International environmental agreements with asymmetric countries: climate clubs vs. global cooperation

Achim Hagen ∗†, Klaus Eisenack *

November 13, 2014

Abstract

We investigate whether global cooperation for emission abatement can be improved if asymmetric countries can sign different parallel environmental agreements. The analysis assumes a two-stage game theoretical model with two country types. In the first stage, countries choose whether to sign one of two international environmental agreements (IEAs) or to be a non-signatory. In the second stage, the coalitions decide about their emission abatement, with each coalition acting as a unitary actor in a non-cooperative Nash game between both coalitions and the non-signatories. Conditions for self-enforcing sets of agreements and the resulting total emission abatement are determined. The results are sensitive to our assumptions on the benefits from abatement. For constant marginal benefits, the possibility of multiple agreements increases the number of cooperating countries and total abatement (compared to the standard case with a single agreement). For decreasing marginal benefits, total emissions are independent of the number of admitted agreements. The paper thus contributes to the emerging discussion on the scope and limits of climate clubs.

∗Department for Economics and Statistics, Carl von Ossietzky University Oldenburg, Germany
†achim.hagen@uni-oldenburg.de
1 Introduction

Most studies on international environmental agreements (IEAs) assume that there is (at most) one self-enforcing coalition that abates emissions. Most of the theoretical literature further assumes symmetric countries. In light of the slow progress in international climate negotiations, however, the idea of climate clubs is getting increasing attention (e.g. Weischer et al., 2012; Ostrom, 2012; Widerberg and Stenson, 2013). Multiple parallel agreements of subsets of nation states might promise more contributions to the global public good. In particular, asymmetric countries may sort into clubs with similar or complementary properties, and thereby increase stability of these clubs. Existing studies with asymmetric countries (but at most one coalition) have shown that in some cases global cooperation can be improved (Barrett, 2001; Eisenack and Kähler, 2012; Heugues, 2012). To our knowledge, the idea of climate clubs has not been generally analyzed in the IEA literature so far.

Our theoretical paper explores the potential of climate clubs by assuming two types of countries and allowing for two disjoint IEAs. Each country of either type can chose whether to join an agreement with other countries of the same type or to sign none of them. Each agreement is framed as a (stable or unstable) coalition, and its members act cooperatively. The Nash game between the two coalitions and the non-signatories is non-cooperative.

Our research thus extends the seminal work of Barrett (1994) and Carraro and Siniscalco (1993) within the latter stream in that it analyses the internal and external stability of coalitions (D’Aspremont et al., 1983) in a setting with simultaneous play. There is no Stackelberg leadership of one or the other coalition.

We only know of a small literature that investigates the case of multiple IEAs. Asheim et al. (2006) model the case of symmetric countries and two agreements. The countries are partitioned in two regions, and can chose whether they sign an agreement for that region or not. Marginal benefits of abatement are constant. The model is solved as an infinitely repeated game under different institutional assump-
tions and renegotiation-proof agreements are identified. For two coexisting agreements, a larger number of cooperating signatories can be sustained, compared to the standard case of a single IEA. Also Osmani and Tol (2010) allow for two coalitions, but additionally consider two country types. They assume a three-stage sequence of play of the coalitions and the non-signatories. Their paper mostly focuses on procedures to compute stable coalitions numerically. They show, by computing some numerical examples, that the possibility of two coalitions can both increase and decrease emission abatement compared to the standard case with one coalition.

The paper starts from a different setting and derives analytical results for two specified classes of benefit functions (constant and decreasing marginal benefits of abatement). In that we are more general than previous work. By choosing a Nash setting we avoid the tedious question about which of the coalitions moves first. In doing so, we confirm but also qualify some of the results from the studies with different assumptions. Inter alia, we find that for constant marginal benefits, total abatement increases, while it remains identical to the standard case for decreasing marginal benefits.

The next section is devoted to the constant marginal benefits case. We first derive the standard case as a benchmark. Then, for two coalitions, we first solve the second stage abatement game, and subsequently the first stage coalition game. The subsequent Section 3 considers the decreasing marginal benefits case. Again, we first derive the standard case as a benchmark, followed by the second stage abatement game for two coalitions. For the first stage game some crucial stability criteria are then derived and interpreted. They are sufficient to deduce the total emissions in a self-enforcing agreement. A summary and discussion concludes the paper.
2 Climate clubs with constant marginal benefits of abatement

In this section we assume the standard two-stage game structure (see Carraro and Siniscalco, 1993) with countries choosing first to be a signatory or non-signatory. In the second stage the signatories choose cooperatively between playing pollute or abate. The model assumptions for the case of one agreement follow Barrett (2001), extended by allowing for two parallel agreements.

There are $N$ countries with $N_1$ type 1 and $N_2$ type 2 countries. If a type $i$ country ($i = 1, 2$) plays abate, it gets the payoff

$$\Pi_i^A = -c + \alpha_i(z_1 + z_2),$$  \hspace{1cm} (1)

with the number of countries of type $i$ that play abate denoted by $z_i$. Countries that play pollute get the payoff

$$\Pi_i^P = \alpha_i(z_1 + z_2).$$  \hspace{1cm} (2)

The additional benefits from one more type 1 country playing abate are equal to the benefits from one more type 2 country playing abate. This refers to the case of a global public good, e.g. a pollutant that has a global impact no matter where it is emitted, as is the case for greenhouse gases. The asymmetry of the countries is expressed by the parameter $\alpha_i$ where $\alpha_2$ is normalized to $\alpha_2 = 1$ and $\alpha_1 \in [0, 1]$. A type 2 country therefore benefits at least as much as a type 1 country from the abatement. It is assumed that the abatement costs $c > 1$, and that the net benefit of the own abatement of each country $i$, $-c + \alpha_i$, is therefore negative. Thus, playing pollute is the dominant strategy if there is no IEA. The Nash equilibrium of a one shot emission game would consequently be unique with all countries playing pollute.
2.1 The standard case with one agreement

As usual, we proceed by backward induction, solving the second stage of the game first. In the case of one agreement with \( k_1 \) type 1 signatories and \( k_2 \) type 2 signatories, the non-signatories all play pollute as a dominant strategy. The aggregate payoff of the signatories is

\[
\Pi^S = -cz_1 + \alpha_1 k_1(z_1 + z_2) - cz_2 + k_2(z_1 + z_2).
\]  

(3)

The signatories maximize their payoff \( \Pi^S \) cooperatively with respect to \( z_i \). The linear payoff function implies the corner solution \( z_i^* = k_i \) if

\[
\alpha_1 k_1 + k_2 > c.
\]  

(4)

Here and in the following, variable referring to the game equilibrium with one agreement are denoted with an \(*\). All signatories of either type play pollute (\( z_i^* = 0 \)) if

\[
\alpha_1 k_1 + k_2 < c.
\]  

(5)

To solve the first stage of the game the criteria of internal and external stability (following D’Aspremont et al. (1983)) are applied. In accordance with these, an agreement is stable if no signatory has an incentive to leave the agreement (internal stability) and no non-signatory wants to join the existing agreement (external stability). Formally, an abating coalition is internally stable if

\[
\Pi_i^A(k_i) > \Pi_i^P(k_i - 1).
\]  

(6)

As playing pollute is a dominant strategy for non-signatories, this condition is only fulfilled if the signatories choose to abate and would decide to pollute if one signatory would leave the agreement. This leads to 'linchpin' equilibria as the withdrawal
of one country would change the decision of all the other signatories from abate to pollute. From condition (4) and (5) we see that \((k_1^*, k_2^*)\) represents a stable and abating coalition if condition (4) holds and if

\[
\begin{align*}
c > \alpha_1 (k_1^* - 1) + k_2^*, \quad (7) \\
c > \alpha_1 k_1^* + (k_2^* - 1). \quad (8)
\end{align*}
\]

Condition (7) implies condition (8) so that internal stability for a single coalition is given if

\[
c + \alpha_1 > \alpha_1 k_1^* + k_2^* > c. \quad (9)
\]

The criterion of external stability is implied by this condition because playing pollute is a dominant strategy for non-signatories and therefore a non-signatory has no incentive to join an abating and internally stable agreement.

It would be interesting to know conditions where only countries of the same type would sign the same agreement. This can be seen by setting one coalition to size zero in (9). An abating coalition with only type 1 signatories \((z_2^* = k_2 = 0)\) is thus possible for

\[
c + \alpha_1 > \alpha_1 k_1^* > c, \quad (10)
\]

and with only type 2 countries \((z_1^* = k_1 = 0)\) if

\[
c + 1 > k_2^* > c. \quad (11)
\]

Conditions (9) to (11) show that a stable agreement could either consist of both types of countries or of countries of only one type. If all of these three types of stable agreements are possible, the Nash equilibrium of the complete game is not unique.
2.2 Abatement decisions and stable coalitions with two agreements

We now assume the possibility of two parallel agreements that take their abatement decisions independently but cooperate internally. Again we proceed by backward-induction, solving the second stage of the game first. Agreement 1 consists of \( k_1 \) type 1 countries and agreement 2 of \( k_2 \) type 2 countries. The aggregate payoff of agreement 1 is thus

\[
\Pi^S_1 = -cz_1 + \alpha_1 k_1 (z_1 + z_2) .
\]  

(12)

Maximization of \( \Pi^S_1 \) leads to the corner solutions

\[
z_1 = \begin{cases} 
  k_1 \text{ (abate) if } \alpha_1 k_1 > c, \\
  0 \text{ (pollute) if } \alpha_1 k_1 < c.
\end{cases}
\]  

(13)

By analogy, the \( k_2 \) signatories of agreement 2 play

\[
z_2 = \begin{cases} 
  k_2 \text{ (Abate) if } k_2 > c, \\
  0 \text{ (Pollute) if } k_2 < c.
\end{cases}
\]  

(14)

We see that the decisions of each agreement \( i \) depend on the number of its signatories \( k_i \) and on the abatement costs \( c \) but are mutually independent.

The first stage of the game is now solved by applying the criteria of internal and external stability in analogy to the case of one agreement. As the abatement decisions of the two agreements are mutually independent, the conditions for both agreements to be internally stable can be reduced to (10) and (11). Like in the case of one agreement, the criterion of external stability is always satisfied because playing pollute is a dominant strategy for non-signatories of an abating agreement so that they have no incentive to join an abating agreement.
We now compare the game equilibrium in the standard case with that of two agreements. A set of stable and abating agreements is denoted by \((k_1^{**}, k_2^{**})\). Here and in the following, variables referring to the game equilibrium with two agreements are denoted with \(**\). By adding (10) and (11), we find that

\[
2c + \alpha_1 + 1 > \alpha_1 k_1^{**} + k_2^{**} > 2c,
\]

holds. This allows to compare with the case of one agreement. For convenience, we use the notation \(K^{**} := \alpha_1 k_1^{**} + k_2^{**}\) to represent a measure for the total abatement by all coalitions. We see from (9) and (15) that \(K^{**} > K^*\), such that the total number of abating countries in the case of two agreements is greater than in the case of one agreement. Thus, the main conclusion of this section is that global cooperation benefits from parallel agreements if marginal benefits from abatement are constant\(^1\).

### 3 Country clubs with decreasing marginal benefits of abatement

To get a broader picture of the possible outcomes with two parallel agreements, this section assumes a payoff-function with decreasing marginal benefits from abatement. The costs of abatement \(c\) for a country are still held to be constant. The

\(^1\) We may ask whether countries in a set of two abating coalitions have an incentive to swap their agreement. This question can be answered as follows: If one signatory country of type \(i\) would change the agreement, the number of signatories \(k_i\) would decrease by one, while the number of signatories in the other agreement would increase. We saw that a decrease in the number of signatories \(k_i\) would change the decision of the remaining signatories from abate to pollute. As a consequence the total number of abating countries would decrease from \(k_i^{**} + k_j^{**}\) to \(k_j^{**} + 1\). Thus, the profit of every country would decrease. For this reason, no signatory-country has an incentive to change the agreement and thereby reduce its own profit.
payoff for a country playing pollute is

$$\Pi^P = \alpha_i[(z_1 + z_2) - d(z_1 + z_2)^2],$$

(16)

whereas a country playing abate gets the payoff

$$\Pi^A = -c + \alpha_i[(z_1 + z_2) - d(z_1 + z_2)^2].$$

(17)

Once again, \(\alpha_2\) is normalized to unity, and \(\alpha_1 \in [0, 1]\) so a type 2 country benefits at least as much as a type 1 country from abatement. The parameter \(d\) is set to \(0 \leq d \leq 1/2\). It is again assumed that abatement costs \(c > 1\), so that the net benefit of a single country playing abatement is negative. Thus, playing pollute is again the dominant strategy for non-signatories. As before, the signatories of either agreement decide jointly about their abatement in the second game stage.

### 3.1 The standard case with one agreement

In the second stage, the signatories cooperatively choose their abatement strategy in order to maximize their aggregated payoff

$$\Pi^s = -c(z_1 + z_2) + (\alpha_1 k_1 + k_2)[(z_1 + z_2) - d(z_1 + z_2)^2],$$

(18)

with respect to \(z_1, z_2\). As before, the agreement consists of \(k_1\) type 1 signatories and \(k_2\) type 2 signatories. Since non-signatories play pollute as a dominant strategy, the total number of countries playing abate is \(Z = z_1 + z_2\), and \(Z^*\) denotes the optimum. It follows from the first order conditions that

$$Z^* = \frac{1}{2d} \left(1 - \frac{c}{\alpha_1 k_1 + k_2}\right).$$

(19)
To solve the first stage, we need to specify a cost sharing rule for the signatories in order to determine the profit for each signatory of a mixed agreement. We assume that each signatory gets the part of the agreement’s total profit that is proportional to its own contribution to this total profit. Thus, countries with lower benefits from the abatement also bear a smaller share of the costs. The signatories’ profit per country $\Pi_{spc}^i$ is consequently

$$
\Pi_{spc}^1 = \alpha_1 \frac{\Pi^S}{\alpha_1 k_1 + k_2} = \Pi_1^P (Z) - \alpha_1 \left( \frac{c \left( \frac{1}{2d} \left( 1 - \frac{c}{\alpha_1 k_1 + k_2} \right) \right)}{\alpha_1 k_1 + k_2} \right), 
$$

(20)

for type 1 countries and

$$
\Pi_{spc}^2 = \alpha_1 \frac{\Pi^S}{\alpha_1 k_1 + k_2} = \Pi_2^P (Z) - \left( \frac{c \left( \frac{1}{2d} \left( 1 - \frac{c}{\alpha_1 k_1 + k_2} \right) \right)}{\alpha_1 k_1 + k_2} \right), 
$$

(21)

for type 2 countries.

From equation (19) we see that the total number of abating countries $Z$ is a function $f(K)$ of $K = \alpha_1 k_1 + k_2$ with $f(K) = \frac{1}{2d} \left( 1 - \frac{c}{K} \right)$. Internal stability is given if the following conditions are fulfilled:

$$
\Pi_{spc}^1 (Z^*) > \Pi_1^P (f(K^* - \alpha_1)), 
$$

(22)

$$
\Pi_{spc}^2 (Z^*) > \Pi_2^P (f(K^* - 1)). 
$$

(23)

The agreement is externally stable if the following conditions hold:

$$
\Pi_1^P (Z^*) > \Pi_{spc}^1 (f(K^* + \alpha_1)),
$$

(24)

$$
\Pi_2^P (Z^*) > \Pi_{spc}^2 (f(K^* + 1)).
$$

(25)

These conditions will be used later to compare the case without clubs with the case of two parallel agreements.
3.2 Abatement decisions with two agreements

For the analysis of the abatement decisions with two agreements we assume again that agreement 1 consists of \( k_1 \) type 1 countries and agreement 2 of \( k_2 \) type two countries. This section analyzes the second stage of the game while the first stage will be analyzed in the subsequent section. The two agreements take their abatement decisions simultaneously and cooperate internally. The aggregated payoff \( \Pi^S_1 \) of the \( k_1 \) signatories of agreement 1 is

\[
\Pi^S_1 = -cz_1 + \alpha_1 k_1 [(z_1 + z_2) - d(z_1 + z_2)^2].
\] (26)

Maximization of \( \Pi^S_1 \) with respect to \( z_1 \) gives the number \( z_1^* \) of signatories that will be chosen to abate. Thus

\[
\frac{\partial \Pi^S_1}{\partial z_1} = -c + \alpha_1 k_1 [1 - 2d(z_1 + z_2)],
\] (27)

yields the reaction function

\[
z_1 = R_1(z_2) = \frac{1}{2d} \left( 1 - \frac{c}{\alpha_1 k_1} \right) - z_2,
\] (28)

as a best response of agreement 1 to the abatement decision \( z_2 \) of agreement 2. Note that it is sufficient to represent to reaction with just the other agreements decisions as argument, as non-signatories play pollute as dominant strategy. The aggregate payoff of the signatories of agreement 2 is

\[
\Pi^S_2 = -cz_2 + k_2 [(z_1 + z_2) - d(z_1 + z_2)^2],
\] (29)

which yields the reaction function

\[
z_2 = R_2(z_1) = \frac{1}{2d} \left( 1 - \frac{c}{k_2} \right) - z_1.
\] (30)
An internal Nash equilibrium, derived from the mutual best responses (28) and (30), consequently exists if

$$\alpha_1 k_1 = k_2.$$  \hspace{1cm} (31)

If this condition holds, every possible combination of abating countries on the reaction functions is a Nash equilibrium. Then $R_1$ is the inverse function of $R_2$ and the total number of countries playing abate in the two parallel agreements is

$$Z^{**} = \frac{1}{2d} \left( 1 - \frac{c}{\alpha_1 k_1} \right) = \frac{1}{2d} \left( 1 - \frac{c}{k_2} \right). \hspace{1cm} (32)$$

If condition (31) does not hold, then there exists no combination of mutually best responses where both agreements choose to abate. Therefore, only countries of one agreement abate. The resulting corner solutions imply that the total number of countries playing abate $Z^{**}$ is either

$$z_1^{**} = 0 \text{ and } z_2^{**} = \frac{1}{2d} \left( 1 - \frac{c}{k_2} \right) = Z^{**}, \hspace{1cm} (33)$$

or

$$z_2^{**} = 0 \text{ and } z_1^{**} = \frac{1}{2d} \left( 1 - \frac{c}{\alpha_1 k_2} \right) = Z^{**}. \hspace{1cm} (34)$$

In both of these cases only one of the two agreements abates, while the other can be regarded as a pseudo-agreement (there is no difference between being a non-signatory or a signatory of the pseudo-agreement). This is due to the parallel reaction functions $R_1$ and $R_2$. We see that

$$\text{if } \alpha_1 k_1 > k_2 \text{ then } z_2^{**} = 0 \text{ and only agreement 1 abates,} \hspace{1cm} (35)$$

$$\text{if } \alpha_1 k_1 < k_2 \text{ then } z_1^{**} = 0 \text{ and only agreement 2 abates.} \hspace{1cm} (36)$$
3.3 Criteria for coalition stability of two agreements and comparison to the case without clubs

We now turn to the stability criteria of the two parallel agreements. If a country of type $i$ leaves its agreement (so that $k_i = k^{**} - 1$) its profit will be at least $\Pi^p_1 (f(k^{**} - 1)$ because if $\alpha_i(k_i^{**} - 1) \geq \alpha_j k_j^{**}$, agreement $i$ will still abate, so that $Z = f(k_i^{**} - 1)$. Otherwise $\alpha_i(k_i^{**} - 1) < \alpha_j k_j^{**}$, and agreement $j$ will abate with $Z > f(k_i^{**} - 1)$. Consequently, the internal stability of one agreement is not affected by the existence of a second agreement.

Thus, the condition for agreement 1 to be internally stable is

$$\Pi^{spc}_1 (z_1^{**}) > \Pi^p_1 (f(k^{**} - 1)),$$  \hspace{1cm} (37)

while external stability is met if

$$\Pi^p(z_1^{**}) > \Pi^{spc}_1 (f(k^{**} + 1)).$$  \hspace{1cm} (38)

Agreement 2 is internally stable if

$$\Pi^{spc}_2 (z_2^{**}) > \Pi^p_2 (f(k^{**} - 1)),$$  \hspace{1cm} (39)

and externally stable if

$$\Pi^p(z_2^{**}) > \Pi^{spc}_2 (f(k^{**} + 1)).$$  \hspace{1cm} (40)

We now use these results to infer how total abatement is changed when multiple coalitions are admitted.

First consider the case where only the $z_1$ members of agreement 1 abate and agreement 2 is a pseudo-agreement (if $\alpha_1 k_1 > k_2$ and $z_2 = 0$). Then, the total number of abating countries is $Z = z_1$. Since $\alpha_1 k_1 + k_2 = K$, it follows that
\[ f(k_1^{**} - 1) = f(K^{**} - \alpha_1), \]
and we see that condition for internal stability in the standard case (22) is the same condition as (37).

In the case that only \( z_2 \) member-countries of agreement 2 abate and agreement 1 is a pseudo-agreement (if \( \alpha_1 k_1 < k_2 \) and \( z_1 = 0 \)), then \( Z = z_2 \). If follows from \( \alpha_1 k_1 + k_2 = K \) that \( f(k_2^{**} - 1) = f(K^{**} - 1) \), and we see again that the condition for internal stability in the standard case (23) and in the case with two agreements (39) are identical.

Analogously to the conditions for internal stability it can be shown that the conditions for external stability (24) and (38) as well as conditions (25) and (40) are the same for the case on one agreement and of two agreements.

We can thus conclude that \( K^{**} = K^* \) and \( Z^{**} = Z^* \). The total number of signatories as well as the measure of total abatements are the same in the case with one agreement and the case with two agreements. This differs substantially from the results with constant marginal benefits of abatement. This is due to the fact that the reaction functions of the two agreements due not intersect in an internal equilibrium, so that only one of the two agreements decides to abate emissions.

## 4 Conclusions

Our paper has analyzed the effects of allowing for two parallel international environmental agreements (IEAs) when there are countries of two types. In a two-stage game, countries first chose whether they sign one agreement, or to be a non-signatory. In the second stage, each coalition acts as a unitary actor in a non-cooperative Nash game between both coalitions and the non-signatories. We compare emissions abatement and coalition stability in the two IEAs case with the standard case where not more than one IEA is possible. We investigate this for constant marginal benefits of abatement and for decreasing marginal benefits.

For constant marginal benefits, two IEAs lead to more total abatement and to a larger number of cooperating countries. Interestingly, total abatement does not
depend on the shares of the country types within the set of all countries in the game. These results follow from the linchpin character of the game equilibrium. One IEA is self-enforcing if all countries would chose to pollute, supposed one more country is leaving the IEA. This effect is replicated for each IEA. Thus two coalitions are stabilized with more abatement than just one.

When marginal benefits decrease linearly, this picture changes. The coalitions’ reaction functions in the second game stage do not intersect, so that corner solutions prevail. One coalition free-rides on the abatement efforts of the other coalition. Such a coalition may be called a pseudo-agreement. So, it does not matter whether more countries participate in any agreement. There is no additional abatement compared to the case with only one agreement.

The comparison of the two cases shows that the effect of climate clubs substantially depends on qualitative properties of abatement benefit functions, even if they are quite simple. In this general sense, our results are in line with the ambiguity results in the examples of Osmani and Tol (2010). In contrast, however, we can generally show for our assumptions that climate clubs are at least not detrimental to global cooperation. It would require further consideration whether the positive effects shown by Asheim et al. (2006) mostly stem from the constant marginal benefits assumption.

Nevertheless, our results need to be taken with precaution. Although our analysis is more general than single numerical examples, it sticks to specific (quadratic) cost and benefit functions. This requests for further generalisation. Also the welfare effects and the comparative statics require more attention. It would further be helpful to determine intercoalition stability (Osmani and Tol, 2010) more explicitly. On the other hand, the paper already shows how different assumptions lead to different effects of country clubs. We think that our analysis is thus a consequent stepping stone towards a more detailed understanding of the determinants for beneficial or detrimental effects of climate clubs.
Further analysis would profit from analyzing more carefully how countries would voluntarily sort into parallel agreements. Do countries of the same type join, or would types mix in IEAs? In any case, we need to conclude that the idea that climate clubs do benefit global climate protection has to be taken with precaution, but that it deserves more analytical attention.

References


