May 2021

Working Paper

016.2021

An Adaptation-Mitigation Game: Does Adaptation Promote Participation in International Environmental Agreements?

Miguel Borrero, Santiago J. Rubio

An Adaptation-Mitigation Game: Does Adaptation Promote Participation in International Environmental Agreements?

By Miguel Borrero, University of Valencia
Santiago J. Rubio, University of Valencia

Summary

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 their levels of emissions. Two types of agreements are studied, a complete agreement for which countries coordinate their decisions on adaptation and emissions, and an adaptation agreement for which there is only coordination when countries decide their levels of adaptation. In both cases, 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 results show that the grand coalition could be stable for both types of agreement, but 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 for the 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

JEL Classification: D62, F53, H41, Q54

Address for correspondence:
Santiago J. Rubio
Facultad de Economìa
University of Valencia
Avda. de los Naranjos s/n
46022 Valencia
Spain
E-mail: santiago.rubio@uv.es

An Adaptation-Mitigation Game: Does Adaptation Promote Participation in International Environmental Agreements? *

Miguel Borrero and Santiago J. Rubio†
Department of Economic Analysis and ERI-CES
University of Valencia

April 8, 2021

*The authors want to thank participants at the 2020 Italian Association of Environmental and Resource Economists Conference (Brescia), the Economic Analysis Department Seminar of the Autonomous University of Madrid (March, 2020), and the 2020 Annual Conference of the European Association of Environmental and Resource Economists (on-line, Berlin) for stimulating discussion. Santiago J. Rubio gratefully acknowledges financial support from the Spanish Ministry of Science, Innovation and Universities under project PID2019-107895RB-I00, and from the Valencian Generality under project PROMETEO/2019/095. Miguel Borrero acknowledges financial support from the FPU doctorate programe from the Spanish Ministry of Science and Education.

†Address: Facultad de Economía, Avda. de los Naranjos s/n, 46022 Valencia (Spain). E-mail address: santiago.rubio@uv.es.

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 their levels of emissions. Two types of agreements are studied, a complete agreement for which countries coordinate their decisions on adaptation and emissions, and an adaptation agreement for which there is only coordination when countries decide their levels of adaptation. In both cases, 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 results show that the grand coalition could be stable for both types of agreement, but 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 for the 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

JEL Classification System: D62, F53, H41, Q54

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. 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.[1] 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. When this does not occur, a technology agreement does not perform much better than an emission agreement.

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

[1]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.

levels of adaptation as in Breton and Sbragia's (2019) paper.This approach has been followed by others authors as Zehaie (2009), Masoudi and Zaccour (2017), Breton and Sbragia (2017) for analyzing different issues. In these papers, it is assumed that countries take a decision on adaptation in anticipation of mitigation policies that imply a commitment on adaptation before deciding on mitigation.[2] Examples of such measures include building infrastructures for water management (dykes, dams, canal systems), change in land use and housing planning, I&D of resistant crops and investing to improve forecasting and monitoring. Thus, adaptation can be interpreted as an investment countries do to avoid or reduce damages coming from future emissions, and in this case the adaptation stage must occur before the emissions stage. In fact, we could consider the adaptation stage as an investment stage as done explicitly in Masoudi and Zaccour (2017) and the adaptation agreement as a technology agreement where the investment stage comes first than the emission stage.

We use a model with linear damages where the marginal damages represent the vulnerability of the country to pollution, but what is new in our analysis is that we assume that the investment in adaptation can reduce the marginal damages, but not bellow a positive lower bound. 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.[3] 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 two types of agreements: a complete agreement and adaptation agreement [4]. For a complete agreement, in the first stage, countries decide on their participation in the agreeement. In the second stage, signatory countries decide on their levels of adaptation as to maximize the agreement net benefits whereas non-signatory countries select

[2]Harstad et al. (2019) also assume that countries decide on investment before they select the level of emissions, but the focus of the paper is on compliance of an IEA in a repeated game framework. The authors distinguish between adaptation, brown and green technologies and find that the best equilibrium requires countries to overinvest in technologies that are green, but to underinvest in adaptation and brown technologies.

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

[4]The authors have also analyzed the case when there is no agreement in adaptation investment obtaining the same result as in the complete agreement. i.e grand coalition is possible but only for extremely high values of technological effectiveness, otherwise an agreement with only three countries can be stable. Due to the extension of the analytical analysis this agreement type has not been included to avoid making the paper excessively long. Nevertheless, analysis is available upon request to the authors.

their levels of adaptation individually. Finally, in the last stage, acting in the same way countries decide on emissions. The game is solved by backward induction. However, in the adaptation agreement there is no cooperation in the third stage, which means that each country decides noncooperatively its emission levels. One could think that it is not interesting to analyze an agreement on national goods, but there are two reasons to do it. Firstly, Breton and Sbragia (2019) include it in their analysis of the impact of adaptation in IEA and we were interested in knowing which would be the results in our framework. Secondly, as it is assumed that countries decide on adaptation before deciding on emissions, it is easy to check that the investment in adaptation indirectly, i.e. through their influence on emissions, creates negative international externalities. Thus, although adaptation is a national good, because of the timing of the game the decision on adaptation in the second stage generates indirectly international externalities through its influence on emissions in the third stage. In this case countries can find profitable to coordinate their decisions on a national good through an international agreement since the level of emissions will depend on their decision on adaptation. For this reason, cooperation in selecting the level of a national good makes sense in this kind of models because adaptation is decided before countries take their decision on mitigation.

A first thing we would like to highlight from our analysis is that for both types of agreement, the properties of the adaptation subgame played in the second stage coincide with the properties of the model without adaptation. Emissions decrease with the number of signatories, the non-signatories' net benefits are larger than the signatories' net benefits for all level of participation, there are positive spillovers coming from cooperation, i.e. the non-signatories' net benefits increase with membership and the difference in net benefits also increases with membership. Moreover, the two models present the property of full cohesiveness. 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.

The issue of whether the models present strategic substitutes or strategic complements and how this affects the relationship between emissions and adaptation has been analyzed by different authors as Zehaie (2009), Marrouch and Ray Chaudhuri (2011), Ebert and Welsch (2011,2012), Eisenack and Kahler (2016) and Breton and Sbragia (2017). In our model, complementarity is explained by the linearity of the damage function with respect total emissions that yields an equilibrium in dominant strategies in the third stage of the game. This has two consequences, first we find a positive relationship between emissions and adaptation in the third stage of the game, and second we also find a relationship of complementarity between the adaptation of the different countries in the second stage of the game in the two types of agreements we study in this paper provided that the second order conditions for the maximization of net benefits are satisfied.

However, it seems that the 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 effectiveness of adaptation. This is our main contribution to this literature. We define the way to link the effectiveness of adaptation with 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 that if this is not the case the levels of participation are low. For instance, for the complete 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 complete agreement, only an agreement consisting of three countries can be stable as occurs in the standard model without adaptation. For the adaptation agreement, participation can be higher but no agreements consisting of more than six countries can be stable. Moreover, in this case it is easy to check that when the degree of effectiveness of adaptation decreases, the participation also decreases. Thus, for our model we can conclude that adaptation does not promote participation in an IEA.

Another result to highlight, is that damages can increase or decrease when the number of signatories augments. The reason is that participation reduces total emissions, but, on the other hand, increases marginal damages because adaptation also decreases with the number of signatories. We find that if the vulnerability of the country is equal of larger than 0.5, damages are decreasing with the participation. Bellow this critical value the effect depends on the parameter values. But, what is clear from our analysis is that an increase in the number of signatories not necessarily reduces environmental damages. However, this does not modify the standard result that net benefits both for signatories and non-signatories augment with cooperation.

Finally, we could point out that the result requiring an extremely high value of the degree of effectiveness of adaptation or in other words a very low level of vulnerability for the stability of the grand coalition is consistent with the standard result derived by Barrett (1994) on large but shallow coalitions, i.e. an IEA will be signed by a lot of countries but only when the IEA increases global net benefits by very little compared with the fully noncooperative outcome. We find that when the gains of full cooperation are large, the grand coalition is unstable in both types of agreements studied in this paper. The majority of countries do not cooperate when they could obtain large gains from this cooperation, instead only a few number of countries will sign the agreement in this case. Thus, we find a link between the effectiveness of adaptation and the degree of cooperation that is consistent with the paradox of cooperation established by Barrett (1994). We could also offer an alternative interpretation of our results consistent with the puzzle of small coalitions established by Carraro and Siniscalco (1993). If we assume that there exists an upper bound for the effectiveness of adaptation that is lower than the threshold value above which the grand coalition is stable, the only equilibrium for the complete agreement is an agreement consisting of three countries and for the adaptation agreement the participation cannot be higher than six countries or lower than three depending of the degree of effectiveness of adaptation.

It is difficult to compare our analysis with the one developed by Breton and Sbragia (2019) because the only thing we share with this paper is the timing of the game and that adaptation is a binding commitment. The two papers use different specifications for the objective function. Breton and Sbragia's (2019) objective function is given by the addition of the environmental damages plus mitigation costs plus adaptation costs all of them being quadratic functions, and in this paper countries maximize the net benefit of emissions with different specifications of the benefit and damage functions. The benefit function is linear-quadratic and the damage function is linear in total emission so that vulnerability in our model is given by the marginal damages whereas in Breton and Sbragia (2019) is given by the difference between total emissions and adaptation that is the argument of the damage function. Moreover, they do not impose any positive lower bound for marginal damages whereas we do. Finally, they solve their model numerically whereas we solve it analytically. But the main novelty of our analysis is that we study the relationship between the degree of effectiveness in adaptation and the participation in an IEA that is not in Breton and Sbragia' (2019) paper. In both models the grand coalition can be theoretically stable, but we show that this requires an extremely high degree of effectiveness in adaptation. In practice, this means that the grand coalition is not an outcome of the game we should expect if more realistic degrees of effectiveness in adaptation are assumed. Moreover, if the grand coalition is not an option, the levels of participation are low in the two types of agreements we study. Thus, we can conclude that for our model adaptation really does not help to get higher levels of participation in IEAs. However, in Breton and Sbragia (2019) the participation can be high even if the grand coalition is not stable depending of the parameter values. They offer a more optimistic results about the role that adaptation plays for promoting participation than the ones we obtain in this paper. Nevertheless, our conclusion is clear, we may claim that cooperation in adaptation is not a *sufficient condition* to obtain more participation in an environmental agreement. We need something else in the model. Asymmetric knowledge spillovers as in Masoudi and Zaccour (2017) or maybe high effectiveness of adaptation levels as in Breton and Sbragia (2019) where the lower bound of marginal damages is zero.

1.1 Literature Review

The stability analysis we present in the next sections enrols in a large strand of 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).[5]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. This list of papers includes Barrett (2008), Marrouch and Ray Chaudhuri (2011), Lazkano et al. (2016), Benchekroun et al. (2017), Bayramoglu et al. (2017, 2018) and Breton and Sbragia (2019). Barrett (2008) examines a model in which adaptation and mitigation are both binary actions, and where a subset of poor countries are unable to adapt. His results show that adaptation improves the prospects for a cooperative agreement but the numerical exercise he develops suggests that this positive effect is limited when the potential gains from cooperation are large. 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. Our model is different in several aspects including the damage function, and analytically concludes that there are only two possible equilibria for the game: the grand coalition if the degree of effectiveness of adaptation is extremely large or a small coalition consisting only of three countries for the rest of values of the degree of effectiveness

[5]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).

of adaptation. Moreover, we find that if the gains from full cooperation are large, the grand coalition is unstable. Lazkano et al. (2016) as Barrett (2008) also look at the effects that differences in adaptation costs have on participation incentives. They present conditions under which adaptation can strengthen or weaken free-riding incentives. 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.[6] Bayramoglu et al.(2017, 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.[7] 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. Moreover, our results are somewhat more pessimistic.

As far as we know, the unique paper that addresses the stability of an adaptation agreement when the level of adaptation is selected before emissions, and adaptation is considered a private/national good is Breton and Sbragia (2019). Breton and Sbragia's (2019) numerical simulations shows that the cooperation in adaptation can boost participation. Even the grand coalition can be stable for some parameter values as occurs in our investigation. We also addresses this issue, but for a different specification of the net benefit function, obtaining a different result on participation as we have just explained above because our assumption about the upper limit that the degree of effectiveness of adaptation can take. We show analytically that the participation cannot be higher than three countries for the complete agreement although could reach six countries for the case of the adaptation agreement.

[6]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.

[7]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.

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. From this list, we would like to highlight the paper by El-Sayed and Rubio (2014). In their analysis, it is assumed that signatories not only coordinate their levels of investment but also pool them so as to fully internalize the spillover effects of their investments, i.e. investment is considered as a club good. However, the results show, that even assuming a strong asymmetry between signatories and non-signatories as regards the effective investment, the participation in the technology agreement is low. Interestingly, the maximum level of participation consists of six countries.[8]

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 complete agreement. Section 4 analyzes the case of an adaptation agreement, and Section 5 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

[8]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.

pollution and the national private good nature of adaptation.[9] 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.

Each country's net benefits consists 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[10]

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

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 $C(a_i) = ca_i^2/2, \ c > 0$.[11] 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 A complete agreement formation game

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

[9] 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.

[10] 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. Lazkano el al. (2016) have used it to analyze the consequences that differences in adaptation costs have on the incentives to participate in an IEA. We assume that the slope of the benefit function is the unity. However, this assumption has no qualitative effects on the results obtained in this paper.

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

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. Thus, emissions are provided by the *partial agreement Nash equilibrium* (PANE) with respect to a coalition defined by Chander and Tulkens (1995). In the second stage, the adaptation subgame, countries act as in the third stage, but now they decide on investment in adaptation. 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]

3.1 The third stage: an equilibrium in dominant strategies

As we have supposed that 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

[12]As countries that cooperate in the second stage are the same countries that cooperate in the third stage, we could obtain the same solution than the one we derive in this section modeling the IEA formation game as a two-stage game with countries deciding on emissions and adaptation at the same time in the second stage. However, we have decided to keep this structure for the game because it facilities the analysis of the adaptation agreement we present in the second part of the paper where countries do not cooperate when they decide on emissions.

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' adaptation

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

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

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

and as follows for signatories

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

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* as the following derivatives show

$$\frac{\partial W_i^f}{\partial a_j^s} = -(d - a_i^f)\frac{n}{\gamma} < 0, \ \frac{\partial W_j^s}{\partial a_k^s} = \frac{1}{\gamma}(-A^s + a_j^s n) = \frac{n}{\gamma}(a_j^s - \bar{a}^s), \ j, k = 1, ..., n, \ j \neq k, \ i = 1, ..., N - n, \tag{9}$$

$$\frac{\partial W_j^s}{\partial a_i^f} = -(d - a_j^s)\frac{1}{\gamma} < 0, \ \frac{\partial W_i^f}{\partial a_l^f} = -(d - a_i^f)\frac{1}{\gamma} < 0, \ i, l = 1, ..., N - n, \ i \neq l, \ j = 1, ..., n, \tag{10}$$

where $\bar{a}^s$ is the average of the distribution of signatories' adaptation.

3.2 The second stage: the PANE of the adaptation game

In this subsection, we solve stage two assuming that in the first stage n countries, with $n \geq 1$, have signed the agreement.[13] 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

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

where the LHS 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. Taking into account (6), the condition (11) implicitly defines the non-signatory reaction function. Applying the implicit function theorem we obtain that

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

[13] 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.

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. If we focus on the symmetric solution, according to (9), we have that $\partial W_j^s / \partial a_k^s = 0$. In this case, the FOC for the maximization of the agreement net benefits is

$$n\frac{\alpha}{\gamma} - \frac{1}{\gamma}(\alpha - nd + na^s)n + E - (d - a^s)\frac{\partial E}{\partial a^s} - ca^s = 0,$$

that taking into account that for the symmetric solution $\partial E/\partial a^s$ is equal to n^2/γ yields

$$n\frac{\alpha}{\gamma} - \frac{1}{\gamma}(\alpha - nd + na^s)n - (d - a^s)\frac{n^2}{\gamma} + E - ca^s = 0,$$

where the the first three terms cancel according to the FOC (4) of the third stage yielding finally the following condition

$$E = ca^s. \tag{12}$$

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

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

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 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 that is given by

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

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

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

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

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

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. 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[14]

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 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

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

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 on participation. However, it is not so clear what occurs with damages. On 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

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

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.

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

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 1 *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 of 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.

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

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

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

and we can conclude that

Proposition 2 *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

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

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.*[15] 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

[15] These two properties are showed in the Appendix.

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 of the aggregate net benefits for all levels of cooperation. Thus, we can conclude that this result is not an artifact of the 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.

3.3 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 complete agreement that includes cooperation on the levels of adaptation. 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.[16]

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

$$f_3 = 2\gamma c - 2N - 7 > 0,$$

$$f_2 = 8\gamma c - 4N - 3 > 0,$$

$$f_1 = (\gamma c)^2 - 2(N+3)\gamma c + (N-2)N > 0,$$

$$f_0 = 3(\gamma c)^2 - 2(N+2)\gamma c - N^2 > 0,$$

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

Proposition 3 *For interior solutions and* $N \geq 7$, *if the degree of effectiveness of adaptation* $1-\beta$ *is larger 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 provided that* $\gamma c \geq \max[(N^2 + 3N)/(N-1),\ (N^2 - \beta N)/(1-\beta)]$.

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

[16] 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.

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 the 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.

If the effectiveness of adaptation is not extremely high, the gains from cooperation will be low because only three countries will sign the agreement. Nevertheless, it would be interesting to investigate whether our result requiring low vulnerability for the stability of the grand coalition is consistent with the 'paradox of cooperation' that establishes that cooperation appears when we do not need it because the gains of cooperation are low. To address this issue we calculate using expressions (7) and (8) the gains of full cooperation

$$W^s(N) - W^f(1) = \frac{c(\gamma c d - N\alpha)^2 (N-1)^2}{2(\gamma c - N^2)(\gamma c - N)^2} > 0, \tag{23}$$

that are positive since according to Assumption 1 $c\gamma > N^2$. From the proof of Prop. 3 we know that for a given value of γ, the grand coalition will be stable if c is lower than $\frac{N^2}{\gamma} + \frac{4(N-1)}{\gamma(N-3)}$. In other words, we can define an upper bound on c such that if c is larger than this upper bound, the grand coalition is unstable. Thus, if we show that the gains coming from full cooperation increase with respect to c we could conclude that for c larger than the upper bound defined above, the grand coalition yields larger gains but is then unstable.

To find the effect that an increase in c has on the difference (23), we calculate the first derivative with respect to this parameter that gives the following expression

$$\frac{\partial(W^s(N) - W^f(1))}{\partial c} = \frac{N(\gamma c - N)G(\gamma c)}{((\gamma c)^3 - N(N+2)(\gamma c)^2 + N^2(2N+1)\gamma c - N^4)^2}, \tag{24}$$

where

$$G(\gamma c) = G_1(\gamma c)^3 + G_2(\gamma c)^2 + G_3 \gamma c + N^4 \alpha^2,$$

with

$$
\begin{aligned}
G_1 &= d(2\alpha - (N+2)d) > 0, \\
G_2 &= N(3Nd^2 + 2d\alpha - 2\alpha^2) < 0, \\
G_3 &= N^3\alpha(\alpha - 4d) > 0,
\end{aligned}
$$

$N > 4$ provided that $d < \alpha/N$ according to Assumption 1. As the coefficients of the polynomial $G(\gamma c)$ in the numerator change their signs twice, we can conclude according to Descartes' rule of signs that the polynomial equation could have a maximum of two positive roots. Moreover, as the independent term is positive if the polynomial equation has two roots, the polynomial must be positive for values of γc 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 this is the case since for $\gamma c = N^2$,the polynomial yields a negative value:$G(N^2) = -N^4(\alpha - Nd)^2(N-1)$, so that we can conclude that the smallest root is lower than $N^2$and the largest root is higher than N^2. It is also easy to check by substitution that the largest root is $\gamma c = \alpha N/d$. Then as γcmust be higher than $\alpha N/d$for having a positive marginal damage, we have that $G(\gamma c)$ is positive for all $\gamma c > \alpha N/d$what implies that the derivative (24) is positive and we can conclude that the gains of full cooperation increase with c and that for $c > \frac{N^2}{\gamma} + \frac{4(N-1)}{\gamma(N-3)}$, countries would obtain more gains from full cooperation but then the agreement would be unstable.

4 An adaptation agreement formation game

In this section we focus on an agreement on investment in adaptation. As we have assumed that the investment in adaptation is a national good, we could think that this kind of agreements have no influence in the national decisions because there are not international spillovers. However, if the countries decide on adaptation before they choose their levels of emissions, the investment in adaptation will affect the net benefits of the rest of countries through the influence that adaptation has on emissions. 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)

$$\alpha - \gamma e_i = d - a_i,\ i = 1, ..., N. \tag{25}$$

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 are given by

$$e_i = \frac{1}{\gamma}(\alpha - d + a_i). \tag{26}$$

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

$$E = \frac{N(\alpha - d) + A}{\gamma}, \tag{27}$$

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

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

$$W_i = \frac{\alpha}{\gamma}(\alpha - d + a_i) - \frac{1}{2\gamma}(\alpha - d + a_i)^2 - (d - a_i)E - \frac{c}{2}(a_i)^2, \tag{28}$$

where total emissions are given by (27).

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

$$\frac{\partial W_i}{\partial a_j} = -\frac{1}{\gamma}(d - a_i) < 0,\ i, j = 1, ..., N,\ i \neq j. \tag{29}$$

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.

The FOCs for non-signatories are

$$\frac{\alpha}{\gamma} - \frac{1}{\gamma}(\alpha - d + a_i^f) + \frac{N(\alpha - d) + A}{\gamma} = \frac{d - a_i^f}{\gamma} + ca_i^f, \tag{30}$$

where

$$A = \sum_{i=1}^{N-n} a_i^f + \sum_{j=1}^{n} a_j^s,$$

However, as expected if we take into account the FOCs of the third stage, (30) simplifies yielding

$$E = \frac{N(\alpha - d) + A}{\gamma} = ca_i^f, \tag{31}$$

where the LHS stands for the marginal benefit of adaptation given by the reduction in damages because the decrease in the marginal damages caused by adaptation and the RHS stands for the marginal costs of adaptation. Thus, there are no differences with the FOCs obtained for non-signatories in the case of a complete agreement.

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

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

Looking for the symmetric solution, the FOC gives

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

that taking into account the FOC of the first stage yields

$$E = \frac{N(\alpha - d) + A}{\gamma} = \frac{(n-1)}{\gamma}(d - a^s) + ca^s, \tag{32}$$

where the LHS is, as in condition (31), 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 (31) and (32), we may obtain the level of adaptation of the Partial Agreement Nash Equilibrium (PANE) of the second stage

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

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

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 (33) is also positive provided that $\alpha N/(2N-1) > d$.[17] 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

$$d - a^s = \frac{\gamma c(\gamma c d - \alpha N)}{(\gamma c)^2 - (N+n-1)\gamma c + (n-1)(N-n)}. \tag{35}$$

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.

$$d - a^f = \frac{(\gamma c - n + 1)(\gamma c d - \alpha N)}{(\gamma c)^2 - (N+n-1)\gamma c + (n-1)(N-n)}. \tag{36}$$

Now, using (33) and (34) we can calculate the emissions for each type of country

$$e^s = \frac{1}{\gamma}\frac{(\gamma c - N)\alpha + (\alpha - d)(\gamma c)^2 - (\gamma c - 1 - N)\alpha n - \alpha n^2}{(\gamma c)^2 - (N+n-1)\gamma c + (n-1)(N-n)}, \tag{37}$$

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

and adding for all countries we obtain total emissions

$$E = ne^s + (N-n)e^f = \frac{c(N(1+\gamma c)(\alpha - d) + (d - N(\alpha - d))n - dn^2)}{(\gamma c)^2 - (N+n-1)\gamma c + (n-1)(N-n)}. \tag{39}$$

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 $a^f > a^s$.

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

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

that imposes a lower bound on d

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

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.[18] 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

$$W^f = \frac{w_4^f n^4 + w_3^f n^3 + w_2^f n^2 + w_1^f n + w_0^f}{2((\gamma c)^2 - (N+n-1)\gamma c + (n-1)(N-n))^2}, \quad (40)$$

where

$$w_4^f = \alpha^2 - \gamma c d^2,$$

$$w_3^f = 2\left(\gamma c d^2 - \alpha^2\right)(N - \gamma c + 1),$$

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

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

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

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

and the following expression for signatories countries

$$W^s = \frac{w_4^s n^4 + w_3^s n^3 + w_2^s n^2 + w_1^s n + w_0^s}{2((\gamma c)^2 - (N+n-1)\gamma c + (n-1)(N-n))^2}, \tag{41}$$

where

$$w_4^s = \alpha^2 - \gamma c d^2,$$

$$w_3^s = 2\left(\gamma c d^2 - \alpha^2\right)(N - \gamma c + 1),$$

$$w_2^s = \left(\gamma c d^2 - \alpha^2\right)\left((\gamma c)^2 + 4\gamma c - N^2 - 4N - 1\right),$$

$$w_1^s = -2\left((\gamma c)^3\left(2Nd^2 - 2Nd\alpha + \alpha^2\right) - (\gamma c)^2\left((N^2 - 1 - N)d^2 - N(N+2)\alpha^2\right)\right.$$
$$\left. - \gamma c\left(N(N+1)d^2 - (N^2 - 1 - N)\alpha^2\right) + N\alpha^2(N+1)\right)$$

$$w_0^s = (\gamma c + 1)\left(-(\gamma c)^3(\alpha - d)((2N-1)d - \alpha) + (\gamma c)^2\left(+2Nd^2 - N^2 d^2 + (N^2 - 2N + 1)\alpha^2\right)\right.$$
$$\left. - N\gamma c\left(+Nd^2 - (N-2)\alpha^2\right) + N^2\alpha^2\right).$$

Next, we calculate the difference in net benefits using (40) and (41)

$$W^f - W^s = \frac{(n-1)(1 + \gamma c + n(\gamma c - 1))(\gamma c d - \alpha N)^2}{2((\gamma c)^2 - (N+n-1)\gamma c + (n-1)(N-n))^2}, \tag{42}$$

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 complete 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.[19] 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

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

where the first term of the RHS is zero according to FOC (25) 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 (26). Thus, we obtain the following expression for the derivative of signatories' net benefits with respect to n

$$\frac{\partial W^s}{\partial n} = (d - a^s) \left(e^f - e^s - (N-n) \frac{\partial e^f}{\partial n} \right) > 0, \tag{43}$$

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

$$\frac{\partial W^f}{\partial n} = (d - a^f)(e^f - e^s + n \frac{\partial e^s}{\partial n} + (N - n - 1) \frac{\partial e^f}{\partial n}) > 0, \tag{44}$$

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

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

given by (42) also increases with the participation. Lastly, we claim that the game presents the property of *full cohesiveness*.[20] 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 complete 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

$$S(n) = -\frac{(n-1)(\gamma cd - \alpha N)^2 F(n)}{((\gamma c)^2 - (n+N-2)\gamma c + (N+1-n)(n-2))^2((\gamma c)^2 + (N+n-1)\gamma c - (n-1)(N-n))^2}, \tag{45}$$

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

$$F(n) = f_5 n^5 + f_4 n^4 + f_3 n^3 + f_2 n^2 + f_1 n + f_0, \tag{46}$$

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 $\gamma c > 2N - 1$.[21] Analyzing this polynomial, we can conclude that

Proposition 4 *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 is 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.*

[20] 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.

[21] We study the sign of these coefficients in the Appendix.

Proof. See Appendix 5. ■

As occurs for the complete 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 complete 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 complete agreement allows, six instead of three.[22] But, in practical terms this is not a big 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 c

$$\begin{aligned} F(\gamma c; n &= 6) = 3(\gamma c)^5 - (\gamma c)^4 (10N + 13) + (\gamma c)^3 \left(7N^2 + 74N - 224\right) - (\gamma c)^2 \left(-172N + 61N^2 - 812\right) \\ &\quad +8\gamma c \left(-184N + 19N^2 + 408\right) - 80 (N-6)^2 , \end{aligned}$$

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

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

[22] Interestingly, El-Sayed and Rubio (2014) also obtain for a technological agreement that the maximum participation consists of six countries and that the participation decreases as spillovers efffects increase until a minimum of three countries. In their model, the investment reduces the abatement costs and the marginal damages are linear.

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

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

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, $F(6) < 0$ and consequently $S(6) > 0$ and $n=6$ is a stable agreement.[23] However, if γc is higher than the fifth root, $F(6) > 0$ and $n=6$ becomes an unstable agreement. But, as $((2-\beta)N-1)/(1-\beta)$ is an increasing strictly convex function of β, we can conclude that the the threshold value of β associated to the $\gamma c = 2N-1$ is lower than the one corresponding to the fifth root of $F(6) = 0$. Thus, as β is inversely related with the degree of effectiveness of adaptation, we can conclude that if the effectiveness of adaptation is bellow 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 $n=6$ bellow which this agreement cannot be stable.[24]

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

$\Rightarrow$ Figure 1 $\Leftarrow$

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

[23]Notice that in the proof of Proposition 4, we show that $S(7) < 0$.

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

related to the vulnerability of the country to the environmental problem. In the graph, $F(n)$ is positive above the curves and negative bellow 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 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.

Finally, we study the relationship between the gains coming from full cooperation and the level of participation in the agreement. For an adaptation agreement the gains from full cooperation are given by the following expression

$$W^s(N) - W^f(1) = \frac{(N-1)^2(\gamma cd - \alpha N)^2}{2\gamma(\gamma c - 2N + 1)(\gamma c - N)^2} > 0, \tag{47}$$

that is positive since, according to Assumption 2, $\gamma c > 2N - 1$. As occurs with the complete agreement, we know from the proof of Proposition 4 that for any given value of γ, the grand coalition will be stable if

$$c \leq \frac{(N-2)((N-1) + \sqrt{(N^2 - 4N + 7)})}{\gamma(N-3)}. \tag{48}$$

Then, if we obtain that the gains from full cooperation increase with respect to c we could conclude that when the gains from full cooperation are large, the cooperation is unstable.

To find the effect that an increase in c has on (47), we calculate the first derivative with respect to this parameter that yields the following expression

$$\frac{\partial(W^s(N) - W^f(1))}{\partial c} = -\frac{(N-1)^2(\gamma cd - \alpha N)G(\gamma c)}{2(\gamma c - 2N + 1)^2(\gamma c - N)^3}, \tag{49}$$

where

$$G(\gamma c) = d(\gamma c)^2 - N(3\alpha - d)\gamma c + N(\alpha(5N - 2) - d(4N - 2)). \tag{50}$$

The denominator is positive if $\gamma c > 2N - 1$ and if the marginal damages are positive γcd must be greater than αN . If the polynomial equation has two roots, it is easy to check that the first root is not relevant for the analysis because is lower than $2N - 1$. Substituting in (50) we obtain that

$$G(2N-1) = d(2N-1)^2 - N(3\alpha - d)(2N-1) + N(\alpha(5N-2) - d(4N-2))$$

$$= -(N-1)(N\alpha - (2N-1)d) < 0,$$

according to Assumption 2. If the polynomial equation has two roots, as the independent term is positive, the polynomial is positive between zero and the first root, negative between the first and the second root, and again positive on the right on the second root. Thus, if $G(2N-1) < 0$, we can conclude that the first root is lower than $2N-1$. Then the only relevant root is the second one, that defines a maximum for (47). On the other hand, the threshold value $\gamma c = \alpha N/d$ is also a stationary point of (47). This stationary point is a minimum given that

$$\left.\frac{\partial^2(W^s(N) - W^f(1))}{\partial(\gamma c)^2}\right|_{\frac{\alpha N}{d}} = \frac{\alpha d^5(N-1)^2}{(\alpha - d)^2 N^2(\alpha N - (2N-1)d)}$$

which is always positive because $d < \alpha N/(2N-1)$ according to Assumption 2. Moreover, this minimum must be on the left of the second root of (50) since the gains of cooperation must decrease for enough high values of c. Notice that for the gains of cooperation (47), the denominator is cubic in c and cuartic in γ whereas for the numerator the expression is quadratic in c and γ. Moreover, for this reason the polynomial equation $G(\gamma c) = 0$ must have two roots because with $G(\gamma c) > 0$ for all γc, (49) would be negative on the right of the minimum that is a contradiction. Thus, as the stability of the grand coalition requires that γc is below the upper bound defined by (48) we will have that γc is also very close of the minimum of the difference in gains with a net marginal damage close to zero because the effectiveness of adaptation is very large. Therefore this means that as we increase γc we move towards the maximum of gains from full cooperation but then the grand coalition is not stable. So up to now we have the usual large but shallow coalition, however we should mention that after the maximum, the function is decreasing and theoretically we could have smaller coalition with similar or even lower gains from cooperation than the grand coalition. Nevertheless, this is only a small difference with the previous case that we want to mention for completeness because this will occur for enough high values of γc even in this case the differences in gains would be insignificant compared to the differences in gains when small agreements correspond to values of γc close to the maximum. Therefore we can claim that we still observe the cited paradox for the adaptation agreement.

5 Conclusions

This paper analyzes the stability of an IEA when countries invest in adaptation before they take their decisions on emissions. We consider two types of agreements with linear damages: a complete agreement for which signatories coordinate their levels of adaptation and emissions, and an adaptation agreement where signatories only cooperate when they decide on investment in adaptation. Moreover, we assume that this investment can reduce the vulnerability of the country, but not bellow 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 whether or not to sign the IEA. Then, in the second stage, signatories (playing together) and non-signatories (playing individually) select their levels of adaptation. In the third stage, each country decides on its emissions noncooperatively for the adaptation agreement and signatories cooperate in the complete agreement. We solve this game by backward induction.

The analysis shows that for both types of agreements, the properties of the adaptation subgame played in the second stage coincide with the properties of the model without adaptation: Emissions decrease with participation; non-signatories' net benefits are larger than the signatories' net benefits for all levels of participation: there are positive spillovers coming from cooperation, i.e. the non-signatories' net benefits increase with membership and the difference in net benefits also increases with membership. Moreover, both models present the property of full cohesiveness. However, we know that in the model without adaptation, emissions are strategic substitutes, but with adaptation the strategic relationship changes and the investments in adaptation are strategic complements.

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 complete 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%. For these quantities the stability is sustained because the increase in adaptation costs that the country exiting the coalition has to support turns the exit unprofitable. However, with more realistic values for the effectiveness of adaptation, the model yields low levels of participation. Three countries for a 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 one of the main conclusions of this paper is 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 of 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 towards 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.

References

[1] Barrett, Scott (1994). "Self-Enforcing International Environmental Agreements." *Oxford Economic Papers* 46, 878-894.

[2] Barrett, Scott (2006). "Climate Treaties and "Breakthrough" Technologies." *American Economic Review* 96, 22-25.

[3] Barrett, Scott (2008). "Dikes versus Windmills: Climate Treaties and Adaptation." Johns Hopkins University, School of Advanced International Studies.

[4] Battaglini, Marco and Bard Harstad (2016). "Participation and Duration of Environmental Agreements." *Journal of Political Economy* 124, 160-124.

[5] Bayramoglu, Basak, Michael Finus, and Jean-Francois Jacques (2018). "Climate Agreements in a Mitigation-Adaptation Game." *Journal of Public Economics* 165, 101-113.

[6] Bayramoglu, Basak, Jean-François Jacques, et Michael Finus (2017). "L'Adaptation Est-Elle un Frein aux Accords Climatiques?" *Revue Française d'Économie* 32, 135-159.

[7] Benchekroun, Hassan, Walid Marrouch, and Amrita Ray Chaudhuri (2017). "Adaptation Technology and Free-Riding Incentives in International Environmental Agreements." In *Economics of International Environmental Agreements*, edited by M. Özgür Kayalica, Selim Çağatay, and Hakan Mihçi, Chapter 11, Routledge.

[8] Breton, Michèle, and Lucia Sbragia (2017). "Adaptation to Climate Change: Commitment and Timing Issues." *Environmental and Resource Economics* 68,975-995.

[9] Breton, Michèle, and Lucia Sbragia (2019). "The Impact of Adaptation on the Stability of International Environmental Agreements." *Environmental and Resource Economics* 74,697-725.

[10] Caparrós, Alejandro (2018). "A Model of the (Incomplete) Paris Agreement on Climate Change." Mimeo.

[11] Carraro, Carlo, and Domenico Siniscalco (1993). "Strategies for the International Protection of the Environment." *Journal of Public Economics* 52, 309-328.

[12] Chander, Parkash, and Henry Tulkens (1995). "A Core-Theoretic Solution for the Design of Cooperative Agreements on Transfrontier Pollution." *International Tax and Public Finance* 2, 279-294.

[13] d'Aspremont, Claude, and Alexis Jacquemin (1988). "Cooperative and Noncooperative R&D in Duopoly with Spillovers." *American Economic Review* 78, 1133-1137.

[14] d'Aspremont, Claude, Alexis Jacquemin, Jean Jaskold Gadszeweiz, and John A. Weymark (1983). "On the Stability of Collusive Price Leadership." *Canadian Journal of Economics* 16, 17-25.

[15] Ebert, Udo, and Heinz Welsch (2011). "Optimal Response Functions in Global Pollution Problems Can Be Upward-Sloping: Acccounting for Adaptation." *Environmental Economics and Policy Studies* 13, 129-138.

[16] Ebert, Udo, and Heinz Welsch (2012). "Adaptation and Mitigation in Global Pollution Problems: Economic Impact of Productivity, Sensitivity, and Adaptive Capacity." *Environmental and Resource Economics* 52, 49-64.

[17] Eisenack, Klaus, and Leonhard Kähler (2016). "Adaptation to Climate Change Can Support Unilateral Emissions Reductions." *Oxford Economic Papers* 68, 258-278.

[18] El-Sayed, Abeer, and Santiago J. Rubio (2014). "Sharing R&D Investments in Cleaner Technologies to Mitigate Climate Change." *Resource and Energy Economics* 38, 168-180.

[19] Finus, Michael, and Alejandro Caparrós (2015). *Game Theory and International Environmental Cooperation: Essential Readings,* Edward Elgar.

[20] Goeschl, Timo, and Grisch Perino (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.

[21] Harstad, Bard (2012). "Climate Contracts: A Game of Emissions, Investments, Negotiations, and Renegotiations." *Review of Economic Studies* 79, 1527-1557.

[22] Harstad, Bard (2016). "The Dynamic of Climate Agreements." *Journal of the European Economic Association* 14, 719-752.

[23] Harstad, Bard, Francesco Lancia, and Alessia Russo (2019). "Compliance Technology and Self-Enforcing Agreements." *Journal of The European Economic Association* 17, 1-29.

[24] Helm, Carsten, and Robert C. Schmidt (2015). "Climate Cooperation with Technology Investments and Border Carbon Adjustment." *European Economic Review* 75, 112-130.

[25] Hoel, Michael, and Aart de Zeeuw (2010). "Can a Focus on Breakthrough Technologies Improve the Performance of International Environmental Agreements?" *Environmental and Resource Economics* 47, 395-406.

[26] Hong, Fuhai, and Larry Karp (2012). "International Environmental Agreements with Mixed Strategies and Investment." *Journal of Public Economics* 96, 685-697.

[27] Lazkano, Itziar, Walid Marrouch, and Bruno Nkuiya (2016). "Adaptation to Climate Change: How Does Heterogeneity in Adaptation Costs Affect Climate Coalitions?" *Environment and Development Economics* 21, 812-838.

[28] Li, Hongxiu, and Horatiu A. Rus (2018). "Climate Change Adaptation and International Mitigation Agreements with Heterogeneous Countries." *Journal of the Association of Environmental and Resource Economists* 6, 503-530.

[29] Marrouch, Walid, and Amrita Ray Chaudhuri. (2011). "International Environmental Agreements in the Presence of Adaptation." FEEM Working Paper, 35.2011.

[30] Marrouch, Walid, and Amrita Ray Chaudhuri (2016). "International Environmental Agreements: Doomed to Fail or Destined to Succeed?" *International Review of Environmental and Resource Economics* 9, 245-319.

[31] Masoudi, Nahid, and Georges Zaccour (2017). "Adapting to Climate Change: Is Cooperation Good for the Environment?" *Economics Letters* 153, 1-5.

[32] Masoudi, Nahid, and Georges Zaccour (2018). "Adaptation and International Environmental Agreements." *Environmental and Resource Economics* 71, 1-21.

[33] Montero, Juan-Pablo (2002). "Permits, Standards, and Technology Innovation." *Journal of Environmental Economics and Management* 44, 23-44.

[34] Rubio, Santiago J. (2017). "Sharing R&D Investments in Breakthrough Technologies to Control Climate Change." *Oxford Economic Papers* 69, 496-521.

[35] Zehaie, Ficre (2009). "The Timing and Strategic Role of Self-Protection." *Environmental and Resource Economics* 44, 337-350.

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

Given that we have previously shown that $W^f > W^s$, lets 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

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

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

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

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

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

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

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

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

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 3

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, we can establish the following relationships

$$\frac{N^2+3N}{N-1} \left\{ \begin{array}{c} > \\ = \\ < \end{array} \right\} \frac{N^2 - \beta N}{1-\beta} \text{ for } \beta \left\{ \begin{array}{c} < \\ = \\ > \end{array} \right\} \beta_3 = \frac{3}{2+N}.$$

Therefore an agreement consisting of three countries is *internally stable* if $\gamma c \geq \max[(N^2 + 3N)/(N-1), (N^2 - \beta N)/(1-\beta)]$. Notice that if we impose an arbitrary upper bound on γc, as $\lim_{\beta \to 1}(N^2 - \beta N)/(1-\beta) = +\infty$ the agreement consisting of three countries could be unstable. But, this will occur for very low values of the degree of effectiveness of adaptation that we do not take into account in this analysis. 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.[25] 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 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

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

right of N.[26] 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 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

[26] 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.

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

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

Let's suppose that

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

which implies that

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

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

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

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

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

which implies that

$$288\left(-28N + 3N^2 + 222\right) - (192N + 144)^2$$

$$= -36\,000N^2 - 63\,360N + 43\,200 \geq 0,$$

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 4

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

$$F(\gamma c; n = 3) = -2(2\gamma c - 1)G(\gamma c), \tag{51}$$

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

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

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

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

$$+65(2N^2-23N+63)\gamma c-75(N-7)^2), \quad (52)$$

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.[27] Moreover, as the coefficients of the second

[27] 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 negatve, $F(n)=0$ would have only one positve root in the interval $(3,7)$ as we have just showed.

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

$$F(N) = (\gamma c)^2 (\gamma c + 1 - 2N) ((\gamma c)^2 (N-3) - 2\gamma c (N-1)(N-2) + 2(N-2)^2). \tag{53}$$

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?

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

$$(N^2 - 4N + 1)^2 - (N-2)^2 \left(N^2 - 4N + 7\right)$$

$$= -9N^2 + 36N - 27 \geq 0,$$

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

$$2N - 1 \in \left(\frac{(N-2)((N-1) - \sqrt{(N^2 - 4N + 7)})}{N-3}, \frac{(N-2)((N-1) + \sqrt{(N^2 - 4N + 7)})}{N-3} \right), \tag{54}$$

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 (54). The result os 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.[28] 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

[28]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.

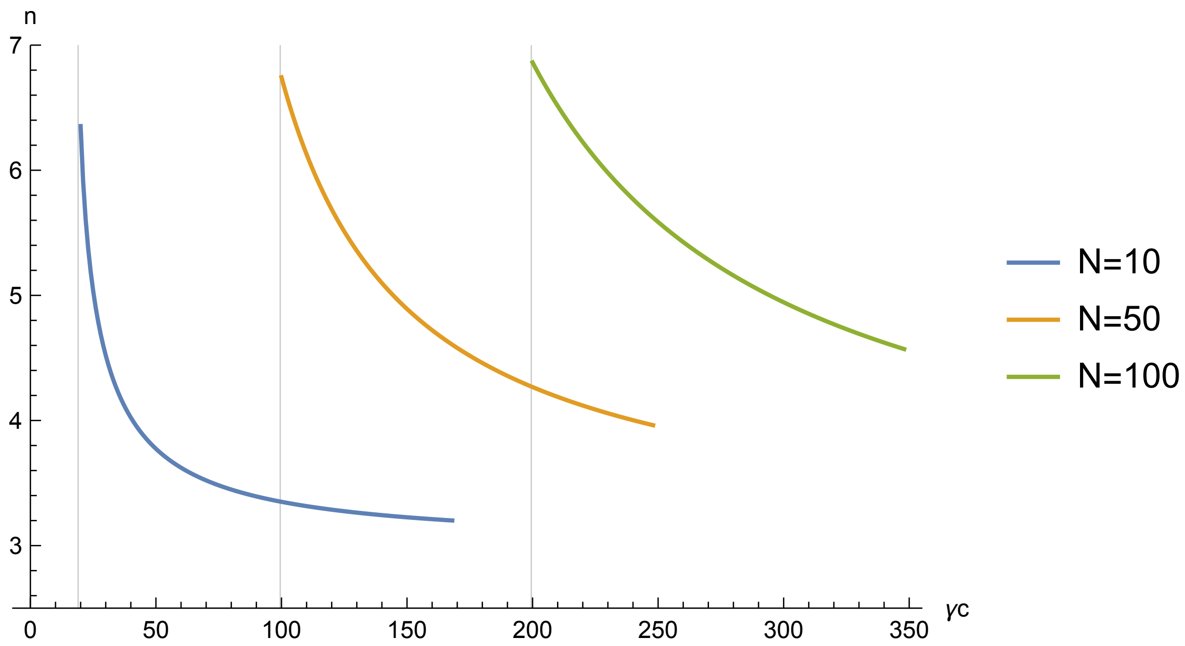
first derivative of $F(n)$ at $n = N$ we obtain the following polynomial in γc, $F'(N) = \gamma c H(\gamma c)$ where

$$\begin{aligned} H(\gamma c) &= (\gamma c)^4 - (6N-11)(\gamma c)^3 + (6N^2-14N+4)(\gamma c)^2 + (4N^3-20N^2+32N-16)\gamma c \\ &\quad -4N^3+18N^2-24N+8, \end{aligned}$$

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

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

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.


n
7
6
5
4
3
0
50
100
150
200
250
300
350
γc
N=10
N=50
N=100

FONDAZIONE ENI ENRICO MATTEI WORKING PAPER SERIES
"NOTE DI LAVORO"
Our Working Papers are available on the Internet at the following address:
http://www.feem.it/getpage.aspx?id=73&sez=Publications&padre=20&tab=1

"NOTE DI LAVORO" PUBLISHED IN 2021

1. 2021, Alberto Arcagni, Laura Cavalli, Marco Fattore, Partial order algorithms for the assessment of italian cities sustainability
2. 2021, Jean J. Gabszewicz, Marco A. Marini, Skerdilajda Zanaj, Random Encounters and Information Diffusion about Product Quality
3. 2021, Christian Gollier, The welfare cost of ignoring the beta
4. 2021, Richard S.J. Tol, The economic impact of weather and climate
5. 2021, Giacomo Falchetta, Nicolò Golinucci, Michel Noussan and Matteo Vincenzo Rocco, Environmental and energy implications of meat consumption pathways in sub-Saharan Africa
6. 2021, Carlo Andrea Bollino, Marzio Galeotti, On the water-energy-food nexus: Is there multivariate convergence?
7. 2021, Federica Cappelli, Gianni Guastella, Stefano Pareglio, Urban sprawl and air quality in European Cities: an empirical assessment
8. 2021, Paolo Maranzano, Joao Paulo Cerdeira Bento, Matteo Manera, The Role of Education and Income Inequality on Environmental Quality. A Panel Data Analysis of the EKC Hypothesis on OECD
9. 2021, Iwan Bos, Marco A. Marini, Riccardo D. Saulle, Myopic Oligopoly Pricing
10. 2021, Samir Cedic, Alwan Mahmoud, Matteo Manera, Gazi Salah Uddin, Information Diffusion and Spillover Dynamics in Renewable Energy Markets
11. 2021, Samir Cedic, Alwan Mahmoud, Matteo Manera, Gazi Salah Uddin, Uncertainty and Stock Returns in Energy Markets: A Quantile Regression Approach
12. 2021, Sergio Tavella, Michel Noussan, The potential role of hydrogen towards a low-carbon residential heating in Italy
13. 2021, Maryam Ahmadi, Matteo Manera, Oil Prices Shock and Economic Growth in Oil Exporting Countries
14. 2021, Antonin Pottier, Emmanuel Combet, Jean-Michel Cayla, Simona de Lauretis, Franck Nadaud, Who emits CO_2? Landscape of ecological inequalities in France from a critical perspective
15. 2021, Ville Korpela, Michele Lombardi, Riccardo D. Saulle, An Implementation Approach to Rotation Programs
16. 2021, Miguel Borrero, Santiago J. Rubio, An Adaptation-Mitigation Game: Does Adaptation Promote Participation in International Environmental Agreements?

Fondazione Eni Enrico Mattei
Corso Magenta 63, Milano – Italia

Tel. +39 02.520.36934
Fax. +39.02.520.36946

E-mail: letter@feem.it
www.feem.it

FEEM
FONDAZIONE ENI
ENRICO MATTEI