

Self-Enforcing International Environmental Agreements: Adaptation and Complementarity*

Santiago J. Rubio[†]

Department of Economic Analysis and ERI-CES

University of Valencia

August 6, 2018

*The author would like to thank participants at the 2018 AERNA Conference-Workshop (Madrid) for stimulating discussion. He also gratefully acknowledges financial support from the Spanish Ministry of Economics and Competitiveness under project ECO2016-77589-R.

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

Abstract

This paper studies the impact of adaptation on the stability of an international emission agreement. To address this issue we solve a three-stage coalition formation game where in the first stage countries decide whether or not to sign the agreement. Then, in the second stage, signatories (playing together) and non-signatories (playing individually) select their levels of emissions. Finally, in the third stage, each country decides on its level of adaptation noncooperatively. We solve this game for two models. For both, it is assumed that damages are linear with respect to emissions which guarantees that emissions are strategic *complements* in the second stage of the game. However, for the first model adaptation reduces the marginal damages of emissions in a *multiplicative* way whereas for the second model the reduction occurs in an *additive* way. Our analysis shows that the models yield different predictions in terms of participation. In the first case, we find that the larger the gains of full cooperation, the larger the cooperation. However, in the second case, the unique stable agreement we find consists of three countries regardless of the gains of full cooperation. These results suggest that complementarity can play in favor of cooperation but that it is not a sufficient condition to obtain more participation in an emission agreement. Finally, we would like to point out that our research indicates that the way adaptation reduces damages plays a critical role over the outcome of the coalition formation game.

Keywords: international environmental agreements, mitigation-adaptation game, strategic complements

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 goods characteristics while adaptation has private/national goods characteristics. Some examples of adaptation are liming, flood protection (like building dykes), housing planning (air-conditioning devices against heat), public water facilities, monitoring and forecasting of environmental problems and other improvements in infrastructures. This distinction 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 the international cooperation in mitigation. The literature indicates that with adaptation emissions can be strategic complements which can promote cooperation yielding a higher degree of participation than the one predicted by the models without adaptation where emissions are strategic substitutes.¹ In this paper, we advance in the analysis of this issue proposing two models that allows us to assess the impact of complementarity on cooperation. The models differ in the way adaptation reduces damages. In the first model, following Montero's (2002) approach on technology innovation, the investment in adaptation reduces damages in a multiplicative way. To simplify the writing, we name it the multiplicative model. The second model, that we call the additive or linear-quadratic model, is the one proposed by Lazkano et al. (2016) where the reduction on damages occurs in an additive way. Both models admit the possibility that emissions can be substitutes or complements depending on parameter values, but we find that the linearity of damages with respect to emissions is a sufficient condition for getting complementarity.

To evaluate the impact of complementarity on cooperation we solve for the two models a three-stage coalition formation game where countries can invest in adaptation once they

¹The idea that strategic complementarities play in favor of larger cooperation can be found for instance in Potters and Suetens (2009). The authors have conducted a laboratory experiment aimed at examining whether strategic substitutability and strategic complementarity have an impact on cooperation in finitely repeated two-player games with a Pareto-inefficient Nash equilibrium. They have obtained that there is significantly more cooperation when actions exhibit strategic complementarities than in the case of strategic substitutes. More recently, Bayramoglu et al. (2018) have showed that adaptation can lead to larger self-enforcing agreements if it causes countries' mitigation to be no longer strategic substitutes but complements.

have decided their level of emissions according to their decision about the participation in an international emission agreement. In both cases, we assume that damages are linear with respect to total emissions which guarantees that emissions are strategic complements in the second stage of the game. In this paper, we omit the stability analysis when emissions can be strategic substitutes because in this case the game with adaptation practically coincides with the mitigation game used in the seminal papers by Carraro and Siniscalco (1993) and Barrett (1994), and we should not expect any substantial change in terms of participation.

Interestingly, we find that the properties of the emission game played by the countries in the second stage are the same for the two models and coincide with the properties of the model without adaptation except in one point: the difference between the net benefit of non-signatories and the net benefit of signatories. This difference is positive in both cases, but for the additive model we find that it increases with participation at an increasing rate. However, for the multiplicative model this difference increases at a decreasing rate and presents a maximum for a level of participation lower than the grand coalition. This difference has important consequences in terms of participation because affects the incentives to enter/exit into/from the agreement. For the additive model, the incentives to enter into the agreement disappears quickly and the unique stable agreement we find consists of three countries. But for the multiplicative model, the incentives only disappear when the maximum difference in net benefits is reached and this can occur for high levels of participation. Moreover, a numerical example shows a positive relationship between the participation and the gains of full cooperation. Thus, we obtain not only that complementarity promotes cooperation but also that the larger the gains of full cooperation, the larger the participation. Cooperation appears when it is more profitable.

These results lead to an interesting corollary that is the main contribution of the paper. The corollary is that complementarity does not seem a *sufficient condition* for getting a larger participation in an international emission agreement. The multiplicative model supports the idea that complementarity plays in favor of cooperation but this is not the case for the linear-quadratic model. In other words, it seems that we need something more than complementarity to obtain a high level of participation in an IEA. Our analysis suggests that the way adaptation reduces damages could have a strong influence in the results we obtain about cooperation.

1.1 Literature Review

The analysis of the relationship between adaptation and complementarity developed in this paper follows the line of research initiated by Ebert and Welsch (2011, 2012) and continued by Farnham and Kennedy (2015), Heuson et al. (2015), Eisenack and Kähler (2016), Auerswald et al. (2018) and Vardar and Zaccour (2018) that focus on the characterization of the noncooperative Nash equilibrium of a global pollution game where adaptation reduces the environmental damages of emissions and presents national/private goods features.² Ebert and Welsch (2011, 2012) and Eisenack and Kähler (2016) study under what conditions the countries' emissions will be strategic complements for a model with two countries where damages are strictly convex with respect to adaptation and emissions. In this paper, we complete this analysis including the possibility of linear damages with respect to one of the arguments of the function, and extend it considering more than two countries and developing the stability analysis. Farnham and Kennedy (2015) also extend the analysis to consider multiple countries. The focus of the paper is on the comparison between the noncooperative Nash equilibrium when all countries decides simultaneously on adaptation and mitigation and the first-best solution. Heuson et al. (2015), in a framework of two countries, analyze the effects of different timings for adaptation, mitigation and technological investment that reduces the abatement costs. They demonstrate for symmetric countries that the strategic effects of advanced adaptation and investment support each other. Consequently, the total level of mitigation is lower for early and intermediate adaptation than in the case of late adaptation, which leads to a further deviation from the efficient solution. Also in a framework with two countries Auerswald et al. (2018) characterize the decision of risk averse governments on mitigation and adaptation policies and Vardar and Zaccour (2018) extend the analysis to a dynamic setting where damages depend on the accumulated stock.

The stability analysis we present in the next sections enrolls 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).³ Surprisingly, in spite of the huge number of paper published on this topic, only a few papers have addressed the effects of adaptation on the participation in an international emission agreement.

²We could add to this list the papers by Zehaie (2009), Buob and Stephan (2011) and Breton and Sbragia (2017a) clarifying that in these papers it is assumed that mitigation and adaptation are perfect substitutes.

³A nice collection of the most influential papers in the field has been published by Finus and Caparrós (2015).

This list of papers includes Barrett (2008), Lazkano et al. (2016), Benchekroun et al. (2017), Bayramoglu et al. (2017, 2018) and Breton and Sbragia (2017b).⁴ 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. Lazkano et al. (2016) 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 identical countries where countries' emissions are strategic substitutes 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. Bayramoglu et al. (2018) obtain using specific payoff functions that if countries' emissions are strategic complements because of the investment in adaptation, stable coalitions will be (weakly) larger with adaptation than without adaptation. They emphasize that this result does not depend on whether adaptation increases or decreases the marginal benefit of total abatement but on the magnitude of this effect. Their analysis shows that the grand coalition could be stable for some parameter values. However, if it is assumed that mitigation and adaptation are substitutes, that is the standard assumption in this literature, Bayramoglu et al. (2017) obtain for a numerical exercise the same result that that obtained by Barrett (2008) that is known as the paradox of cooperation: participation is large (low) when the potential gains from full cooperation are low (large). In this paper, we present a model that overcomes this paradox. Moreover, we find that complementarity seems to be a necessary condition to obtain a larger participation in an emission agreement but not a sufficient condition. In other words, we need something more than complementarity. As our analysis shows, the way adaptation reduces the damages plays a critical role in the result

⁴We could also mention the papers by Masoudi and Zaccour (2016, 2017) and Harstad et al. (2018). Masoudi and Zaccour (2016, 2017) are interested in a type of investment in adaptation that presents *imperfect international/public goods* characteristics. Assuming that countries choose investment before emissions and that signatories pool their investments to fully internalize the spillovers, they find using a numerical exercise that a higher adaptability leads to lower coalition size but higher welfare. Harstad et al. (2018) use a repeated extensive-form game to show that when the discount factor is so small that the first best cannot be achieved, countries are motivated to comply with an agreement and emit less only if they have previously invested less (more) in an (a) adaptation (green) technology.

we obtain about the participation with complementarity. Finally, Breton and Sbragia (2017b) uses the model proposed by Breton and Sbragia (2017a) to evaluate the impact of adaptation on the stability of IEAs. As in their model countries' mitigation are strategic substitutes when adaptation is chosen in the last stage of the game, they find a low level of participation with adaptation. Interestingly, a numerical exercise shows that the results are the opposite if adaptation is chosen before mitigation.

To end this review of the literature, we would like to mention that besides the investment in adaptation other papers have studied the impact of investment in green technologies that reduces the abatement costs on 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), and Rubio (2017). One of the issues examined by this literature is to know whether a technology agreement could be a good alternative to an emission agreement. Solving a three-stage technology agreement formation game, 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. If this is not the case, a technology agreement does not perform much better than an emission agreement. In this kind of games, countries select their level of investment before they take their decisions on emissions whereas in this paper we are assuming that adaptation occurs after countries have decided on emissions.⁵

The paper is organized in four sections. In the next section, Section 2, we solve the coalition formation game for the multiplicative model and in Section 3 for the additive model. Section 4 closes the paper with the conclusions and the presentation of different issues for future research.

⁵We 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. Contingent to this realization countries bargain repeatedly over abatement and associated transfer levels, potentially signing short-term agreements. The author obtains that an incomplete agreement implements the optimal level of abatement but at the cost of a lower investment.

2 The multiplicative model

We present a static model with N countries that pollute the environment and negotiate an IEA to control emissions. Each country i benefits from emissions e_i , $i = 1, \dots, N$. For this first model, we assume the benefit is represented by

$$B(e_i) = \frac{\alpha}{\beta} e_i^\beta, \quad \alpha > 0, \quad \beta \in (0, 1).$$

Let E denote the sum of all countries' emissions and assume that total emissions imply damage in the country considered but that this can be reduced by investment in adaptation, a_i . Following Montero's (2002) approach to model technology innovation, we assume that adaptation reduces damages in a *multiplicative* way, i.e. $D(a_i, E) = f(a_i)D(E)$ where $f(0) = 1$, $\lim_{a_i \rightarrow \infty} f(a_i) = 0$, $f' < 0$ and $f'' > 0$.⁶ For giving more structure to the model we assume that $f(a_i) = 1/(1+a_i)$ and $D(E) = dE^\eta/\eta$, where $d > 0$ and $\eta \geq 1$. This specification of the damage function has the following properties $D_a < 0$, $D_E > 0$, $D_{aa} > 0$ and $D_{aE} = D_{Ea} < 0$ provided that $E > 0$.⁷ Moreover,

$$D_{EE} = \frac{(\eta - 1)dE^{\eta-2}}{1 + a_i} \geq 0, \quad \eta \geq 1,$$

and the determinant of the Hessian matrix is

$$D_{aa}D_{EE} - D_{aE}D_{Ea} = \frac{\eta - 2}{\eta} \frac{d^2 E^{2(\eta-1)}}{(1 + a_i)^4}$$

so that if $\eta > 2$ then the damage function is strictly convex.

Finally, we assume linear investment costs $C(a_i) = ca_i$, with $c > 0$, and the net benefit is given by

$$W_i(e_i, E_{-i}, a_i) = \frac{\alpha}{\beta} e_i^\beta - \frac{d(e_i + E_{-i})^\eta}{\eta(1 + a_i)} - ca_i, \quad (1)$$

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

⁶The convexity of this function reflects that the resources invested in adaptation present decreasing returns.

⁷The sign of the cross derivative implies that adaptation increases with global emissions as was clarified by Ingham et al. (2013). However, as Ebert and Welsch (2011, 2012), Eisenack and Kähler (2016) and Bayramoglu et al. (2018) have showed the strategic relationship between countries' emissions does not on the sign of this cross derivative but on its magnitude.

2.1 The fully noncooperative equilibrium

Before addressing the stability analysis of the agreement, we will calculate the Nash equilibrium of the game with the aim of investigating whether adaptation changes the strategic relationship between emission in comparison with the game without adaptation where emissions are strategic substitutes. Following Ebert and Welsch (2012), we assume that countries adapt after emissions are carried out. In this case, the calculation of the fully noncooperative equilibrium has two-stage. In the first stage, we assume that countries decide unilaterally their emissions taking as given the emissions of the rest of countries. In the second stage, given the level of emissions decided in the first stage, countries decide also unilaterally their level of investment in adaptation taking as given the decision of all other countries. The equilibrium is computed by backward induction.⁸

As in the second stage countries decide on adaptation taking as given emissions, the optimization problem to be solved in this stage can be written as a total costs minimization problem

$$\min_{\{a_i\}} TC_i(E, a_i) = \frac{dE^\eta}{\eta(1 + a_i)} + ca_i,$$

where the total costs are defined as the sum of damages and adaptation costs. The FOC yields

$$\frac{dE^\eta}{\eta(1 + a_i)^2} = c,$$

that requires that the marginal reduction in damage because of the investment in adaptation must be equal to its marginal cost.⁹ This condition allows us to write adaptation as a function of total emissions

$$a_i = \left(\frac{d}{\eta c} \right)^{1/2} E^{\frac{\eta}{2}} - 1. \quad (2)$$

As expected the investment in adaptation increases with total emissions and is independent of the adaptation selected by the other countries.

Thus, the minimized total cost function is

⁸Given that the investment in adaptation does not create any international externality we could also calculate the equilibrium assuming that countries decide on emissions and adaptation simultaneously. Nevertheless, we define a two-stage game because in this way we can connect the analysis developed in this subsection with the stability analysis presented later on. A formal proof of this equivalence is given by Eisenack and Kähler (2016).

⁹The SOC is satisfied because the damage function is strictly convex with respect to adaptation.

$$TC_i(E) = 2 \left(\frac{cd}{\eta} \right)^{\frac{1}{2}} E^{\frac{\eta}{2}} - c, \quad (3)$$

The second derivative of this function is

$$\frac{\partial^2 TC_i}{\partial E^2} = (cd\eta)^{1/2} \frac{\eta - 2}{2} E^{\frac{\eta}{2} - 2}.$$

According to this expression, total costs are strictly convex if $\eta > 2$, linear if $\eta = 2$ and strictly concave if $\eta < 2$. Thus, $\eta > 2$ not only guarantees the strict convexity of damage functions but also the strict convexity of the total costs of emissions.

In the first stage, the net benefit is maximized with respect to the country's emissions taking as given the emissions of the rest of countries.

$$\max_{\{e_i\}} W_i(e_i, E_{-i}) = \frac{\alpha}{\beta} e_i^\beta - 2 \left(\frac{cd}{\eta} \right)^{\frac{1}{2}} (e_i + E_{-i})^{\frac{\eta}{2}} + c. \quad (4)$$

The optimal emissions must satisfy the following condition

$$\frac{\alpha}{e_i^{1-\beta}} = (cd\eta)^{1/2} (e_i + E_{-i})^{\frac{\eta}{2} - 1}, \quad (5)$$

that establishes that the marginal benefit of emissions must be equal to its marginal total costs. Moreover, it must be satisfied as well the second-order condition (SOC)

$$\frac{\partial^2 W_i}{\partial e_i^2} = \alpha(\beta - 1)e_i^{\beta-2} - (cd\eta)^{1/2} \frac{\eta - 2}{2} (e_i + E_{-i})^{\frac{\eta}{2} - 2} < 0. \quad (6)$$

Observe that is $\eta \geq 2$, the net benefit function is strictly concave and the SOC holds for the optimal value defined by the FOC. However, if $\eta \in [1, 2)$ the fulfillment of the SOC is not guaranteed and additional constraints on the parameters could be necessary for the existence of a solution for the maximization of net benefit.

Nest, we apply the implicit function theorem to (5) to obtain the slope of the best response function for emissions

$$\frac{\partial e_i}{\partial e_j} = \frac{(cd\eta)^{1/2} \frac{\eta-2}{2} E^{\frac{\eta}{2}-2}}{\alpha(\beta-1)e_i^{\beta-2} - (cd\eta)^{1/2} \frac{\eta-2}{2} E^{\frac{\eta}{2}-2}}, \quad i, j = 1, \dots, N, \quad i \neq j. \quad (7)$$

This expression says us that the strategic relationship between emissions depends on the curvature of the total cost function since the denominator is negative if the SOC for the maximization of net benefit holds, so that we can conclude that

Proposition 1 *If the total cost function of emissions is convex (concave) the emissions are strategic substitutes (complements). If the function is linear the emissions are independent.*

It is immediate from (7) that if emissions are strategic substitutes, the absolute value of the slope of the best response function is lower than one.¹⁰

Finally, we study the optimal response of adaptation to a variation of the other countries' emissions. Using (2) we obtain that

$$\frac{\partial a_i}{\partial e_j} = \left(\frac{d}{\eta c}\right)^{1/2} \frac{\eta}{2} E^{\frac{\eta}{2}-1} \left(\frac{\partial e_i}{\partial e_j} + 1\right),$$

that substituting $\partial e_i/\partial e_j$ by (7) yields

$$\frac{\partial a_i}{\partial e_j} = \left(\frac{d}{\eta c}\right)^{1/2} \frac{\eta}{2} E^{\frac{\eta}{2}-1} \left(\frac{\alpha(\beta-1)e_i^{\beta-2}}{\alpha(\beta-1)e_i^{\beta-2} - (cd\eta)^{1/2} \frac{\eta-2}{2} E^{\frac{\eta}{2}-2}}\right) > 0, \quad (8)$$

$i, j = 1, \dots, N$, $i \neq j$, which allows us to conclude that

Proposition 2 *The investment in adaptation of a country is a strategic complement of the other countries' emissions.*

Observe that in (8) the denominator is negative according to the SOC (6) and the numerator is also negative because $\beta < 1$. The result is that investment in adaptation is a strategic complement of other countries' emissions regardless of the properties of the total costs function, i.e. independently of whether emissions are strategic substitutes or complements.

With $\eta \geq 2$, the net benefit function (4) presents the same properties that the model of global pollution without adaptation and we should not expect that an IEA with a high degree of participation when the countries decide to invest in adaptation. In the next section, we assume that $\eta = 1$ and investigate whether the change of the strategic relationship between countries' emissions leads to different results.

2.2 Participation with complementarity

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

¹⁰Ebert and Welsch (2011, 2012) claim that the complementarity appears when the adaptation capacity is bigger than the sensitivity to global pollution. Our analysis shows that in this case, the total costs function is concave and consequently the marginal total costs of emissions are decreasing. Thus, a high adaptation capacity will explain the strategic complementarity of emissions provided that it yields decreasing marginal total costs.

Given the number of signatories and the emissions of all countries, in the third stage, the adaptation game, each country simultaneously selects its national investment in adaptation acting unilaterally and taking the investment of all other countries as given. In the second stage, the emission game, the signatories choose their emissions to maximize the agreement's net benefit taking as given the emissions of non-signatories. These choose their emissions acting noncooperatively and taking the emissions of all other countries as given to maximize their national net benefit. Both signatories and non-signatories select their emissions simultaneously. Thus, emissions are given by the partial agreement Nash equilibrium (PANE) with respect to a coalition defined by Chander and Tulkens (1995). In the first stage, it is assumed that 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. The agreement is formed by all countries that have chosen to sign. Under open membership, any country is free to join the agreement if it is interested. Finally, we assume that the signing of the agreement plays as a commitment device on signatories. Thus, the signing of the agreement implies that the signatories will honor their obligations and no deviation will occur after countries sign the agreement, i.e. they will coordinate the levels of emissions to maximize the net benefit of the agreement. The game finishes when the adaptation game is over.

Third stage

As the investment in adaptation is a national good, the third stage of the agreement formation game coincides with the second stage of total costs minimization analyzed above and according to (2) we will have that

$$a_i^s = a_j^f = \left(\frac{dE}{c} \right)^{1/2} - 1, \quad (9)$$

where

$$E = \sum_{i=1}^n e_i^s + \sum_{j=1}^{N-n} e_j^f, \quad i = 1, \dots, n; \quad j = 1, \dots, N - n,$$

and n stands for the number of signatories, s for a signatory country and f for a non-signatory. Both types of countries will select the same level of adaptation because we have assumed that countries are *a priori* symmetric. For these levels of adaptation, the minimized total costs function is

$$TC_i(E) = 2 (cd)^{\frac{1}{2}} E^{\frac{1}{2}} - c.$$

Using this costs function, the net benefit can be written as a function of emissions

$$W_i(e_i, E_{-i}) = \frac{\alpha}{\beta} e_i^\beta - 2(cd)^{\frac{1}{2}} (e_i + E_{-i})^{\frac{1}{2}} + c.$$

Second stage: the PANE of the emission game

We solve stage two assuming that in the first stage n countries, with $n \geq 2$, have signed the agreement. In this stage, non-signatories select their emissions acting noncooperatively and taking the emissions of the rest of countries, signatories and non-signatories, as given in order to maximize their national net benefit. Under this assumption, the maximization problem for the non-signatories coincides with the maximization problem of the first stage studied in the previous subsection and the optimal emissions must satisfy the following condition

$$\frac{\alpha}{(e_j^f)^{1-\beta}} = \left(\frac{cd}{E}\right)^{\frac{1}{2}}.$$

Notice that the marginal total costs are decreasing with respect to total emissions. Using this condition we can write the optimal non-signatories's emissions as a function of total emissions

$$e_j^f = \alpha^{\frac{1}{1-\beta}} \left(\frac{E}{cd}\right)^{\frac{1}{2(1-\beta)}}. \quad (10)$$

Moreover, the SOC (6) must be rewritten as follows

$$\frac{\partial^2 W_j^f}{\partial (e_j^f)^2} = \alpha(\beta - 1)(e_j^f)^{\beta-2} + \frac{1}{2} \frac{(cd)^{\frac{1}{2}}}{E^{\frac{3}{2}}} < 0. \quad (11)$$

On the other hand, signatory countries coordinate their emissions to maximize the net benefits of the agreement taking as given the emissions of non-signatories

$$\max_{\{e_1^s, \dots, e_n^s\}} W_A = \sum_{i=1}^n W_i^s = \sum_{i=1}^n \left\{ \frac{\alpha}{\beta} (e_i^s)^\beta - 2(cdE)^{\frac{1}{2}} + c \right\}.$$

Focusing on a symmetric solution, the FOC yields

$$\frac{\alpha}{(e^s)^{1-\beta}} = n \left(\frac{cd}{E}\right)^{\frac{1}{2}},$$

that establishes that the marginal benefit of national emissions must be equal to the marginal total costs for the agreement. This condition allows us to write the signatories' emissions as a function of total emissions

$$e^s = \left(\frac{\alpha}{n}\right)^{\frac{1}{1-\beta}} \left(\frac{E}{cd}\right)^{\frac{1}{2(1-\beta)}}, \quad (12)$$

where $E = ne^s + (N - n)e^f$. The SOC for the maximization of the agreement net benefits is¹¹

$$\frac{\partial^2 W_A}{\partial (e^s)^2} = n \left\{ \alpha(\beta - 1)(e^s)^{\beta-2} + \frac{n^2 (cd)^{\frac{1}{2}}}{2 E^{\frac{3}{2}}} \right\} < 0. \quad (13)$$

The solution of this stage is easy to calculate substituting (10) and (12) in the expression that yields total emissions

$$E = ne^s + (N - n)e^f = \left(\frac{\alpha}{(cd)^{\frac{1}{2}}} \right)^{\frac{2}{1-2\beta}} \left(\frac{1}{n^{\frac{\beta}{1-\beta}}} + N - n \right)^{\frac{2(1-\beta)}{1-2\beta}}. \quad (14)$$

Once we have the total emissions by substitution in (10), (12) and (9) we obtain the national emissions for both types of countries and the investment in adaptation.

Before calculating the payoffs for the countries, we check whether the SOC and the non-negativity constraint are satisfied by this solution. Substituting e^f by (10) in (11) the SOC reads

$$\frac{\partial^2 W^f}{\partial (e^f)^2} = \frac{\beta - 1}{\alpha^{\frac{1}{1-\beta}}} \frac{(cd)^{\frac{2-\beta}{2(1-\beta)}}}{E^{\frac{2-\beta}{2(1-\beta)}}} + \frac{(cd)^{\frac{1}{2}}}{2E^{\frac{3}{2}}} < 0.$$

Next, eliminating E using (14) and taking common factor $(cd)^{\frac{1}{2}}$ we obtain the following condition

$$(cd)^{\frac{1}{2}} \left(2 \left(\frac{1}{n^{\frac{\beta}{1-\beta}}} + N - n \right) (\beta - 1) + 1 \right) < 0,$$

that yields a constraint over β that can be written as follows

$$1 - \frac{1}{2} \frac{1}{\frac{1}{n^{\frac{\beta}{1-\beta}}} + N - n} > \beta. \quad (15)$$

As the left-hand side (LHS) of the inequality is a decreasing function with respect to n we will have that the lowest upper bound of β occurs for $n = N - 1$

$$1 - \frac{1}{2} \frac{1}{\frac{1}{(N-1)^{\frac{\beta}{1-\beta}}} + 1} > \beta.$$

Therefore, if this condition holds the SOC will be satisfied for any n in the interval $[1, N - 1]$. From this expression is straightforward to conclude that $\beta \leq 1/2$ is a sufficient condition for the fulfillment of the SOC for any level of membership.

Proceeding in the same way for signatories, the following condition over β is obtained

¹¹If the SOC is satisfied it is easy to check that emissions are strategic complements in this stage of the game. We will omit the details to avoid reiterations and shorten the text.

$$1 - \frac{1}{2} \frac{1}{1 + (N - n)n^{\frac{\beta}{1-\beta}}} > \beta. \quad (16)$$

The first derivative of the LHS with respect to n gives

$$\frac{n^{\frac{\beta}{1-\beta}}}{2} \frac{N\beta - n}{n(1 - \beta)(1 + (N - n)n^{\frac{\beta}{1-\beta}})^2}.$$

Thus, the LHS is increasing for $n < N\beta$ and decreasing for $n > N\beta$ and consequently the minimum value for the LHS will correspond with $n = 2$ or $n = N$. Comparing the LHS for these two values we obtain that

$$1 - \frac{1}{2} \frac{1}{1 + (N - 2)2^{\frac{\beta}{1-\beta}}} > \frac{1}{2},$$

where the LHS of the inequality corresponds to the value we obtain for $n = 2$ and $1/2$ to the value we obtain for $n = N$. Therefore, the minimum value that the LHS of (16) will take in the interval $[2, N]$ will be $1/2$. Then, if $\beta < 1/2$, the SOC will be satisfied for all $n \in [2, N]$. Moreover, this condition guarantees that SOC also holds for non-signatories. For this reason, we establish the following assumption

Assumption 2.1: $\beta < 1/2$.

Next, we check the non-negativity constraint on emissions and adaptation. It is straightforward that the emissions are positive since total emissions given by (14) are positive for any level of participation. However, the non-negativity of the investment in adaptation is not guaranteed. Substituting (14) in (9) we obtain

$$a = \left(\frac{\alpha}{c^{1-\beta} d^\beta} \right)^{\frac{1}{1-2\beta}} \left(\frac{1}{n^{\frac{\beta}{1-\beta}}} + N - n \right)^{\frac{1-\beta}{1-2\beta}} - 1. \quad (17)$$

From this expression we can conclude that the non-negativity constraint would be satisfied for enough low values for d , in particular if

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

The minimum value for the LHS of this inequality occurs for $n = N$ that defines a sufficient condition for the non-negativity of adaptation that we assume is satisfied by the model.

Assumption 2.2: $\left(\frac{\alpha}{c^{1-\beta}} \right)^{\frac{1}{\beta}} \frac{1}{N} \geq d$.

Thus, assumptions 2.1 and 2.2 guarantee that both emissions and adaptation are non-negative and that the SOCs are satisfied for both types of countries.

Finally, eliminating national emissions from the net benefit function we obtain the following expression

$$W^f = \frac{\alpha^{\frac{1}{1-\beta}}}{\beta} \left(\frac{E}{cd} \right)^{\frac{\beta}{2(1-\beta)}} - 2(cdE)^{\frac{1}{2}} + c, \quad (18)$$

for the non-signatories' net benefit and

$$W^s = \frac{\alpha^{\frac{1}{1-\beta}}}{\beta} \frac{1}{n^{\frac{\beta}{1-\beta}}} \left(\frac{E}{cd} \right)^{\frac{\beta}{2(1-\beta)}} - 2(cdE)^{\frac{1}{2}} + c, \quad (19)$$

for the signatories's net benefit.

Comparing (10) and (12) and the above expressions we can conclude that

Proposition 3 *The signatories' emissions are lower than the non-signatories' emissions for any level of participation whereas both types of countries invest the same amount of resources in adaptation. Thus, the signatories' net benefit will be lower than the non-signatories's net benefit.*

Next, we study the effects of a change in membership on the different variables of the model. According to (14) it is immediate that total emissions will decrease with n and the same will occur for national emissions according to expressions (10) and (12) and for adaptation according to expression (9). For evaluating the impact of a change in n on non-signatories' net benefit we take the first derivative of (18) that gives

$$\begin{aligned} \frac{\partial W^f}{\partial n} &= \frac{\alpha^{\frac{1}{1-\beta}}}{2(1-\beta)} \left(\frac{E}{cd} \right)^{\frac{\beta}{2(1-\beta)}-1} \frac{1}{cd} \frac{\partial E}{\partial n} - \left(\frac{cd}{E} \right)^{\frac{1}{2}} \frac{\partial E}{\partial n} \\ &= \frac{1}{(cd)^{\frac{\beta}{2(1-\beta)}} E^{\frac{1}{2}}} \left(\frac{\alpha^{\frac{1}{1-\beta}}}{2(1-\beta) E^{\frac{1-2\beta}{2(1-\beta)}}} - (cd)^{\frac{1}{2(1-\beta)}} \right) \frac{\partial E}{\partial n}. \end{aligned}$$

Substituting E by (14) in the parenthesis we obtain

$$\frac{\alpha^{\frac{1}{1-\beta}}}{2(1-\beta) E^{\frac{1-2\beta}{2(1-\beta)}}} - (cd)^{\frac{1}{2(1-\beta)}} = (cd)^{\frac{1}{2(1-\beta)}} \left(\frac{1}{2(1-\beta)} \frac{1}{n^{\frac{\beta}{1-\beta}}} + N - n - 1 \right).$$

This expression is negative if the SOC for the maximization of net benefit holds. Let's suppose that the parenthesis is positive or zero

$$\frac{1}{2(1-\beta)} \frac{1}{\frac{1}{n^{\frac{\beta}{1-\beta}}} + N - n} - 1 \geq 0,$$

in that case it must be satisfied that

$$\beta \geq 1 - \frac{1}{2} \frac{1}{\frac{1}{n^{\frac{\beta}{1-\beta}}} + N - n},$$

what contradicts (15) that requires that β be lower than this expression. Therefore, it can be concluded that $\partial W^f / \partial n > 0$ since $\partial E / \partial n < 0$ and the model presents *positive spillovers* coming from cooperation. We can also show although it is not so immediate that the signatories' net benefit also increases with the number of signatories¹². Summarizing,

Proposition 4 *Both the signatories' emissions and the non-signatories' emissions decrease with the participation. The total emissions and the investment in adaptation also decrease with the number of signatories whereas that the net benefit augments for both types of countries.*

Next, we evaluate the effect of cooperation on the difference between the net benefit of signatories and non-signatories using the following expression we get from (18) and (19)

$$W^s - W^f = \frac{\alpha^{\frac{1}{1-\beta}}}{\beta} \left(\frac{1}{n^{\frac{\beta}{1-\beta}}} - 1 \right) \left(\frac{E}{cd} \right)^{\frac{\beta}{2(1-\beta)}}. \quad (20)$$

The first derivative of this expression with respect to n says us that

Proposition 5 *There exists a threshold value for N such that is N is larger than this threshold value, the absolute value of the difference $W^s - W^f$ has a maximum in the interval $[2, N - 1]$.*

Proof. See Appendix A.2. ■

This is a key result of the paper. It establishes that for large enough levels of participation, the gap between the net benefit of signatories and non-signatories decreases as the agreement expands what, as we will see in the next subsection, can stabilize an agreement with a high degree of participation. This is a property of the game we will not find in the additive model.

Finally, we evaluate the effect of cooperation on the aggregate net benefits using again expressions (18) and (19)

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

¹²This proof is relegated to Appendix A.1.

$$= \frac{\alpha^{\frac{1}{1-\beta}}}{\beta} \left(n^{\frac{1-2\beta}{1-\beta}} + N - n \right) \left(\frac{E}{cd} \right)^{\frac{\beta}{2(1-\beta)}} - 2N(cdE)^{\frac{1}{2}} + Nc. \quad (21)$$

The analysis of this expression allows us to conclude that

Proposition 6 *The aggregate net benefit increases with the participation. The game presents the property of full cohesiveness.*

Proof. See Appendix A.3. ■

This result is indicating that the aggregate net benefit increases when the agreement is enlarged gradually and obtain its maximum in the grand coalition. Full cohesiveness justifies the search for large stable coalitions.

The first stage: the stability analysis

Next, we investigate whether there exists a stable agreement. First, we present the concept of coalitional stability used in this paper that was proposed by d'Aspremont et al. (1983) and has been extensively applied in the literature on IEAs.

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

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 a non-signatory as 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 auxiliary 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. Substituting total emissions by (14) in (18) and (19) and taking common factor we obtain the following expression for the auxiliary function

$$S(n) = \left(\frac{\alpha}{(cd)^\beta} \right)^{\frac{1}{1-2\beta}} \left(\left(\frac{1}{n^{\frac{\beta}{1-\beta}}} + N - n \right)^{\frac{\beta}{1-2\beta}} \left(\frac{1}{\beta n^{\frac{\beta}{1-\beta}}} - 2 \left(\frac{1}{n^{\frac{\beta}{1-\beta}}} + N - n \right) \right) - \left(\frac{1}{(n-1)^{\frac{\beta}{1-\beta}}} + N - n + 1 \right)^{\frac{\beta}{1-2\beta}} \left(\frac{1}{\beta} - 2 \left(\frac{1}{(n-1)^{\frac{\beta}{1-\beta}}} + N - n + 1 \right) \right) \right). \quad (22)$$

Instead of looking for a solution for $S(n) = 0$, we will begin the stability analysis investigating whether the grand coalition can be stable. If this is the case, $S(N)$ should be positive or zero. Doing $n = N$ in the previous expression we get

$$S(N) = \left(\frac{\alpha}{(cd)^\beta} \right)^{\frac{1}{1-2\beta}} \left(\left(\frac{1}{N^{\frac{\beta}{1-\beta}}} \right)^{\frac{\beta}{1-2\beta}} \left(\frac{1}{\beta N^{\frac{\beta}{1-\beta}}} - 2 \left(\frac{1}{N^{\frac{\beta}{1-\beta}}} \right) \right) - \left(\frac{1}{(N-1)^{\frac{\beta}{1-\beta}}} + 1 \right)^{\frac{\beta}{1-2\beta}} \left(\frac{1}{\beta} - 2 \left(\frac{1}{(N-1)^{\frac{\beta}{1-\beta}}} + 1 \right) \right) \right),$$

where the second parenthesis can be written as follows

$$= \frac{1}{N^{\frac{\beta}{1-2\beta}}} \frac{1-2\beta}{\beta} - \left(\frac{1}{(N-1)^{\frac{\beta}{1-\beta}}} + 1 \right)^{\frac{\beta}{1-2\beta}} \left(\frac{1}{\beta} - 2 \left(\frac{1}{(N-1)^{\frac{\beta}{1-\beta}}} + 1 \right) \right), \quad (23)$$

so that

$$\frac{1}{\beta} - 2 \left(\frac{1}{(N-1)^{\frac{\beta}{1-\beta}}} + 1 \right) \leq 0$$

is a sufficient condition to give a stable agreement consisting of all countries. When this condition is satisfied as an equality we can write it as follows

$$N = 1 + \left(\frac{2\beta}{1-2\beta} \right)^{\frac{1-\beta}{\beta}}, \quad (24)$$

that gives a value for β equal to 0.25 for $N = 2$ and whose graphical representation in the interval $[0.25, 0.5)$ is

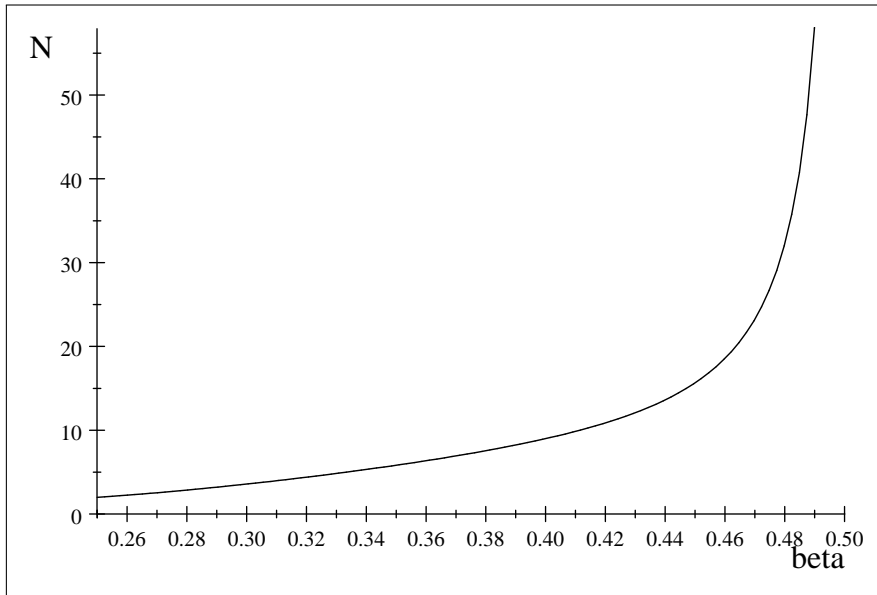


FIGURE 1. Threshold values for β .

Then for all $N > 2$ there will exist a value and only a value for β , $\hat{\beta}$, implicitly defined by (24) such that if β is equal to or larger than $\hat{\beta}$, (23) is positive what implies that $S(N) > 0$ and it can be concluded that

Proposition 7 *For all $N > 2$ there exists a threshold value $\hat{\beta}$ that increases with N and is defined in an implicit way by (24) in the interval $(0.25, 0.5)$ such that if $\beta \in [\hat{\beta}, 0.5)$ the grand coalition is stable.*

Observe that this is a sufficient condition, i.e., the grand coalition could be stable for values of β lower than $\hat{\beta}$. This result suggests that cooperation increases with β but unfortunately it is not possible to determine analytically this relationship. For this reason, in the next subsection we solve a numerical example with the aim of investigating whether this relationship occurs and which is the link between the potential gains of cooperation and the participation in the agreement.

2.3 Numerical example

As the sign of $S(n)$ only depends on β and N we will consider only variations in the value of these parameters. In this numerical exercise we give the following values to these two parameters

$$\beta = \{0.05, 0.10, 0.15, 0.20, 0.25\}, \quad N = \{10, 30, 50, 70, 90\}$$

for $\alpha = 35$, $c = 20$ and $d = 2$. These values have been fixed taking into account Assumption 2.2 so that for all possible combinations the investment in adaptation satisfy the non-negativity constraint. The values of β are in the interval $[0.05, 0.25]$ for different reasons. The first one is that we already know according to the previous proposition that the grand coalition can be stable for values larger than 0.25. A second reason is that as β tends to 0.50 the model gives huge variations between the fully noncooperative equilibrium and the fully cooperative solution (grand coalition). Finally, we have considered that below 0.05, the benefit of emissions would have a very small weight in the payoffs.

The first table shows the number of signatories for each combination (β, N) and between parenthesis the participation in relative terms: number of signatories over the total of countries. The numerical simulation indicates that the membership increases with β but that in relative terms decreases with the number of countries. In fact, for $N \geq 50$, we need a value for β equal to 0.25 for having a participation above 50%.

$\beta \backslash N$	10	30	50	70	90
0.05	6 (60.0)	8 (26.7)	8 (16.0)	8 (11.4)	8 (8.9)
0.10	7 (70.0)	9 (30.0)	9 (18.0)	9 (12.9)	9 (10.0)
0.15	8 (80.0)	12 (40.0)	12 (24.0)	13 (18.6)	13 (14.4)
0.20	8 (80.0)	18 (60.0)	20 (40.0)	20 (28.6)	21 (23.3)
0.25	9 (90.0)	26 (86.7)	42 (84.0)	52 (74.3)	57 (63.3)

TABLA 1. $\alpha=35, c=20, d=2$.

The first figure in the second table stands for the national gains coming from full cooperation, i.e., the difference between the net benefit the countries would have if the grand coalition ($n = N$) is formed and the net benefit corresponding to the fully noncooperative equilibrium ($n = 1$). Between parenthesis we give the closing the gap index as was defined by Eyckmans and Finus (2006). This index shows the gains of the stable cooperation as a percentage of the gains of full cooperation. The numerical analysis indicates clearly that the gains of full cooperation increase with β . Thus, if we take into account that in Table 1 we find a positive relationship between the number of signatories and β , we can conclude that the larger the gains of full cooperation, the larger the incentives the countries have to sign an international agreement to control emissions. The figures in Table 2 indicate that in the 88% of the cases the IEA allows to achieve a figure higher than 10% of the gains that the grand coalition would achieve and that in the 72% of the cases this figure is equal to or higher than 20%.

$\beta \backslash N$	10	30	50	70	90
0.05	671 (57.8)	2676 (20.3)	4821 (13.9)	7033 (9.7)	9289 (7.5)
0.10	1038 (71.7)	4387 (29.3)	8142 (17.1)	12121 (12.1)	16216 (9.3)
0.15	1799 (84.6)	8218 (43.3)	15873 (25.8)	24282 (20.0)	33255 (15.5)
0.20	3691 (86.8)	18820 (69.0)	38413 (47.6)	61004 (34.5)	85951 (28.4)
0.25	9922 (96.3)	59480 (94.9)	131540 (93.4)	220440 (86.7)	323440 (77.4)

TABLA 2. $\alpha=35, c=20, d=2$.

Next, we represent graphically the difference in net benefits as a function of the number of signatories for $\beta = 0.15$ and $N = 10$. Figure 2 illustrates Prop. 5. We can see in this figure that the difference in net benefits increases in a concave way with the participation until a maximum

is reached between 8 and 9 to decrease afterwards.

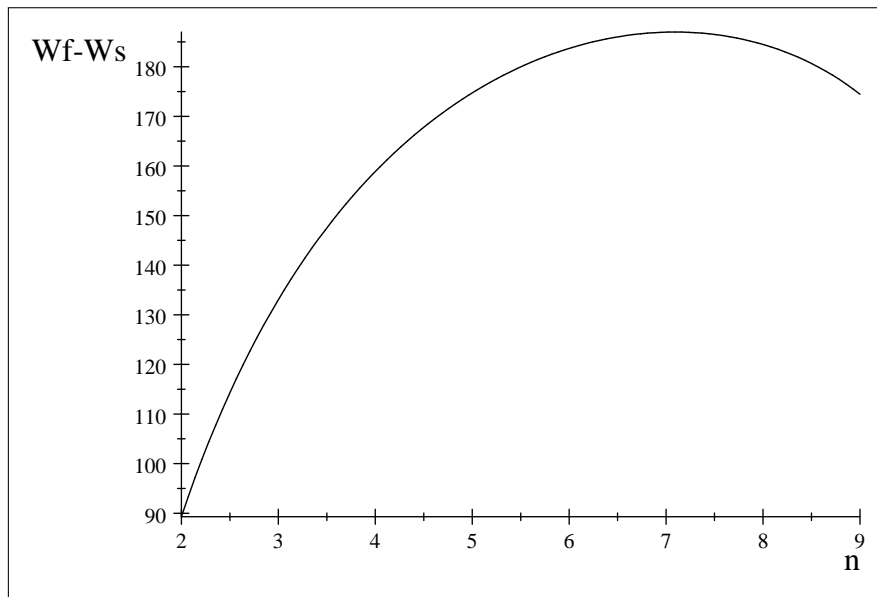


FIGURE 2. Difference in the net benefit between signatories and non-signatories.

Finally, we plot the auxiliary function $S(n)$ for the same parameter values. As $S(8)$ is positive and $S(9)$ is negative, the agreement consisting of 8 countries is stable. In fact, it is the unique stable agreement the game yields. The figure says us that the countries find profitable to enter into the agreement until an agreement of 8 countries is formed. The difference in net benefits increases but at a decreasing rate that makes positive the difference $W^s(n) - W^f(n-1)$ in the interval $[2,8]$. Notice that the agreement stabilizes just for the first integer at the right of the maximum in Fig. 2.

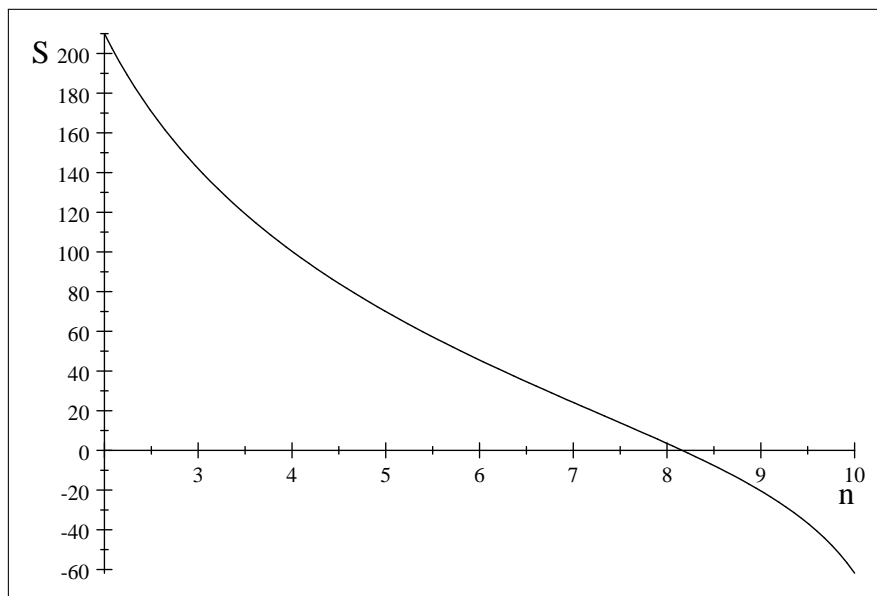


FIGURE 3. The auxiliary function $S(n)$ for $\beta=0.15$ and $N=10$.

Next, we will present an alternative model that has exactly the same features we find for the multiplicative model except that the difference in net benefits is not a concave function but an increasing convex function with respect to the participation, and we will check that although the emissions are also strategic complements the number of signatories is not larger than three.

3 The additive model

In this section we use the symmetric version of the model proposed by Lazkano et al. (2016) to evaluate the effects of adaptation on participation in an IEA. For this model, the economic benefit of emissions is given by

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

and the environmental damage by

$$D(a_i, E) = \frac{1}{\eta} (d - a_i) E^\eta, \quad \alpha > d > a_i > 0, \quad \eta \geq 1.$$

For this specification of the damage function $D_a < 0$, $D_E > 0$, $D_{aa} = 0$ and $D_{aE} = D_{Ea} < 0$ for $E > 0$ whereas

$$D_{EE} = (\eta - 1)(d - a_i)E^{\eta-2} \geq 0, \quad \eta \geq 1.$$

Now, a_i stands for the reduction in marginal damage in particular for the reduction in d that can be achieved investing resource in adaptation. In comparison with the previous model, we assume now that the reduction in marginal damages occurs in an *additive* way. We also assume that the environmental damage cannot be completely eliminated through adaptation. The cost of reducing the marginal damage is increasing and is given by $C(a_i) = ca_i^2/2$, $c > 0$.¹³ Thus, the net benefit is

$$W_i(a_i, e_i, E_{-i}) = \alpha e_i - \beta \frac{e_i^2}{2} - \frac{1}{\eta} (d - a_i) (e_i + E_{-i})^\eta - \frac{c}{2} a_i^2. \quad (25)$$

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

3.1 The fully noncooperative equilibrium

As we did in the previous section, we calculate the fully noncooperative equilibrium in two stages. For this model, total costs are given by the following expression

$$TC_i(a_i, E) = \frac{1}{\eta}(d - a_i)E^\eta + \frac{c}{2}a_i^2.$$

The FOC can be written as follows: $E^\eta/\eta = ca_i$, where the LHS of the condition is the marginal reduction in damage because of adaptation (marginal benefit of adaptation) and the right-hand side (RHS) represents the marginal cost. Using this condition we can write adaptation as a function of total emissions

$$a_i = \frac{E^\eta}{c\eta}. \quad (26)$$

As occurs for the multiplicative model, adaptation increases with total emissions and is independent of the level of adaptation chosen by the other countries. The SOC for the minimization of total costs is satisfied because the adaptation cost is strictly convex.

Substituting (26) in the total cost function gives

$$TC_i(E) = \frac{E^\eta}{\eta} \left(d - \frac{E^\eta}{2\eta c} \right). \quad (27)$$

Then, the marginal total cost of emissions will be

$$MTC_i(E) = E^{\eta-1} \left(d - \frac{E^\eta}{\eta c} \right) \quad (28)$$

and the total cost function is increasing until a maximum is reached for

$$E^* = (\eta cd)^{\frac{1}{\eta}}. \quad (29)$$

To complete the characterization of the total cost function we calculate its second derivative

$$\frac{dMTC_i}{dE} = \frac{d^2TC_i}{dE^2} = E^{\eta-2} \left((\eta - 1)d - (2\eta - 1)\frac{E^\eta}{\eta c} \right). \quad (30)$$

This derivative is zero if $\eta > 1$ for

$$\bar{E} = \left(\frac{(\eta - 1)\eta cd}{2\eta - 1} \right)^{\frac{1}{\eta}} \quad (31)$$

that is lower than E^* , and negative if $\eta = 1$ for all $E \geq 0$. Thus, if $\eta > 1$ the total costs are strictly convex for $E \in (0, \bar{E})$ and strictly concave for $E \in (\bar{E}, E^*]$ and if $\eta = 1$ the function is strictly concave for all $E \in [0, E^*]$. In this case, the marginal total costs are decreasing.

In the first stage, the net benefit is maximized with respect to the country's emissions taking as given the emissions of all other countries

$$\max_{\{e_i\}} W_i(e_i, E_{-i}) = \alpha e_i - \beta \frac{e_i^2}{2} - \frac{(e_i + E_{-i})^\eta}{\eta} \left(d - \frac{(e_i + E_{-i})^\eta}{2\eta c} \right).$$

Then, the optimal emissions must fulfill the following condition

$$\alpha - \beta e_i = (e_i + E_{-i})^{\eta-1} \left(d - \frac{(e_i + E_{-i})^\eta}{\eta c} \right), \quad (32)$$

so that the marginal benefit of emissions must be equal to its marginal total costs. Moreover, the optimal emissions should satisfy the SOC

$$\frac{d^2 W_i}{de_i^2} = -\beta - (e_i + E_{-i})^{\eta-2} \left((\eta-1)d - (2\eta-1) \frac{(e_i + E_{-i})^\eta}{\eta c} \right) < 0. \quad (33)$$

Next, we apply the implicit function theorem to (32) to calculate the slope of the best response function for emissions

$$\frac{\partial e_i}{\partial e_j} = \frac{E^{\eta-2} \left((\eta-1)d - (2\eta-1) \frac{E^\eta}{\eta c} \right)}{-\beta - E^{\eta-2} \left((\eta-1)d - (2\eta-1) \frac{E^\eta}{\eta c} \right)}, \quad i, j = 1, \dots, N, \quad i \neq j, \quad (34)$$

where the numerator is the second derivative of the total cost function. Thus, for the additive model Prop. 1 can be rewritten as follows

Proposition 8 *If the maximization problem of net benefit has a solution that satisfies the SOC and $\eta > 1$, the total costs are convex for $E \in (0, \bar{E})$ and then the emissions are strategic substitutes whereas the total costs are concave for $E \in (\bar{E}, E^*]$ and then the emissions are strategic complements. E^* and \bar{E} are given by (29) and (31) respectively. However, if $\eta = 1$ the total costs are concave and emissions are strategic complements for $E \in (\bar{E}, E^*]$.*

It is also straightforward from (34) that if emissions are strategic substitutes the absolute value of $\partial e_i / \partial e_j$ is lower than one. The main difference with the previous model is that now for $\eta > 1$ the convexity or concavity of the total costs function depends on the level of aggregate emissions. For low values, the function is convex but for high values is concave.

Finally, we analyze the optimal response of adaptation to a change in the other countries' emissions. Using (26) we get that

$$\frac{\partial a_i}{\partial e_j} = \frac{E^{\eta-1}}{c} \left(\frac{\partial e_i}{\partial e_j} + 1 \right),$$

that substituting $\partial e_i/\partial e_j$ by (34) yields

$$\frac{\partial a_i}{\partial e_j} = \frac{E^{\eta-1}}{c} \left(\frac{-\beta}{-\beta - E^{\eta-2} \left((\eta-1)d - (2\eta-1)\frac{E^\eta}{\eta c} \right)} \right) > 0,$$

$i, j = 1, \dots, N$, $i \neq j$. Prop. 2 is also satisfied for the additive model.

In the next section, we investigate the stability of an IEA assuming that $\eta = 1$ what guarantees that emissions are strategic complements.

3.2 Participation in the linear-quadratic model

As the agreement formation game has been introduced in Section 2 we directly present the outcome of the third stage.

Third stage

According to expression (26) the level of adaptation selected by countries is

$$a_i^s = a_j^f = \frac{E}{c}. \quad (35)$$

For these levels of adaptation, the minimized total costs are given by a linear-quadratic function

$$TC_i(E) = dE - \frac{E^2}{2c},$$

with a maximum for $E^* = cd$ that gives a marginal costs function that is *linear*. Using this cost function we obtain also a linear-quadratic specification for the net benefit function

$$W_i(e_i, E_{-i}) = \alpha e_i - \beta \frac{e_i^2}{2} - \left(d(e_i + E_{-i}) - \frac{(e_i + E_{-i})^2}{2c} \right)$$

Second stage: the PANE of the emission game

Then, for non-signatories the optimal emissions must satisfy the following condition

$$\alpha - \beta e_j^f = d - \frac{E}{c}.$$

Observe that the marginal total cost of emissions on the RHS of this condition is decreasing with respect total emissions. Using this condition we can write the non-signatories' emissions as a function of total emissions

$$e_j^f = \frac{1}{\beta} \left(\alpha - \left(d - \frac{E}{c} \right) \right). \quad (36)$$

Moreover, the SOC (33) yields

$$\frac{\partial^2 W_j^f}{\partial (e_j^f)^2} = -\beta + \frac{1}{c} < 0 \rightarrow 1 < \beta c. \quad (37)$$

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

$$\max_{\{e_1^s, \dots, e_n^s\}} W_A = \sum_{i=1}^n W_i^s = \sum_{i=1}^n \left\{ \alpha e_i^s - \frac{\beta}{2} (e_i^s)^2 - E \left(d - \frac{E}{2c} \right) \right\}.$$

Looking for the symmetric solution, the FOC gives

$$\alpha - \beta e^s = n \left(d - \frac{E}{c} \right),$$

that requires that the marginal benefit of national emissions be equal to the marginal total costs for the agreement. From this condition we obtain that

$$e^s = \frac{1}{\beta} \left(\alpha - n \left(d - \frac{E}{c} \right) \right), \quad (38)$$

where $E = ne^s + (N - n)e^f$.

The SOC for the maximization of the agreement net benefits yields

$$\frac{\partial^2 W_A}{\partial (e^s)^2} = -\beta + \frac{n^2}{c} < 0 \rightarrow n^2 < \beta c.$$

Hence, if we assume that $N^2 < \beta c$, the SOC will hold for any level of participation including the fully cooperative solution and the SOC for the maximization of non-signatories' net benefits (37) will be also satisfied.¹⁴

The calculation of total emissions is immediate using (36) and (38)

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

Next, substituting in (35), (36) and (38) the level of adaptation and the emissions for non-signatories and signatories are obtained

$$a = \frac{N\alpha - (n^2 + N - n)d}{c\beta - (n^2 + N - n)}, \quad (40)$$

$$e^f = \frac{c\beta(\alpha - d) - n(n - 1)\alpha}{c\beta - (n^2 + N - n)}, \quad e^s = \frac{c\beta(\alpha - nd) + \alpha(N - n)(n - 1)}{c\beta - (n^2 + N - n)}. \quad (41)$$

¹⁴It is easy to check that if the SOC's hold, emissions are strategic complements in this stage of the game.

Observe that condition $N^2 < \beta c$ guarantees that the denominator of these expressions is positive. On the other hand, as $n^2 + N - n$ increases with n , $\alpha/N > d$ will give a positive numerator for a and e^s . Moreover, this condition also guarantees that the numerator of e^f is positive since $e^s < e^f$ according to (36) and (38). As was established in the presentation of the model, we also require that for the equilibrium quantities the marginal damage be positive. Using (40) the marginal damage can be written as follows

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

that is positive for $d > N\alpha/c\beta$. It is easy to check that the interval $(N\alpha/c\beta, \alpha/N)$ is not empty if $N^2 < \beta c$ so that we can summarize all these conditions in the following assumption

Assumption 3.1: $N^2 < \beta c$ and $d \in (N\alpha/c\beta, \alpha/N)$.

Thus, if this assumption holds, the level of adaptation is positive and the emissions are positive both for signatories and non-signatories. Moreover, the marginal damage and the marginal total costs are positive. Observe that the marginal total costs are given by $d - \frac{E}{c}$ what implies that is equal to $d - a$ according the expression (35). Then, if d is larger than a the marginal total costs and hence the marginal benefit are positive.

Finally, eliminating national emissions from the net benefit function using (36) we get the following expression

$$W^f = \frac{\alpha^2 - d^2}{2\beta} - \left(\frac{\beta c - 1}{\beta c}\right) E(d - \frac{E}{2c}), \quad (42)$$

for the non-signatories' net benefit, and using (38) we obtain the following expression

$$W^s = \frac{\alpha^2 - n^2 d^2}{2\beta} - \left(\frac{\beta c - n^2}{\beta c}\right) E(d - \frac{E}{2c}), \quad (43)$$

for the signatories' net benefit.

Next, we could show that Props. 3, 4 and 6 of the multiplicative model also hold for the linear-quadratic model, but for avoiding reiteration we will skip them for focusing on the result that makes a difference with the previous model: the behaviour of the difference between the signatories' net benefit and the non-signatories' net benefit.

For the linear-quadratic model, we obtain that

Proposition 9 *The absolute value of the difference $W^s - W^f$ increases with the number of signatories for all $n \in [2, N - 1]$.*

Proof. Using (42) and (43) we can calculate the difference in the net benefits

$$W^s - W^f = -\frac{n^2 - 1}{\beta} \left(\frac{d^2}{2} - \frac{1}{c} E \left(d - \frac{E}{2c} \right) \right),$$

where the second factor is positive what establishes that $W^s < W^f$. Notice that for $\eta = 1$, total cost is given by $E \left(d - \frac{E}{2c} \right)$ so that the difference between parenthesis will be minimum when total cost is maximum for $E^* = cd$, but for this level of emissions it is easy to check that the difference in the parenthesis is zero. Thus, as the equilibrium must occur for a level of emissions that yields positive marginal total costs according to the FOCs for the maximization of net benefits, i.e. for a level of emissions lower than cd , the parenthesis is positive for the quantities that maximize the net benefits. Next, we take the derivative of this difference with respect to n

$$\frac{\partial(W^s - W^f)}{\partial n} = - \left(\frac{2n}{\beta} \left(\frac{d^2}{2} - \frac{1}{c} E \left(d - \frac{E}{2c} \right) \right) - \frac{n^2 - 1}{\beta c} \left(d - \frac{E}{c} \right) \frac{\partial E}{\partial n} \right) \quad (44)$$

and calculate

$$\frac{\partial E}{\partial n} = \frac{c(N\alpha - cd\beta)(2n - 1)}{(c\beta - (n^2 + N - n))^2},$$

that is negative according to Assumption 3.1. The result is that (44) is negative. Take into account that $d - (E/c)$ are the marginal total costs that must be positive. ■

The first stage: the stability analysis

For the linear-quadratic specification of the net benefit function, the auxiliary function $S(n)$ reads as follows

$$\begin{aligned} S(n) = & -\frac{d^2(n^2 - 1)}{2\beta} \\ & + \left(\frac{\beta c - 1}{\beta c} \right) \left(\frac{cd(N\alpha - ((n-1)^2 + N - n + 1)d)}{c\beta - ((n-1)^2 + N - n + 1)} - \frac{c(N\alpha - ((n-1)^2 + N - n + 1)d)^2}{2(c\beta - ((n-1)^2 + N - n + 1))^2} \right) \\ & - \left(\frac{\beta c - n^2}{\beta c} \right) \left(\frac{cd(N\alpha - (n^2 + N - n)d)}{c\beta - (n^2 + N - n)} - \frac{c(N\alpha - (n^2 + N - n)d)^2}{2(c\beta - (n^2 + N - n))^2} \right). \end{aligned} \quad (45)$$

Using this expression it can be shown that

Proposition 10 *An agreement consisting of three countries is stable if $N \geq 7$.*

Proof. Doing $n = 3$ in (45) the auxiliary function yields

$$S(3) = \frac{4(N\alpha - cd\beta)^2((N-1)c\beta - N^2 - 3N)}{\beta(c\beta - (N+2))^2(c\beta - (N+6))^2}$$

that is positive if $c\beta > N^2$ for $N \geq 3$. Thus, an agreement formed by three countries is internally stable if $N \geq 3$. On the other hand, $n = 4$ yields

$$S(4) = -\frac{(N\alpha - cd\beta)^2((c\beta)^2 - 3(2N + 10)c\beta + 5N^2 + 56N + 144)}{2\beta(c\beta - (N + 6))^2(c\beta - (N + 12))^2}$$

that takes a negative value if $c\beta > N^2$ for $N \geq 7$ what establishes that an agreement consisting of three countries is also externally stable. Take into account that $S(4) < 0$ implies that $W^s(4) < W^f(3)$ what indicates that if there is an agreement signed by three countries, a non-signatory gets a larger net benefit being a non-signatory than joining the agreement. ■

Although it is easy to show the stability of agreement consisting of three countries, it is more difficult to show that this is the only stable agreement. For this reason in the next subsection we solve a numerical exercise that suggests that only an agreement with three countries is stable.

3.3 Numerical example

In the example we give parameter α , β and N the values 1000, 2 and 10 respectively and for parameters c and d we propose the following values

$$c = \{80.0, 82.5, 87.5, 87.5, 90.0\}, \quad d = \{65.0, 72.5, 80.0, 87.5, 95.0\},$$

taking into account that all resulting combinations (c, d) must satisfy Assumption 3.1. For the twenty-five combinations studied only an agreement consisting of three countries is stable.

Next, we show graphically the difference in net benefits as a function of the number of signatories for $c = 82.5$ and $d = 72.5$. As can be seen in Fig. 4, this difference increase in a convex way giving huge differences in net benefits when the number of signatories is high.

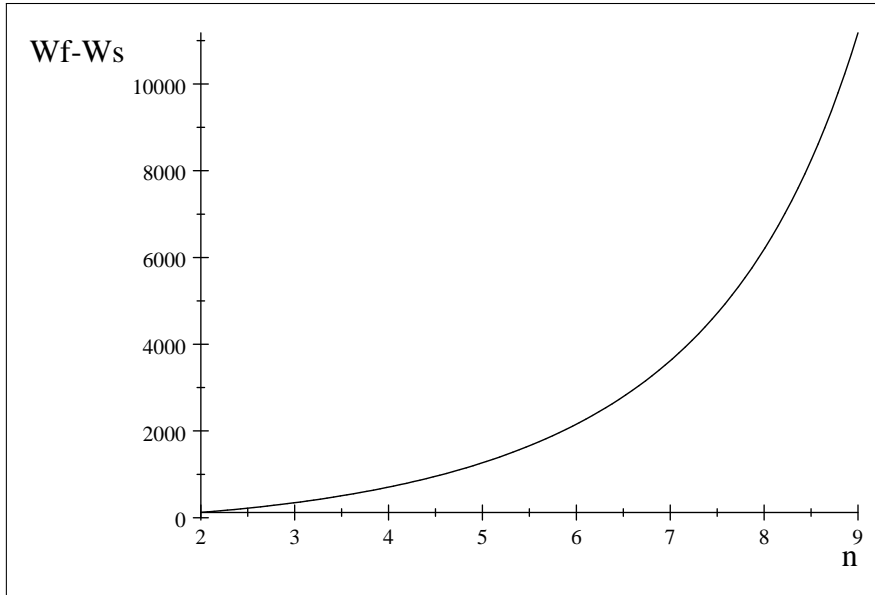


FIGURE 4. Difference in the net benefit between signatories and non-signatories.

Finally, we plot the auxiliary function $S(n)$ for the same parameter values. Fig. 5a is defined in the interval $[2,10]$ and Fig. 5b in the interval $[2,4]$. As Fig. 5b shows clearly the unique stable agreement consist of 3 countries. The figures show that the incentive to enter into the agreement disappears quickly. The difference in net benefits increases at an increases rate that makes negative the difference $W^s(n) - W^f(n - 1)$ for $n > 3$.

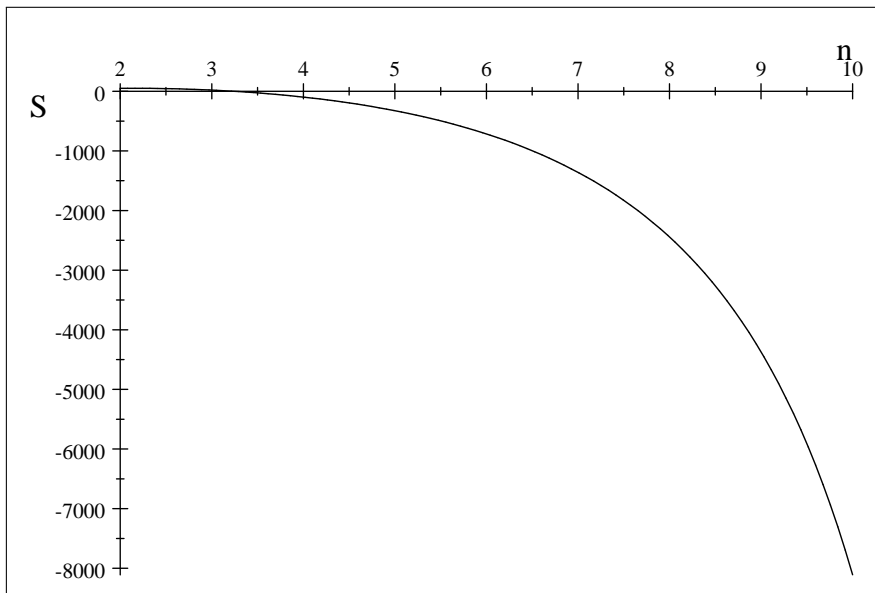


FIGURE 5a. The auxiliary function $S(N)$ for $c=82.5$ and $d=72.5$

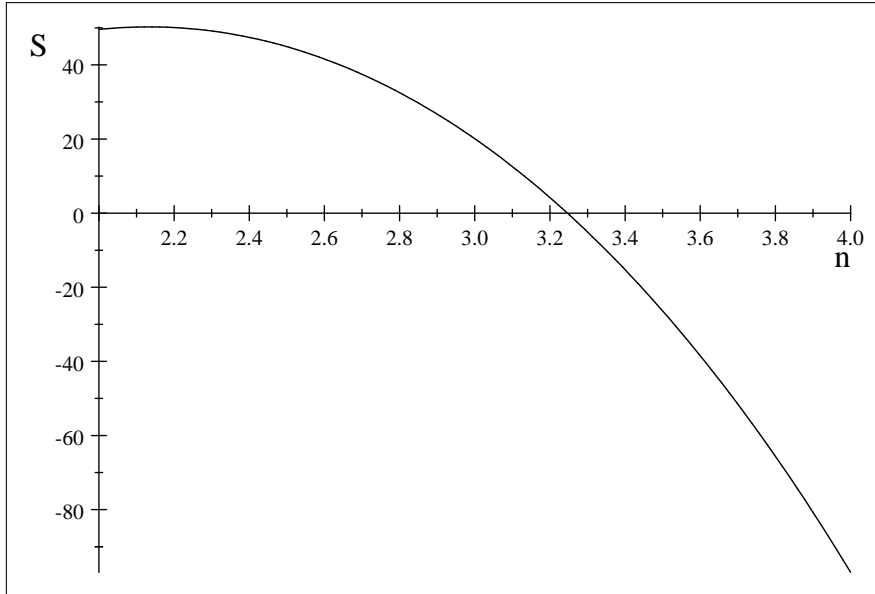


FIGURE 5b. The auxiliary function $S(N)$ for $c=82.5$ and $d=72.5$ in the interval $[2,4]$

Comparing these results with those obtained for the multiplicative model, we can conclude that the difference in the way the participation affects the net benefits is behind the difference in the outcome the game gives on participation. It should be highlighted that this is the unique difference we find between the two models. The question now is what is explaining this difference and our hypothesis is the difference in the way adaptation reduces damages. However, the clear conclusion we can derive of the comparison between the two models is that the strategic complementarity does not seem a sufficient condition to get higher levels of cooperation. For both models, countries' emissions are complements but they yield different predictions on participation.

4 Conclusions

This paper analyzes the impact of adaptation on the stability of an IEA. To address this issue 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 emissions. Finally, in the third stage, each country decides on its level of adaptation noncooperatively. We solve this game for two different models assuming that damages are linear with respect to emissions. This assumption guarantees that emission are complements in the second stage of the game for both cases. For the first model, that we have named the multiplicative model, adaptation reduces damages

in a *multiplicative* way. However, for the second model, that we have named the additive or linear-quadratic game, the reduction in damages occurs in an additive way. The analysis shows that the properties of the emission game played in the second stage are the same for the two models and coincide with the properties of the model without adaptation except in regards the difference between the net benefit of non-signatories and the net benefit of signatories. This difference is positive in both cases, but for the additive model we find that it increases at an increasing rate, however for the multiplicative model it increases at a decreasing rate and presents a maximum for a level of participation lower than the grand coalition. This different behaviour in the difference between net benefits is enough to lead to different outcomes in the first stage of the game. We obtain for the additive model that complementarity does not modify the results obtained by the models without adaptation where emissions are strategic substitutes, but it promotes cooperation when adaptation reduces damages in a multiplicative way. In fact, we find that in this case the larger the gains of full cooperation, the larger the participation. Comparing the results obtained for each model, we can conclude that complementarity plays in favor of cooperation but it is not a sufficient condition to obtain more participation in an IEA. Our research suggests that the way adaptation reduces damages plays a critical role over the outcome of the coalition formation game. Nevertheless, we could claim also that the differences in the benefit function could affect the outcome of the game. Although this is a claim that deserves to be written down in the research agenda, our conjecture is that the different specifications of the benefit function are not behind the difference we obtain about cooperation. The reason is that for both models the benefit function has the same basic properties. The benefit increases with emissions and the marginal benefit is decreasing, i.e. the benefit function is an increasing concave function for both specifications.

There are two obvious extensions for the game analyzed in this paper that could be addressed in future research. The first one refers to the timing of the game and the role of adaptation as a commitment device. It is easy to show that our results hold if emissions and adaptation are chosen simultaneously, but if the investment in adaptation is chosen before countries decide about their levels of emissions both the fully noncooperative equilibrium and the outcome of the coalition formation game will change as has been pointed out by Zehaie (2009), Heuson et al. (2015) and Breton and Sbragia (2017b). Notice that in the game analyzed in this paper, the investment in adaptation of a country is a strategic complement of the other countries' emissions but is independent of the other countries' investments. A change in the timing of

the game could modify these strategic relationships between the investments in adaptation and surely would yield different results in terms of participation.

A second straightforward 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 and introduce this difference in a game where countries decide first on investment in adaptation. 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] Auerswald, Heike, Kai A. Konrad, and Marcel Thum (2018). "Adaptation, Mitigation and Risk-Taking in Climate Policy." *Journal of Economics* 124, 269-287.
- [2] Barrett, Scott (1994). "Self-Enforcing International Environmental Agreements." *Oxford Economic Papers* 46, 878-894.
- [3] Barrett, Scott (2006). "Climate Treaties and "Breakthrough" Technologies." *American Economic Review* 96, 22-25.
- [4] Barrett, Scott (2008). "Dikes versus Windmills: Climate Treaties and Adaptation." Johns Hopkins University, School of Advanced International Studies.
- [5] Battaglini, Marco and Bard Harstad (2016). "Participation and Duration of Environmental Agreements." *Journal of Political Economy* 124, 160-124.
- [6] Bayramoglu, Basak, Michael Finus, and Jean-Francois Jacques (2016). "Climate Agreements in a Mitigation-Adaptation Game." *Journal of Public Economics* 165, 101-113.
- [7] 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.
- [8] 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.

- [9] Breton, Michèle, and Lucia Sbragia (2017a). “Adaptation to Climate Change: Commitment and Timing Issues.” *Environmental and Resource Economics* 68, 975-995.
- [10] Breton, Michèle, and Lucia Sbragia (2017b). “The Impact of Adaptation on the Stability of International Environmental Agreements.” Les Cahiers du GERAD, G-2017-66.
- [11] Buob, Seraina, and Gunter Stephan (2011). “To Mitigate or To Adapt: How to Confront Global Climate Change.” *European Journal of Political Economy* 27, 1-16.
- [12] Caparrós, Alejandro (2018). “A Model of the (Incomplete) Paris Agreement on Climate Change.” Mimeo.
- [13] Carraro, Carlo, and Domenico Siniscalco (1993). “Strategies for the International Protection of the Environment.” *Journal of Public Economics* 52, 309-328.
- [14] 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.
- [15] d’Aspremont, Claude, Alexis Jacquemin, Jean Jaskold Gadszewicz, and John A. Weymark (1983). “On the Stability of Collusive Price Leadership.” *Canadian Journal of Economics* 16, 17-25.
- [16] Ebert, Udo, and Heinz Welsch (2011). “Optimal Response Functions in Global Pollution Problems Can Be Upward-Sloping: Accounting for Adaptation.” *Environmental Economics and Policy Studies* 13, 129-138.
- [17] Ebert, Udo, and Heinz Welsch (2012). “Adaptation and Mitigation in Global Pollution Problems: Economic Impacts of Productivity, Sensitivity, and Adaptive Capacity.” *Environmental and Resource Economics* 52, 49-64.
- [18] Eisenack, Klaus, and Leonhard Kähler (2016). “Adaptation to Climate Change Can Support Unilateral Emissions Reductions.” *Oxford Economic Papers* 68, 258-278.
- [19] 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.

- [20] Eyckmans, Johan, and Michael Finus (2006). “New Roads to International Environmental Agreements: The Case of Global Warming.” *Environmental Economics and Policy Studies*, 7, 391-414.
- [21] Farnham, Martin, and Peter Kennedy (2015). “Adapting to Climate Change: Equilibrium Welfare Implications for Large and Small Economies.” *Environmental and Resource Economics* 61, 345-363.
- [22] Finus, Michael, and Alejandro Caparrós (2015). *Game Theory and International Environmental Cooperation: Essential Readings*, Edward Elgar.
- [23] 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.
- [24] Harstad, Bard (2012). “Climate Contracts: A Game of Emissions, Investments, Negotiations, and Renegotiations.” *Review of Economic Studies* 79, 1527-1557.
- [25] Harstad, Bard (2016). “The Dynamic of Climate Agreements.” *Journal of the European Economic Association* 14, 719-752.
- [26] Harstad, Bard, Francesco Lancia, and Alessia Russo (2018). “Compliance Technology and Self-Enforcing Agreements.” Mimeo.
- [27] Helm, Carsten, and Robert C. Schmidt (2015). “Climate Cooperation with Technology Investments and Border Carbon Adjustment.” *European Economic Review* 75, 112-130.
- [28] Heuson, Clemens, Wolfgang Peters, Reimund Schwarze, and Anna-Katharina Topp (2015). “Investment and Adaptation as Commitment Devices in Climate Policies.” *Environmental and Resource Economics* 62, 769-790.
- [29] 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.
- [30] Hong, Fuhai, and Larry Karp (2012). “International Environmental Agreements with Mixed Strategies and Investment.” *Journal of Public Economics* 96, 685-697.

- [31] Ingham, Alan, Jie Ma, and Alistair M. Ulph (2013). “Can Adaptation and Mitigation Be Complements?” *Climatic Change* 120, 39-53.
- [32] 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.
- [33] Masoudi, Nahid, and Georges Zaccour (2016). “Adaptation and International Environmental Agreements.” *Environmental and Resource Economics*, forthcoming.
- [34] Masoudi, Nahid, and Georges Zaccour (2017). “Adapting to Climate Change: Is Cooperation Good for the Environment?” *Economics Letters* 153, 1-5.
- [35] Montero, Juan-Pablo (2002). “Permits, Standards, and Technology Innovation.” *Journal of Environmental Economics and Management* 44, 23-44.
- [36] Potters, Jan, and Sigrid Suetens (2009). “Cooperation in Experimental Games of Strategic and Substitutes.” *Review of Economic Studies* 76, 1125-1147.
- [37] Rubio, Santiago J. (2017). “Sharing R&D Investments in Breakthrough Technologies to Control Climate Change.” *Oxford Economic Papers* 69, 496-521.
- [38] Vardar, Baris, and Georges Zaccour (2018). “The Strategic Impact of Adaptation in a Transboundary Pollution Dynamic Game.” *Environmental Modeling & Assessment*, forthcoming.
- [39] Zehaie, Ficare (2009). “The Timing and Strategic Role of Self-Protection.” *Environmental and Resource Economics* 44, 337-350.

A.1 The signatories’ net benefit increases with the participation (multiplicative model).

Substituting the aggregate emissions in (19) by (14) the following expression for the signatories’ net benefit is obtained

$$W^s = \frac{\alpha^{\frac{1}{1-2\beta}}}{(cd)^{\frac{\beta}{1-2\beta}}} \left(\frac{1}{\beta n^{\frac{\beta}{1-\beta}}} \left(\frac{1}{n^{\frac{\beta}{1-\beta}}} + N - n \right)^{\frac{\beta}{1-2\beta}} - 2 \left(\frac{1}{n^{\frac{\beta}{1-\beta}}} + N - n \right)^{\frac{1-\beta}{1-2\beta}} \right) + c. \quad (46)$$

Next, we calculate the derivative of the net benefit with respect to n

$$\begin{aligned} \frac{\partial W^s}{\partial n} &= \frac{\alpha^{\frac{1}{1-2\beta}}}{(cd)^{\frac{\beta}{1-2\beta}}} \left(-\frac{1}{1-\beta} \frac{1}{n^{\frac{1}{1-\beta}}} \left(\frac{1}{n^{\frac{\beta}{1-\beta}}} + N - n \right)^{\frac{\beta}{1-2\beta}} \right. \\ &\quad - \frac{1}{n^{\frac{\beta}{1-\beta}}} \frac{1}{1-2\beta} \left(\frac{1}{n^{\frac{\beta}{1-\beta}}} + N - n \right)^{\frac{\beta}{1-2\beta}-1} \left(\frac{\beta}{1-\beta} \frac{1}{n^{\frac{1}{1-\beta}}} + 1 \right) \\ &\quad \left. + 2 \frac{1-\beta}{1-2\beta} \left(\frac{1}{n^{\frac{\beta}{1-\beta}}} + N - n \right)^{\frac{1-\beta}{1-2\beta}-1} \left(\frac{\beta}{1-\beta} \frac{1}{n^{\frac{1}{1-\beta}}} + 1 \right) \right), \end{aligned}$$

that taking common factor yields

$$\begin{aligned} &= \frac{\alpha^{\frac{1}{1-2\beta}}}{(cd)^{\frac{\beta}{1-2\beta}}} \frac{1}{(1-\beta)n^{\frac{1}{1-\beta}}} \left(\frac{1}{n^{\frac{\beta}{1-\beta}}} + N - n \right)^{\frac{\beta}{1-2\beta}} \\ &\times \left(-1 - \frac{1}{n^{\frac{\beta}{1-\beta}}} \frac{1}{1-2\beta} \left(\frac{1}{n^{\frac{\beta}{1-\beta}}} + N - n \right)^{-1} (\beta + (1-\beta)n^{\frac{1}{1-\beta}}) + \frac{2(1-\beta)}{1-2\beta} (\beta + (1-\beta)n^{\frac{1}{1-\beta}}) \right). \end{aligned}$$

Developing the second parenthesis the derivative can be written as follows

$$\begin{aligned} \frac{\partial W^s}{\partial n} &= \frac{\alpha^{\frac{1}{1-2\beta}}}{(cd)^{\frac{\beta}{1-2\beta}}} \frac{1}{(1-\beta)n^{\frac{1}{1-\beta}}} \left(\frac{1}{n^{\frac{\beta}{1-\beta}}} + N - n \right)^{\frac{\beta}{1-2\beta}} \\ &\times \left(-1 + \frac{1-2\beta + 2(1-\beta)(N-n)n^{\frac{\beta}{1-\beta}} (\beta + (1-\beta)n^{\frac{1}{1-\beta}})}{1 + (N-n)n^{\frac{\beta}{1-\beta}} (1-2\beta)} \right). \end{aligned} \quad (47)$$

As the first factor is positive, the derivative will be positive if the second factor is positive.

Let's suppose the contrary

$$\frac{1-2\beta + 2(1-\beta)(N-n)n^{\frac{\beta}{1-\beta}} (\beta + (1-\beta)n^{\frac{1}{1-\beta}})}{1 + (N-n)n^{\frac{\beta}{1-\beta}} (1-2\beta)} \leq 1.$$

Developing this expression and doing the following variable change $x = n^{\frac{1}{1-\beta}}$ we obtain that

$$-2\beta^2 + 3\beta - 1 - (2\beta^2 - 4\beta + 1)Nx^\beta + (4\beta^2 - 7\beta + 2)x + 2(1-\beta)^2Nx^{1+\beta} - 2(1-\beta)^2x^2 \leq 0. \quad (48)$$

It is immediate that for $n = 1$ the LHS of the inequality is positive and that for $n = 0$ is negative if $\beta < 0.5$. Observe that the independent term is negative for values of β in the interval $(0, 0.5)$.

Moreover, it is easy to check that for this interval, the expression is also positive for $n = N$.

Substituting x by $N^{\frac{1}{1-\beta}}$, the expression gives

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

$$-2(1 - \beta)^2(N^{\frac{1}{1-\beta}})^2 = \left(N^{\frac{1}{1-\beta}} - 1\right) (2\beta^2 - 3\beta + 1) > 0 \text{ for } \beta \in (0, 0.5).$$

Thus, (48) yields a contradiction for $n = 1$ and $n = N$. Next, we show that a contradiction is also obtained for $n \in (1, N)$ so that it can be concluded that the signatories' net benefit increases with the cooperation. For obtaining this contradiction we write the LHS of (48) as follows

$$f(\beta) - g(\beta)Nx^\beta + h(\beta)x + 2(1 - \beta)^2Nx^{1+\beta} - 2(1 - \beta)^2x^2, \quad (49)$$

where $f(\beta) = -2\beta^2 + 3\beta - 1$ is negative for $\beta \in (0, 0.5)$, $g(\beta) = 2\beta^2 - 4\beta + 1$ is positive for $\beta \in (0, 0.29289)$ and negative for $\beta \in (0.29289, 0.5)$, and $h(\beta) = 4\beta^2 - 7\beta + 2$ is positive for $\beta \in (0, 0.35961)$ and negative for $\beta \in (0.35961, 0.5)$.

Firstly, let's consider the case for which $g(\beta)$ and $h(\beta)$ are both positive, i.e., when $\beta \in (0, 0.29289)$. For these values of β , the first derivative of (49) with respect to x is

$$-\beta g(\beta)Nx^{\beta-1} + h(\beta) + 2(1 + \beta)(1 - \beta)^2Nx^\beta - 4(1 - \beta)^2x, \quad (50)$$

that will take a zero value for the roots of the following equation

$$h(\beta)x^{1-\beta} + 2(1 + \beta)(1 - \beta)^2Nx = \beta g(\beta)N + 4(1 - \beta)^2x^{2-\beta}, \quad (51)$$

where the LHS is an increasing concave function equal to zero for $x = 0$ and the RHS is an increasing convex function equal to a positive value given by $\beta g(\beta)N$ for $x = 0$. As the derivative of (49) for $x = 0$ is $-\infty$ and the function is negative for $n = 0$ and positive for $n = 1$, the equation (51) has at least one root that defines a minimum in the interval $[0, 1]$. In that case, the function will be increasing for $n > 1$ what establishes that it will be positive in the interval $(1, N)$. However, the equation (51) could have two solutions. The lowest one would give the minimum we have just characterized and the greatest would define a maximum. In this case, on the right of this maximum there would exist a value for n above which the function will be negative. But as we know that for $n = N$ the expression is positive this critical value would be higher than N and also in this case (49) would be positive in the interval $(1, N)$.

Let's suppose next that $\beta = 0.26289$. For this value we have that $g(\beta) = 0$ and $h(\beta) > 0$. Now, the first derivative of (49) is

$$h(\beta) + 2(1 + \beta)(1 - \beta)^2Nx^\beta - 4(1 - \beta)^2x,$$

and their extremes will be given by the following equation

$$h(\beta) + 2(1 + \beta)(1 - \beta)^2 Nx^\beta = 4(1 - \beta)^2 x.$$

As the LHS is an increasing concave function that takes a positive value equal to $h(\beta)$ for $x = 0$ whereas the RHS is an increasing linear function that is zero for $x = 0$, this equation will have only one root. Moreover, on the left of the root, the first derivative is positive and consequently the root is a maximum and the argument we have just presented above applies for concluding that (49) is positive in the interval $(1, N)$.

Next, we analyze the sign of (49) for values of β in the interval $(0.26289, 0.35961)$. For this interval, the first coefficient, $g(\beta)$, is negative and the second one, $h(\beta)$, positive and the extremes of the expression are given by

$$-\beta g(\beta)N + h(\beta)x^{1-\beta} + 2(1 + \beta)(1 - \beta)^2 Nx = 4(1 - \beta)^2 x^{2-\beta},$$

where the LHS is an increasing concave function of x that takes a positive value equal to $-\beta g(\beta)N$ when $x = 0$ and the RHS is an increasing convex function that is equal to zero for $x = 0$. Thus, the function (49) has only one extreme. On the left of the extreme, the first derivative is positive and the extreme will be a maximum and as occurs for the previous cases the function will be positive in the interval $(1, N)$.

If $\beta = 0.35961$ then $h(\beta) = 0$ and $g(\beta) < 0$. In this case the first derivative is

$$-\beta g(\beta)Nx^{\beta-1} + 2(1 + \beta)(1 - \beta)^2 Nx^\beta - 4(1 - \beta)^2 x,$$

and the extremes of the function are given by the following equation

$$-\beta g(\beta)N + 2(1 + \beta)(1 - \beta)^2 Nx = 4(1 - \beta)^2 x^{2-\beta}.$$

The LHS of the equation is an increasing linear function with an intersection point with the vertical axis equal to the positive value $-\beta g(\beta)N$, whereas the RHS is an increasing convex function that is zero for $x = 0$ therefore the equation has only one positive root. On the left of the extreme, the first derivative is positive and consequently the extreme is a maximum and the same argument we have used in the previous cases applies.

Finally, we study the sign of the function (49) for values of β in the interval $(0.35961, 0.5)$ for which $g(\beta)$ and $h(\beta)$ are negative. For these values the equation that gives us the extremes is equal to

$$-\beta g(\beta)N + 2(1 + \beta)(1 - \beta)^2 Nx = -h(\beta)x^{1-\beta} + 4(1 - \beta)^2 x^{2-\beta}, \quad (52)$$

where the LHS is an increasing linear function that takes a positive values equal to $-\beta g(\beta)N$ for $x = 0$ and the RHS is an increasing function that presents an inflection point for the following value of x

$$x = -\frac{\beta h(\beta)}{4(2 - \beta)(1 - \beta)^2} > 0. \quad (53)$$

On the right of this value, the second derivative is negative and on the left positive. Therefore, the function is initially (strictly) concave to become (strictly) convex on the right of the inflection point. Thus, the equation could have until three roots, but it is easy to define a sufficient condition for uniqueness.

We are going to calculate first the value of the RHS of (52) corresponding to the inflection point. Afterwards, we will equalize that value with the intersection point with the vertical axis of the LHS what will allow us to define a threshold value for N such that for any value of N equal to or larger than the threshold value, the equation (52) will have only one positive root. Observe that if the value of the intersection point with the vertical axis of the LHS coincides with the value of the RHS for the inflection point, no solution for equation (52) could be in the concave section of the RHS, and the equation will present only one root for the convex section of the RHS. Substituting (53) in the RHS of (52) we obtain the following expression

$$= -h(\beta) \left(-\frac{\beta h(\beta)}{4(2 - \beta)(1 - \beta)^2} \right)^{1-\beta} + 4(1 - \beta)^2 \left(-\frac{\beta h(\beta)}{4(2 - \beta)(1 - \beta)^2} \right)^{2-\beta}.$$

Doing this expression equal to $-\beta g(\beta)N$, the following threshold value for N is obtained

$$\tilde{N} = \frac{h(\beta)}{\beta g(\beta)} \left(-\frac{\beta h(\beta)}{4(2 - \beta)(1 - \beta)^2} \right)^{1-\beta} - \frac{4(1 - \beta)^2}{\beta g(\beta)} \left(-\frac{\beta h(\beta)}{4(2 - \beta)(1 - \beta)^2} \right)^{2-\beta}.$$

The graphical representation of this expression in the interval $(0.35961, 0.5)$ shows that the threshold value is increasing with β , that is equal to zero for the lower extreme of the interval and takes the value 1.08 for the upper extreme. Thus, for any $N > 2$, the equation (52) will have a unique positive root that defines a maximum and it can be concluded that the expression (49) is positive in the interval $(1, N)$. Summarizing, it has been shown that for all β in the interval $(0, 0.5)$ and all n in the interval $[1, N]$, (48) will yield a contradiction for $N > 2$ and we can

conclude that the derivative (47) is positive or in other words that the signatories' net benefit increases with the number of signatories.

A.2 Proof of Proposition 5

Taking the derivative with respect to n of the difference in the net benefit given by (20) we obtain

$$\frac{\partial(W^s - W^f)}{\partial n} = \frac{\alpha^{\frac{1}{1-\beta}}}{\beta} \frac{\beta}{1-\beta} \left(\frac{E}{cd}\right)^{\frac{\beta}{2(1-\beta)}} \frac{1}{n^{\frac{\beta}{1-\beta}}} \left(-\frac{1}{n} + \frac{1-n^{\frac{\beta}{1-\beta}}}{2E} \frac{\partial E}{\partial n}\right). \quad (54)$$

We will analyze the sign of the parenthesis since the rest of terms are positive. Firstly, we calculate $\partial E/\partial n$ using (14)

$$\frac{\partial E}{\partial n} = -\left(\frac{\alpha}{(cd)^{\frac{1}{2}}}\right)^{\frac{2}{1-2\beta}} \frac{2(1-\beta)}{1-2\beta} \left(\frac{1}{n^{\frac{\beta}{1-\beta}}} + N - n\right)^{\frac{2(1-\beta)}{1-2\beta}-1} \left(\frac{\beta}{1-\beta} n^{\frac{\beta}{1-\beta}-1} + 1\right),$$

so that the parenthesis of (54) can be written as follows

$$\begin{aligned} & -\frac{1}{n} + \frac{1-n^{\frac{\beta}{1-\beta}}}{2E} \frac{\partial E}{\partial n} \\ &= -\frac{1}{n} \left(1 + \frac{(1-n^{\frac{\beta}{1-\beta}})(\beta n^{\frac{\beta}{1-\beta}} + (1-\beta)n)n^{\frac{\beta}{1-\beta}}}{(1-2\beta)(1+(N-n)n^{\frac{\beta}{1-\beta}})}\right). \end{aligned} \quad (55)$$

Next, we investigate if there exists a value for n that makes zero this parenthesis. As (54) would be also zero in this case we will have localized an extreme for the difference in the net benefits. Thus, this extreme should satisfy that

$$\frac{(n^{\frac{\beta}{1-\beta}} - 1)(\beta n^{\frac{\beta}{1-\beta}} + (1-\beta)n)n^{\frac{\beta}{1-\beta}}}{(1-2\beta)(1+(N-n)n^{\frac{\beta}{1-\beta}})} = 1.$$

A condition that can be rewritten as follows

$$(n^{\frac{\beta}{1-\beta}} - 1)(\beta n^{\frac{\beta}{1-\beta}} + (1-\beta)n)n^{\frac{\beta}{1-\beta}} = (1-2\beta)(1+(N-n)n^{\frac{\beta}{1-\beta}}).$$

Developing both sides of the equation and reordering terms we obtain the following expression

$$\beta n^{\frac{3\beta}{1-\beta}} + (1-\beta)n^{\frac{1+\beta}{1-\beta}} = 1-2\beta + (1-2\beta)Nn^{\frac{\beta}{1-\beta}} + \beta n^{\frac{1}{1-\beta}} + \beta n^{\frac{2\beta}{1-\beta}}. \quad (56)$$

In order to study whether this equation has a solution we propose a variable change $n^{\frac{1}{1-\beta}} = x$

$$\beta x^{3\beta} + (1-\beta)x^{1+\beta} = (1-2\beta) + (1-2\beta)Nx^\beta + \beta x + \beta x^{2\beta}. \quad (57)$$

The LHS of this equation is an increasing convex function that takes a value of zero for $x = 0$ if $\beta \geq 1/3$ and presents an inflection point for

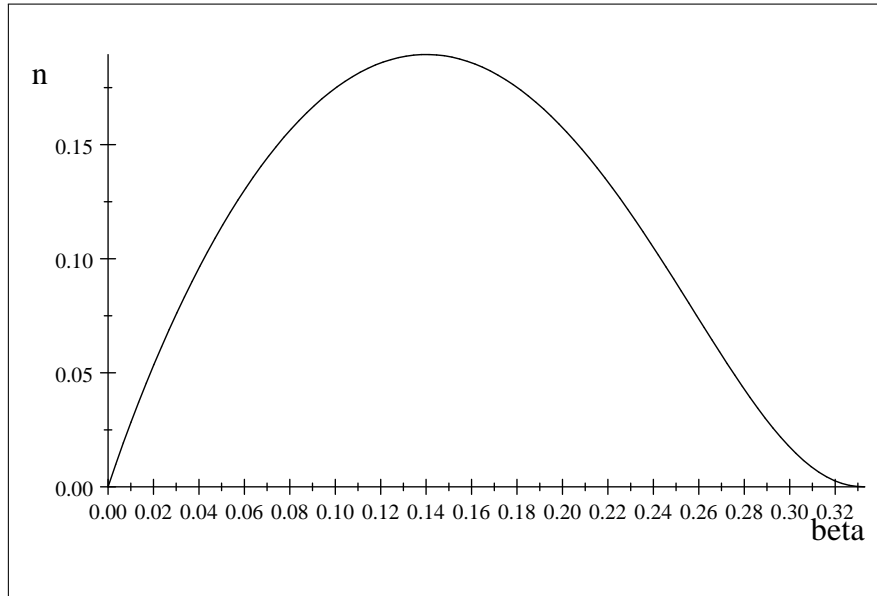
$$x = \left(\frac{3(1-3\beta)\beta}{(1+\beta)(1-\beta)} \right)^{\frac{1}{1-2\beta}}$$

if $\beta < 1/3$. On the left of the inflection point, the second derivative is negative and the LHS is a concave function of x . Then on the right of the inflection point, the second derivative is positive and the function is convex.

Taking into account the variable change, the value of n for the inflection point depends on β as follows

$$n = \left(\frac{3(1-3\beta)\beta}{(1+\beta)(1-\beta)} \right)^{\frac{1-\beta}{1-2\beta}}.$$

The graphical representation of this expression in the interval $(0, \frac{1}{3})$ is the following



Therefore, the inflection point occurs for a value of n lower than one. Moreover, it is easy to check that for $n = 1$, the RHS of (56) is larger than the LHS.

Next, we characterize the properties of the RHS of (57). It is easy to check that is an increasing concave function of x that takes the value $1 - 2\beta > 0$ for $x = 0$. Therefore, as the RHS is an increasing concave function larger than the LHS for $n = 1$ and the the LHS is an increasing convex function for $n > 1$, the equation has only one positive root larger than 1.

Thus, on the left of the root we will have that

$$\beta x^{3\beta} + (1-\beta)x^{1+\beta} < (1-2\beta) + (1-2\beta)Nx^\beta + \beta x + \beta x^{2\beta},$$

what implies that

$$(n^{\frac{\beta}{1-\beta}} - 1)(\beta n^{\frac{\beta}{1-\beta}} + (1 - \beta)n)n^{\frac{\beta}{1-\beta}} < (1 - 2\beta) \left(1 + (N - n)n^{\frac{\beta}{1-\beta}}\right),$$

so that

$$\frac{(n^{\frac{\beta}{1-\beta}} - 1)(\beta n^{\frac{\beta}{1-\beta}} + (1 - \beta)n)n^{\frac{\beta}{1-\beta}}}{(1 - 2\beta) \left(1 + (N - n)n^{\frac{\beta}{1-\beta}}\right)} < 1,$$

and

$$0 < 1 + \frac{(1 - n^{\frac{\beta}{1-\beta}})(\beta n^{\frac{\beta}{1-\beta}} + (1 - \beta)n)n^{\frac{\beta}{1-\beta}}}{(1 - 2\beta) \left(1 + (N - n)n^{\frac{\beta}{1-\beta}}\right)},$$

yielding a negative value for (55) what implies that

$$\frac{\partial(W^s - W^f)}{\partial n} < 0.$$

As the difference $W^s - W^f$ is negative, the sign of this derivative is saying us that the absolute value of the difference is increasing, that it will reach a maximum for the positive root of (56) to decrease afterwards.

Finally, we try to determine whether this root is larger or lower than N . Using (56) we write N as a function of n

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

and calculate the difference

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

where the subtrahend is the 45° line. Then, if this difference is positive we will have that $n < N(n)$ and the difference in the net benefit will be maximum for a level of participation lower than N . In order to determine the sign of this difference we are going to investigate first if there exists a value for n for which the difference is zero

$$\frac{\beta n^{\frac{3\beta}{1-\beta}} + (1 - \beta)n^{\frac{1+\beta}{1-\beta}} - (1 - 2\beta) - \beta n^{\frac{1}{1-\beta}} - \beta n^{\frac{2\beta}{1-\beta}}}{(1 - 2\beta)n^{\frac{\beta}{1-\beta}}} - n = 0.$$

Developing this expression and ordering terms we get the following equation

$$\beta n^{\frac{3\beta}{1-\beta}} + (1 - \beta)n^{\frac{1+\beta}{1-\beta}} = (1 - 2\beta) + (1 - \beta)n^{\frac{1}{1-\beta}} + \beta n^{\frac{2\beta}{1-\beta}}. \quad (58)$$

This is an equation very similar to equation (56) we have just analyzed that has also only one positive root. Moreover, for values of n lower than the root, the RHS of the equation is larger than the LHS

$$\beta n^{\frac{3\beta}{1-\beta}} + (1-\beta)n^{\frac{1+\beta}{1-\beta}} < (1-2\beta) + (1-\beta)n^{\frac{1}{1-\beta}} + \beta n^{\frac{2\beta}{1-\beta}},$$

what implies that

$$\frac{\beta n^{\frac{3\beta}{1-\beta}} + (1-\beta)n^{\frac{1+\beta}{1-\beta}} - (1-2\beta) - \beta n^{\frac{1}{1-\beta}} - \beta n^{\frac{2\beta}{1-\beta}}}{(1-2\beta)n^{\frac{\beta}{1-\beta}}} < n.$$

Thus, for values of n lower than the root $N(n) < n$ and for values larger than the root $N(n) > n$. In this case, the root that maximizes the difference in the net benefit is lower than N . Therefore, it can be concluded that for all $\beta \in (0, 0.5)$ and all $N > 2$, there exists a value for n , n^* , such that $N(n^*) = n^*$ so that if $N > n^*$, the value for n that satisfies (56) will be lower than N and the difference in the net benefit between signatories and non-signatories ultimately will decrease with the participation.

A.3 Proof of Proposition 6 (Full cohesiveness)

Firstly, we substitute E by (14) in (21) obtaining the following expression for the net benefit

$$W = \frac{\alpha^{\frac{1}{1-2\beta}}}{(cd)^{\frac{\beta}{1-2\beta}}} \left(\frac{1}{\beta} \left(\frac{1}{n^{\frac{\beta}{1-\beta}}} + N - n \right)^{\frac{\beta}{1-2\beta}} \left(n^{\frac{1-2\beta}{1-\beta}} + N - n \right) - 2N \left(\frac{1}{n^{\frac{\beta}{1-\beta}}} + N - n \right)^{\frac{1-\beta}{1-2\beta}} \right) + Nc. \quad (59)$$

Taking common factor the derivative with respect to n is

$$\begin{aligned} \frac{\partial W}{\partial n} &= \frac{\alpha^{\frac{1}{1-2\beta}}}{(cd)^{\frac{\beta}{1-2\beta}}} \frac{1}{\beta(1-\beta)(1-2\beta)n^{\frac{1}{1-\beta}}} \left(\frac{1}{n^{\frac{\beta}{1-\beta}}} + N - n \right)^{\frac{\beta}{1-2\beta}} \\ &\times \left(-\beta \left(\frac{1}{n^{\frac{\beta}{1-\beta}}} + N - n \right)^{-1} (\beta + (1-\beta)n^{\frac{1}{1-\beta}}) \left(n^{\frac{1-2\beta}{1-\beta}} + N - n \right) \right. \\ &\left. + (1-2\beta)((1-2\beta)n - (1-\beta)n^{\frac{1}{1-\beta}}) + 2N\beta(1-\beta)(\beta + (1-\beta)n^{\frac{1}{1-\beta}}) \right), \quad (60) \end{aligned}$$

where the sign of the first factor (first line) is positive. Next, we focus on finding out the sign of the parenthesis (second and third lines).

The development of the parenthesis of (60) yields

$$\begin{aligned} & -\frac{\beta^2 n^{\frac{\beta}{1-\beta}} + \beta(1-\beta)n^{\frac{1+\beta}{1-\beta}}}{1 + (N-n)n^{\frac{\beta}{1-\beta}}} \left(n^{\frac{1-2\beta}{1-\beta}} + N - n \right) \\ & + (1-2\beta)^2 n + 2N(1-\beta)\beta^2 + (1-\beta)(2N\beta(1-\beta) - (1-2\beta))n^{\frac{1}{1-\beta}}. \end{aligned} \quad (61)$$

Thus, this expression should be positive to get that the aggregate net benefits increase with the participation. It is immediate that for $n = 0$, (61) is positive. On the other hand, for $n = N$ (61) gives

$$(1-\beta)(1-2\beta)N(\beta N^{\frac{1}{1-\beta}} + 1 - \beta - N^{\frac{\beta}{1-\beta}}).$$

This expression is positive if $\beta N^{\frac{1}{1-\beta}} + 1 - \beta - N^{\frac{\beta}{1-\beta}}$ is positive. Calculating the first derivative with respect to N it is checked that $\beta N^{\frac{1}{1-\beta}} + 1 - \beta - N^{\frac{\beta}{1-\beta}}$ has a minimum for $N = 1$ and that for this minimum the value of this expression is zero. Accordingly, it can be concluded that for $N \geq 2$ the expression is positive and (61) is positive for $n = N$.

Let's suppose next that (61) is negative or zero in the interval $(0, N)$.

$$\begin{aligned} & -\frac{\beta^2 n^{\frac{\beta}{1-\beta}} + \beta(1-\beta)n^{\frac{1+\beta}{1-\beta}}}{1 + (N-n)n^{\frac{\beta}{1-\beta}}} \left(n^{\frac{1-2\beta}{1-\beta}} + N - n \right) \\ & + (1-2\beta)^2 n + 2N(1-\beta)\beta^2 + (1-\beta)(2N\beta(1-\beta) - (1-2\beta))n^{\frac{1}{1-\beta}} \leq 0, \end{aligned}$$

that would require that

$$\begin{aligned} & \left((1-2\beta)^2 n + 2N(1-\beta)\beta^2 + (1-\beta)(2N\beta(1-\beta) - (1-2\beta))n^{\frac{1}{1-\beta}} \right) \left(1 + Nn^{\frac{\beta}{1-\beta}} - n^{\frac{1}{1-\beta}} \right) \\ & \leq \left(\beta^2 n^{\frac{\beta}{1-\beta}} + \beta(1-\beta)n^{\frac{1+\beta}{1-\beta}} \right) \left(n^{\frac{1-2\beta}{1-\beta}} + N - n \right). \end{aligned} \quad (62)$$

Developing both sides of the inequality and grouping terms we obtain that

$$\begin{aligned} & 2N(1-\beta)\beta^2 + (1-3\beta)(1-\beta)n + (4N\beta^3 - (1+2N)\beta^2 - (2N-3)\beta + N-1)n^{\frac{1}{1-\beta}} \\ & + N\beta^2(2N(1-\beta)-1)n^{\frac{\beta}{1-\beta}} + N(\beta-1)^2(2N\beta-1)n^{\frac{1+\beta}{1-\beta}} - (3\beta^2-3\beta+1)n^{\frac{2-\beta}{1-\beta}} \\ & - (1-\beta)^2(2N\beta-1)n^{\frac{2}{1-\beta}} \leq 0, \end{aligned} \quad (63)$$

where the coefficients $4N\beta^3 - (1+2N)\beta^2 - (2N-3)\beta + N-1$, $2N(1-\beta)-1$ and $3\beta^2-3\beta+1$ are positive for $\beta \in (0, 0.5)$ and $2N\beta-1 = 0$ for $\beta = 1/2N$. Therefore, we have two threshold

values for $\beta : 1/2N$ and $1/3$ and consequently for studying the sign of (63) we need to consider the following cases: i) $\beta \in (0, \frac{1}{2N})$; ii) $\beta = \frac{1}{2N}$; iii) $\beta \in (\frac{1}{2N}, \frac{1}{3})$; iv) $\beta = \frac{1}{3}$; v) $\beta \in (\frac{1}{3}, \frac{1}{2})$.

First case $\beta \in (0, \frac{1}{2N})$

First, we investigate whether there exists a value for n that makes (63) zero. Reordering terms we obtain the following equation for n

$$\begin{aligned} & 2N(1-\beta)\beta^2 + (1-3\beta)(1-\beta)n + N\beta^2(2N(1-\beta)-1)n^{\frac{\beta}{1-\beta}} \\ &= N(\beta-1)^2(1-2N\beta)n^{\frac{1+\beta}{1-\beta}} + (3\beta^2-3\beta+1)n^{\frac{2-\beta}{1-\beta}} \\ & - (4N\beta^3 - (1+2N)\beta^2 - (2N-3)\beta + N-1)n^{\frac{1}{1-\beta}} - (1-\beta)^2(1-2N\beta)n^{\frac{2}{1-\beta}}. \end{aligned} \quad (64)$$

The LHS of this equation is an increasing concave function that takes a positive value equal to $2N(1-\beta)\beta^2$ for $n=0$. However, the RHS, that is equal to zero for $n=0$, could have an extreme given by the solution to the following equation

$$\begin{aligned} & N(1-\beta)(1-2N\beta)(1+\beta)n^{\frac{\beta}{1-\beta}} + (3\beta^2-3\beta+1)\frac{2-\beta}{1-\beta}n \\ &= \frac{1}{1-\beta}(4N\beta^3 - (1+2N)\beta^2 - (2N-3)\beta + N-1) + 2(1-\beta)(1-2N\beta)n^{\frac{1}{1-\beta}}. \end{aligned} \quad (65)$$

As the LHS of this equation is an increasing concave function equal to zero for $n=0$ and the RHS is an increasing convex function that takes a positive value for $n=0$ we could face three possibilities:

1) The convex function (the RHS) is above the concave function (the LHS) for all $n \geq 0$ and there does not exist an extreme for the RHS of (64). Then, the first derivative of the RHS of (64) is negative and the RHS of (64) is decreasing. As its initial value is zero, the RHS of (64) is negative for all n and therefore the LHS is larger than the RHS and (63) yields a contradiction.

2) The convex function (the RHS) is tangent to the concave function (the LHS). In this case, the first derivative of the RHS of (64) is negative and the RHS of (64) is decreasing except for one value of n that defines an inflection point. Again we obtain a contradiction.

3) The equation (65) has two solutions. In this case, the RHS of (64) presents first a minimum that yields a negative value and afterwards a maximum. Now, four subcases should be considered:

3a) The maximum value of the RHS of (64) is negative or zero, then the same argument that in the previous cases applies.

3b) The maximum value of the RHS of (64) is positive but the LHS of (64) is larger than the RHS for all n . Also in this case we obtain a contradiction for (63).

3c) The equation (64) has only one root that necessarily is on the left of the maximum of the RHS of (64). In this case (63) is positive except for a value of n that defines a inflection point in the increasing function of aggregate net benefits. On the left of this point, this function is concave.

3d) The equation (64) has two positive roots. On the left of the smaller root and on the right of the higher root (63) is positive. As has been established above that for $n = N$ (61) is positive, we need to find out if N is lower than the smaller root or higher than the larger root. If the slope of the RHS of (64) was positive for $n = N$ then N would be lower than the smaller root since this must be in the increasing section of the RHS of (64) on the left of the maximum and again (63) would yield a contradiction for $n \in (0, N)$.

Doing $n = N$ in the first derivative of the RHS of (64) and taking common factor we obtain the following expression

$$\frac{N^{\frac{\beta}{1-\beta}}}{1-\beta} \left(-(1-\beta)^3 N^{\frac{1}{1-\beta}} + 2(1-\beta)^3 \beta N^{\frac{2-\beta}{1-\beta}} + (1-\beta) (7\beta^2 - 4\beta + 1) N + \beta^2 - 3\beta + 1 \right), \quad (66)$$

where $7\beta^2 - 4\beta + 1$ and $\beta^2 - 3\beta + 1$ are positive for $\beta \in (0, 1/2N)$ and $N > 2$. Therefore, the first derivative of the RHS of (64) will be positive for $n = N$ if the parenthesis of (66) is positive. For determining the sign of the parenthesis, first we investigate whether there exists a value for N for which the parenthesis is zero studying whether the following equation has a solution

$$\beta^2 - 3\beta + 1 = (1-\beta)^3 N^{\frac{1}{1-\beta}} - 2(1-\beta)^3 \beta N^{\frac{2-\beta}{1-\beta}} - (1-\beta) (7\beta^2 - 4\beta + 1) N. \quad (67)$$

In this equation the LHS is a constant with respect to N and the RHS is zero for $N = 0$ and its first derivative with respect to N is

$$(1-\beta)^2 N^{\frac{\beta}{1-\beta}} - 2(1-\beta)^2 \beta (2-\beta) N^{\frac{1}{1-\beta}} - (1-\beta) (7\beta^2 - 4\beta + 1), \quad (68)$$

so that the the RHS of (67) could present an extreme if the following equation has a solution

$$(1-\beta)^2 N^{\frac{\beta}{1-\beta}} = 2(1-\beta)^2 \beta (2-\beta) N^{\frac{1}{1-\beta}} + (1-\beta) (7\beta^2 - 4\beta + 1). \quad (69)$$

As the LHS of this equation is an increasing concave function that takes a value equal to zero for $N = 0$ and the the RHS is an increasing convex function that takes a positive value for

$N = 0$ we could have that there does not exist a solution, that there is only one root or that there are two solutions.

If there does not exist a solution is because

$$(1 - \beta)^2 N^{\frac{\beta}{1-\beta}} < 2(1 - \beta)^2 \beta (2 - \beta) N^{\frac{1}{1-\beta}} + (1 - \beta) (7\beta^2 - 4\beta + 1)$$

for all N , and in this case (68) would be negative and the RHS of (67) would be a decreasing function taking negative values for all n . Thus, (66) would be positive for all N and the first derivative of the RHS of (64) would be positive for $n = N$ as we wanted to show.

If there exists only one solution again we would have that

$$(1 - \beta)^2 N^{\frac{\beta}{1-\beta}} < 2(1 - \beta)^2 \beta (2 - \beta) N^{\frac{1}{1-\beta}} + (1 - \beta) (7\beta^2 - 4\beta + 1)$$

for all N except for the value of N for which (69) is satisfied. Again the RHS of (67) would be a decreasing function taking negative values for all N with the unique difference with respect to the previous case that it will present an inflection point for the value of N given by the equation (69).

If there exist two roots we would have that

$$(1 - \beta)^2 N^{\frac{\beta}{1-\beta}} < 2(1 - \beta)^2 \beta (2 - \beta) N^{\frac{1}{1-\beta}} + (1 - \beta) (7\beta^2 - 4\beta + 1)$$

on the left of the smaller solution, that

$$(1 - \beta)^2 N^{\frac{\beta}{1-\beta}} > 2(1 - \beta)^2 \beta (2 - \beta) N^{\frac{1}{1-\beta}} + (1 - \beta) (7\beta^2 - 4\beta + 1)$$

between the smaller root and the higher root, and again that

$$(1 - \beta)^2 N^{\frac{\beta}{1-\beta}} < 2(1 - \beta)^2 \beta (2 - \beta) N^{\frac{1}{1-\beta}} + (1 - \beta) (7\beta^2 - 4\beta + 1)$$

on the right of the higher root so that we can conclude that the smaller root is a minimum and the higher root is a maximum for the RHS of (67). If the maximum gives a value for the RHS of (67) lower than the constant on the LHS of (67), the expression (66) will be positive for all N . On the other hand, if the maximum value for the RHS of (67) is larger than the constant on the LHS of (67), this equation will have two roots. In that case, on the left of the smaller root we will have that (66) is positive, that between the two roots (66) is negative and that on the right of the higher root it is positive again.

It is easy to check that for $N = 1$, (66) is positive. Now, we would have to determine if $N = 1$ is on the left of the smaller root or on the right of the higher root. Firstly, we calculate

the value of the RHS of (67) for $N = 1$ that gives us the negative value $2\beta^2 (\beta^2 - 1)$. Secondly, we calculate the slope of the RHS of (67) for $N = 1$. Substituting in (68) it is obtained that the slope is equal to $\beta(2\beta^3 - \beta^2 - 1)$ that takes negative values for $\beta \in (0, 1)$. Summarizing, for $N = 1$ the RHS of (67) and its slope are negative so that to place $N = 1$ on the left of the smaller root or on the right of the higher root we will have to find out the sign of the second derivative of the RHS of (67) to know whether $N = 1$ is in the initial convex section or the final concave section. The second derivative of the RHS of (67) is

$$(1 - \beta) \beta N^{-\frac{1-2\beta}{1-\beta}} - 2(1 - \beta)\beta(2 - \beta)N^{\frac{\beta}{1-\beta}},$$

that doing $N = 1$ yields $-\beta(2\beta^2 - 5\beta + 3)$ that takes a negative value for $\beta \in (0, 1)$. Therefore, the second derivative is negative and $N = 1$ is in the section of the function that is concave and decreasing and consequently on the right of the higher root of (67). Then (66) is positive for all $N \geq 1$ and N is lower than the smaller root given by the equation (64) and we obtain from (63) a contradiction. Thus, it can be concluded that (63) is positive in the interval $[0, N]$ that means that the aggregate net benefits increase with the participation for $\beta \in (0, \frac{1}{2N})$ and $n \in (1, N)$.

Second case $\beta = \frac{1}{2N}$.

Substituting β by $1/2N$ in (64) the following expression is obtained

$$\begin{aligned} & (2N - 1) + (2N - 1)(2N - 3)n + 2N(N - 1)n^{\frac{1}{2N-1}} \\ &= (4N^2 - 6N + 3)n^{\frac{4N-1}{2N-1}} - (4N^3 - 8N^2 + 4N + 1)n^{\frac{2N}{2N-1}}, \end{aligned} \quad (70)$$

where $4N^2 - 6N + 3$ and $4N^3 - 8N^2 + 4N + 1$ are positive for $N > 2$. The LHS is an increasing concave function that takes a positive value for $n = 0$. Whereas the RHS, that is equal to zero for $n = 0$, presents a minimum for

$$n = \frac{(4N^3 - 8N^2 + 4N + 1)2N}{(4N^2 - 6N + 3)(4N - 1)}$$

so that the RHS of (70) is an increasing convex function on the right of the minimum hence the equation (70) presents a unique positive root. On the right of this root, (63) is positive and on the left negative. As at the beginning of the proof, we have checked that the expression (61) is positive for $n = N$ we can conclude that the value for which (63) is zero is larger than N and hence that in the interval $(0, N)$, (63) is positive resulting in a contradiction. Thus, we

can state that (63) is positive in the interval $[0, N]$ and consequently the aggregate net benefits increase with the participation for $\beta = 1/2N$.

The proof that (63) yields a contradiction for $n \in (1, N)$ for the third and fourth cases is very similar to the one we have just presented for $\beta = 1/2N$. For this reason and with the aim of reducing the extension of the Appendix we omit the details.

Fifth case $\beta \in (\frac{1}{3}, \frac{1}{2})$.

Taking into account the changes in the coefficient signs in (64) the following expression is obtained

$$\begin{aligned}
& 2N(1-\beta)\beta^2 - n(3\beta-1)(1-\beta) + N\beta^2(2N(1-\beta)-1)n^{\frac{\beta}{1-\beta}} \\
& = -N(\beta-1)^2(2N\beta-1)n^{\frac{1+\beta}{1-\beta}} + (3\beta^2-3\beta+1)n^{\frac{2-\beta}{1-\beta}} \\
& -(4N\beta^3 - (1+2N)\beta^2 - (2N-3)\beta + N-1)n^{\frac{1}{1-\beta}} + (1-\beta)^2(2N\beta-1)n^{\frac{2}{1-\beta}}. \quad (71)
\end{aligned}$$

The LHS, that takes a positive value for $n = 0$, has a maximum for

$$n = \left(\frac{N\beta^3(2N(1-\beta)-1)}{(3\beta-1)(1-\beta)^2} \right)^{\frac{1-\beta}{1-2\beta}}.$$

On the other hand, it is easy to check that the RHS, that is equal to zero for $n = 0$, presents a minimum for a positive value of n . Calculating the first derivative of the RHS of (71) and taking common factor $n^{\frac{\beta}{1-\beta}}$ we obtain the following expression

$$\begin{aligned}
& n^{\frac{\beta}{1-\beta}} \left(-N(1-\beta)(2N\beta-1)(1+\beta)n^{\frac{\beta}{1-\beta}} + (3\beta^2-3\beta+1)\frac{2-\beta}{1-\beta}n \right. \\
& \left. - \frac{1}{1-\beta}(4N\beta^3 - (1+2N)\beta^2 - (2N-3)\beta + N-1) + 2(1-\beta)(2N\beta-1)n^{\frac{1}{1-\beta}} \right),
\end{aligned}$$

that is zero if the following equation has a solution

$$\begin{aligned}
& (3\beta^2-3\beta+1)\frac{2-\beta}{1-\beta}n + 2(1-\beta)(2N\beta-1)n^{\frac{1}{1-\beta}} \\
& = N(1-\beta)(2N\beta-1)(1+\beta)n^{\frac{\beta}{1-\beta}} + \frac{1}{1-\beta}(4N\beta^3 - (1+2N)\beta^2 - (2N-3)\beta + N-1).
\end{aligned}$$

The LHS of this equation, that is zero for $n = 0$, is an increasing convex function and the RHS, that also is zero for $n = 0$, is an increasing concave function so that the equation has only one root that defines a minimum. Then, the equation (71) has a unique root and on the left of this root, the LHS is larger than the RHS and therefore (63) is positive. As has already been shown

that for $n = N$ (63) is positive we obtain that the derivative of the aggregate net benefits with respect to n is also positive in the interval $(1, N)$ for $\beta \in (\frac{1}{3}, \frac{1}{2})$.

Summarizing, we have seen that for $\beta \in (0, 0.5)$, (63) is positive what yields a contradiction and it can be concluded that the aggregate net benefits increase with the number of signatories.