a_A^s)/\gamma$ that also gives non-signatories' emissions in the mitigation agreement: $e_M^f = (\alpha - d + a_M)/\gamma$. Then, as we have just established that $a_A^s > a_M$ we can conclude that $e_A^s > e_M^f$ and hence given that within agreements, signatories always emit less than non-signatories finally we obtain that $e_A^f > e_A^s > e_M^f > e_M^s$. Under the same rationale as in the aggregate adaptation comparison we can conclude that aggregate emissions in the adaptation agreement will be larger than in the mitigation agreement for all n.

Now we investigate which agreement would allow countries to achieve the largest net benefit when all countries decide to participate in the agreement. Substituting the level of adaptation and emissions when the grand coalition is formed in the net benefit functions given by (8) for the mitigation agreement and by (26) for the adaptation agreement, we obtain that the difference in net benefits is given by the following expression

$$W_M^s(N) - W_A^s(N) = \frac{(\gamma cd - N\alpha)^2 (N-1)^2 ((c\gamma - 2N)c\gamma + N^2)}{2\gamma(\gamma c - N)^2(\gamma c - N^2)(\gamma c - 2N + 1)}$$

---

[30] In the expression the subscript A stands for the adaptation agreement and M for the mitigation agreement.
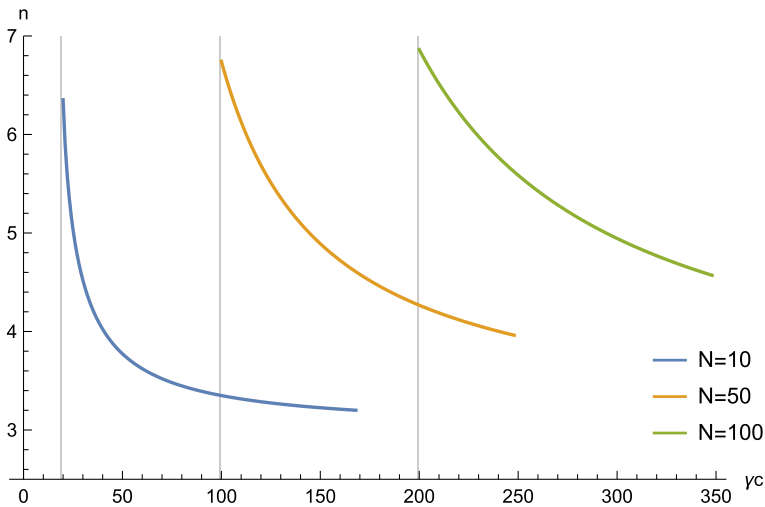
Fig. 1 .

that is positive for $\gamma c > N^2$. Thus, although an adaptation agreement allows countries to reduce the average/marginal damages of total emissions in a larger amount than the one they achieve when a mitigation agreement is formed, the increase in total emission and adaptation costs eliminate this positive effect of the adaptation agreement yielding finally a larger net benefits for signatories when they cooperate only when they take their decisions on emissions. However, as our previous results establish, it does not matter which is the level of net benefits the grand coalition can achieve because for both types of agreements, the scope of cooperation is very limited if we think that N is not a small number. The grand coalition is unstable for realistic levels of vulnerability for both types of agreements, and for realistic values the participation in a complete/mitigation agreement is only of three countries and of a maximum of six countries for the adaptation agreement. Thus, although both type of agreements have different effects in the gains coming from full cooperation, there is not a big difference in terms of coalition stability.

6 Conclusions

This paper analyzes the stability of an IEA when countries invest in adaptation before they take their decisions on emissions. We consider a model with linear damages and we use it to evaluate the stability of three types of agreements: a complete agreement, a mitigation agreement and an adaptation agreement. Moreover, we assume that the investment in adaptation can reduce the vulnerability of the country, but not below a positive lower bound. This means that we are bounding from above the effectiveness of adaptation technology to include, in our opinion, a more realistic modeling of its possible effects. To address the issue of stability we propose a three-stage coalition formation game where in the first stage countries decide non-cooperatively whether or not to sign the IEA. Then, in the second stage, they select the levels of adaptation. For the complete agreement and the adaptation agreement, signatories coordinate their actions at this stage. Finally, countries choose their emissions. At this stage, countries coordinate

their actions for the complete agreement and the mitigation agreement. We solve this game by backward induction.

The analysis shows that to cooperate in the second stage when countries also cooperate when they decide on their emissions does not add anything to the mitigation agreement, in other words, the complete agreement coincides with the mitigation agreement. We find that when signatories decide on investments, the net effects that the decision of a signatory has on the net benefits of the rest of signatories through the effects that adaptation has on emissions is equal to the first order condition for the maximization of agreement's net benefits with respect to emissions. Thus, these net effects are zero doing the cooperation unnecessary.

We also find that for the adaptation agreement and the mitigation agreement, the properties of the adaptation subgame played in the second stage coincide with the properties of the model without adaptation. The unique difference is that in the model without adaptation, emissions are strategic substitutes, but with adaptation we have that the levels of adaptation in the second stage of the game are strategic complements. Moreover, for the model with adaptation we obtain, that environmental damages can increase or decrease with the number of signatories, the reason is that although total emissions decrease when the agreement expands, the adaptation also decreases increasing the marginal damages. Therefore there exists a non-trivial trade-off.

Our findings predict that the grand coalition can be stable for both types of agreements, but for unrealistic levels of the degree of effectiveness of adaptation. For instance, a grand coalition formed by one hundred countries requires a degree of effectiveness of adaptation of 99.96% for the mitigation agreement, and 99.95% for an adaptation agreement. This means that the investment in adaptation must be able to reduce the marginal damages in a percentage larger than 99%. However, with more realistic values for the effectiveness of adaptation, the model yields low levels of participation. Three countries for a mitigation/complete agreement and no more than six for an adaptation agreement. Thus, we could also conclude that it is clear that the complementarity between the investment in adaptation does not have any relevant consequence in the incentive countries have to sign an IEA.

Therefore the main conclusions of this paper are that, we agree with the newly claimed results that the incursion of adaptation in IEAs may mathematically allow an enhance of participation; however, here we show analytically that this requires an extremely high reduction in vulnerability through adaptation. For this reason, we believe that under any realistic assumption regarding the scope of adaptation technology, the inclusion of adaptation does not enhance participation and therefore it should not be considered as a policy solution toward larger agreements.

There are two obvious extensions for the game analyzed in this paper that could be addressed in future research. The first one is developing the stability analysis for a quadratic damage function. The difficulty with the development of this analysis is that the model has not an explicit solution and the analysis will have to be based on numerical methods. Another interesting extension is to drop the assumption of symmetry. In the line of Lazkano et al. (2016) paper, we could consider that countries have different adaptation costs. In this framework, it would be interesting to investigate the role of cooperation in adaptation taking into account the possibility of transfers between countries with different adaptation costs.

Appendix

A.1 Increasing $W^f - W^s$ difference with participation

Given that we have previously shown that $W^f > W^s$, let us analyze how does this difference vary with the level of participation. First, using (7) and (8) we obtain that the difference in net benefits is

$$W^f - W^s = \frac{\alpha}{\gamma}(n-1)(d - a(n)) - \frac{1}{2\gamma}(\alpha - d + a(n))^2 + \frac{1}{2\gamma}(\alpha - nd + na(n))^2$$

$$= \frac{1}{2\gamma}(a(n) - d)^2 (n^2 - 1).$$

Taking the first derivative with respect to n gives the following expression

$$\frac{\partial(W^f - W^s)}{\partial n} = -\frac{1}{\gamma}(d - a(n))\frac{\partial a}{\partial n}(n^2 - 1) + (a(n) - d)^2 n,$$

that is positive because adaptation decreases with respect to n and the net marginal damage, $d - a$, is assumed positive.

A.2 Full cohesiveness

First, using that $E = ca$ for both signatories and non-signatories, net benefits read

$$W^f = \frac{1}{2\gamma}((\alpha^2 - d^2) - (\gamma c - 1)(2da(n) - a(n)^2))$$

for non-signatories and

$$W^s = \frac{1}{2\gamma}((\alpha^2 - d^2 n^2) - (\gamma c - n^2)(2da(n) - a(n)^2)),$$

for signatories. Then the aggregate net benefits can be written as follows

$$W = nW^s + (N - n)W^f$$

$$= n\frac{1}{2\gamma}((\alpha^2 - d^2 n^2) - (\gamma c - n^2)(2da(n) - a(n)^2)) + (N - n)\frac{1}{2\gamma}((\alpha^2 - d^2) - (\gamma c - 1)(2da(n) - a(n)^2))$$

$$= \frac{1}{2\gamma}((N\alpha^2 - (n^3 + N - n)d^2) + (n^3 - n - (\gamma c - 1)N)(2da(n) - a(n)^2)).$$

Substituting $a(n)$ by (13) and taking the first derivative we obtain the following expression

$$\frac{\partial W}{\partial n} = \frac{(N\alpha - \gamma cd)^2 P(n)}{2\gamma (n^2 - n + N - \gamma c)^3},$$

where the denominator is negative provided that $\gamma c > N^2$ and

$$P(n) = n^4 + n^3 + 3(\gamma c - 1 - N)n^2 - (4N(\gamma c - 1) - 1)n + (2N - 1)\gamma c - N.$$

For this polynomial $3(\gamma c - 1 - N)$, $4N(\gamma c - 1) - 1$ and $(2N - 1)\gamma c - N$ are positive if $\gamma c > N^2$. Then, according to Descartes' rule of signs, $P(n) = 0$ could have a maximum of two positive real roots. As the independent term is positive if this is the case, the polynomial must be positive for values of n lower than the smallest root and higher than the largest root and negative between the two roots. In fact, it is easy to check that polynomial equation has two positive real roots since for $n = 1$, the polynomial gives a negative value: $P(1) = -2\gamma c(N - 1)$. Moreover, we also obtain a negative value for $n = N$, $P(N) = -(\gamma c - N^2)(N - 1)^2$, so that we can conclude that for all $n \in [1, N]$, $P(n)$ must be negative and consequently $\partial W / \partial n > 0$ since the denominator is also negative for $\gamma c > N^2$.

A.3 Proof of Proposition 5

We begin this proof showing that the polynomial equation, $F(n) = 0$, defined by the polynomial (22) has at least one positive real root that is in the interval (3, 4). Calculating the values of the polynomial for the extremes of this interval we obtain that for $n = 3$

$$F(3) = 4((N - 1)\gamma c - N^2 - 3N) \geq 0,$$

for $\gamma c \geq (N^2 + 3N)/(N - 1)$. If we compare this lower bound for γc with the one defined in Assumption 1, $\gamma c > N^2$, it is easy to check that $N^2 > (N^2 + 3N)/(N - 1)$ and consequently we can conclude that an agreement consisting of three countries is *internally stable*.

Next, we check whether the agreement is externally stable. For $n = 4$ we have that

$$F(4) = -(\gamma c)^2 + (20 + 6N)\gamma c - 5N^2 - 56N - 144 < 0,$$

for $\gamma c = N^2$ and $N \geq 7$. Moreover, considering the product γc as the argument of the polynomial, the first derivative with respect to γc evaluated at $\gamma c = N^2$ is also negative so that we can establish that N^2 is larger than the highest root of $F(4) = 0$ and consequently that $F(4) < 0$ for $\gamma c > N^2$, and also for $\gamma c > (N^2 - \beta N)/(1 - \beta)$ since $(N^2 - \beta N)/(1 - \beta) > N^2$. Then, $S(4) = W^s(4) - W^f(3)$ is also negative and $W^s(4) < W^f(3)$ so that an agreement consisting of three countries is also *externally stable*. Thus, we can conclude that there exists at least one stable agreement consisting of three countries.

However, according to Descartes' rules of signs $F(n) = 0$ could have three positive real roots. Taking into account that the coefficients of the first derivative change the sign twice, we can conclude that the function has two extremes. Moreover, as the coefficients of the the second derivative also change the sign twice, we can conclude that the function has two inflection points. As the independent term and the leading coefficient of this polynomial are negative we know that until the first inflection point the function is concave, between the first and the second inflection point is convex and on the right of the second inflection point is again concave. Then, the first extreme will be a minimum and the second extreme a maximum and the function will be positive between zero and the first root, negative between the first root and the second root, positive again between the second root and the third root, and finally negative on the right of the third root.

Next, we investigate whether the grand coalition could be stable. If this is the case, $F(N)$ must be positive or zero.[31] A first straightforward conclusion is that then N must be higher

---

[31] Notice that for the grand coalition, the agreement is stable if it is internally stable.

than the second root and lower than or equal to the third root, since one of the three roots is between 3 and 4 and its slope is negative. This means that the lowest root of $F(n) = 0$ is in this interval. Next, we evaluate $F(n)$ at N. The result is

$$F(N) = -(N - 3)(\gamma c)^2 + (N^3 - 3N^2 + 2N - 2)\gamma c - N^2(N - 2)(N^2 - N + 2).$$

Doing $F(N) = 0$, we obtain a second degree equation for c that has two positive roots

$$(\gamma c)_1 = N^2 < (\gamma c)_2 = N^2 + \frac{4(N - 1)}{N - 3},$$

so that for $\gamma c \in (N^2, N^2 + \frac{4(N-1)}{N-3}]$, $F(N) \geq 0$ and the grand coalition is stable. However, according to Assumption 1, γc must be larger than $(N^2 - \beta N)/(1 - \beta) > N^2$. As this lower bound for γc is increasing with β and is equal to N^2 for $\beta = 0$, we can calculate the critical value for β that defines the set of values for this parameter for which the grand coalition is stable solving the following equation

$$N^2 + \frac{4(N - 1)}{N - 3} - \frac{N^2 - \beta N}{1 - \beta} = 0,$$

that yields

$$\beta_N = \frac{4}{N^2 - 3N + 4},$$

and a degree of effectiveness equal to

$$1 - \beta_N = \frac{N^2 - 3N}{N^2 - 3N + 4}.$$

Thus, if $\gamma c \in (N^2, N^2 + \frac{4(N-1)}{N-3}]$ the grand coalition is stable for $\beta \in (0, \beta_N]$. In this case, we have a first root between 3 and 4, a second root between 4 and N, and a third root on the right of N.[32] However, if $\gamma c > N^2 + (4(N - 1))/(N - 3)$, $F(N) < 0$ and the grand coalition is not stable. Then, it is easy to check that N is on the left of the second root of $F(n) = 0$ and that consequently there is only one stable agreement consisting of three countries. Notice that if $\gamma c = N^2 + (4(N - 1))/(N - 3)$, $F(N) = 0$, and N is a root of $F(n) = 0$. However, it could be the second or the third root. To solve this question, we need to check the slope of the root because the second one has a positive slope and the slope of the third one is negative. Evaluating the first derivative of $F(n)$ at $n = N$ we obtain the following expression

$$F'(N) = -(\gamma c)^2 + (6N^2 - 14N + 6)\gamma c - 5N^4 + 14N^3 - 14N^2 + 8N,$$

that is positive for $c = N^2 + (4(N - 1))/(N - 3)$

$$F'\left(N; \gamma c = N^2 + \frac{4(N - 1)}{N - 3}\right) = 8\frac{(N - 1)^2}{(N - 3)^2}(N^2 - 6N + 7) > 0.$$

Thus, we can conclude that for $\gamma c = N^2 + (4(N - 1))/(N - 3)$, N is the second root of $F(n) = 0$, so that for $\gamma c > N^2 + (4(N - 1))/(N - 3)$, $F(N)$ must be negative, because on the

---

[32] Notice that between 3 and 4 with $F(3) > 0$ and $F(4) < 0$ we cannot have more than one root. Theoretically, we could have three root, but then $F(N)$ could not be positive.

right of the second root the function takes positive values. Thus, N will be between the first and the second root and the only stable agreement consists of three countries.

A.4 Signs of the coefficients of $F(n)$ for the adaptation agreement

It is trivial under assumption $\gamma c > 2N - 1$ that $f_5 > 0$. f_4 is a quadratic function of γc with positive $(\gamma c)^2$ coefficient. Its roots are:

$$\frac{1}{4}(5 + 2N \pm \sqrt{4N^2 + 4N - 15})$$

It is easy to check that positive root which is the largest root is smaller than $2N - 1$ for any positive value of N. Let's suppose the contrary

$$\frac{1}{4}(5 + 2N + \sqrt{4N^2 + 4N - 15}) \geq 2N - 1,$$

what implies that

$$4N^2 + 4N - 15 - (6N - 9)^2 = -32N^2 + 112N - 96 \geq 0,$$

but this is a contradiction for $N \geq 3$. Therefore we can conclude that $f_4 > 0$ for $\gamma c > 2N - 1$. For f_3 according to Descrates' rule of signs, $f_3(c) = 0$ could have two positive real roots. With negative values on the left of the first root and on the right of the second root. Evaluating f_3 at $\gamma c = 2N - 1$, we obtain a negative value

$$f_3(2N - 1) = -6N^3 - 22N^2 + 46N - 30 < 0 \text{ for } N \geq 2,$$

and also a negative value for the first derivative

$$f_3'(2N - 1) = -11N^2 - 26N + 31 < 0 \text{ for } N \geq 2.$$

Then, taking into account that the function is concave for all $\gamma c > 0$, $2N - 1$ must be larger than the second root and we can conclude that $f_3 < 0$ for all $\gamma c > 2N - 1$. $f_2(\gamma c) = 0$ according to Descartes' rule of signs could have three positive real roots with negative values on the right of the third root. If we show that $2N - 1$ is on the right of the third root we could conclude that f_2 is negative for all $\gamma c > 2N - 1$. We cannot obtain the roots of the polynomial equation, but if we obtain that $f_2(2N - 1)$ is negative and that $2N - 1$ is higher than the highest inflection point of the function, we could conclude that f_2 is negative for all $\gamma c > 2N - 1$. Evaluating the function at $2N - 1$ we obtain the following expression

$$f_2(2N - 1) = -8N^4 + 38N^3 + 32N^2 - 74N + 38 < 0 \text{ for } N \geq 6.$$

Now, as the second derivative of f_2 is a quadratic function we can calculate the inflection points. The largest is

$$(\gamma c)^i = \frac{6(4N - 1) + (12(72N + 16N^2 + 307))^{1/2}}{48}.$$

Let's suppose that

$$(\gamma c)^i = \frac{6(4N-1) + (12(72N + 16N^2 + 307))^{1/2}}{48} \geq 2N - 1,$$

which implies that

$$12(72N + 16N^2 + 307) - (72N - 42)^2$$
$$= -4992N^2 + 6912N + 1920 \geq 0,$$

that is a contradiction for $N \geq 2$. Then, $\gamma c = 2N - 1$ is greater than the second inflection point and must be on the right of the third root of $f_2(\gamma c) = 0$, so that for $\gamma c > 2N - 1$, $f_2 < 0$. For f_1 we follow the same argument we have used for f_2 except that now the function takes positive values on the right of the third root. Evaluating the function at $2N - 1$ we obtain the following result

$$f_1(2N - 1) = 8N^5 + 56N^4 - 178N^3 + 94N^2 - 18N - 2 > 0 \text{ for } N \geq 2.$$

Next, we calculate the second derivative

$$f_1'' = 20(\gamma c)^3 - 12(2N - 11)(\gamma c)^2 + 6(N^2 - 12N + 4)\gamma c + 2(N^2 - 6N - 16),$$

so that $f_1''(\gamma c) = 0$ will give the two positive inflection points the function has. On the left of the first inflection point, f_1'' is positive and the function is convex, between the the two inflection points f_1'' is negative and the function is concave and, finally, on the right of the second inflection point f_1'' is again positive and the function is again convex. If we evaluate the second derivative at $2N - 1$ we obtain a positive value

$$f_1''(2N - 1) = 76N^3 + 236N^2 - 324N + 56 > 0 \text{ for } N \geq 2.$$

But, $2N - 1$ could be on the left of the first inflection point or on the right of the second inflection point. To find out which is the case, we need to calculate the third derivative

$$f_1''' = 60(\gamma c)^2 - 24(2N - 11)\gamma c + 6(N^2 - 12N + 4).$$

This derivative is zero for the following values

$$\gamma c = \frac{24(2N - 11) \pm \sqrt{288(-28N + 3N^2 + 222)}}{120},$$

that define two extremes for the second derivative. The lowest value is a maximum and the highest value is a minimum. It is easy to check that $2N - 1$ is higher than the minimum.

Let's suppose that

$$\frac{24(2N - 11) + \sqrt{288(-28N + 3N^2 + 222)}}{120} \geq 2N - 1,$$

which implies that

$$288(-28N + 3N^2 + 222) - (192N + 144)^2$$
$$= -36\,000N^2 - 63\,360N + 43\,200 \geq 0,$$

that is a contradiction for $N \geq 2$. Then as $f_1''(2N - 1)$ is positive, $2N - 1$ must be higher than the highest inflection point and as $f(2N - 1)$ is also positive, $2N - 1$ must be higher than the third root of $f_1(\gamma c) = 0$ and we can conclude that $f_1 > 0$ for all $\gamma c > 2N - 1$. Finally, $f_0(\gamma c) = 0$ has a unique positive real root with a leading coefficient negative. So, if $f_0(2N - 1)$ is negative, f_0 will be negative for $\gamma c > 2N - 1$ as is the case

$$f_0(2N - 1) = 2N\left(-28N^4 + 44N^3 + 5N^2 - 26N + 9\right) < 0 \text{ for } N \geq 2.$$

A.5 Proof of Proposition 6

Notice firstly that $S(n) = 0$ only if $F(n) = 0$. Thus, we can focus on the analysis of polynomial equation $F(n) = 0$. As the leading coefficient of $F(n)$ is positive and its independent term is negative, the polynomial equation has at least one positive root. Next, we show that if there exists only one positive root, it is larger than 2 and lower than 7. To claim this, we need to show that $F(3)$ is negative and $F(7)$ is positive. For $n = 3$, $F(n)$ yields the following polynomial in γc

$$F(\gamma c; n = 3) = -2(2\gamma c - 1)G(\gamma c), \tag{43}$$

where

$$G(\gamma c) = (N - 2)(\gamma c)^3 - (N^2 - 4)(\gamma c)^2 + (2N^2 - 7N + 3)(\gamma c) - (N - 3)^2.$$

Therefore to show that $F(\gamma c)$ is negative we need to show that $G(\gamma c)$ is positive for $\gamma c > 2N - 1$. $G(\gamma c)$ presents three changes of sign for the coefficients and according to the Descartes' rule of signs, the polynomial equation could have a maximum of three positive real roots. With positive values on the right of the third root since the leading coefficient is positive. If we show that $2N - 1$ is on the right of the third root, we could conclude that $G(\gamma c)$ is positive for all $\gamma c > 2N - 1$. However, we cannot calculate the roots of the polynomial equation, but if we obtain that $G(2N - 1)$ is positive and that $2N - 1$ is higher than the highest inflection point of the function, we could conclude that $G(\gamma c)$ is positive for all $\gamma c > 2N - 1$ and consequently $F(\gamma c; n = 3)$ negative. Evaluating the function at $2N - 1$ we obtain the following expression

$$G(2N - 1) = 4N^4 - 20N^3 + 28N^2 - 10N - 6 > 0 \text{ for } N \geq 4.$$

Next, as the second derivative de $G(\gamma c)$ is a quadratic function we can calculate the inflection points. The largest is

$$(\gamma c)^i = \frac{N^2 - 4 + \sqrt{(N - 2)\left(17N - 4N^2 + N^3 - 17\right)}}{3(N - 2)}$$

$$(\gamma c)^i = \frac{N^2 - 4 + \sqrt{(N - 2)\left(17N - 4N^2 + N^3 - 17\right)}}{3(N - 2)}$$

Let's suppose that

$$(\gamma c)^i = \frac{N^2 - 4 + \sqrt{(N - 2)\left(17N - 4N^2 + N^3 - 17\right)}}{3(N - 2)} \geq 2N - 1,$$

which implies that

$$(N-2)\left(17N - 4N^2 + N^3 - 17\right) - (5N^2 - 15N + 10)^2$$
$$= -24N^4 + 144N^3 - 300N^2 + 249N - 66 \geq 0,$$

that is a contradiction for $N \geq 3$. So that, $\gamma c = 2N - 1$ is greater than the second inflection point and must be on the right of the third root of $G(\gamma c) = 0$ since $G(2N - 1)$ is positive. Then, we can conclude that for $\gamma c > 2N - 1$, $G(\gamma c) > 0$ and $S(3) > 0$. Obviously, if $G(\gamma c) = 0$ has only one positive root, since the independent term is negative we obtain the same conclusion.

Next, we show that $F(n)$ is positive for $n = 7$. For $n = 7$, $F(n)$ yields the following polynomial in γc

$$F(\gamma c; n = 7) = 2(2(\gamma c)^5 - 2(7 + 3N)(\gamma c)^4 + (4N^2 + 57N - 180)(\gamma c)^3 - (43N^2 - 122N - 928)(\gamma c)^2$$

$$+65(2N^2 - 23N + 63)\gamma c - 75(N - 7)^2), \tag{44}$$

where all polynomials in N are positive for $N > 8$. $F(\gamma c; n = 7)$ presents five changes in the sign of coefficients and according to the Descartes' rule of signs, the polynomial equation $F(\gamma c; n = 7) = 0$ could have five, three or one positive real root. However, regardless of the equation has five or three roots if we have that if $F(2N - 1; n = 7)$ is positive and $2N - 1$ is larger than the highest inflection point as the leading coefficient of (52) is positive, we could conclude as in the previous case that $F(\gamma c; n = 7) > 0$ for $\gamma c > 2N - 1$. For $\gamma c = 2N - 1$, we have the following expression

$$F(2N - 1; n = 7) = 88N^4 - 1432N^3 + 4232N^2 + 12\,268N - 13\,356 > 0 \text{ for } N \geq 11.$$

Now, we calculate the second derivative

$$F''(\gamma c; n = 7) = 2(40c^3 - 24(7 + 3N)c^2 + 6(4N^2 + 57N - 180)c - 2(43N^2 - 122N - 928)).$$

Thus, the function could have three or only one inflection point. If the function has three inflection point, the function is convex between the first inflection point and the second inflection point and on the right of the third inflection point. If we evaluate the second derivative at $2N - 1$ we obtain a positive value

$$F''(2N - 1; n = 7) = 160N^3 - 580N^2 - 2836N + 5456 > 0 \text{ for } N \geq 6.$$

But then, $2N - 1$ could be between the the first inflection point and the second inflection point or on the right of the third inflection point. To progress in the argumentation, we need to calculate the third derivative

$$F'''(\gamma c; n = 7) = 2(120(\gamma c)^2 - 48(7 + 3N)\gamma c + 6(4N^2 + 57N - 180)).$$

This derivative is zero for the following values

$$(\gamma c)^i = \frac{2(7 + 3N) \pm \sqrt{-117N + 16N^2 + 1096}}{10},$$

that define two extremes for the second derivative. The lowest value is a maximum and the largest value is a minimum. It is easy to check that $2N - 1$ is higher than the minimum.

Let's suppose that the highest value of γc is larger than $2N - 1$

$$(\gamma c)^i = \frac{2(7 + 3N) + \sqrt{-117N + 16N^2 + 1096}}{10} \geq 2N - 1,$$

which implies that

$$-117N + 16N^2 + 1096 - (14N - 24)^2$$
$$= -180N^2 + 555N + 520 \geq 0,$$

that is a contradiction for $N \geq 4$. Then as $2N - 1$ is higher than the minimum of the second derivative and moreover $F''(2N - 1; n = 7)$ is positive, $2N - 1$ must be larger than the highest inflection point. In this case, as $F(2N - 1; n = 7)$ is also positive, $2N - 1$ must be higher than the the highest root of $F(\gamma c; n = 7) = 0$ and we can conclude that $F(\gamma c; n = 7) > 0$ for all $\gamma c > 2N - 1$. Finally, if $F(\gamma c; n = 7) = 0$ has only one positive root as the leading coefficient is positive and the independent term negative if $F(2N - 1; n = 7)$ is positive, it is also positive for all $\gamma c > 2N - 1$. Thus, if $S(3) > 0$ and $S(7) < 0$, then there exists at least a value for n, n^* such that $S(n^*) = 0$ and $S'(n^*) < 0$. If the root is a natural number, it defines the participation in the stable agreement, if this is not the case, the stable agreement is given by the first natural number on the left of n^*.

However, according to Descartes' rule of signs $F(n) = 0$ could have three positive real roots. Taking into account that the coefficient of the first derivative change the sign twice, we can conclude that the function has two extremes.[33] Moreover, as the coefficients of the second derivative only change the sign once, we can conclude that the function has one inflection point. Moreover, as the independent term of the second derivative is negative and the leading coefficient is positive, the function is first concave and on the right of the inflection point convex. Then, the first extreme will be a maximum and the second extreme a minimum and the function will be negative between zero and the first root, positive between the first root and the second root, negative again between the second root and the third root, and finally negative on the right of the third root.

Next, we investigate whether the grand coalition could be stable. If this is the case, $F(N)$ must be negative or zero. A first straightforward conclusion is that then N must be higher than the second root and lower than or equal to the third root, since one of the three roots is between 3 and 7 and its slope for $F(n)$ is negative. Then if $N > 7$, the lowest root of $F(n) = 0$ must be in this interval. Next, we evaluate $F(n)$ at N yielding the following expression

$$F(N) = (\gamma c)^2 (\gamma c + 1 - 2N)((\gamma c)^2 (N - 3) - 2\gamma c(N - 1)(N - 2) + 2(N - 2)^2). \quad (45)$$

The sign of this expression depends of a second degree equation for γc that has two positive roots

$$\gamma c = \frac{(N - 2)((N - 1) \pm \sqrt{(N^2 - 4N + 7)})}{N - 3},$$

so that for γc in the close interval defined by these two roots, $F(N) \leq 0$ and the grand coalition could be stable. Does $2N - 1$ belong to this interval?

---

[33] The function could be increasing for all $n > 0$, but in this case as the leading coefficient of $F(n)$ is positive and the independent tern negative, $F(n) = 0$ would have only one positive root in the interval $(3, 7)$ as we have just showed.

Let's suppose that

$$\frac{(N-2)((N-1)-\sqrt{(N^2-4N+7)})}{N-3} \geq 2N-1,$$

which implies that

$$-N^2+4N-1 \geq (N-2)\sqrt{(N^2-4N+7)} > 0,$$

that is a contradiction for $N \geq 4$.

Next, let's suppose that

$$2N-1 \geq \frac{(N-2)((N-1)+\sqrt{(N^2-4N+7)})}{N-3},$$

which implies that

$$(N^2-4N+1)^2-(N-2)^2(N^2-4N+7)$$
$$= -9N^2+36N-27 \geq 0,$$

that is a contradiction for $N \geq 4$. Thus, we can conclude that

$$2N-1 \in \left(\frac{(N-2)((N-1)-\sqrt{N^2-4N+7})}{N-3}, \frac{(N-2)((N-1)+\sqrt{N^2-4N+7})}{N-3} \right), \tag{46}$$

However, according to Assumption 2, γc must be larger than $(((2-\beta)N-1)/(1-\beta) > 2N-1$. As this lower bound for γc is increasing with β and is equal to $2N-1$ for $\beta = 0$. We can calculate the critical value for β that defines the set of values for this parameter for which the grand coalition is stable solving the following equation

$$\frac{(N-2)((N-1)+\sqrt{N^2-4N+7})}{N-3} - \frac{(2-\beta)N-1}{1-\beta} = 0,$$

where the first term is the upper limit of the interval (46). The result of this equation is

$$\beta_N = -\frac{N^2-4N+1-(N-2)\sqrt{N^2-4N+7}}{2+(N-2)\sqrt{(N^2-4N+7)}},$$

and a degree of effectiveness of adaptation equal to

$$1-\beta_N = \frac{N^2-4N+3}{2+(N-2)\sqrt{N^2-4N+7}}.$$

Thus, if

$$\gamma c \in \left(2N - 1, \frac{(N-2)((N-1) + \sqrt{N^2 - 4N + 7})}{N - 3} \right]$$

the grand coalition is stable for $\beta \in (0, \beta_N]$. In this case, we have a first root between 3 and 7, a second root between 7 and N, and a third root on the right of N.[34] However, if $\gamma c > (N-2)((N-1) + \sqrt{(N^2 - 4N + 7)})/(N-3)$, $F(N) > 0$ and the grand coalition is not stable. But, in this case it is easy to check that N will be on the left of the second root of $F(n) = 0$ and that consequently there will be only one stable agreement with a number of signatories between 3 and 6. Notice that if $\gamma c = (N-2)((N-1) + \sqrt{(N^2 - 4N + 7)})/(N-3)$, $F(N) = 0$, and N is a root of equation $F(n) = 0$. However, it could be the second or the third root. To find out which is the case, the only thing we have to do is to check the slope of the root because the second one has a negative slope whereas the slope of the third root is positive. Evaluating the first derivative of $F(n)$ at $n = N$ we obtain the following polynomial in γc, $F'(N) = \gamma c H(\gamma c)$ where

$$H(\gamma c) = (\gamma c)^4 - (6N - 11)(\gamma c)^3 + (6N^2 - 14N + 4)(\gamma c)^2 + (4N^3 - 20N^2 + 32N - 16)\gamma c$$
$$- 4N^3 + 18N^2 - 24N + 8,$$

that is negative for $\gamma c = (N-2)((N-1) + \sqrt{(N^2 - 4N + 7)})/(N-3)$,

$$F'\left(N; c = \frac{(N-2)((N-1) + \sqrt{(N^2 - 4N + 7)})}{N - 3} \right)$$
$$= -\frac{(N-2)^2}{(N-3)^4} \left(2(4N^4 - 51N^3 + 191N^2 - 313N + 209) \right.$$
$$+ \left(\sqrt{-4N + N^2 + 7} \right)^3 (N-2)(-17N + 2N^2 + 25)$$
$$\left. - \sqrt{-4N + N^2 + 7}(N-2)(-21N + 2N^2 + 53)(N-1)^2 \right) < 0 \text{ for } N \geq 2.$$

Then, we can conclude that when $\gamma c = (N-2)((N-1) + \sqrt{(N^2 - 4N + 7)})/(N-3)$, N is the second root of $F(n) = 0$. This implies that for $c > (N-2)((N-1) + \sqrt{(N^2 - 4N + 7)})/(N-3)$, N must be on the left of the second root because on the right, the function takes negative values. Thus, N will be between the first and the second root and in this case there is only one stable agreement with a maximum of participation of six countries.

---

[34] Notice that between 3 and 7 with $F(3) < 0$ and $F(7) > 0$ we cannot have more than one root. Theoretically we could have three root, but then $F(N)$ could not be negative.

References

Barrett, S. (1994). Self-enforcing international environmental agreements. *Oxford Economic Papers, 46*, 878–894. https://doi.org/10.4324/9781315202310-13

Barrett, S. (2006). Climate treaties and "breakthrough" technologies. *American Economic Review, 96*, 22–25. https://doi.org/10.1257/000282806777212332.

Barrett, S. (2020). Dikes versus windmills: Climate treaties and adaptation. *Climate Change Economics, 11*(04), 2040005. https://doi.org/10.1142/s2010007820400059

Battaglini, M., & Harstad, B. (2016). Participation and duration of environmental agreements. *Journal of Political Economy, 124*, 160–124. https://doi.org/10.1086/684478

Bayramoglu, B., Finus, M., & Jacques, J.-F. (2018). Climate agreements in a mitigation-adaptation game. *Journal of Public Economics, 165*, 101–113. https://doi.org/10.1016/j.jpubeco.2018.07.005

Bayramoglu, B., Jacques, J.-F., & Finus, M. (2017). L'adaptation est-elle un frein aux accords climatiques? *Revue Française d' Économie, 32*, 135–159.

Benchekroun, H., Marrouch, W., & Ray Chaudhuri, A. (2017). Adaptation technology and free-riding incentives in international environmental agreements. In *Economics of International Environmental Agreements*, edited by M. Özgür Kayalica, Selim Ça ğatay, and Hakan Mihçi, Chapter 11, Routledge.

Breton, M., & Sbragia, L. (2017). Adaptation to climate change: Commitment and timing issues. *Environmental and Resource Economics, 68*, 975–995. https://doi.org/10.1007/s10640-016-0056-9

Breton, M., & Sbragia, L. (2019). The impact of adaptation on the stability of international environmental agreements. *Environmental and Resource Economics, 74*, 697–725. https://doi.org/10.1007/s10640-019-00341-y

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

d'Aspremont, C., & Jacquemin, A. (1988). Cooperative and noncooperative R&D in duopoly with spillovers. *American Economic Review, 78*, 1133–1137.

d'Aspremont, C., Jacquemin, A., Gadszeweiz, J. J., & Weymark, J. A. (1983). On the stability of collusive price leadership. *Canadian Journal of Economics, 16*, 17–25. https://doi.org/10.2307/134972

El-Sayed, A., & Rubio, S. J. (2014). Sharing R&D investments in cleaner technologies to mitigate climate change. *Resource and Energy Economics, 38*, 168–180.

Finus, M., & Caparrós, A. (2015). *Game theory and international environmental cooperation: essential readings*. United Kingdom: Edward Elgar.

Finus, M., Furini, F., & Rohrer, A. V. (2021). The efficiency of international environmental agreements when adaptation matters: Nash-Cournot vs Stackelberg leadership. *Journal of Environmental Economics and Management, forthcoming.*https://doi.org/10.1016/j.jeem.2021.102461

Goeschl, T., & Perino, G. (2017). The climate policy hold-up: Green technologies, intellectual property rights, and the abatement incentives of international agreements. *Scandinavian Journal of Economics, 119*, 709–732. https://doi.org/10.1111/sjoe.12179

Harstad, B. (2012). Climate contracts: A game of emissions, investments, negotiations, and renegotiations. *Review of Economic Studies, 79*, 1527–1557. https://doi.org/10.1093/restud/rds011

Harstad, B. (2016). The dynamics of climate agreements. *Journal of the European Economic Association, 14*, 719–752.

Harstad, B., Lancia, F., & Russo, A. (2019). Compliance technology and self-enforcing agreements. *Journal of the European Economic Association, 17*, 1–29. https://doi.org/10.1111/jeea.12138

Helm, C., & Schmidt, R. C. (2015). Climate cooperation with technology investments and border carbon adjustment. *European Economic Review, 75*, 112–130. https://doi.org/10.1016/j.euroecorev.2015.01.007

Hoel, M., & de Zeeuw, A. (2010). Can a focus on breakthrough technologies improve the performance of international environmental agreements? *Environmental and Resource Economics, 47*, 395–406. https://doi.org/10.1007/s10640-010-9384-3

Hong, F., & Karp, L. (2012). International environmental agreements with mixed strategies and investment. *Journal of Public Economics, 96*, 685–697. https://doi.org/10.1016/j.jpubeco.2012.05.003

Lazkano, I., Marrouch, W., & Nkuiya, B. (2016). Adaptation to climate change: How does heterogeneity in adaptation costs affect climate coalitions? *Environment and Development Economics, 21*, 812–838. https://doi.org/10.1017/s1355770x16000097

Li, H., & Rus, H. A. (2018). Climate change adaptation and international mitigation agreements with heterogeneous countries. *Journal of the Association of Environmental and Resource Economists, 6*, 503–530. https://doi.org/10.1086/702644

Marrouch, W., & Ray Chaudhuri, A. (2016). International environmental agreements: Doomed to fail or destined to succeed? *International Review of Environmental and Resource Economics, 9*, 245–319. https://doi.org/10.1561/101.00000078

Masoudi, N., & Zaccour, G. (2017). Adapting to climate change: Is cooperation good for the environment? *Economics Letters, 153*, 1–5. https://doi.org/10.1016/j.econlet.2017.01.018

Masoudi, N., & Zaccour, G. (2018). Adaptation and international environmental agreements. *Environmental and Resource Economics, 71*, 1–21.

Montero, J.-P. (2002). Permits, standards, and technology innovation. *Journal of Environmental Economics and Management, 44*, 23–44. https://doi.org/10.1006/jeem.2001.1194

Rubio, S. J. (2017). Sharing R&D investments in breakthrough technologies to control climate change. *Oxford Economic Papers, 69*, 496–521. https://doi.org/10.1093/oep/gpw067

Zehaie, F. (2009). The timing and strategic role of self-protection. *Environmental and Resource Economics, 44*, 337–350. https://doi.org/10.1007/s10640-009-9289-1

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Miguel Borrero[1] · Santiago J. Rubio[1]

✉ Miguel Borrero
miguel.borrero@uv.es

[1] Department of Economic Analysis and ERI-CES, University of Valencia, Facultad de Economía, Avda. de los Naranjos s/n, 46022 Valencia, Spain