00/20

International Environmental Negotiations: Does Coalition Size Matter?

Samuel Rutz and Thomas Borek

October 2000

Institut für Wirtschaftsforschung

Eidgenössische Technische Hochschule Zürich
Weinbergstrasse 35
CH-8006 Zürich

Telefon: 0041-1-632 41 28
Telefax: 0041-1-632 12 18
email: library@wif.gess.ethz.ch

Postadresse:

Institut für Wirtschaftsforschung
ETH Zentrum WEH
CH-8092 Zürich

International Environmental Negotiations: Does Coaltion Size Matter?[1]

By **Samuel Rutz*** and **Thomas Borek****

Abstract: This paper reconsiders a widely used game of coalition formation in international environmental negotiations. Due to the mathematical problems of giving a full characterization of the solution, up to now most of the work on this subject rested on numerical simulations to derive results. In this paper we show for a general class of payoff functions that when the game is approximated by assuming a continuum of players, a solution can be found. Using this result as a "benchmark solution", we further show that gains from cooperation resulting in simulations are due to an "integer effect", i.e. coalition size being treated as a discrete variable.

Keywords: transboundary pollution, international environmental agreements, coalition formation, partial cooperation

*Center for Economic Research, Swiss Federal Institute of Technology, CH-8952 Zurich, e-mail rutz@wif.gess.ethz.ch

**Socioeconomic Institute, University of Zurich, Hottingerstrasse 10, CH-8006 Zurich, e-mail tobo@acess.unizh.ch

October, 2000

[1] The authors are grateful to Martin Brown, Simon Gächter, Matthias Gysler, Armin Schmutzler, Renate Schubert and participants at the EAERE-FEEM-VIU Postgraduate Summer School in Environmental and Resource Economics 2000 in Venice for helpful comments and suggestions.

1 Introduction

In recent years the problem of transboundary pollution has attracted the attention of many environmental economists. Transboundary pollution refers to pollution which is released in one country but causes damage in at least one other country. Such damages can affect a limited number of countries, like the acid rain problem or the pollution of the Rhine. They can however also occur on a global scale, as in the case of the depletion of the ozone layer or the problem of climate change due to CO_2-emissions.

The reduction of transboundary pollution is inherently a public good. Due to the lack of a supranational authority however, transboundary pollution problems cannot be solved with common strategies for public good provision. Command and control regulation, Pigou taxes or emission certificates for example cannot be enforced between sovereign countries. The reduction of transboundary pollution is therefore subject to diplomatic negotiations, which can only be successful if resulting agreements are profitable for all affected parties.

Game theory provides a promising framework to analyze negotiations between sovereign states on transboundary environmental problems. As a result the past few years have seen a growing game-theoretic literature (cooperative and non-cooperative) on transboundary pollution (for an overview see e.g. F. Missfeldt [14]).

A lot of this previous work focuses on static, non-cooperative games to reduce CO_2-emissions. Such games are usually modeled as two-stage games. At the first stage, countries decide non-cooperatively whether or not to sign an agreement given a burden-sharing rule which is adopted by the signatory countries. At the second stage, both signatory and non-signatory countries set their abatement levels. Stable coalitions of signatory countries in such games have been determined for the case of homogeneous countries by Barrett [2, 3], Carraro and Siniscalco [9] and Hoel [13] and by Barrett [6] and Botteon and Carraro [8] for the case of heterogeneous countries.

Two important results have emerged from this literature. First, there are stable, but only partial coalitions. Therefore, in the equilibrium of the game there are two groups of countries, signatories and non-signatories. Second, the coalition generally involves only a very small fraction of the negotiating countries and efficiency gains are usually modest.

These results have however mainly been derived considering numerical simulations and can therefore not claim general validity. In this paper we provide an analytical solution to a general class of such games for the case of identical countries and non-orthogonal reaction functions.[2] We first derive a "benchmark solution" by approximating these games with a continuum of players and show that the establishment of a profitable coalition then is impossible, i.e. the equilibrium of the approximated games is characterized by no gains from cooperation and an abatement level equal to the situation where there is no cooperation between countries at all. Using this "benchmark solution", it is then possible to show, that gains from cooperation resulting in numerical simulations are solely the result of an "integer effect", i.e. coalition size being a discrete variable. It is therefore not surprising that these gains are typically extremely small and that the overall effect on the environment, compared to a situation without cooperation, is near to zero in most simulations.

These results have several implications. First, in the framework of these models the size of the stable coalition is irrelevant to gains from cooperation since stability, independent of the specification of the payoff function, is always bought at the price of giving away these gains. Further, since only minor gains from cooperation can be realized, there hardly exist possibilities to increase coalition size and reduce environmental pollution, as e.g. Carraro and Siniscalco [9] suggest.

Second, since in the equilibrium of these games no country has an incentive to commit to an improvement of environmental quality, it seems doubtful whether the model is an appropriate theoretical foundation for further research as several authors suggest. Bohm [7] and Hoel [13] e.g. evaluate policy measures available to a coalition of cooperating countries trying to reduce aggregate demand and/or supply for fossil fuels. Escapa and Gutiérrez [12] assume in their work that an agreement involving all the affected parties has been reached and analyze different burden-sharing rules. Although it seems *per se* reasonable to assume that there

[2] Non-orthogonal reaction functions refer to the scenario where non-signatory countries will increase their emissions whenever the signatory countries reduce their own, while in the scenario of orthogonal reaction functions there is no such leakage, i.e. the non-signatory countries do not offset the abatement efforts of the signatory countries by increasing their polluting activities.

are groups of countries willing to cooperate and improve environmental quality, such an assumption cannot theoretically be founded with the above models.

The structure of the paper is as follows: In part 2 we describe the standard game, which will be analyzed in two versions (a sequential and a simultaneous one) in the following sections. In part 3 we first focus on international environmental negotiations modeled as a sequential game. We shortly reconsider the standard solution of this game, as can be found in the literature and summarize results from numerical simulations. Then, we derive the "benchmark solution" when the game is approximated by a continuum of players and draw conclusions. In part 4 we focus on international environmental negotiations modeled as a simultaneous game. Again, a "benchmark solution" for the continuous version of this game will be presented. In part 5 we analyze the above mentioned "integer effect" and show that the transition from the continuous to the discrete version of the game is of minor importance to efficiency gains. Finally, some concluding remarks and propositions for further research are contained in part 6.

2 The Standard Model

The complete information two-stage game of international environmental negotiations we analyze in the next two sections has the following structure. To keep matters simple, we assume that all countries are identical.[3]

> *Stage 1:*
>
> There are N identical countries, which decide simultaneously whether or not to sign an agreement.[4]

[3] In the following index s stands for signatory countries and index n for non-signatory countries.

[4] In the case of heterogeneous countries one needs to specify a burden-sharing rule to be adopted by the signatory countries. Such a burden-sharing rule defines how costs of abatement are split between signatory countries and can be any of the rules derived from cooperative game theory, e.g. the Nash bargaining solution or the Shapley value. In the case of identical countries however, any of the mentioned rules leads to an equal split of the cost of abatement, i.e. the burden-sharing rule does not matter.

Stage 2:

The k signatory countries choose their aggregate abatement level $Q_s = \sum_{i=1}^{k} q_{s_i}$, where $q_{s_i} \geq 0$ denotes the individual abatement level of signatory country i, by maximizing their collective net benefits.[5]

The $(N-k)$ non-signatory countries simultaneously choose their individual abatement level $q_{n_j} \geq 0$ to maximize their individual net benefit.

The net benefit to non-signatory country j is then given by the payoff function $\Pi_{n_j} = B(Q) - C(q_{n_j})$, where $Q = Q_s + Q_n$ and $Q_n = \sum_{j=k+1}^{N} q_{n_j}$, while the net benefit to a signatory country can be expressed as $\Pi_{s_i} = B(Q) - C(q_{s_i})$. Inherent to the transboundary pollution problem, the benefits of abatement $B(\cdot)$ typically depend not only on the individual adopted abatement level, but on the abatement undertaken by all countries. The costs of abatement $C(\cdot)$, however, are always incurred by the individual country. This specification of the payoff function mirrors the public good character of transboundary pollution: a country cannot be excluded from the benefits of abatement even if it undertakes no abatement effort at all and hence incurs no abatement costs.

Further, two additional assumptions are worth mentioning:

1. The cost functions are independent, in particular we do not account for economies of scale in abatement technology.

2. Countries are proposed to sign a single agreement, i.e. countries not signing the initial agreement cannot negotiate a different agreement.

Following the literature, we assume that $B(\cdot)$ and $C(\cdot)$ are twice differentiable with $B_Q > 0$, $B_{QQ} < 0$, $C_q > 0$, $C_{qq} > 0$ and hence $\Pi_{qq} < 0$, where a single subscript stands for the first and a double subscript stands for the second derivative. These assumptions are necessary to ensure the existence of the solution of the optimization problem at the second stage of the game.

[5] This implies that the chosen abatement level of the signatory countries meets the "Samuelson-condition" for the provision of public goods.

In what follows, we will analyze two versions of the above described game. In the sequential version of the game signatory countries choose their abatement level first. Observing this, the non-signatory countries then choose their own abatement level. The structure of the game implies a "first-mover" advantage for the signatory countries, as they know that the non-signatory countries will condition their decision on the chosen abatement level of the coalition. In the simultaneous version of the game signatory and non-signatory countries choose abatement levels at the same time.

3 International Environmental Negotiations Modeled as a Sequential Game

3.1 The Standard Solution

In this section we shortly discuss the solution to the sequential version of the above described game, as presented in Barrett [4]. The game is solved by backward induction, i.e. we start by looking at the optimal choice of the abatement level of the non-signatory countries at the second stage of the game.

When the non-signatory countries get to move at the second stage, each of them faces the following optimization problem, where $Q_{-j} = Q - q_{n_j}$ denotes every possible aggregate abatement level all countries except non-signatory country j can choose:

$$(1) \qquad \max_{q_{n_j}} B(Q_{-j} + q_{n_j}) - C(q_{n_j}).$$

We denote the solution of this optimization problem by $R_{n_j}(k, Q_{-j})$, the reaction function of non-signatory country j. Note that the best reply of country j is also a function of the number of signatory countries k. The solution then requires that in the optimum the marginal benefits from abatement should equal the marginal costs of abatement, i.e.

$$(2) \qquad B_Q\left(Q_{-j} + R_{n_j}(k, Q_{-j})\right) = C_q\left(R_{n_j}(k, Q_{-j})\right).$$

Given the assumption that all countries are identical, we know however that in the equilibrium of the subgame all non-signatory countries choose the same individual abatement level q_n. The aggregate abatement level chosen by the non-signatory countries can then be expressed as $R_n(k, Q_s) = (N-k)R_{n_j}(k, Q_{-j})$.

The signatory countries maximize their collective net benefits, anticipating the aggregate abatement level $R_n(k, Q_s)$ of the non-signatory countries. Maximizing the collective net benefits implies that a signatory country not only takes into account the individual benefit arising from aggregate abatement, but the benefits arising to all the k members of the coalition. The costs of abatement are however still incurred by the individual country. The optimization problem of signatory country i then amounts to

$$(3) \qquad q^*_{s_i}(k) = \underset{q_{s_i}}{\arg\max}\ kB\left(\sum_{i=1}^{k} q_{s_i} + R_n(k, Q_s)\right) - \sum_{i=1}^{k} C(q_{s_i}).$$

Again, given the assumption of identical countries, in the equilibrium of the subgame all signatory countries will choose the same individual abatement level and the optimization problem of signatory country i can be simplified to

$$(4) \qquad q^*_{s_i}(k) = \underset{q_{s_i}}{\arg\max}\ kB\left(Q_{s_{-i}} + q_{s_i} + R_n(k, Q_s)\right) - C(q_{s_i}),$$

where $Q_{s_{-i}} = (k-1)q_{s_i}$ denotes the aggregate abatement level of all signatory countries except country i. Note that this formulation of the optimization problem captures the above mentioned "first-mover" advantage. Since the signatory countries know that the non-signatory countries condition their decision on the chosen abatement level Q_s, this reaction enters the optimization problem.

The solution of the optimization problem for signatory country i then requires that it sets an abatement level $q^*_{s_i}$, such that it does not wish to revise its choice after observing the actual chosen abatement level of the non-signatory countries, i.e.

$$(5) \qquad kB_Q\left(Q_{s_{-i}}(k) + q^*_{s_i}(k) + R_n(k, Q_s)\right)\left(1 + \tfrac{\partial R_n(k,Q_s)}{\partial Q_s}\right) = C_q\left(q^*_{s_i}(k)\right).$$

Exploiting the homogeneity of the countries once more and expressing the aggregate abatement level as $Q = kq_s + (N-k)q_n$, the equilibrium conditions (2) and (5) substantially simplify. Hence, for a given number of signatory countries k, we can now state the equilibrium conditions, which have to be satisfied simultaneously at the second stage of the game as follows:

$$(2a) \qquad B_Q\left(kq_s^*(k) + (N-k)q_n^*(k)\right) = C_q\left(q_n^*(k)\right)$$

$$(5a) \qquad kB_Q\left(kq_s^*(k) + (N-k)q_n^*(k)\right)\left(1 + (N-k)\frac{\partial q_n(k,Q_s^*)}{\partial Q_s}\right) = C_q\left(q_s^*(k)\right).$$

A subgame perfect equilibrium of this two-stage game further requires that all decisions made at the first stage (to join or not to join the coalition) are mutual best responses for all countries. This equilibrium condition is usually termed as the "stability condition". A coalition k is said to be "stable" if the following two conditions are met: First, there is no incentive for a signatory country to leave the coalition, i.e. no signatory country can increase its payoff by leaving the coalition. Second, there is no incentive for a non-signatory country to join the coalition, i.e. a non-signatory country cannot increase its payoff by joining the coalition.[6] Formally these two conditions can be expressed as follows, where k^* stands for the size of the stable coalition:

$$(6) \qquad \Pi_s(k^*) \geq \Pi_n(k^*-1) \text{ and } \Pi_n(k^*) \geq \Pi_s(k^*+1).$$

3.2 Results from Numerical Simulations

Due to the fact that it is usually not possible to solve Eq. (2a) and (5a) analytically for $q_s^*(k)$ and $q_n^*(k)$, an explicit characterization of the solution seems impossible.[7] Therefore, solutions are typically derived by numerical simulations.[8] Two important results emerging from these simulations are emphasized in the literature (see e.g. Carraro and Siniscalco [10]):

1. In general there exist stable, but only partial coalitions. Therefore, the equilibrium is not characterized by "no cooperation", but there are two groups of countries, signatory and non-signatory countries.

[6] This concept was developed for the analysis of cartels by d'Aspremont et al. [1]. Some authors refer to the first part of the stability condition as "internal stability" and to the second part as "external stability".

[7] Barrett [3, 4] derives explicit solutions for a few specifications of the payoff function.

[8] For simulations with identical countries see e.g. Barrett [3], simulations with asymmetric countries can be found in Barrett [6] or Botteon and Carraro [8].

2. Stable coalitions generally involve only a very small fraction of the negotiating countries, and gains from cooperation (in terms of social welfare and environment quality) are usually small.

The small coalition size found in most of the simulations is seen as a particularly unsatisfying result. Two ways to deal with this problem have emerged in the literature. The first way is to alter the structure of the game by means of appropriate policy measures, like e.g. issue linkage or trade sanctions. In this paper we will however not pursue this strand of research any further.[9] The second way is to alter the assumptions about the cost and benefit functions. As e.g. Barrett [3] shows in his simulations, for certain specific functional forms of the payoff function large coalitions can exist. But even in these cases, the resulting gains from cooperation, again in terms of social welfare and environmental quality, are very small. Our following analysis shows that this is no surprise, as gains from cooperation in the above game are merely the result of an "integer effect".

3.3 The "Benchmark Solution"

As we have explained above, due to the resulting mathematical problems to solve equilibrium conditions explicitly for the optimal abatement levels and the size of the stable coalition, most results in the literature are derived by numerical simulations. Such results can however not claim general validity and might be sensitive to the specification of the payoff function and the calibration of the model. To understand the underlying logic of these games more deeply, it is therefore desirable to derive a more general solution. In what follows, we will present such a solution for a general class of payoff functions, having the properties introduced in section 2. In order to do so, we approximate the game in a first step by assuming a continuum of players in the range $(0, N]$, which then implies that coalition size k can take any value in this range. The structure of the game is otherwise maintained as described above, i.e. the approximated game satisfies all the assumptions of the discrete game except that players are treated as a continuum. The solution of this

[9] For the subject of issue linkage see e.g. Cesar and De Zeeuw [11]. The strategy of trade sanction is e.g. discussed in Barrett [5].

slightly modified version of the game will proof as a helpful benchmark to derive general results for the discrete version of the game.

The crucial step that allows us to solve this modified game is the resulting simplification of the stability condition (6), when coalition size k is treated as a continuous variable. The first part of Eq. (6) then simplifies to $\Pi_s(k^*) \geq \Pi_n(k^*)$ and the second part to $\Pi_n(k^*) \geq \Pi_s(k^*)$. This together implies $\Pi_s(k^*) = \Pi_n(k^*)$, i.e. in a stable coalition k^* the payoffs of signatory and non-signatory countries must be equal.

A proposition about the existence and uniqueness of the solution of the modified game is considered in the following. We first show that the (simplified) stability condition requires that all countries set the same abatement level. Then we show that the only subgame perfect equilibrium is for each country to set the non-cooperative abatement level q_{nc}, i.e. the same abatement level which results in a situation without cooperation between countries. It follows then that in the resulting equilibrium the effect on the environmental quality is zero, i.e. aggregate payoff and abatement are the same as in a non-cooperative situation. This result does however not imply that no coalition of signatory countries will be formed. Rather, the equilibrium is characterized by a stable coalition involving k countries setting the non-cooperative abatement level q_{nc}. We shall denote such a coalition as k_0.

Proposition 1 *For the approximated game, there exists a unique subgame perfect Nash equilibrium in pure strategies. In this equilibrium there are no gains from cooperation. Although there are no gains from cooperation in the resulting equilibrium, there exists a stable coalition involving k_0 signatory countries.*

Proof. A subgame perfect equilibrium requires at the first stage of the game that $\Pi_s(k^*) = \Pi_n(k^*)$. This condition can as well be written as:

$$(7) \qquad B\left(Q^*(k^*)\right) - C\left(q_s^*(k^*)\right) = B\left(Q^*(k^*)\right) - C\left(q_n^*(k^*)\right).$$

Due to the homogeneity of the countries, benefits from abatement for signatory and non-signatory countries will always be equal. Therefore their costs must be equal in equilibrium. Since we assumed monotonicity of the cost function

($C_q > 0$) it follows immediately that in any equilibrium of the approximated game $q_n^*(k^*) = q_s^*(k^*)$ must hold, i.e. all countries must set the same abatement level in equilibrium.

In what follows, we show that there cannot be an equilibrium in which all countries set the same abatement level and this abatement level is not the non-cooperative abatement level q_{nc}. For this purpose we start by looking at the optimal choice of the abatement level of non-signatory country j. These decisions are governed by the reaction function $R_{n_j}(k, Q_{-j})$. Although it is not possible to derive an explicit expression for $R_{n_j}(k, Q_{-j})$, it is easy to see that non-signatory country j conditions its choice of the abatement level on Q_{-j}, the aggregate abatement level of all other countries except country j. Further it is possible to determine the slope of the reaction function. Totally differentiating and rearranging Eq. (2) yields

$$(8) \qquad \frac{\partial R_{n_j}(k, Q_{-j})}{\partial Q_{-j}} = \frac{B_{QQ}(Q(k))}{C_{qq}\big(q_{n_j}(k)\big) - B_{QQ}(Q(k))} < 0.$$

Since $B_{QQ} < 0$ and $C_{qq} > 0$ the slope of the reaction function is strictly negative, i.e. when the aggregate abatement level of all other countries except country j increases, non-signatory country j decreases its abatement level and vice versa.

Noting that the non-cooperative abatement level is defined as $q_{nc} = R_{n_j}(k, Q_{nc_{-j}})$, where $Q_{nc_{-j}} = (N-1)q_{nc}$, it is now straight forward to show that the only subgame perfect equilibrium of the game is $q_n^*(k^*) = q_s^*(k^*) = q_{nc}$.

Assume:

1. $Q_{-j} = (N-1)q_{nc}$: By definition it is a best response for country j to set the non-cooperative abatement level q_{nc} and hence $q_n^*(k^*) = q_s^*(k^*) = q_{nc}$.
2. $Q_{-j} > (N-1)q_{nc}$: Due to the negative slope of the reaction function, it is a best response for country j to set $q_j > q_{nc}$ and hence $q_n^*(k^*) \neq q_s^*(k^*)$.
3. $Q_{-j} < (N-1)q_{nc}$: Due to the negative slope of the reaction function, it is a best response for country j to set $q_j > q_{nc}$ and hence $q_n^*(k^*) \neq q_s^*(k^*)$.

The remaining issue is to determine the size of the stable coalition. This can simply be done by deriving an explicit expression for k_0, the stable coalition associated with a situation where there are no gains from cooperation.

Totally differentiating and rearranging Eq. (2a), the equilibrium condition for the non-signatory countries at the second stage of the game, yields an explicit expression for the slope of the reaction function $q_n(k, Q_s^*)$:

$$(9) \qquad \frac{\partial q_n(k,Q_s^*)}{\partial Q_s} = \frac{B_{QQ}(Q^*(k))}{C_{qq}(q_n^*(k))-(N-k)B_{QQ}(Q^*(k))}.$$

Substituting Eq. (2a) and (9) into Eq. (5a), the equilibrium condition for the signatory countries at the second stage, rearranging terms and evaluating the resulting expression at $k^* = k_0$ yields

$$(10) \qquad k_0 \left(1 + (N-k)\frac{B_{QQ}(Q_{nc})}{C_{qq}(q_{nc})-(N-k_0)B_{QQ}(Q_{nc})}\right) = 1$$

Solving (10) for k_0 then yields:

$$(11) \qquad k_0 = \frac{C_{qq}(q_{nc})-NB_{QQ}(Q_{nc})}{C_{qq}(q_{nc})-B_{QQ}(Q_{nc})}.$$

Since $B_{QQ} < 0$ and $C_{qq} > 0$, it follows immediately that $k_0 > 0$. Simple rearrangements of (11) show that $k_0 < N$. ■

The intuition behind the first part of the proposition can be illustrated in the (q_{n_j}, Q_{-j})-space as depicted in figure 1. The ray from the origin represents the equilibrium condition of the first stage of the game. Only on this ray the chosen abatement levels of all countries are equal: $Q_{-j}^*(k^*) = (N-1)q_{n_j}^*(k^*)$. Further, the reaction function of non-signatory country *j*, $R_{n_j}(k, Q_{-j})$, is depicted in figure 1. As we showed above, the slope of $R_{n_j}(k, Q_{-j})$ is negative.

insert figure 1 here

Now assume that all the other countries except non-signatory country j choose an aggregate abatement level $Q_{-j} > Q_{nc_{-j}}$. For the equilibrium condition of the first stage of the game to be satisfied, non-signatory country j must choose an abatement level $q_j > q_{nc}$. According to country j's reaction function it will however be optimal to choose an abatement level $q_j < q_{nc}$ and we can conclude that there cannot be an equilibrium with $Q_{-j} > Q_{nc_{-j}}$.

If all the other countries except non-signatory country j choose an aggregate abatement level $Q_{-j} < Q_{nc_{-j}}$, the same logic applies. For the equilibrium condition of the first stage of the game to be satisfied, non-signatory country j must choose an abatement level $q_j < q_{nc}$ but according to its reaction function it will choose an abatement level $q_j > q_{nc}$. Consequently, any situation with $Q_{-j} < Q_{nc_{-j}}$ cannot be an equilibrium either. Hence, the only stable situation is when all countries set the non-cooperative abatement level q_{nc}.

The second part of the proposition states that although there are no gains from cooperation in equilibrium, a positive number of countries will form a coalition. The intuition behind this result is the following: The "first-mover" advantage for a signatory country consists in having the opportunity to set an abatement level before all other countries. This advantage is exploited by the signatory countries by setting lower abatement levels than the non-signatory countries (as long as $k < k_0$), which implies higher payoffs to the members of the coalition then to the non-signatory countries. In fact, as we show in the appendix, the first countries forming a coalition will choose a lower abatement effort then the non-cooperative abatement level q_{nc} and the optimal reaction for the non-signatory countries is then to choose an abatement level higher then q_{nc}. In such a situation the signatory countries take a free-ride on the abatement effort of the non-signatory countries.[10] But as long as the payoff of the members of the coalition is higher than the payoff of the non-signatory countries, there is always an incentive for non-signatory countries to join the coalition. An increasing coalition size however implies a declining "first-mover" advantage and hence a rising abatement level of the coalition. As a result, the number of countries joining the coalition is exactly the fraction it takes to push up the abatement effort of the signatory countries to the non-cooperative abatement level q_{nc}.

As we have shown, there are no gains from cooperation when the game is played with identical countries and coalition size is treated as a continuous variable. Since

[10] From this point of view the expression "coalition for the environment" is misleading, as all that the signatory countries do, is maximizing their payoff irrespective of the effect on the environment.

the resulting coalition size is in general no integer number, gains from cooperation resulting in simulations must consequently be due to the fact that coalition size is treated as a discrete variable. The connection between the equilibrium in the continuous and the discrete version of the game will be the subject of section 5.

4 International Environmental Negotiations Modeled as a Simultaneous Game

As mentioned in the introduction, the second stage of the game cannot only be modeled as a sequential game, but also as a simultaneous game. The difference to the sequential version of the game consists in signatory and non-signatory countries choosing their abatement level simultaneously. This implies that signatory countries no longer possess a "first-mover" advantage. Consequently, to find the equilibrium conditions of the second stage of the game, we need to compute the reaction function of the non-signatory and the signatory countries. While the optimization problem of the non-signatory countries is the same as in the sequential version of the game (see Eq. (1) and (2)), the optimization problem of each signatory country i now amounts to

$$(12) \qquad \max_{q_{s_i}} kB(Q_{-i} + q_{s_i}) - C(q_{s_i}),$$

where $Q_{-i} = Q - q_{s_i}$ denotes the aggregate abatement level of all countries except country i. We denote the solution of this optimization problem by $R_{s_i}(k, Q_{-i})$, the reaction function of signatory country i. The optimal choice of the abatement level $q^*_{s_i}$ then requires that the marginal benefits from abatement equal the marginal costs of abatement, i.e.

$$(13) \qquad kB_Q\left(Q_{-i} + R_{s_i}(k, Q_{-i})\right) = C_q\left(R_{s_i}(k, Q_{-i})\right).$$

Given the assumption that all countries are identical and thus all signatory countries choose the same individual abatement level q_s and all non-signatory countries choose the same individual abatement level q_n, these equilibrium conditions can be simplified in the same way as shown for the sequential version of the game. Hence, at the second stage the following two conditions have to be satisfied simultaneously:

$$(2a) \qquad B_Q\left(kq_s^*(k) + (N-k)q_n^*(k)\right) = C_q\left(q_n^*(k)\right)$$

$$(13a) \qquad kB_Q\left(kq_s^*(k) + (N-k)q_n^*(k)\right) = C_q\left(q_s^*(k)\right).$$

These two equations can further be simplified to a single equilibrium condition by substituting Eq. (2a) into (13a):

$$(14) \qquad kC_q\left(q_n^*(k)\right) = C_q\left(q_s^*(k)\right).$$

The equilibrium condition of the first stage of the game again requires that the resulting coalition k needs to be "stable". It is the same as in the sequential version of the game (see Eq. (6)):

$$(15) \qquad \Pi_s(k^*) \geq \Pi_n(k^*-1) \text{ and } \Pi_n(k^*) \geq \Pi_s(k^*+1).$$

As in the sequential version of the game, it is usually not possible to solve the resulting equilibrium conditions explicitly for the optimal abatement levels and the size of the stable coalition. Therefore, in order to derive a general solution, we approximate the above game again by assuming a continuum of players in the range $(0, N]$ and compute our "benchmark solution". Technically, the same logic as in the sequential version of the game applies then. We can treat coalition size k as a continuous variable and stability condition (15) simplifies to $\Pi_s(k^*) = \Pi_n(k^*)$. As in section 3.3 (see Eq. (7)), this condition can be written as

$$(16) \qquad B\left(Q^*(k^*)\right) - C\left(q_s^*(k^*)\right) = B\left(Q^*(k^*)\right) - C\left(q_n^*(k^*)\right).$$

Since we assumed monotonicity of the cost function ($C_q > 0$) it follows immediately that in any equilibrium $q_n^*(k^*) = q_s^*(k^*)$ must hold. Thus, as signatory and non-signatory countries must choose the same abatement level in the equilibrium of the game, the only k that satisfies Eq. (14) is $k^* = 1$, i.e. there is only one country in the coalition. But as one can easily see, this "signatory country" then maximizes the same payoff function as each non-signatory country, i.e. it behaves like a non-cooperating country. This implies that in the equilibrium there is no cooperation at all between countries and therefore no gains from cooperation can be realized.

Consequently, the conclusions are the same as in the case of environmental negotiations modeled as a sequential game: the resulting payoffs and abatement levels in the equilibrium are the same as if there was no cooperation between countries. Further, resulting gains from cooperation in the discrete version of this game can only be due to an "integer effect".

5 The "Integer Effect"

The fact that gains from cooperation in numerical simulations are always very small seems to suggest that the resulting coalition size k^* cannot be far from our "benchmark solution" k_0. A reasonable conjecture would thus be that k^* is always equal to k_0 rounded up to the next integer number: $k^* = [k_0] + 1$. Barrett [4] however notes: "*For all the simulations I have run (see Barrett* [3]*),* $k_0 < k^* < k_0 + 2$". [11] This indicates that the stable coalition in the discrete type game can differ by more than one integer from k_0. But if the stable coalition can differ by more then one integer from k_0, then there is *a priori* no argument why the difference between k_0 and k^* cannot amount to more then two integers. As we will show in the following, this indeed is the case. The crucial point of the following discussion is, however, that the difference between k_0 and k^* is of little importance, since coalition size is of minor relevance to gains from cooperation.

To show this, we start by looking at a slightly modified stability condition. The only difference to Eq. (6) is that the integer number 1 is replaced by ε:

$$(17) \qquad \Pi_s(k^*) \geq \Pi_n(k^* - \varepsilon) \text{ and } \Pi_s(k^* + \varepsilon) \leq \Pi_n(k^*).$$

[11] The notation is adapted to our notation.

As one can easily see, $\varepsilon = 1$ corresponds the discrete type model and $\varepsilon = 0$ corresponds the above discussed continuous type model. Instead of considering both parts of the stability condition, it is however sufficient to look at the function

$$(18) \qquad f(k,\varepsilon) \equiv \Pi_s(k^* + \varepsilon) - \Pi_n(k^*)$$

and search for the stable coalition k^*_ε satisfying $f(k^*,\varepsilon) = 0$. Obviously such a coalition k^*_ε satisfies the right part of the stability condition with equality. As we proof in the appendix, such a coalition does however as well satisfy the left part of the stability condition. Totally differentiating $f(k^*,\varepsilon) = 0$ and rearranging terms then yields

$$(19) \qquad \frac{dk^*}{d\varepsilon} = \frac{1}{\frac{\Pi_n^{'}(k^*_\varepsilon)}{\Pi_s^{'}(k^*_\varepsilon+\varepsilon)} - 1} > 0,$$

where $\Pi_{n,s}^{'}(\cdot)$ stands for the derivative of the payoff function with respect to k. Intuitively, Eq. (19) tells us what will happen to the size of the stable coalition, when we go from the continuous to the discrete version of the game. More precise, Eq. (19) allows to measure the difference between k_0, which corresponds to the continuous situation with $\varepsilon = 0$ and the stable coalition k^* with $\varepsilon = 1$.

The fact that $\frac{dk^*}{d\varepsilon}$ is positive has the following intuition: For any coalition size k smaller then k_0, the payoff to the non-signatory countries is smaller then in a situation without cooperation between countries. Therefore, in such a situation there is always an incentive for non-signatory countries to join the coalition and consequently the stable coalition in the continuous and the discrete version of the game cannot be smaller then k_0. A proof of this suggestion can be found in the appendix.

It is now possible to distinguish two cases:

$$\begin{cases} \left(\frac{dk^*}{d\varepsilon}\right)_{\varepsilon_0} \approx 0 \quad \forall \varepsilon_0 \in [0,1] \\ \\ \left(\frac{dk^*}{d\varepsilon}\right)_{\varepsilon_0} >> 0 \;\; \forall \varepsilon_0 \in [0,1]\,. \end{cases}$$

The realization of these two polar cases obviously depends on the ratio of the slope of the two payoff functions. In what follows, we will show that both cases imply negligible gains from cooperation in the equilibrium of the game.

In the first case, when $\frac{dk^*}{d\varepsilon} \approx 0$, nothing happens when we go from the continuous to the discrete version of the game. That is, in such a situation the size of the stable coalition k^* cannot depart by more then one integer from k_0 and consequently gains from cooperation are minor. Intuitively, this case describes a situation where there is a strong incentive for non-signatory countries to free ride on the abatement effort of the coalition.

The second case is the more interesting one. When $\frac{dk^*}{d\varepsilon} >> 0$, the size of the stable coalition in the discrete version of the game can get substantially bigger then in the continuous version. This happens when $\frac{\Pi_n^{'}}{\Pi_s^{'}} \rightarrow 1$, i.e. the payoff function of signatory and non-signatory countries have approximately the same slope and thus run more or less parallel. Since the stable coalition k_ε^* however still satisfies Eq. (18), $\Pi_s(k^* + \varepsilon) = \Pi_n(k^*)$, this implies that the two payoff functions are very near to each other and thus $\Pi_s(k^*) \approx \Pi_n(k^*)$. Consequently, the abatement level chosen by signatory and non-signatory countries must be similar. According to the reaction function of a non-signatory country this can however only be the case, when the chosen abatement levels are close to the non-cooperative abatement level q_{nc}. Intuitively, this case corresponds to a situation where there is not much to gain from an environmental agreement, i.e. efficiency gains are minor even if all affected countries would cooperate.

To summarize, the coalition in the discrete version of the game can get substantially bigger then in the continuous version. This is however merely an "integer effect", since a bigger coalition does not enhance efficiency gains, i.e. the size of the stable coalition in the equilibrium of the game is irrelevant to gains from cooperation.

6 Summary and Outlook

We have reconsidered two versions of a widely used model of international environmental negotiations on the reduction of CO_2-emissions. Due to the mathematical problems of giving a full characterization of the solution of these games, up to now most of the work on this subject rested on numerical simulations to derive results. We have shown for a general class of payoff functions that when these models are approximated by assuming a continuum of players, an analytical solution can be found. In both versions of the model, there exist stable coalitions but there are no gains from cooperation at all. Using the solution of this modified version of the game, we have further shown that gains from cooperation in simulations result from an "integer effect", i.e. from the fact that coalition size is treated as a discrete variable. However, the stable coalition in the discrete version of the game needs not necessarily to be near to the stable coalition in the continuous version. Since a bigger coalition in the discrete game then in the continuous game does not imply any substantial efficiency gains it is, however, not surprising that these gains are extremely small in virtually all simulations.

In the light of these results it seems doubtful, whether it is possible to enlarge coalition size and improve environmental quality in the above model, as suggested by some authors. Carraro and Siniscalco [9] e.g. consider the possibility to use resulting gains from cooperation to induce non-signatory countries to join the coalition. But, since there are hardly any gains from cooperation which could be redistributed, such a strategy seems very limited. Further, from a theoretical point of view, this model cannot be used to motivate the existence of coalitions as a starting point for further research as several authors suggest. In most of this work, see e.g. Bohm [7] or Hoel [13], the process of coalition formation is neglected and different policy measures available to a group of cooperating countries trying to reduce environmental pollution are analyzed. Although the assumption that there might be groups of countries willing to cooperate *per se* seems reasonable, it cannot be theoretically founded with the discussed model, since in the equilibrium of this game no country has an incentive to contribute to a reduction of environmental pollution.

To summarize, one can conclude that the structure of the discussed model is not well suited to explain international environmental cooperation as observed in the real world. The unsatisfying results of the model are however strongly driven by the assumption of identical countries, which implies that in the equilibrium of the game all countries must set similar abatement levels. This of course is not a very realistic assumption when studying negotiations on the reduction of CO_2-

emissions. According to Barrett [6], the United States, the European Union and the former USSR account for about one-half of global CO_2-emissions. Therefore, an interesting question is, whether the above results are still likely to hold, when heterogeneity between countries is introduced. Several authors, see e.g. Barrett [6] or Botteon and Carraro [10], have run simulations with heterogeneous countries, but their results do not substantially differ from the ones obtained in simulations with identical countries. However, from a theoretical point of view, the effect of heterogeneity on the equilibrium of the game seems not entirely clear. Generally, there are two crucial differences between the case with identical and asymmetric countries. First and trivially, it will no longer be true that in the equilibrium all signatory and all non-signatory countries choose the same abatement level. Second, with asymmetric countries there is the possibility of multiple equilibria i.e. there might exist several compositions of the coalition which are stable. If resulting gains from cooperation in simulations with asymmetric countries are real efficiency gains or if they are, as in the case of identical countries, just the result of an "integer effect", is subject to further research.

7 Appendix

In section 5, we introduced a slightly modified stability condition (see Eq. (17)) and claimed that it is sufficient to consider the function $f(k_\varepsilon^*, \varepsilon)$, see Eq. (18), to show that resulting gains from cooperation in numerical simulation are due to an "integer effect". We further established the result that k_ε^*, the solution to the problem $f(k_\varepsilon^*, \varepsilon) = 0$, satisfies the right part of the stability condition (17) with equality. In this appendix we show that k_ε^* satisfies as well the left part of stability condition (17).

To proof this, we first have to show that there cannot exist a stable coalition smaller then k_0, i.e. $k_\varepsilon^* > k_0$. In order to do so, two propositions are considered. The first establishes the fact that for any coalition size smaller then k_0, the payoff to non-signatory countries is smaller then in a situation without cooperation at all. The second proposition states that the payoff function of the signatory countries has a minimum at k_0. This two results together allow the desired conclusion that $k_\varepsilon^* > k_0$. It is then only a small step to proof that k_ε^* satisfies the left part of the stability condition.

Proposition 2 *For any coalition size smaller then k_0, the payoff to the non-signatory countries is smaller then the non-cooperative payoff Π_{nc}, i.e. the pay-off resulting in a situation without any cooperation between countries: $\Pi_n(k) < \Pi_{nc}\ \forall k < k_0$.*

Proof. As we proved in section 3.3, the stable coalition k_0, with signatory and non-signatory countries choosing the non-cooperative abatement level q_{nc}, is the unique subgame perfect equilibria in the continuous version of the game. This implies that $\forall k \neq k_0$ signatory and non-signatory countries choose different abatement levels. More precise, the uniqueness of the equilibria implies that $\forall k \neq k_0$ either $q_s^*(k) < q_n^*(k)$ or $q_s^*(k) > q_n^*(k)$.

Now consider a coalition $k < \min\{1, k_0\}$. Substituting Eq. (2a), the equilibrium condition of the non-signatory countries, into Eq. (5a), the equilibrium condition of the signatory countries, yields

$$\text{(A1)} \qquad k\left(1 + (N-k)\frac{\partial q_n(k, Q_s^*)}{\partial Q_s}\right) C_q\left(q_n^*(k)\right) = C_q\left(q_s^*(k)\right).$$

Since the reaction function has a negative slope, the term in the bracket is always positive but smaller then one and since $k < \min\{1, k_0\}$, it follows that $\forall k < k_0$ the marginal abatement costs of the signatory countries are smaller then the marginal abatement costs of the non-signatory countries. Further, since we assumed monotony of the cost function ($C_q > 0$), it follows immediately that $\forall k < k_0$: $q_s^*(k) < q_n^*(k)$. This implies then by definition of the reaction function: $q_s^*(k) < q_{nc} < q_n^*(k)\ \forall k < k_0$. Using Eq. (2a), the equilibrium condition of the non-signatory countries, again, we can thus conclude that

$$\text{(A2)} \qquad B_Q\left(Q^*(k)\right) = C_q\left(q_n^*(k)\right) > C_q\left(q_{nc}\right) = B_Q\left(Q_{nc}\right)\ \ \forall k < k_0,$$

(since $C_{qq} > 0$) and hence $Q^*(k) < Q_{nc}\ \forall k < k_0$ (since $B_{QQ} < 0$). This establishes the desired result that $\Pi_n(k) < \Pi_{nc}\ \forall k < k_0$. ■

Proposition 3 *For any other coalition size then k_0, the payoff to the signatory countries is bigger then the non-cooperative payoff Π_{nc}: $\Pi_s(k) > \Pi_{nc}\ \forall k \neq k_0$. That is, k_0 is a minimum of $\Pi_s(k)$.*

Proof. To proof proposition 2 it is necessary to show that the payoff function $\Pi_s(k)$ has a minimum at k_0. To show this we need to compute a interpretable expression for $\frac{d\Pi_s(k)}{dk}$.[12]

As a preliminary step, we compute two derivatives that will proof helpful in the following. Differentiating the reaction function of the non-signatory countries with respect to k yields the following expression:

$$\text{(A3)} \qquad \frac{dq_n^*}{dk} = \frac{dR_n(k,Q_s^*)}{dk} = \frac{\partial R_n(k,Q_s^*)}{\partial k} + \frac{\partial R_n(k,Q_s^*)}{\partial Q_s}\frac{dQ_s^*}{dk}.$$

By totally differentiating Eq. (2a), the equilibrium condition for the non-signatory countries, the first term on the right hand side of Eq. (A3) can be split up further to

$$\text{(A4)} \qquad \frac{\partial R_n(k,Q_s^*)}{\partial k} = \frac{B_{QQ}(q_s-q_n)}{C_{qq}-(N-k)B_{QQ}} = (q_s-q_n)\frac{\partial R_n(k,Q_s)}{\partial Q_s},$$

since $\frac{\partial R_n(k,Q_s^*)}{\partial Q_s} = \frac{B_{QQ}}{C_{qq}-(N-k)B_{QQ}}$ (see Eq. (9)).

We then start by differentiating the payoff function of a signatory country and substituting Eq. (A3) and (A4) in succession. Note further that $\frac{dQ_s^*}{dk} = k\frac{dq_s^*}{dk}$.

$$\text{(A5)} \qquad \frac{d\Pi_s(k)}{dk} = B_Q(Q^*)\frac{dQ^*}{dk} - C_q(q_s^*)\frac{dq_s^*}{dk}$$

$$= B_Q(Q^*)\left(q_s^* - q_n^* + k\frac{dq_s^*}{dk} + (N-k)k\frac{dq_n^*}{dk}\right) - C_q(q_s^*)\frac{dq_s^*}{dk}$$

$$= B_Q(Q^*)\left[q_s^* - q_n^* + k\frac{dq_s^*}{dk} + (N-k)\left(\frac{\partial R_n(k,Q_s^*)}{\partial k} + \frac{\partial R_n(k,Q_s^*)}{\partial Q_s}\frac{dQ_s^*}{dk}\right)\right]$$
$$-C_q(q_s^*)\frac{dq_s^*}{dk}$$

$$= \frac{dq_s^*}{dk}\left[kB_Q(Q^*)\left(1 + (N-k)\frac{\partial R_n(k,Q_s^*)}{\partial Q_s}\right) - C_q(q_s^*)\right] +$$
$$B_Q(Q^*)\left(q_s^* - q_n^* + (N-k)(q_s^* - q_n^*)\frac{\partial R_n(k,Q_s^*)}{\partial k}\right)$$

It is easy to see that the expression in the first bracket is exactly the equilibrium condition for a signatory country (see Eq. (5a)), which has to be equal to zero in the optimum. Eq. (A5) thus simplifies to:

[12] For ease of notation, the dependence of abatement levels and payoffs on the coalition size k will be omitted in the following.

(A5.1) $\frac{d\Pi_s(k)}{dk} = (q_s^* - q_n^*)B_Q(Q^*)\left(1 + (N-k)\frac{\partial R_n(k,Q_s^*)}{\partial k}\right)$.

Using Eq.(5a) again, this expression simplifies further to:

(A5.2) $\frac{d\Pi_s(k)}{dk} = \frac{1}{k}(q_s^* - q_n^*)C_q(q_s^*)$.

From Eq. (A5.2) it now easy to see that $\Pi_s(k)$ has a minimum at k_0:

$$\begin{cases} \frac{d\Pi_s(k)}{dk} < 0 \ \forall k < k_0 \quad \text{since } q_s^* < q_{nc} < q_n^* \\ \\ \frac{d\Pi_s(k)}{dk} = 0 \quad k = k_0 \quad \text{since } q_s^* = q_{nc} = q_n^* \\ \\ \frac{d\Pi_s(k)}{dk} > 0 \ \forall k > k_0 \quad \text{since } q_s^* > q_{nc} > q_n^*. \end{cases}$$

■

Proposition 1 and 2 together imply that $\Pi_s(k) > \Pi_n(k)\ \forall k < k_0$. Further, since $\Pi_s(k) > \Pi_{nc}\ \forall k \neq k_0$ it follows $\forall \varepsilon > 0$ that $\Pi_s(k+\varepsilon) > \Pi_n(k)\ \forall k < k_0$, i.e. the right part of the stability condition (17) cannot be satisfied for a $k < k_0$. Therefore we can conclude that there cannot exist a stable coalition involving less then k_0 countries and consequently: $k_\varepsilon^* > k_0\ \forall \varepsilon > 0$.

With this result it is now easy to show that k_ε^*, the solution to the problem $f(k^*, \varepsilon) = 0$, satisfies the right and the left part of stability condition (17). The left part of the stability condition requires $\Pi_s(k^*) - \Pi_n(k^* - \varepsilon) \geq 0$. This is however equivalent to demanding $f(k, \varepsilon) > 0$ for $k = k_\varepsilon^* - \varepsilon$. Note that substituting $k = k_\varepsilon^* - \varepsilon$ into the right part of the stability condition yields exactly the left part of the stability condition. Now, since k_ε^* satisfies $f(k^*, \varepsilon) = 0$, this function is either strictly positive or negative for any coalition size k smaller then k_ε^*. It is therefore not necessary to check if $f(k, \varepsilon) > 0$ exactly at $k = k_\varepsilon^* - \varepsilon$, but it is sufficient to check this condition for any given $k < k_\varepsilon^*$. Our candidate of course is k_0, which is smaller then k_ε^*, as we proved above. Using the fact that $\Pi_s(k) > \Pi_{nc}\ \forall k \neq k_0$ and $\Pi_n(k_0) = \Pi_{nc}$, it is straight forward to conclude that $\Pi_s(k_0 + \varepsilon) - \Pi_n(k_0) > 0$ and thus $f(k, \varepsilon) > 0\ \forall k < k_\varepsilon^*$. Hence, coalition size k_ε^* satisfies the left and the right part of the stability condition (17).

References

[1] C.A. d'Aspremont, J. Jacquemin, J. Gabszeweiz and J.A. Weymark, On the stability of collusive price leadership, *Can. J. Econom.* **16**, 17-25 (1983).

[2] S. Barrett, International Environmental Agreements as Games, *in* "Conflicts and Cooperation in Managing Environmental Resources" (R. Pething, Ed.), Springer-Verlag, Berlin (1992).

[3] S. Barrett, Self-enforcing international environmental agreements, *Oxford Econom. Papers* **46**, 878-894 (1994).

[4] S. Barrett, Towards a theory of international environmental cooperation, *in*: "New directions in the economic theory of the environment" (C. Carraro and D. Siniscalco, Ed.), Cambridge University Press (1997).

[5] S. Barrett, The strategy of trade sanctions in international environmental agreements, *Resource and Energy Economics* **19**, 345-361 (1997).

[6] S. Barrett, Heterogeneous international environmental agreements, *in*: "International Environmental Agreements: Strategic Policy Issues" (C. Carraro, Ed.), E. Elgar, Cheltenham (1998).

[7] P. Bohm, Incomplete International Cooperation to Reduce CO2 Emissions: Alternative Policies, *J. Env. Econom. Management* **24**, 258-271(1993).

[8] M. Botteon and C. Carraro, Burden sharing and coalition stability in environmental negotiations with asymmetric countries, *in*: "International Environmental Agreements: Strategic Policy Issues" (C. Carraro, Ed.), E. Elgar, Cheltenham (1998).

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

[10] C. Carraro and D. Siniscalco, International environmental agreements: Incentives and political economy, *European Econom. Rev.* **42**, 561-572 (1998).

[11] H. Cesar and A. De Zeeuw, Issue linkage in global environmental problems, *in* "Economic Policy for the Environment and Natural Resources" (A. Xepapadeas, Ed.), E. Elgar, Cheltenham (1994).

[12] M. Escapa and M.J. Gutiérrez, Distribution of Potential Gains from International Environmental Agreements: The Case of the Greenhouse Effect, *J. Env. Econom. Management* **33**, 1-16 (1997).

[13] M. Hoel, Efficient Climate Policy in the Presence of Free Riders, *J. Env. Econom. Management* **27**, 259-274 (1994).

[14] F. Missfeldt, Game-theoretic Modelling of Transboundary Pollution, *J. Econom. Surveys* **13**, 287-321 (1999).

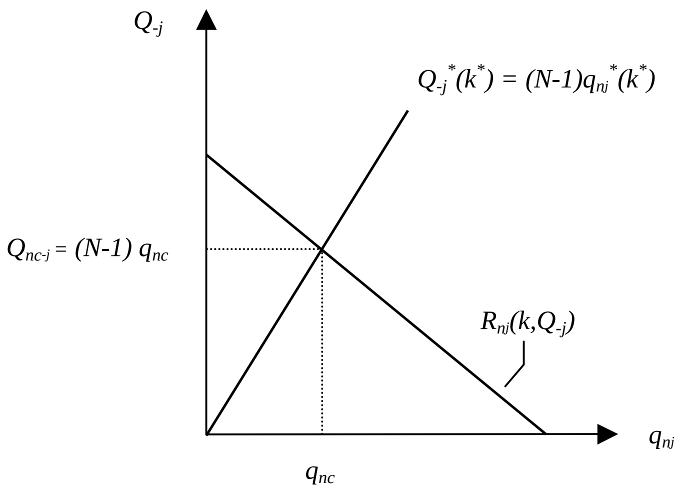

Figure 1