

# Sharing R&D Investments in Breakthrough Technologies to Control Climate Change\*

Abeer El-Sayed and Santiago J. Rubio<sup>†</sup>

Department of Economic Analysis and ERICES

University of Valencia, Spain

August 2011

---

\*A very preliminary version of the paper was presented at the 2nd Valencia-Louvain Workshop on Game Theory and Economic Behavior, Louvain-la-Neuve (Belgium), 28 May, 2010. The present version was presented at the Environmental Protection and Sustainability Forum, Exeter (United Kingdom), 13-15 April, 2011, the 18th Annual Conference of the European Association of Environmental and Resource Economists, Rome (Italy), 29 June - 2 July, 2011 and the 8th Game Theory Practice Meeting Dedicated to Global and International Issues, Riverside (California), 11-12 July, 2011. We are grateful to the participants in these conferences and in particular to Rafael Moner for their very useful comments. The authors acknowledge funding from the *Ministerio de Ciencia e Innovación* under grant ECO2010-21539.

<sup>†</sup>Corresponding author: Department of Economic Analysis, University of Valencia, Edificio Departamental Oriental, Avda. de los Naranjos s/n, 46022 Valencia, Spain. E-mail address: santiago.rubio@uv.es

## Abstract

This paper examines international cooperation on technological development as an alternative to international cooperation on greenhouse gas emission reductions. It is assumed that when countries cooperate they coordinate their R&D investments so as to minimize the agreement costs of controlling emissions and that they also pool their R&D efforts so as to fully internalize the spillover effects of their investments in R&D. In order to analyze the scope of cooperation, an agreement formation game is solved in three stages. First, countries decide whether or not to sign the agreement. Then, in the second stage, signatories (playing together) and non-signatories (playing individually) select their investment in R&D. Finally, in the third stage, each country decides its level of emissions non-cooperatively. For linear environmental damages, our findings show that the grand coalition is the only self-enforcing IEA if marginal damages are sufficiently significant. This occurs because when all countries share their R&D investments, if one country leaves the agreement it must face larger abatement and investment costs because of the sharp reduction in its effective investment caused by the exit. The result is that countries lose the incentive to act as free-riders within the agreement. If marginal damages are not large enough, the grand coalition can still be stable provided that the scope of R&D spillovers is not very great. Finally, the model for quadratic environmental damages is solved, finding the main results of the paper are robust to this change in the specification of environmental damages.

**Keywords:** International environmental agreements; R&D investment; Technology spillovers; Breakthrough technologies

**JEL Classification System:** D74, F53, H41, Q54, Q55

# 1 Introduction

Climate change is becoming an important issue in human lives. Due to the absence of a supra-national authority that can enforce environmental policies to control greenhouse gas (GHG) emissions on a global scale, countries have had to negotiate an international environmental agreement (IEA), the Kyoto Protocol, to address this problem. The aim of the Kyoto Protocol is to achieve a reduction in GHG emissions of 5% taking as reference the level of 1990 for countries of Annex B in the commitment period 2008-2012. However, there are many doubts about the possibilities of reaching the target of abating GHG emissions for that period. Limited coverage and moderate emission reductions requirements are two limitations that can reduce the effectiveness of the agreement. Moreover, as the last United Nations Climate Change Conferences in Copenhagen (2009) and Cancún (2010) have shown, there is increasing uncertainty about whether there will be any follow-up after 2012.

Because of the doubts about the effectiveness of the Kyoto Protocol, several scholars have asked whether other types of agreements can be designed to achieve large reductions of GHG emissions. One idea would be to focus on technology improvements in order to reduce abatement costs, as this might increase a country's willingness to undertake significant emission reductions. For example, it could be beneficial to supplement a Kyoto-type agreement with technology elements if technological development depends not only on a country's own R&D investment but also on R&D by other countries through cross-country technology spillovers, see for instance, Carraro and Siniscalco (1997). Even with no explicit agreement on emissions, a technology agreement leading to increased R&D in clean technologies, and thus to lower abatement costs, might yield a reduction in emissions. This is the argument behind the proposals of a climate agreement on technology development, see for instance, Barrett (2006).<sup>1</sup>

---

<sup>1</sup>There are several international proposals to promote climate-technology R&D, such as the Carbon Sequestration Leadership Forum (with 21 member countries plus the European Commission), the International Partnership for the Hydrogen Economy (17 countries plus the European Commission) and the ITER (International Thermonuclear Experimental Reactor) project, although the ITER project cannot see exclusively as a climate-technology R&D project. An overview of technology-oriented agreements

The aim of the present paper is to examine international cooperation on technological development as an alternative to international cooperation on GHG emission reductions. Cooperation on technological development may be designed in several ways. This paper follows the approach adopted by Kamien et al. (1992) in their analysis of the effects of R&D cartelization and research joint ventures on oligopolistic competition and assumes that when countries cooperate they coordinate their R&D activities so as to minimize the agreement costs of controlling emissions and they also share R&D investments and avoid duplication of R&D activities. In other words, when countries cooperate they pool their R&D efforts so as to fully internalize the spillover effects of their investment in R&D.<sup>2</sup>

In order to analyze these issues, the model proposed by Golombek and Hoel (2005) to analyze climate policy under technology spillovers is employed. In the model, abatement costs are assumed to depend both on the level of abatement and the technology level of the country and environmental damages are assumed to be linear. The possibility of GHG emissions being completely eliminated is also admitted, that is, fossil fuels could be completely replaced by other non-polluting energies. The complete elimination of GHG will require the development and diffusion of revolutionary, “breakthrough” technologies.<sup>3</sup>

We analyze the formation of an IEA as a three-stage game. In the first stage, countries decide on their participation in the agreement. Then, in the second stage, signatories select investment in R&D to minimize the total costs of the agreement and fully internalize the spillover effects of their investments, whereas non-signatories act unilaterally. Finally, in the third stage, each country decides its level of emissions non-cooperatively.

Our findings show that if marginal damages are sufficiently large, both signatories 

---

 stressing their potential role in addressing the free-riding incentives in climate negotiations can be found in de Coninck et al. (2008).

<sup>2</sup>The idea that the degree of spillovers is different among countries which cooperate than among countries which do not cooperate can be also found in Xepapadeas (1995) and Carraro and Siniscalco (1997). In Xepapadeas (1995), it is assumed that when all countries enter into an international agreement, the level of technology is common to all countries (it is a perfect public good). Carraro and Siniscalco (1997) normalize to zero the spillover effects for non-signatories. The present paper assumes that spillover effects are positive for non-signatories and are fully internalized for signatories.

<sup>3</sup>See Barrett (2009) for a survey of the possibilities of developing these kinds of technologies.

and non-signatories invest in R&D. Nevertheless, the grand coalition is the only self-enforcing IEA. The explanation for this result is given by the *asymmetry* introduced by the technology agreement between signatories and non-signatories. For the same global investment, signatories enjoy a greater level of effective investment than non-signatories, as the agreement implies that signatories fully internalize the spillover effects of their investments in R&D, whereas this is not the case for non-signatories. Thus, if one country abandons the grand coalition, its total costs increase, which makes the grand coalition stable. In other words, when marginal damages are sufficiently large and there is a complete abatement of emissions, the nature of the game changes and cooperation causes *negative spillovers* on non-signatories' total costs. When marginal damages are not sufficiently large, non-signatories leave to invest. However, the grand coalition may remain stable provided that the scope of R&D spillovers is not very great. If this condition is satisfied, a reduction in the global investment of signatories significantly increases the abatement costs of non-signatories making it unprofitable to exit the agreement. Finally, if marginal damages are too low, investment becomes unprofitable for both signatories and non-signatories and no technology agreement will be signed. In order to check the robustness of the results, the model for quadratic environmental damages is also solved, finding that the grand coalition is also the only self-enforcing IEA in this case if marginal damages are sufficiently large.

Our findings also show that when marginal damages are sufficiently large to justify the complete elimination of emissions, the efficient solution of the game can be achieved without cooperation in the third stage. In other words, a technology agreement with full participation is sufficient to implement the efficient solution. The intuition behind this result is that if the countries cooperate when they abate emissions, investments in R&D are going to be lower in the second stage. However, if they do not cooperate, investments are going to be greater but, on the other hand, abatement costs are lower. Then the sign of the difference in total costs for these two options: to cooperate or not to cooperate in the third stage depends on the level of marginal damages. If marginal damages are sufficiently large, the increase in abatement costs when countries cooperate is not compensated by the reduction in investment costs achieved in the second stage and the result is that total costs

are larger when countries cooperate in the third stage. Finally, we find that fully non-cooperative equilibrium investments in R&D exceed the investment levels corresponding to the efficient solution.<sup>4</sup> The reason explaining this result is that if countries only cooperate in the second stage, abatement is the same for the cooperative solution and the fully non-cooperative equilibrium but, as the spillover effects are completely internalized when countries cooperate to select investment, lower investments are required to eliminate the same level of emissions than in the fully non-cooperative equilibrium. Obviously, this yields lower total costs for the cooperative solution making inefficient the non-cooperative equilibrium.

Although the literature on IEAs is very extensive, only a few theoretical papers have addressed the issue studied in the present paper. The first paper worth commenting on is Carraro and Siniscalco (1997).<sup>5</sup> They employ a numerical example to show that one possible way of overcoming the free-ride incentive is to link the unstable emission agreement with an agreement on technological cooperation which is shown to be profitable and stable. This paper completes that analysis by showing analytically but in a simpler model that even with no explicit agreement on emissions, a technology agreement can promote cooperation and achieve substantial reductions in emissions. More recently, Barrett (2006) shows that breakthrough technologies cannot improve the performance of international environmental agreements with the exception of breakthrough technologies that exhibit increasing returns to scale.<sup>6</sup> We obtain the opposite result in the framework of a model that shares the same features. The reason that explains the difference in the results is that while Barrett (2006) assumes global investment in R&D to be a *perfect public good*, this paper assumes that some imperfections exist and introduces an asymmetry between signatories and non-signatories as regards the degree of spillovers. These changes are enough to reverse the nature of the game and yield different results

---

<sup>4</sup>This result is not new, it has been already shown by Heal and Tarui (2010) in the framework of a more general model.

<sup>5</sup>Xepapadeas (1995) does not address the issue of the stability of the IEA. He focuses on the comparison of the fully non-cooperative equilibrium to the internationally optimal solution.

<sup>6</sup>Ruis and de Zeeuw (2010) give support to this idea in the framework of a model with quadratic investment costs and without spillover effects.

from those derived by Barrett (2006). Hoel and de Zeeuw (2010) show that a focus on the R&D phase in the development of breakthrough technologies also changes the result obtained by Barrett (2006). Assuming that the cost of adoption decreases with respect to the level of R&D, they find that even without increasing returns to scale a technology agreement yields better results than those obtained by focusing on abatement targets, although the first best cannot be achieved. Finally, it is worth adding that different empirical papers give support to the idea that supplementing an emission agreement with technology elements or replacing an emission agreement with a technology agreement can have positive effects on the participation into the agreement. See for instance the papers written by Kemfert (2004), Buchner and Carraro (2005) and Lessman and Edenhofer (2011).<sup>7</sup>

The paper is organized as follows. The next section specifies the model. In Section 3, the fully non-cooperative equilibrium is calculated and in Section 4 the efficient outcome. Section 5 presents the analysis of the technology agreement. The analysis of the technology agreement for quadratic environmental damages is presented in Section 6. The conclusions drawn from this research are detailed in Section 7.

## 2 The Model

We develop a static model with  $N$  countries that pollute the atmosphere and negotiate the control of greenhouse gas (GHG) emissions, taking into account the effects of spillovers in R&D from one country to another. The model is based on Golombek and Hoel (2005). It is assumed that the effective investment in a country  $i$ ,  $y_i$ ,  $i = 1, \dots, N$ , depends on the amount invested in R&D in that country,  $x_i$ , and also the investments in R&D undertaken in all other countries. However, technological diffusion is not perfect, only part of the R&D investments undertaken in other countries is beneficial for country  $i$ . Hence, the

---

<sup>7</sup>Nagashima and Dellink (2008) address the effects of asymmetric spillovers, that affect the marginal abatement cost curve, on the participation in an emission agreement. Their results show that spillovers do not substantially increase the success of IEAs. However, in their model the size of the spillovers cannot be controlled by the signatories, as their state of technology is exogenous.

effective investment of country  $i$  is given by

$$y_i = x_i + \gamma X_{-i}, \quad \gamma \in [0, 1], \quad (1)$$

where  $X_{-i} = \sum_{j \neq i} x_j$ . This specification for effective R&D investment was introduced by Spence (1984) and has been recently used by Golombek and Hoel (2005, 2008, 2011) in the analysis of climate policy under technology spillovers.<sup>8</sup> However, following the approach adopted by Kamien et al. (1992) in their analysis of the effects of R&D cartelization and research joint ventures on oligopolistic competition, it is assumed that when countries cooperate they pool their R&D efforts so as to fully internalize spillover effects, which implies that in this case we will assume that  $\gamma = 1$  for signatories.

In the absence of any explicit abatement activities, emissions in each country depend only on the technology level of the country. So, the business as usual emissions (BAU) for a level of effective investment equal to  $y_i$  is defined as  $\bar{E}(y_i) = \delta - \alpha y_i$ , with  $\delta, \alpha > 0$ , and  $\alpha$  representing emission abatement per each unit invested in clean technologies. According to that, we can define the abatement of country  $i$  as  $A_i = \bar{E}(y_i) - E_i = \delta - \alpha y_i - E_i$  where  $E_i$  stands for the current emissions generated by country  $i$ . Thus, abatement costs depend both on the level of abatement and the level of effective investment. Effective R&D investment reduces abatement costs because it reduces the intensity of emissions in the production of goods and services for a country. The greater the effective R&D investment, the lower the ratio of GHG emissions over the GDP of the country and, consequently, the lower the abatement costs. For this specification, there is a critical value for  $y_i$  given by  $\delta/\alpha$  for which GHG emissions are completely eliminated, in other words, fossil fuels could be completely substituted by other non-polluting energies. It is assumed that abatement costs are quadratic

$$C(A_i) = \frac{c}{2} A_i^2 = \frac{c}{2} (\delta - \alpha y_i - E_i)^2, \quad c > 0. \quad (2)$$

Following Golombek and Hoel (2005), the price of R&D investments is normalized to one and investment is irreversible. So, the cost of investing in R&D is  $R(x_i) = x_i$ .

---

<sup>8</sup>Golombek and Hoel (2005, 2008, 2011) analyze the effects on R&D investments, emissions and welfare of different types of agreements implemented by different types of instruments, including a technology agreement implemented by a subsidy, but they do not study the stability of the agreements.

Finally, in each country environmental damages depend on global emissions,  $E = \sum_{i=1}^N E_i$ . Environmental damages are assumed to be linear:  $D(E) = dE$ ,  $d > 0$ . Thus, the total costs of controlling GHG emissions for the representative country can be written as follows

$$TC_i = \frac{c}{2}(\delta - \alpha y_i - E_i)^2 + dE + x_i, \quad (3)$$

where  $y_i = x_i + \gamma X_{-i}$ , with  $\gamma \in [0, 1]$  and  $E = \sum_{i=1}^N E_i$ .

### 3 Fully Non-Cooperative Equilibrium

The fully non-cooperative equilibrium can be calculated as the equilibrium of a two-stage game. In the first stage, countries decide the level of investment in R&D. In the second stage they decide about emissions. In both stages, the Nash equilibrium is calculated. Solving by backward induction, we begin analyzing the equilibrium of the second stage.

For a given technology, optimal emissions can be calculated by minimizing the following total cost function

$$TC_i = \frac{c}{2}(\delta - \alpha y_i - E_i)^2 + dE, \quad i = 1, \dots, N,$$

which yields for the representative country<sup>9</sup>

$$E_i = \bar{\delta} - \alpha y_i. \quad (4)$$

As emissions must be non-negative, the following upper bound must be imposed on effective investment

$$\frac{\bar{\delta}}{\alpha} \geq y_i = x_i + \gamma X_{-i},$$

which can be expressed as a constraint on investment

$$\frac{\bar{\delta}}{\alpha} - \gamma X_{-i} \geq x_i. \quad (5)$$

---

<sup>9</sup>In order to simplify the notation,  $\bar{\delta}$  will be used to represent the difference  $\delta - (d/c)$ . It is assumed that  $\delta$  is high enough to have positive emissions when the effective investment is zero both in the fully non-cooperative equilibrium and the efficient solution.

Using (4), global emissions can be calculated as

$$E = \sum_{i=1}^N E_i = N\bar{\delta} - \alpha Y, \quad (6)$$

where  $Y$  is global effective investment in R&D.

$$Y = \sum_{i=1}^N y_i = \sum_{i=1}^N (x_i + \gamma X_{-i}). \quad (7)$$

Next, using (4) and (6), total costs can be written as

$$TC_i = \frac{d^2}{2c} + dN\bar{\delta} - \alpha dY. \quad (8)$$

Observe that global effective investment in R&D becomes a *public good*. Any investment made by a country reduces the total costs of all countries.

Now we calculate the equilibrium for the first stage. As the cost of investing in R&D is linear, there is a *linear programming problem* defined for the representative country as follows

$$\min_{\{x_i\}} TC_i = \frac{d^2}{2c} + dN\bar{\delta} - \alpha dY + x_i, \quad (9)$$

$$s.t. \frac{\bar{\delta}}{\alpha} - \gamma X_{-i} \geq x_i, \quad (10)$$

$$x_i \geq 0, \quad (11)$$

where  $Y$  is given by (7).

The first-derivative of the total costs is

$$\frac{\partial TC_i}{\partial x_i} = -\alpha d \frac{\partial Y}{\partial x_i} + 1,$$

where  $\partial Y / \partial x_i = 1 + \gamma(N - 1)$ , so that

$$\frac{\partial TC_i}{\partial x_i} = -\alpha d(1 + \gamma(N - 1)) + 1.$$

Thus, there exists a threshold for  $d$  defined by condition  $\partial TC_i / \partial x_i = 0$  equal to

$$\hat{d}^{nc} = \frac{1}{\alpha(1 + \gamma(N - 1))}. \quad (12)$$

Then, when  $d > \hat{d}^{nc}$  total costs are *decreasing* with respect to  $x_i$ , and the countries invest in R&D until emissions are completely eliminated. In this case, constraint (10) is satisfied as an equality and the symmetric equilibrium yields

$$x_i^{nc} = \frac{\bar{\delta}}{\alpha(1 + \gamma(N - 1))}, \quad y_i^{nc} = \frac{\bar{\delta}}{\alpha}. \quad (13)$$

Finally, substituting in the total costs function, the following expression is obtained

$$TC_i^{nc} = \frac{d^2}{2c} + \frac{\bar{\delta}}{\alpha(1 + \gamma(N - 1))}. \quad (14)$$

Observe that although effective investment is independent of the technology diffusion parameter, both investment in R&D and total costs decrease as technology spillovers increase.

However, if  $d \leq \hat{d}^{nc}$ , total costs are independent or *increasing* with respect to  $x_i^{nc}$  and the optimal policy is not to invest. Thus, emissions are  $E_i^{nc} = \bar{\delta}$  and total costs are given by

$$TC_i^{nc} = \frac{d^2}{2c} + dN\bar{\delta}, \quad (15)$$

which corresponds to the outcome of the standard model of emissions abatement for a given technology.

## 4 The Efficient Solution

In order to characterize the efficient solution, the game is solved again in two stages, but on this occasion assuming that countries minimize global total costs in both stages. We begin analyzing the solution of the second stage. Given the technology, countries select emissions to minimize global total costs given by

$$GTC = \sum_{i=1}^N TC_i = \sum_{i=1}^N \left( \frac{c}{2} (\delta - \alpha y_i - E_i)^2 + dE \right).$$

The solution to the optimization problem is

$$E_i = \delta - \alpha y_i - \frac{Nd}{c}, \quad i = 1, \dots, N. \quad (16)$$

As emissions must be non-negative, an upper bound must be imposed on effective investment

$$\frac{1}{\alpha} \left( \delta - \frac{Nd}{c} \right) \geq y_i = x_i + X_{-i},$$

which can be expressed as a constraint on investment<sup>10</sup>

$$\frac{1}{\alpha} \left( \delta - \frac{Nd}{c} \right) - X_{-i} \geq x_i. \quad (17)$$

Using (16), total emissions can be calculated

$$E = \sum_{i=1}^N E_i = N \left( \delta - \frac{Nd}{c} \right) - \alpha Y, \quad (18)$$

where  $Y$  is global effective investment in R&D.

$$Y = \sum_{i=1}^N y_i = \sum_{i=1}^N (x_i + X_{-i}). \quad (19)$$

Using (16) and (18), total costs for the representative country can be written as

$$TC_i = \frac{d^2 N^2}{2c} + dN \left( \delta - \frac{Nd}{c} \right) - \alpha dY.$$

Next, in the first stage, countries select the level of investment to minimize the global total costs of controlling emissions that are given by the following expression

$$\begin{aligned} GTC &= \sum_{i=1}^N TC_i = \sum_{i=1}^N \left( \frac{d^2 N^2}{2c} + dN \left( \delta - \frac{Nd}{c} \right) - \alpha dY + x_i \right) \\ &= \frac{d^2 N^3}{2c} + dN^2 \left( \delta - \frac{Nd}{c} \right) - \alpha dNY + \sum_{i=1}^N x_i. \end{aligned} \quad (20)$$

As the cost of investing in R&D is linear, the choice problem of the countries can be represented as well for the efficient solution as a *linear programming problem*.

$$\min_{\{x_1, \dots, x_N\}} GTC = \frac{d^2 N^3}{2c} + dN^2 \left( \delta - \frac{Nd}{c} \right) - \alpha dNY + \sum_{i=1}^N x_i, \quad (21)$$

$$s.t. \quad \frac{1}{\alpha} \left( \delta - \frac{Nd}{c} \right) - X_{-i} \geq x_i, \quad (22)$$

$$x_i \geq 0, \quad i = 1, \dots, N, \quad (23)$$

---

<sup>10</sup>We assume that when countries cooperate they pool their R&D investment so as to fully internalize the spillover effects, i.e.  $\gamma = 1$  for the efficient solution.

where  $Y$  is given by (19).

The first-derivative of global total costs for the representative country is

$$\frac{\partial GTC}{\partial x_i} = -\alpha d N \frac{\partial Y}{\partial x_i} + 1,$$

where  $\partial Y/\partial x_i = N$ , so that

$$\frac{\partial GTC}{\partial x_i} = -\alpha d N^2 + 1.$$

Thus,  $\partial GTC/\partial x_i = 0$  defines a threshold for  $d$  equal to

$$\hat{d}^e = \frac{1}{\alpha N^2}. \quad (24)$$

Then, when  $d > \hat{d}^e$  total costs are *decreasing* with respect to investment, and the countries invest in R&D until emissions are completely eliminated, as was the case for the non-cooperative equilibrium, but for different values of marginal damages. In this case, constraint (22) is satisfied as equality and the symmetric equilibrium yields

$$x_i^e = \frac{1}{\alpha N} \left( \delta - \frac{Nd}{c} \right), \quad (25)$$

and

$$y_i^e = N x_i^e = \frac{1}{\alpha} \left( \delta - \frac{Nd}{c} \right). \quad (26)$$

Finally, substituting in the total costs function, the following expression is obtained

$$TC_i^e = \frac{d^2 N^2}{2c} + \frac{1}{\alpha N} \left( \delta - \frac{Nd}{c} \right). \quad (27)$$

For  $d \leq \hat{d}^e$ , the total costs are independent or *increasing* with respect to  $x_i$  and the optimal policy is not to invest. Thus emissions are  $E_i^e = \delta - Nd/c$  and total costs are

$$TC_i^e = \frac{d^2 N^2}{2c} + Nd \left( \delta - \frac{Nd}{c} \right). \quad (28)$$

However, it is easy to show that cooperation in the second stage is inefficient when countries invest in the first stage. When there is no cooperation in the second stage, total costs are given by (8), except that  $Y = \sum_{i=1}^N (x_i + X_{-i})$  as countries cooperate in the first stage.

Thus, in the first stage countries select the level of investment to minimize the global total costs of controlling emissions that are given by the following expression

$$\begin{aligned} GTC &= \sum_{i=1}^N TC_i = \sum_{i=1}^N \left( \frac{d^2}{2c} + d(N\bar{\delta} - \alpha Y) + x_i \right) \\ &= \frac{d^2 N}{2c} + dN^2\bar{\delta} - \alpha dNY + \sum_{i=1}^N x_i. \end{aligned}$$

As the cost of investing in R&D is linear, the choice of the countries can be represented again by a *linear programming problem*.

$$\min_{\{x_1, \dots, x_N\}} GTC = \frac{d^2 N}{2c} + dN^2\bar{\delta} - \alpha dNY + \sum_{i=1}^N x_i, \quad (29)$$

$$s.t. \quad \frac{\bar{\delta}}{\alpha} - X_{-i} \geq x_i, \quad (30)$$

$$x_i \geq 0, \quad i = 1, \dots, N, \quad (31)$$

where  $Y$  is given by (19). If the last two terms of global total costs are compared to the last two terms of expression (21), they can be seen to be identical and, therefore, the value for  $d$  which triggers investment in R&D is the same both when countries cooperate in both stages and when they cooperate only in the first stage. However, when countries only cooperate in the first stage, investment is given by (30):

$$x_i = \frac{\bar{\delta}}{\alpha} - X_{-i},$$

and the symmetric equilibrium yields

$$x_i^e = \frac{\bar{\delta}}{\alpha N}, \quad y_i^e = Nx_i^e = \frac{\bar{\delta}}{\alpha}. \quad (32)$$

Finally, substituting in the total costs function, the following expression is obtained

$$TC_i^e = \frac{d^2}{2c} + \frac{\bar{\delta}}{\alpha N}. \quad (33)$$

Now, in order to show that cooperation in both stages is not efficient, the total costs of the representative country when countries cooperate in both stages are compared to the

total costs when they cooperate only in the first stage, yielding the following difference<sup>11</sup>

$$\begin{aligned} TC_i^{2e} - TC_i^{1e} &= \frac{d^2 N^2}{2c} + \frac{1}{\alpha N} \left( \delta - \frac{Nd}{c} \right) - \left( \frac{d^2}{2c} + \frac{\bar{\delta}}{\alpha N} \right) \\ &= \frac{d}{c} \left( \frac{d}{2} (N^2 - 1) - \frac{1}{\alpha N} (N - 1) \right). \end{aligned}$$

Both costs are identical when

$$d = \frac{2}{\alpha N} \frac{N - 1}{N^2 - 1}.$$

But  $d$  must be greater than  $\hat{d}^e = 1/\alpha N^2$  to have a positive investment. Comparing these two critical values yields

$$\frac{1}{\alpha N^2} - \frac{2}{\alpha N} \frac{N - 1}{N^2 - 1} = \frac{(N - 1)^2}{\alpha N^2 (N^2 - 1)} > 0.$$

Then, we have that when  $d$  is greater than  $\hat{d}^e = 1/\alpha N^2$  is also greater than  $2(N - 1)/\alpha N(N^2 - 1)$  and consequently  $TC_i^{2e}$  is greater than  $TC_i^{1e}$ , and we can conclude that

**Proposition 1** *Cooperation in the second stage is not efficient when countries invest in R&D.*

The intuition of this result is that when countries cooperate in the second stage they reduce emissions, for a given technology, in an amount greater than if they do not cooperate so that in the first stage, investment is lower but not enough to compensate for the difference in abatement costs faced in the second stage.

Next, the non-cooperative equilibrium is compared to the efficient outcome. The first step of this comparison is to ascertain the relationship between the thresholds that trigger the investment in R&D in each case. Calculating the difference between  $\hat{d}^{nc}$  and  $\hat{d}^e$  we obtain

$$\hat{d}^{nc} - \hat{d}^e = \frac{1}{\alpha} \left( \frac{(N - 1)(N + 1 - \gamma)}{N^2(1 + \gamma(N - 1))} \right) > 0.$$

Thus, when countries cooperate in the adoption of a new technology, they invest for lower marginal damages than those for which they invest when they do not cooperate. In the

---

<sup>11</sup>Where  $TC_i^{2e}$  stands for the total costs of the representative country when countries cooperate in both stages and  $TC_i^{1e}$  denotes the total costs of the representative country when countries cooperate only in the first stage.

second step of this comparison, it is assumed that  $d$  is greater than  $\hat{d}^{nc}$  and the countries' investments are compared in both cases. Contrary to conventional wisdom, fully non-cooperative equilibrium investments in R&D exceed the investment levels corresponding to the efficient solution

$$x_i^{nc} - x_i^e = \frac{\bar{\delta}}{\alpha} \frac{(1 - \gamma)(N - 1)}{N(1 + \gamma(N - 1))} > 0,$$

although the effective investment is the same in both cases. The reason explaining this result is that if countries does not cooperate in the second stage, abatement is the same than in the fully non-cooperative equilibrium but, as the spillover effects are fully internalize when the countries cooperate in the first stage, lower investment efforts are required to completely eliminate the same level of emissions. Consequently, the difference in total costs of eliminating emissions is given by the difference in investment costs obtained above as there are not differences in abatement costs.<sup>12</sup>

For  $d$  lower than or equal to  $\hat{d}^{nc}$  but higher than  $\hat{d}^e$ , the countries do not invest in R&D when they do not cooperate and emissions are positive. Nevertheless, the costs of controlling emissions again are greater than when they cooperate. For this case the difference in costs can be calculated from (15) and (33)

$$TC_i^{nc} - TC_i^e = \frac{d^2}{2c} + dN\bar{\delta} - \left( \frac{d^2}{2c} + \frac{\bar{\delta}}{\alpha N} \right) = \bar{\delta} \left( \frac{\alpha d N^2 - 1}{\alpha N} \right),$$

which is positive for  $d > \hat{d}^e = 1/\alpha N^2$ . Now, when the countries cooperate they support larger investment costs but environmental damages are completely eliminated. Thus, if marginal damages are sufficiently large, the total costs of controlling emissions are lower than when they do not cooperate. Thus, we can conclude

**Proposition 2** *When countries cooperate in the adoption of a new technology, they invest for lower marginal damages than those for which they invest when they do not cooperate, and they eliminate emissions incurring lower costs.*

Finally, when  $d$  is lower than or equal to  $\hat{d}^e$ , the cooperative outcome is not to invest in R&D and the cooperation in the first stage does not yield any difference with respect to

---

<sup>12</sup>Notice that the price of R&D investments is normalized to one.

the fully non-cooperative equilibrium. Now, cooperation in the second stage is required to obtain the efficient solution. For this case, our model becomes the standard model of pollution control through emission abatement for a given technology and the fully non-cooperative equilibrium becomes inefficient.

## 5 The Technology Agreement

We say that a technology agreement is formed if the countries pool their R&D investments so as to fully internalize the spillover effects and they select the level of investment in order to minimize the total costs of the agreement. Then, if all countries are in the agreement, the effective investment in R&D is given by

$$y_i = X = \sum_{j=1}^N x_j, \quad i = 1, \dots, N.$$

However, if full cooperation is not achieved, the effective investment of signatories is given by

$$y_j^s = X^s + \gamma X^f = \sum_{k=1}^n x_k^s + \gamma \left( \sum_{l=1}^{N-n} x_l^f \right), \quad j = 1, \dots, n, \quad (34)$$

where  $n$  stands for the number of signatories,  $s$  for a signatory country and  $f$  for a non-signatory country. With partial cooperation, there is partial internalization of spillover effects. As non-signatories do not cooperate, their effective investment is given by (1) but now this expression can be written as follows

$$y_i^f = x_i^f + \gamma(X^s + X_{-i}^f) = x_i^f + \gamma \left( \sum_{k=1}^n x_k^s + \sum_{l=1}^{N-n-1} x_l^f \right), \quad i = 1, \dots, N - n. \quad (35)$$

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

Given the level of participation in the agreement and the investment in R&D of all countries, at the third stage, the emission game, each country simultaneously selects its own emissions acting non-cooperatively and taking the emissions of all other countries as

given, i.e. there is assumed to be no cooperation as regards the selection of the level of emissions. At the second stage, the R&D investment game, signatory countries coordinate their R&D activities so as to minimize the sum of agreement costs taking as given the R&D investments of non-signatories. As we have just pointed out, signing a technology agreement implies that countries share their R&D investments so as to fully internalize spillover effects, so that in this case the effective investment for signatories is given by (34). Non-signatories choose their investment in R&D acting non-cooperatively and taking the investments of all other countries as given in order to minimize their own costs of controlling emissions. Signatories and non-signatories choose their R&D investment simultaneously. Thus, R&D investments are provided by the *partial agreement Nash equilibrium* with respect to a coalition defined by Chander and Tulkens (1995). Finally, it is assumed at the first stage 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 and the agreement is formed by all players who have chosen to sign. As usual, the level of participation in the agreement is given by the stability conditions. Under open membership, any country is free to join the agreement if interested. Finally, it is assumed that the signing of the agreement is binding on signatories. They therefore acquire a commitment to stay and implement the agreement during the second stage of the game so that full compliance is achieved. The game finishes when the emission subgame is over.

## 5.1 The Partial Agreement Nash Equilibrium of the Investment Game

In this section, stages two and three are solved by backward induction assuming that in the first stage  $n$  countries, with  $n \geq 2$ , have signed the agreement. As it is assumed there is no cooperation in the emission game, the total costs supported by all countries are given by (8), except that now the global effective investment in R&D is given by

$$Y = \sum_{i=1}^{N-n} y_i^f + \sum_{j=1}^n y_j^s = \sum_{i=1}^{N-n} (x_i^f + \gamma X_{-i}) + \sum_{j=1}^n (X^s + \gamma X^f). \quad (36)$$

Next, the partial agreement Nash equilibrium of the investment game is calculated. As non-signatories do not cooperate at this stage, the analysis is identical to that performed in Section 3. Thus, non-signatories will invest in R&D provided that marginal damages are greater than  $\hat{d}^{nc}$ . For signatories, the choice made by the countries can be represented by a linear programming problem as (29)-(31) except that  $N$  must substitute by  $n$  and the upper bound on investment is now given by

$$\frac{\bar{\delta}}{\alpha} - X_{-j}^s - \gamma X^f \geq x_j^s, \quad j = 1, \dots, n. \quad (37)$$

Then, the first derivative of the agreement total costs is

$$\frac{\partial TC_A}{\partial x_j^s} = -\alpha dn \frac{\partial Y}{\partial x_j^s} + 1,$$

where  $\partial Y / \partial x_j^s = n + \gamma(N - n)$ , so that

$$\frac{\partial TC_A}{\partial x_j^s} = -\alpha dn (n + \gamma(N - n)) + 1.$$

Thus,  $\partial TC_A / \partial x_j^s = 0$  defines a threshold for  $d$  equal to

$$\hat{d}^s(n) = \frac{1}{\alpha n (n + \gamma(N - n))}.$$

Comparing this expression to that obtained for the efficient solution, we now have that the critical value of  $d$  which triggers investment in R&D depends on the number of signatories. It is easy to check that  $\hat{d}^s(n)$  decreases with respect to the number of signatories and takes values between<sup>13</sup>

$$\hat{d}^s(N) = \frac{1}{\alpha N^2} \leq \hat{d}^s(n) \leq \hat{d}^s(2) = \frac{1}{2\alpha(2 + \gamma(N - 2))}. \quad (38)$$

Then, when  $d$  is greater than  $\hat{d}^s(2)$  the total costs of the agreement are *decreasing* with respect to investment independently of the level of participation, and the signatories invest in R&D until emissions are completely eliminated. However, when  $d$  is in the interval  $(\hat{d}^s(N), \hat{d}^s(2)]$ , the total costs of the agreement are decreasing depending on the

---

<sup>13</sup>Notice that  $\hat{d}^s(N) = \hat{d}^e$ , the critical value of  $d$  for the efficient solution and  $\hat{d}^s(1) = \hat{d}^{nc}$ , the critical value of  $d$  for the fully non-cooperative equilibrium. So that the values for  $\hat{d}^s(n)$  are defined in the interval  $[\hat{d}^e, \hat{d}^{nc}]$ .

number of signatories. In this case, for a given value of  $d$ , it is necessary a minimum of cooperation, given by  $d = \hat{d}^s(n)$ , to make the investment in R&D profitable. If this is not the case, the signatories do not invest in clean technologies and cooperation is not enough to promote the replacement of fossil fuels. The condition  $d = \hat{d}^s(n)$  yields the following second-degree equation for  $n$

$$d\alpha(1 - \gamma)\hat{n}^2 + d\alpha N\hat{n} - 1 = 0, \quad (39)$$

which has two real roots, one is positive and the other negative. The critical value  $\hat{n}$  is defined by the positive root. Thus, given  $d$  in the interval  $(\hat{d}^s(N), \hat{d}^s(2)]$ , participation in the agreement must be at least equal to the lowest natural number on the right of  $\hat{n}$  to make it profitable for signatories to invest in R&D. Moreover, as  $\hat{d}^s(n)$  is a decreasing function with respect to  $n$ , the minimum level of participation decreases as marginal damages increase.

As we have just seen, the decision on investing in R&D for both signatories and non-signatories depends critically on the value of marginal environmental damages. Taking into account this result, the optimal decision of the countries can be characterized. If damages are sufficiently large, in particular, if  $\hat{d}^{nc} < d$ , both signatories and non-signatories invest in R&D. If  $d$  belongs to the interval  $(\hat{d}^s(2), \hat{d}^{nc}]$ , only the signatories invest in R&D regardless of the number of signatories. However, if  $d$  belongs to the interval  $(\hat{d}^s(N), \hat{d}^s(2)]$ , signatories will invest in R&D provided that participation is greater than the critical value given by (39). Finally, if  $d$  is equal to or lower than  $\hat{d}^s(N)$ , both signatories and non-signatories will not invest. Figure 1 shows whether the investment is positive or zero, for signatories and non-signatories, depending on marginal damages and the number of signatories.

⇒ FIGURE 1 ⇐

Next, the level of investment is calculated for different values of  $d$ . In fact, three possibilities must be considered. First, both signatories and non-signatories invest in R&D. Second, only signatories invest. Third, both signatories and non-signatories do not find it profitable to invest in R&D.

When  $d > \hat{d}^{nc}$ , both signatories and non-signatories find it profitable to invest in R&D and countries will invest until emissions are zero, i.e. until the upper bounds on investment are reached, which yields

$$x_i^f = \frac{\bar{\delta}}{\alpha} - \gamma X_{-i}, \quad x_j^s = \frac{\bar{\delta}}{\alpha} - X_{-j}^s - \gamma X^f.$$

Assuming symmetry,  $X_{-i}$ ,  $X_{-j}^s$  and  $X^f$  are written as

$$X_{-i} = (N - n - 1)x_i^f + nx_j^s, \quad X_{-j}^s = (n - 1)x_j^s, \quad X^f = (N - n)x_i^f,$$

and we obtain the following pair of reaction functions

$$x_i^f = \frac{\frac{\bar{\delta}}{\alpha} - n\gamma x_j^s}{1 + \gamma(N - n - 1)}, \quad (40)$$

$$x_j^s = \frac{1}{n} \left( \frac{\bar{\delta}}{\alpha} - \gamma(N - n)x_i^f \right), \quad (41)$$

which establishes that investments in R&D are *strategic substitutes*.

The solution to the previous system yields the optimal level of investment for signatories and non-signatories

$$x_i^f = \frac{\bar{\delta}}{\alpha(1 + \gamma(N - n))}, \quad (42)$$

$$x_j^s = \frac{x_i^f}{n} = \frac{\bar{\delta}}{\alpha n(1 + \gamma(N - n))}. \quad (43)$$

Notice that signatories' investment is always lower than non-signatories' investment. Moreover, for non-signatories, investment increases as participation increases. However, for signatories it depends on the number of signatories and the scope of the spillovers. Nevertheless, global investment decreases with the number of signatories because full internalization of R&D investment provides signatories with a larger effective investment for the same level of R&D investment.

As the two types of countries invest until emissions are zero, the effective investment is the same for both types of countries

$$y_i^f = y_j^s = \frac{\bar{\delta}}{\alpha}, \quad Y = \frac{N\bar{\delta}}{\alpha}.$$

Now by substituting the effective investments in the total cost functions, we obtain the following expressions for total costs

$$TC_i^f = \frac{d^2}{2c} + \frac{\bar{\delta}}{\alpha(1 + \gamma(N - n))}, \quad i = 1, \dots, N - n, \quad (44)$$

$$TC_j^s = \frac{d^2}{2c} + \frac{\bar{\delta}}{\alpha n(1 + \gamma(N - n))}, \quad j = 1, \dots, n. \quad (45)$$

Observe, that the signatories' total costs are always lower than the non-signatories' total costs and that there are *negative* spillovers for non-signatories stemming from cooperation, i.e. cooperation increases the cost of non-signatories. Nevertheless, global total costs decrease as cooperation increases.

Next, we analyze the case where marginal damages are low enough to make it unprofitable for non-signatories to invest in R&D.<sup>14</sup> Using the reaction function (41), the optimal investment levels for signatories are

$$x_j^s = \frac{\bar{\delta}}{\alpha n}, \quad y_j^s = X^s = \frac{\bar{\delta}}{\alpha}. \quad (46)$$

The signatories completely eliminate emissions, whereas the non-signatories do not. The effective investment for them is

$$y_i^f = \gamma X^s = \frac{\gamma \bar{\delta}}{\alpha},$$

and emissions are

$$E_i^f = (1 - \gamma)\bar{\delta}.$$

Notice that the greater the technology spillovers, the lower the non-signatories' emissions.

Adding the effective investment for signatories and non-signatories, we obtain that global effective investment is

$$Y = (N - n)y_i^f + ny_j^s = \frac{\bar{\delta}}{\alpha} (n + \gamma(N - n)).$$

Finally, adding the emissions for non-signatories, we obtain that global emissions decrease with cooperation

$$E = (N - n)(1 - \gamma)\bar{\delta}.$$

---

<sup>14</sup>When marginal damages are so low that both signatories and non-signatories find investing in R&D unprofitable, or when the level of participation does not reach the minimum required to justify investment by signatories, the outcome of the game is the fully non-cooperative equilibrium studied in Section 3.

Now by substituting effective investment and emissions in the total cost functions we obtain the following expressions

$$TC_i^f = \frac{d^2}{2c} + d(N-n)(1-\gamma)\bar{\delta}, \quad i = 1, \dots, N-n, \quad (47)$$

$$TC_j^s = \frac{d^2}{2c} + d(N-n)(1-\gamma)\bar{\delta} + \frac{1}{\alpha n}\bar{\delta}, \quad j = 1, \dots, n. \quad (48)$$

Contrary to the previous case, the total costs of signatories are greater than the total costs of non-signatories regardless of the level of cooperation. Moreover, *positive* spillovers now stem from cooperation. This difference in the sign of spillovers from cooperation explains, as we will see in the next section, the different results as regards participation in a self-enforcing agreement depending on the level of marginal damages.

## 5.2 The Nash Equilibrium of the Membership Game

In this section, we use stability conditions to investigate whether a self-enforcing technology agreement exists. First, we present the definition of coalitional stability from d'Aspremont et al. (1983), which has been extensively used in the literature on international environmental agreements.

**Definition 1** *An agreement consisting of  $n$  signatories is self-enforcing if  $TC_j^s(n) \leq TC_i^f(n-1)$  for  $j = 1, \dots, n$  and  $TC_i^f(n) \leq TC_j^s(n+1)$  for  $i = 1, \dots, N-n$ .*

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 decisions. 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 decisions. To check the stability conditions the auxiliary function  $\Omega(n) = TC_j^s(n) - TC_i^f(n-1)$  is used. If  $\Omega(n) = 0$  has a unique positive solution and  $\Omega(n)$  is increasing around this positive solution, then there is a self-enforcing agreement given by the greatest natural number on the left of the positive solution to equation

$\Omega(n) = 0$  provided that this number is equal to or lower than  $N$ . If we represent this number by  $\tilde{n}$ , we have that  $\Omega(\tilde{n})$  is negative and the internal stability condition is satisfied. Moreover, as  $\Omega(n)$  is an increasing function,  $\Omega(\tilde{n} + 1)$ , where  $\tilde{n} + 1$  is the lowest natural number on the right of the positive solution to equation  $\Omega(n) = 0$ , must be positive which means that  $TC_j^s(\tilde{n} + 1)$  is greater than  $TC_i^f(\tilde{n})$  which according to Definition 1 means that an agreement consisting of  $\tilde{n}$  countries is also externally stable.<sup>15</sup> If  $N$  is lower than  $\tilde{n}$ , the grand coalition could be stable provided that  $\Omega(N)$  is negative. If  $\Omega(n) = 0$  has more than one positive solutions, we could have more than one self-enforcing agreement.

Next, the stability analysis is performed to investigate whether there exists a self-enforcing technology agreement. For the first case studied above, when both signatories and non-signatories invest in R&D, we calculate  $\Omega(n)$  using the total costs functions (44) and (45)

$$\Omega(n) = \frac{\bar{\delta}}{\alpha} \left( \frac{1 + \gamma(N - n + 1) - n(1 + \gamma(N - n))}{n(1 + \gamma(N - n))(1 + \gamma(N - n + 1))} \right).$$

As the denominator is positive for any level of cooperation, the solution to equation  $\Omega(n) = 0$  is given by the number that makes the numerator equal zero. Developing the numerator, the following function of  $n$  is obtained

$$f(n) = \gamma n^2 - (1 + \gamma(N + 1))n + 1 + \gamma(N + 1). \quad (49)$$

It is easy to show that  $f(n) = 0$  has two real positive roots provided that  $N$  is equal to or greater than three, and that the function is decreasing around the lowest root and increasing around the greatest root. Then, according to the argument presented above, an agreement consisting of a number of signatories equal to the greatest natural number on the left of the highest root is self-enforcing provided that this number is lower than  $N$ . We call this number  $\tilde{n}$ . In order to ascertain whether this is the case, we only need to substitute  $N$  in (49). The result is that  $f(N)$  is negative for  $N \geq 2$  which means that  $N \leq \tilde{n}$ , and implies that the grand coalition is the only stable agreement. Remember that for the grand coalition it is only necessary to check the internal stability condition

---

<sup>15</sup>If the positive solution to  $\Omega(n) = 0$  is a natural number. The self-enforcing agreement consists of a number of signatories equal to the solution to the equation and the internal stability condition is satisfied as an equality.

to ascertain whether it is self-enforcing. Therefore

**Proposition 3** *If marginal damages are sufficiently large, in particular if  $d$  is bigger than  $\hat{d}^{nc}$ , the grand coalition is the unique self-enforcing technology agreement independently of the degree of spillover effects.*

Next, stability conditions are analyzed when it is not optimal for non-signatories to invest, i.e. when marginal damages are equal to or lower than  $\hat{d}^{nc}$ . Now, the auxiliary function  $\Omega(n)$  is built using the total costs for non-signatories given by (47) and the total costs for signatories given by (48) which means that the total costs of non-signatories are calculated assuming that signatories invest in R&D.

$$\check{\Omega}(n) = \bar{\delta} \left( \frac{1}{\alpha n} - d(1 - \gamma) \right). \quad (50)$$

The solution to equation  $\check{\Omega}(n) = 0$  is

$$\check{n} = \frac{1}{\alpha d(1 - \gamma)}. \quad (51)$$

As the slope of  $\Omega(n)$  is negative when  $\Omega(n) = 0$ , the only stable agreement is the grand coalition, provided that  $N$  is greater than  $\check{n}$ . Thus, the difference

$$N - \check{n} = N - \frac{1}{\alpha d(1 - \gamma)},$$

should be positive or zero for the grand coalition to be stable. The difference is positive when

$$d \geq \check{d} = \frac{1}{\alpha(1 - \gamma)N}. \quad (52)$$

In order to advance in the analysis of stability conditions, the properties of the function  $\check{d}(\gamma)$  defined by the r.h.s. of (52) must be studied. It is easy to show that  $\check{d}(\gamma)$  is an increasing convex function that takes the value  $1/\alpha N$  for  $\gamma = 0$  and tends to infinite when  $\gamma$  tends to one. Moreover,  $\check{d}(\gamma)$  is equal to  $\hat{d}^{nc}$  when  $\gamma = (N - 1)/(2N - 1)$  so that for  $\gamma$  in the interval  $[0, (N - 1)/(2N - 1))$ ,  $\check{d}(\gamma)$  is lower than  $\hat{d}^{nc}$ . Then we can conclude that when  $\gamma$  is lower than or equal to  $(N - 1)/(2N - 1)$ , if the marginal damages are larger or equal to  $\check{d}(\gamma)$ , the grand coalition is the only self-enforcing technology agreement as the internal

stability condition will be satisfied for  $N$ . When this is not the case and the marginal damages are lower than  $\check{d}(\gamma)$  for all values of  $\gamma$ , two cases can be distinguished. First, when marginal damages are lower than  $\check{d}(\gamma)$  and they belong to the interval  $(\hat{d}^s(2), \hat{d}^{nc}]$  which requires that  $\gamma$  is greater than  $(N-4)/(3N-4)$ .<sup>16</sup> In this case, according to the analysis of the previous section, see Figure 1, signatories invest regardless of the number of countries that belong to the agreement and only the grand coalition can be stable, but as condition (52) is not satisfied because the marginal damages are lower than  $\check{d}(\gamma)$ , it must be concluded that there does not exist any stable agreement for these values of marginal damages. The second case to analyze is when marginal damages are lower than  $\check{d}(\gamma)$  and they belong to the interval  $(\hat{d}^s(N), \hat{d}^s(2)]$ .<sup>17</sup> As in the previous case, the grand coalition is not stable. But for these values of the marginal damages, signatories invest in R&D provided that a minimum of participation is reached, see again Figure 1. Otherwise, signatories' investments are zero. Given this difference with the previous case, it should be investigated whether an agreement consisting of a number of signatories lower than  $N$  may be stable. In particular, whether an agreement that satisfies the minimum of participation defined by the curve  $\hat{d}^s(n)$  in Figure 1 may be stable. Notice that if membership moves from the area above the curve  $\hat{d}^s(n)$  to the area below, the signatories will react to the exit by reducing investment to zero and this may be enough to deter the exit and stabilize the agreement.

In order to explore this possibility, the difference in costs must be recalculated as (50) cannot be used to check the internal stability condition in this case. Notice that (50) has been defined using the non-signatories' total costs when signatories invest in R&D. When the signatories' investment is zero, country emissions and total costs are given by the fully non-cooperative equilibrium calculated in Section 3. Thus, now the difference in costs that must be used to check the internal stability condition is given by the difference

---

<sup>16</sup>This critical value is obtained doing  $\check{d}(\gamma)$  equal to  $\hat{d}^s(2)$ . Remember that  $\check{d}(\gamma)$  is an increasing function so that for  $\gamma > (N-4)/(3N-4)$ ,  $\check{d}(\gamma)$  is going to be greater than  $\hat{d}^s(2)$ .

<sup>17</sup>Notice that if  $\gamma$  belongs to the interval  $[0, (N-4)/(3N-4)]$ , then when the marginal damages are lower than  $\check{d}(\gamma)$  they are also lower than  $\hat{d}^s(2)$ . However, this is not true when  $\gamma$  is greater than  $(N-4)/(3N-4)$ . For this reason we have to impose this second condition.

of expression (48) for signatories and expression (15) for non-signatories corresponding to the fully non-cooperative equilibrium

$$\hat{\Omega}(n) = \bar{\delta} \left( d((N - n)(1 - \gamma) - N) + \frac{1}{\alpha n} \right). \quad (53)$$

Using  $\hat{\Omega}(n) = 0$ , the following second-degree equation is obtained

$$\alpha d(1 - \gamma)n^2 + \alpha \gamma d N n - 1 = 0. \quad (54)$$

This equation has two real roots, one negative and the other positive. Moreover, (53) is a decreasing convex function for  $n > 0$ .

Notice that the positive solution to equation (54) is the same as the positive solution to equation (39), i.e. it coincides with  $\hat{n}$ , the value for which the total costs of the agreement are independent of the investment. Thus, the curve  $\hat{d}^s(n)$  in Figure 1 also represents all the values of  $n$  for different values of  $d$  for which (53) is zero.

Finally, to confirm whether the difference in costs (53) can be used to check the internal stability condition, we must ascertain the relative position of functions (50) and (53). The difference between  $\hat{\Omega}(n)$  and  $\check{\Omega}(n)$  is given by the following expression

$$\hat{\Omega}(n) - \check{\Omega}(n) = -d\bar{\delta}((n - 1)(1 - \gamma) + \gamma N) < 0,$$

which is negative for all  $n \geq 2$  so that it can be concluded that  $\hat{\Omega}(n) < \check{\Omega}(n)$  and consequently that  $\hat{n} < \check{n}$ . Let us now call  $\tilde{n}$  to the lowest natural number on the right of curve  $\hat{d}^s(n)$ . Then the following relationship is obtained:  $\hat{n} < \tilde{n} < N < \check{n}$ . According to the function  $\check{\Omega}(n)$ , none of the values from  $\tilde{n} + 1$  to  $N$  satisfy the internal stability condition, but the internal stability condition for  $\tilde{n}$  must be checked using the function  $\hat{\Omega}(n)$ , and as  $\hat{n}$  is lower than  $\tilde{n}$  we find that  $\hat{\Omega}(\tilde{n})$  is negative and  $\tilde{n}$  satisfies the internal stability condition. Moreover, as  $\check{\Omega}(\tilde{n} + 1)$  is positive, the external stability condition is also satisfied and then an agreement consisting of a number of signatories equal to  $\tilde{n}$  is the only self-enforcing technology agreement. Figure 2 illustrates this argument.

⇒ FIGURE 2 ⇐

Thus, it can be concluded that if marginal damages are lower than  $\hat{d}(\gamma)$  for all  $\gamma$  and they belong to the interval  $(\hat{d}^s(N), \hat{d}^s(2)]$ , the lowest natural number greater than  $\hat{n}$ , being  $\hat{n}$  the positive root of equation (54), is the unique self-enforcing technology agreement.

The following proposition summarizes these results that are representing in Figure 3.

**Proposition 4** *If marginal damages are not sufficiently large, in particular if  $d \in (\hat{d}^s(N), \hat{d}^{nc}]$ , the membership of a self-enforcing agreement depends on the level of marginal damages and the scope of spillover effects. Three cases can be distinguished: i) if  $d \geq \check{d}(\gamma)$ , then the grand coalition is the unique self-enforcing technology agreement. This condition can be satisfied only for  $\gamma \in [0, (N-1)/(2N-1)]$ ; ii) if  $d \leq \check{d}(\gamma)$  and  $d \in (\hat{d}^s(2), \hat{d}^{nc}]$ , then there does not exist any self-enforcing technology agreement. These two conditions can be satisfied only for  $\gamma > (N-4)/(3N-4)$ ; iii) if  $d \leq \check{d}(\gamma)$  and  $d \in (\hat{d}^s(N), \hat{d}^s(2)]$ , then the lowest natural number greater than  $\hat{n}$  is the unique self-enforcing technology agreement, being  $\hat{n}$  the positive root of equation (54). These two conditions can be satisfied for all  $\gamma$ .*

⇒ FIGURE 3 ⇐

The previous analysis clearly establish that it is impossible to stabilize the grand coalition if the spillover effects are greater than 1/2 or if the marginal damages are lower than  $1/\alpha N$  when non-signatories do not invest in R&D.

Moreover, as  $\hat{n}$  decreases when marginal damages increase according to function  $\hat{d}^s(n)$  it is obtained that

**Corollary 1** *If  $d$  belongs to the interval  $(\hat{d}^s(N), \hat{d}^s(2)]$  and is lower than  $\check{d}(\gamma)$ , the greater the marginal damages, the lower the level of participation in a technology agreement.*

An standard result in the literature on international environmental agreements. Finally, if both signatories and non-signatories do not invest in R&D because marginal damages are too low, it does not make sense to negotiate any technology agreement. However, as we know from the standard model of emission abatement with linear damages, in this case it makes sense to negotiate an emission agreement although we do not have to expect great results from the negotiation because only an agreement consisting of three countries is self-enforcing.

## 6 The Technology Agreement with Quadratic Environmental Damages

It is well known that with linear environmental damages, the emissions game has an equilibrium in dominant strategies, in other words, the reaction functions of the countries are orthogonal. In order to check whether this assumption is critical for achieving the result that the grand coalition is stable for large marginal environmental damages, in this section we solve the technology agreement for quadratic environmental damages. In this case, the total cost function is given by<sup>18</sup>

$$TC_i = \frac{c}{2}(\delta - \alpha y_i - E_i)^2 + d_0 E + \frac{d_1}{2} E^2 + x_i.$$

The three stages of the game and the levels of effective investments for signatories and non-signatories are defined in the same way as in the technology agreement with linear environmental damages.

### 6.1 The Nash Equilibrium of the Emissions Game

Assuming that in the first stage  $n$  countries (with  $n \geq 2$ ) have signed the agreement and that in the second stage the countries have selected their levels of investment in R&D, the emission for non-signatories and signatories for the emissions game are given by the first-order conditions

$$E_i^f = \bar{\delta} - \alpha y_i^f - \frac{d_1}{c} E, \quad i = 1, \dots, N - n, \quad (55)$$

$$E_j^s = \bar{\delta} - \alpha y_j^s - \frac{d_1}{c} E, \quad j = 1, \dots, n, \quad (56)$$

where  $\bar{\delta} = \delta - (d_0/c)$ . So that the reaction functions for signatories and non-signatories are

$$\begin{aligned} E_i^f &= \frac{c(\bar{\delta} - \alpha y_i^f)}{c + d_1} - \frac{d_1}{c + d_1} E_{-i}, \quad i = 1, \dots, N - n, \\ E_j^s &= \frac{c(\bar{\delta} - \alpha y_j^s)}{c + d_1} - \frac{d_1}{c + d_1} E_{-j}, \quad j = 1, \dots, n, \end{aligned}$$

---

<sup>18</sup>Notice that now marginal environmental damages are increasing and positive for zero emissions as it occurs for the linear case.

where  $E_{-i}$  and  $E_{-j}$  stand for global emissions minus non-signatory  $i$ 's emissions and minus signatory  $j$ 's emissions respectively. Thus, as the reaction functions have a negative slope, emissions are *strategic substitutes*.

Aggregating for (55) and (56), we obtain the total emissions

$$E = \sum_{i=1}^{N-n} E_i^f + \sum_{j=1}^n E_j^s = \sum_{i=1}^{N-n} \left( \bar{\delta} - \alpha y_i^f - \frac{d_1}{c} E \right) + \sum_{j=1}^n \left( \bar{\delta} - \alpha y_j^s - \frac{d_1}{c} E \right),$$

$$E = \frac{c(N\bar{\delta} - \alpha Y)}{c + Nd_1}, \quad (57)$$

where  $Y$  is the global effective investment in R&D.

$$Y = \sum_{i=1}^{N-n} y_i^f + \sum_{j=1}^n y_j^s = \sum_{i=1}^{N-n} (x_i^f + \gamma X_{-i}) + \sum_{j=1}^n (X^s + \gamma X^f).$$

Then, by substitution in (55) and (56), we obtain the emissions for non-signatories and signatories

$$E_i^f = \bar{\delta} - \alpha y_i^f - \frac{d_1(N\bar{\delta} - \alpha Y)}{c + Nd_1}, \quad i = 1, \dots, N - n, \quad (58)$$

$$E_j^s = \bar{\delta} - \alpha y_j^s - \frac{d_1(N\bar{\delta} - \alpha Y)}{c + Nd_1}, \quad j = 1, \dots, n. \quad (59)$$

In order to ascertain the effect of investment on emissions, the derivative of these expressions is used

$$\begin{aligned} \frac{\partial E_i^f}{\partial x_i^f} &= -\alpha \frac{\partial y_i^f}{\partial x_i^f} + \frac{d_1 \alpha}{c + Nd_1} \frac{\partial Y}{\partial x_i^f} = -\alpha + \frac{d_1 \alpha}{c + Nd_1} (1 + (N - 1)\gamma) \\ &= -\alpha \frac{d_1(N - 1)(1 - \gamma) + c}{c + Nd_1} < 0. \end{aligned}$$

Thus, it can be concluded that investment in R&D by one non-signatory reduces its emissions. We find the same result for signatories

$$\begin{aligned} \frac{\partial E_j^s}{\partial x_j^s} &= -\alpha \frac{\partial y_j^s}{\partial x_j^s} + \frac{d_1 \alpha}{c + Nd_1} \frac{\partial Y}{\partial x_j^s} = -\alpha + \frac{d_1 \alpha}{c + Nd_1} ((N - n)\gamma + n) \\ &= -\alpha \frac{d_1(N - n)(1 - \gamma) + c}{c + Nd_1} < 0. \end{aligned}$$

Substituting (57) and (58)-(59) and in light of our assumption that there is no cooperation in the emission game, the total costs faced by all countries will be given by

$$TC_l = \frac{d_0}{2c} + d_0 \left( \frac{c + d_1}{c + Nd_1} \right) (N\bar{\delta} - \alpha Y) + \frac{d_1}{2} \left( \frac{c(c + d_1)}{(c + Nd_1)^2} \right) (N\bar{\delta} - \alpha Y)^2 + x_l, \quad l = 1, \dots, N. \quad (60)$$

## 6.2 The Partial Agreement Nash Equilibrium of the Investment Game

Next, the partial agreement Nash equilibrium of the investment game is calculated using (60). By taking the derivative of the total cost function with respect to the investment in R&D of non-signatories, we obtain

$$\frac{\partial TC_i^f}{\partial x_i^f} = -\alpha d_0 \left( \frac{c + d_1}{c + Nd_1} \right) \frac{\partial Y}{\partial x_i^f} - \alpha d_1 \left( \frac{c(c + d_1)}{(c + Nd_1)^2} \right) (N\bar{\delta} - \alpha Y) \frac{\partial Y}{\partial x_i^f} + 1 = 0,$$

for  $i = 1, \dots, N - n$ . As  $\partial Y / \partial x_i^f = 1 + \gamma(N - 1)$ , the previous condition can be written as

$$\frac{\alpha(1 + (N - 1)\gamma)(c + d_1)}{c + Nd_1} \left( d_0 + d_1 \frac{c(N\bar{\delta} - \alpha Y)}{c + Nd_1} \right) = 1, \quad (61)$$

where the left-hand side represents the marginal benefits of investment for non-signatories.<sup>19</sup>

As the signatories are cooperating in the investment game, then the agreement total cost function becomes

$$TC_A = \sum_{j=1}^n TC_j^s = \frac{nd_0}{2c} + nd_0 \left( \frac{c + d_1}{c + Nd_1} \right) (N\bar{\delta} - \alpha Y) + \frac{nd_1}{2} \left( \frac{c(c + d_1)}{(c + Nd_1)^2} \right) (N\bar{\delta} - \alpha Y)^2 + \sum_{j=1}^n x_j^s. \quad (62)$$

By taking the derivative of (62) with respect to the investment in R&D of the signatories, it is obtained that

$$\frac{\partial TC_A}{\partial x_j^s} = -\alpha nd_0 \left( \frac{c + d_1}{c + Nd_1} \right) \frac{\partial Y}{\partial x_j^s} - \alpha nd_1 \left( \frac{c(c + d_1)}{(c + Nd_1)^2} \right) (N\bar{\delta} - \alpha Y) \frac{\partial Y}{\partial x_j^s} + 1 = 0,$$

<sup>19</sup>Notice that  $(d_0 + d_1 \frac{c(N\bar{\delta} - \alpha Y)}{c + Nd_1}) = d_0 + d_1 E$  stands for marginal environmental damages.

for  $j = 1, \dots, n$ . As  $\partial Y/\partial x_j^s = n + \gamma(N - n)$ , the previous condition yields

$$\frac{\alpha n(n + (N - n)\gamma)(c + d_1)}{c + Nd_1} \left( d_0 + d_1 \frac{c(N\bar{\delta} - \alpha Y)}{c + Nd_1} \right) = 1, \quad (63)$$

where the left-hand side represents the marginal benefits of investment for signatories.

By comparing (61) and (63), it is pretty obvious that the marginal benefits of signatories are higher than the marginal benefits of the non-signatories for all  $n > 1$  for the same level of global effective investment.

Next, we study the conditions to achieve a corner solution, i.e. to completely eliminate emissions. This occurs when the marginal benefits of investment are greater than or equal to the marginal costs for zero global emissions. Then, using (61) and (63) a critical value for  $d_0$  can be obtained for both non-signatories and signatories, which makes the derivative of total costs equal to zero for zero global emissions:

$$\hat{d}_0^f = \frac{c + Nd_1}{\alpha(c + d_1)(1 + \gamma(N - 1))},$$

$$\hat{d}_0^s(n) = \frac{c + Nd_1}{\alpha n(c + d_1)(n + \gamma(N - n))}.$$

Observe that in this case, the critical values of  $d_0$  for both non-signatories and signatories depend on  $c$  and  $d_1$ . Also, as in the linear model, the critical value of  $d_0$  for signatories is decreasing with respect to the level of cooperation and takes values between

$$\hat{d}_0^s(N) = \frac{c + Nd_1}{\alpha N^2(c + d_1)},$$

$$\hat{d}_0^s(2) = \frac{c + Nd_1}{2\alpha(c + d_1)(2 + \gamma(N - 2))}.$$

Moreover, it is obvious that  $\hat{d}_0^f = \hat{d}_0^s(1)$  so that  $\hat{d}_0^s < \hat{d}_0^f$  for all  $n > 1$ . Therefore, for any  $d_0 > \hat{d}_0^f$ , both signatories and non-signatories will invest at the maximum level that leads to the elimination of emissions. Then the levels of effective investment for both non-signatories and signatories will be, according to (58) and (59) exactly the same as in the linear model

$$y_i^f = y_j^s = \frac{\bar{\delta}}{\alpha}, \quad Y = \frac{N\bar{\delta}}{\alpha},$$

except that now  $\bar{\delta} = \delta - (d_0/c)$ , and the same occurs for investment in R&D:

$$x_i^f = \frac{\bar{\delta}}{\alpha(1 + \gamma(N - n))}, \quad x_j^s = \frac{x_i^f}{n} = \frac{\bar{\delta}}{\alpha n(1 + \gamma(N - n))}.$$

Finally, by substituting in the total costs functions, we obtain

$$TC_i^f = \frac{d_0^2}{2c} + \frac{\bar{\delta}}{\alpha(1 + \gamma(N - n))}, \quad i = 1, \dots, N - n, \quad (64)$$

$$TC_j^s = \frac{d_0^2}{2c} + \frac{\bar{\delta}}{\alpha n(1 + \gamma(N - n))}, \quad j = 1, \dots, n. \quad (65)$$

### 6.3 The Nash Equilibrium of the Membership Game

Comparing (64) and (65) to the total costs functions of the linear model (44) and (45), it is clear that the stability analysis, when the level of marginal damages is high enough, will yield the same result as that obtained for linear environmental damages. Thus, the grand coalition is the only self-enforcing agreement.

We would like to clarify that by moving from the high level of marginal damages to the lower level (in other words, moving from the corner solution to the interior one), the solution of the model with quadratic environmental damages might give different results to those obtained in the linear model. Actually, what this paper intends to highlight is the possibility of having an agreement at high levels of marginal damages that completely eliminates completely emissions. Using a linear model, we have shown that the grand coalition is the only self-enforcing agreement and the robustness of this result (the corner solution) has been also shown by introducing environmental damages in a quadratic form.<sup>20</sup>

## 7 Conclusions

This paper aims to study the effects of R&D spillovers on the formation and stability of IEAs by solving a three-stage game where the membership decision is taken in the first

---

<sup>20</sup>We have also analyzed the case when the investment cost is quadratic, obtaining the same result. This case is available from the authors upon request.

stage, the investment game is played in the second stage and the emission game is played in the last stage. It is assumed that the marginal abatement costs of signatory countries are decreased by the sum of signatories' R&D efforts in addition to some spillovers from non-signatories' R&D. We find that for the technology agreement the grand coalition is the only stable agreement if marginal damages are large enough. When all countries share their R&D investments, withdrawing from the agreement entails facing larger abatement and investment costs because of a sharp reduction in the effective investment which eliminates the incentive to act as a free-rider in the agreement. We show the robustness of our linear model at a high level of marginal damages by introducing quadratic environmental damages. Although the results between the linear and the quadratic model at low levels of marginal damages may be different, this does not affect the interesting result of having that the grand coalition is stable if marginal damages are sufficiently large. For this reason, in order to avoid the complexity of quadratic models, we have decided to analyze the stability of a technology agreement using the simpler case of linear environmental damages.

Some extensions of the model are on the agenda for future research. Primarily, we could consider the case of an agreement in which signatories only cooperate to define the level of investment, but do not pool their R&D efforts. Alternatively, it would also be interesting to look at the case where the agreement only consists of fully internalizing the spillovers effects. Finally, it could be of interest to explore the effect of timing on the scope of cooperation reached by a technology agreement, in particular to analyze the model by changing the order of the stages so as to play the investment game in the first stage and the membership game in the second. In this way, the effect of strategic behavior in the investment game on the level of participation may be investigated.

## References

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

- [2] Barrett, Scott (2009). “The Coming Global Climate-Technology Revolution.” *Journal of Economic Perspectives*, 23, 53-75.
- [3] Buchner, Barbara and Carlo Carraro (2005). “Economic and Environmental Effectiveness of a Technology-Based Climate Protocol.” *Climate Policy*, 4, 229-248.
- [4] Carraro, Carlo and Domenico Siniscalco (1997). “R&D Cooperation and the Stability of International Environmental Agreements.” In *International Environmental Negotiations: Strategic Policy Issues*, edited by Carlo Carraro. Edward Elgar Publishing.
- [5] 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.
- [6] D’Aspremont, C.A., Jacqemin, J., Gadszewicz, J., and Weymark, J.A.(1983). “On the Stability of Collusive Price Leadership.” *Canadian Journal of Economics*, 16, 17- 25.
- [7] de Coninck, Heleen, Carolyn Fischer, Richard G. Newell and Takahiro Ueno (2008). “International Technology-Oriented Agreements to Address Climate Change.” *Energy Policy*, 36, 335-356.
- [8] Golombek, Rolf and Michael Hoel (2005). “Climate Policy under Technology Spillovers.” *Environmental and Resource Economics*, 31, 201-227.
- [9] Golombek, Rolf and Michael Hoel (2008). “Endogenous Technology and Tradable Emission Quotas.” *Resource and Energy Economics*, 30, 197-208.
- [10] Golombek, Rolf and Michael Hoel (2011). “International Cooperation on Climate-Friendly Technologies.” *Environmental and Resource Economics*, 49, 473-490.
- [11] Heal, Geoffrey and Nori Tarui (2010). “Investment and Emission Control under Technology and Pollution Externalities.” *Resource and Energy Economics*, 32, 1-14.

- [12] 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.
- [13] Kamien, Morton I., Eitan Muller and Israel Zang (1992). “Research Joint Ventures and R&D Cartels.” *American Economic Review*, 82, 1293-1306.
- [14] Kemfert, Claudia (2004). “Climate Coalitions and International Trade: Assessment of Cooperation Incentives by Issue Linkage.” *Energy Policy*, 32, 455-465.
- [15] Lessman, Kai and Ottmar Edenhofer (2011). “Research Cooperation and International Standards in a Model of Coalition Stability.” *Resource and Energy Economics*, 33, 36-54.
- [16] Nagashima, Miyuki and Rob Dellink (2008). “Technology Spillovers and Stability of International Climate Coalitions.” *International Environmental Agreements: Politics, Law and Economics*, 8, 343-365.
- [17] Ruis, Arjan and Aart de Zeeuw (2010). “International Cooperation to Combat Climate Change.” *Public Finance and Management*, 10, 379-404.
- [18] Spence, Michael (1984). “Cost Reduction, Competition, and Industry Performance.” *Econometrica*, 52, 101-122.
- [19] Xepapadeas, A. (1995). “Induced Technical Change and International Agreements under Greenhouse Warming.” *Resource and Energy Economics*, 17, 1-23.

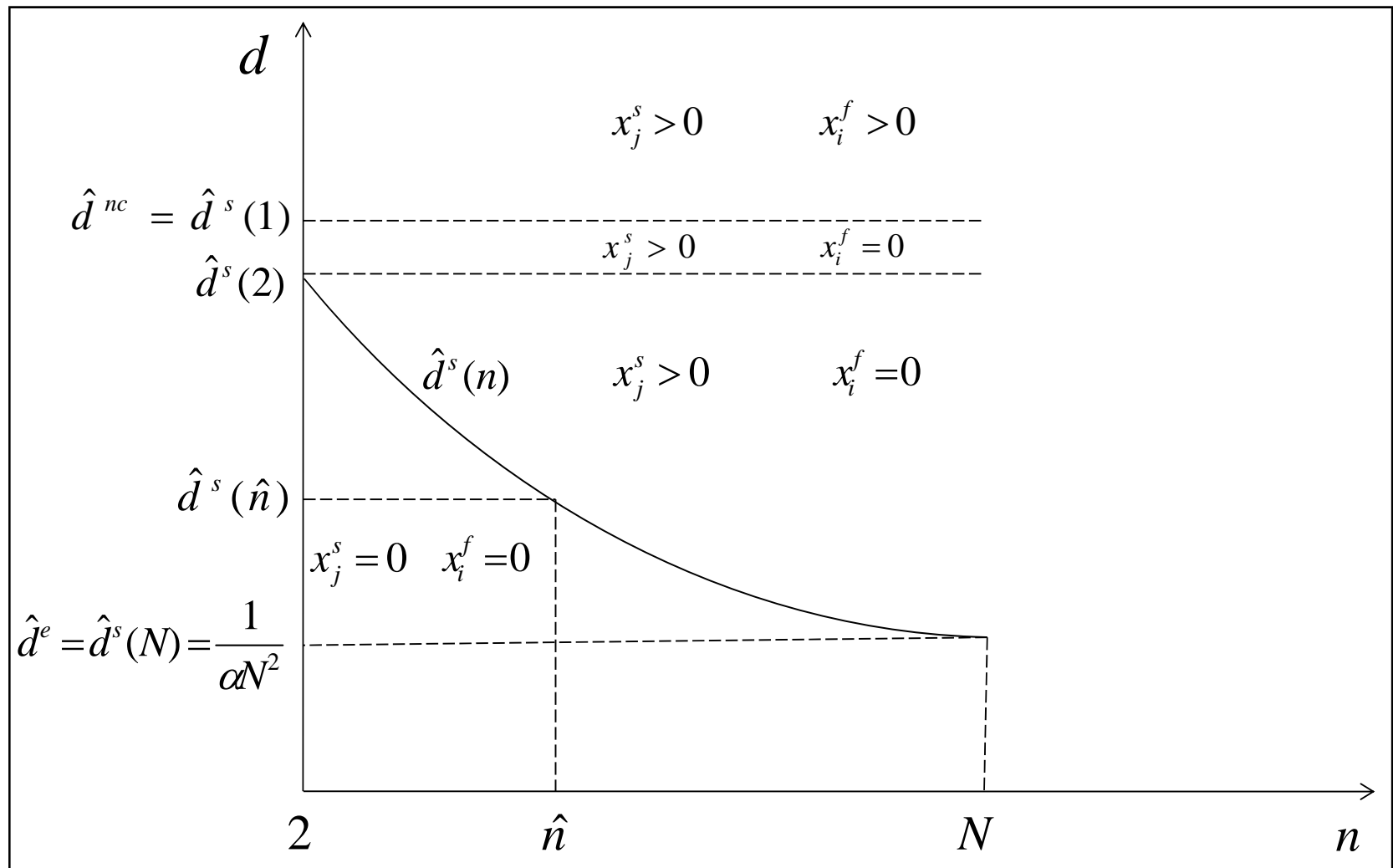


FIGURE 1. R&D investment for signatories and non-signatories

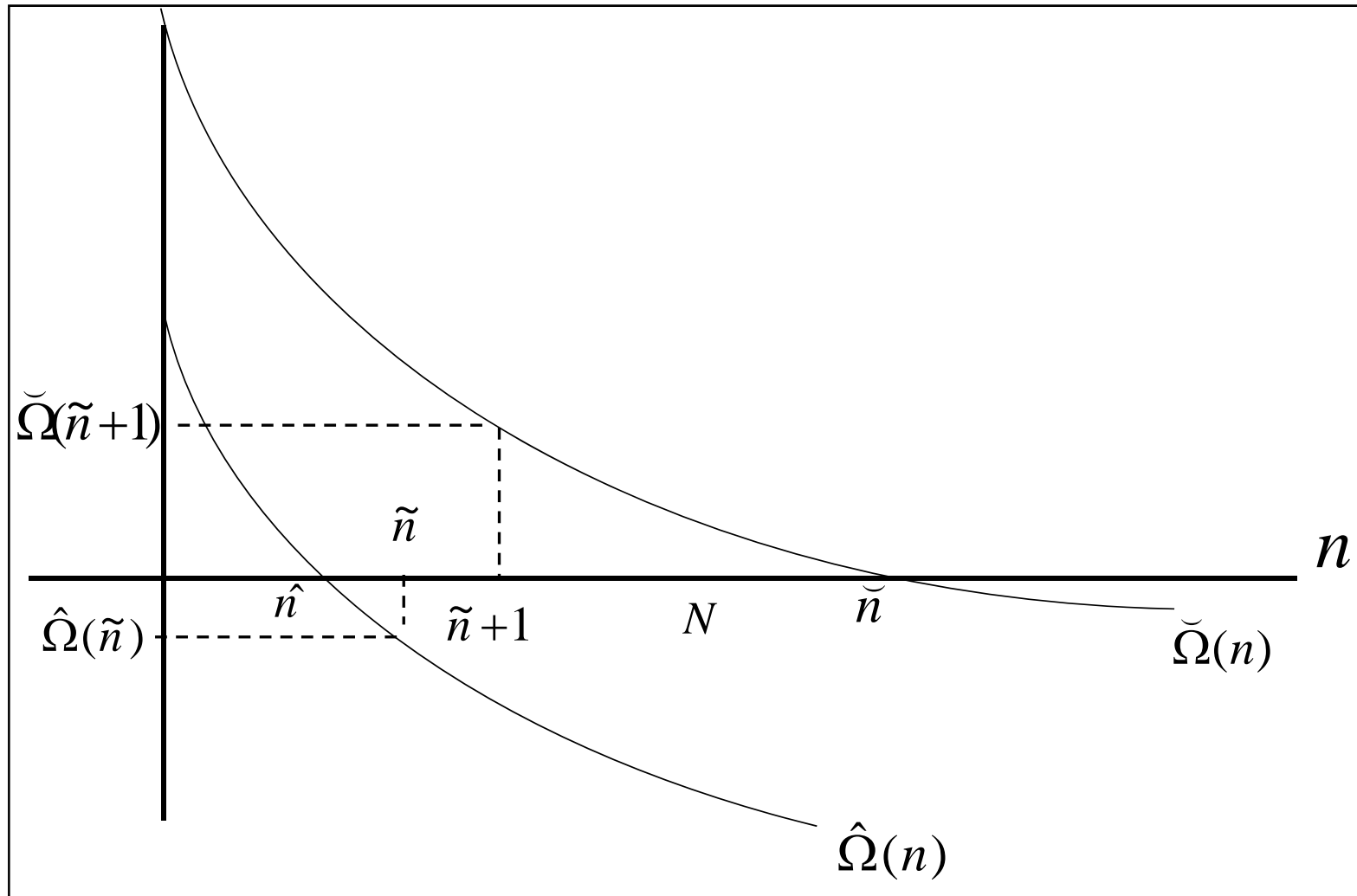


FIGURE 2. Stability conditions

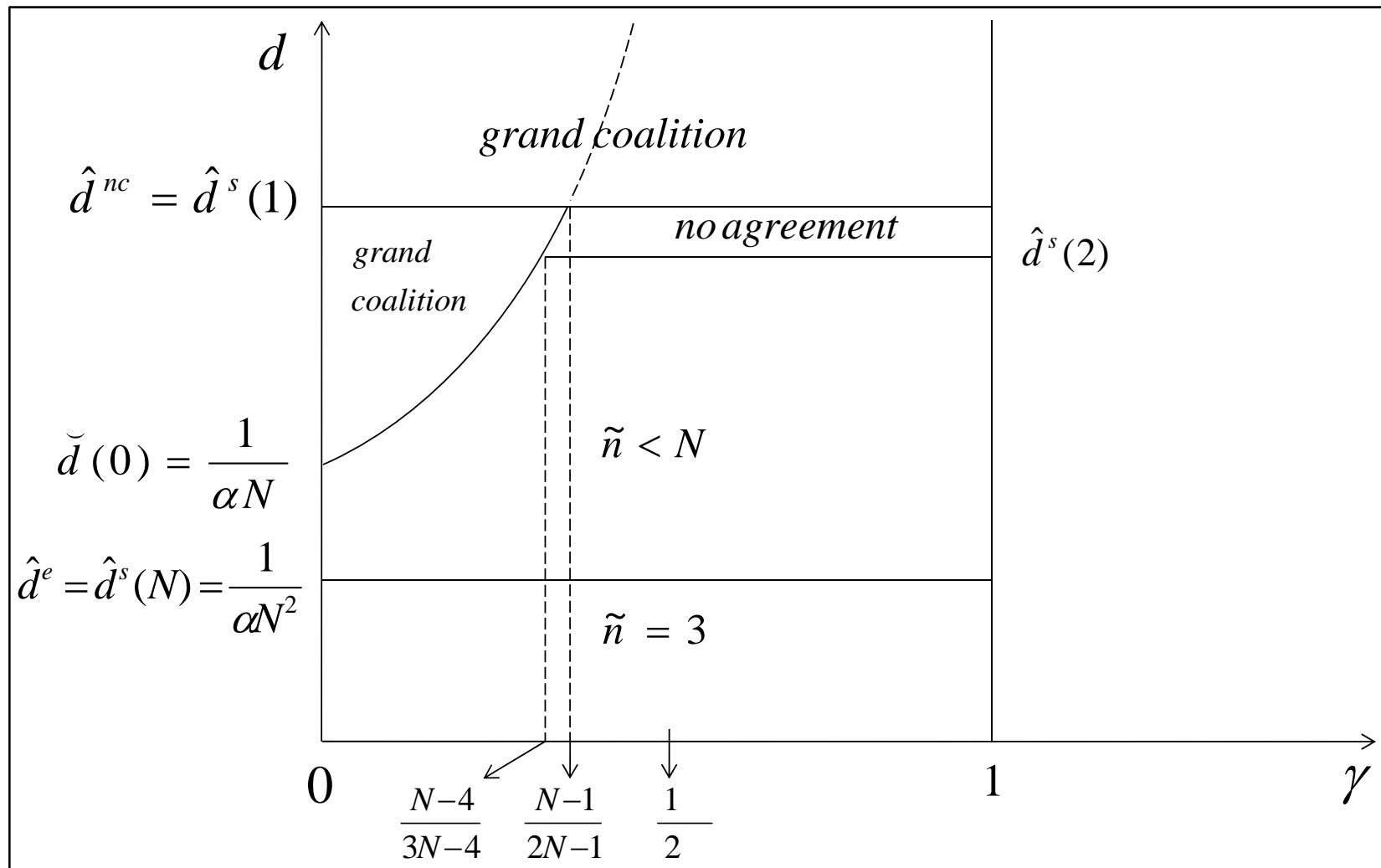


FIGURE 3. Membership of stable agreements