Participation and Duration of Environmental Agreements: Investment lags matter*

Thomas Eichner
Department of Economics, University of Hagen

Gilbert Kollenbach
Department of Economics, University of Hagen

Abstract

This paper analyzes participation in international environmental agreements in a dynamic game in which countries pollute and invest in two types of clean technology that differ in investment lags. If investments are non-contractible, countries underinvest in the long-lag technology in the last period of the contract which leads to a hold-up problem. Countries do not underinvest in the short-lag technology. If the short-lag technology is sufficiently cheap, the hold-up problem becomes irrelevant, and significant participation is not feasible. Our paper supplements Battaglini and Harstad (2016), who point out that the hold-up problem may result in significant participation and even in the first-best outcome, and shows that the assumptions required for significant participation may be more limited than expected.

JEL classification: F55, H87, Q54
Key words: environmental agreements, investment lag, timing, stable coalition

*Eichner: Department of Economics, University of Hagen, Universitaetsstr. 41, 58097 Hagen, Germany, email: thomas.eichner@fernuni-hagen.de; Kollenbach: Department of Economics, University of Hagen, Universitaetsstr. 41, 58097 Hagen, Germany, email: gilbert.kollenbbach@fernuni-hagen.de. We are grateful to Bard Harstad, Marco Battaglini, two anonymous referees and the Co-Editor Stephen Paul Holland for helpful comments. Remaining errors are the authors' sole responsibility.

1 Introduction

The analysis of international environmental agreements (IEAs) has a long tradition in economics. The standard tool for analyzing IEAs is static and dynamic game theory.[1] IEAs are essential for many pollution problems that cross national boundaries such as global warming and climate change. IEAs could help to keep the world mean temperature from rising in the medium term by 2° C above pre-industrial levels and to stabilize the world climate at safe levels through the reduction of global greenhouse gas emissions. For example, the Montreal Protocol was quite successful at limiting ozone-depleting chlorofluorocarbons, but the Kyoto Protocol and related agreements have been less successful at limiting carbon emissions. The economic literature is quite pessimistic about effective IEAs with many participants, due to strong free-riding incentives.

An important exception is Battaglini and Harstad (2016, henceforth B&H) who show in a dynamic game with emissions and investments in a clean technology, that significant participation and even the first-best outcome is feasible in the case of incomplete contracts, i.e. when emissions but not investment are contractible. This result is denoted as B&H's main result. The driving force for the attractiveness of the incomplete contract is as follows: During the last period of the contract, coalition countries underinvest, since investments in this period pay off only in the next period, in which it is uncertain whether a new contract will materialize. A hold-up problem arises that reduces free-riding incentives, such that significant participation is feasible.

The present paper investigates the potential of incomplete contracts, especially of the hold-up problem, to reduce the countries' free-riding incentives in IEAs. For that purpose we analyze a dynamic game with pollution and two types of clean technology which differ in investment lags. The one type builds up new technology in the next period, i.e with an *inter*-period investment lag. The other type builds up new technology in the current period, i.e. with an *intra*-period investment lag. Hereinafter, we refer to these technologies as *long-lag* technology and *short-lag* technology, respectively. In the case of incomplete contracts, coalition countries underinvest in the long-lag technology in the last period of the contract, which leads to a hold-up problem, whereas they do not underinvest in the short-lag technology. If the short-lag technology is sufficiently cheaper than the long-lag technology,

[1] The literature on dynamic games of IEAs has been surveyed by Calvo and Rubio (2012), Benchekroun and van Long (2012) and de Zeeuw (2018). More recently, Karp and Sakamoto (2021) and Kováč and Schmidt (2021) investigate the role of beliefs about the random consequences of (re)opening negotiations and the impact of delays in renegotiations, respectively, on the stability of IEAs in dynamic games.

clean energy is predominantly generated by means of the former technology, the hold-up problem becomes irrelevant and B&H's main result does not hold (see Corollary 2(ii)).

Our game is different from B&H's in two respects. Both differences are consistent with assumptions in Harstad (2012, 2016). First, in our four-stage game, pollution takes place after investment decisions are made. Second, we extend B&H's game by a clean short-lag technology, i.e. in addition to B&H's clean technology with an inter-period investment lag, in our model there is also a technology with an intra-period investment lag. Our game nests B&H's game as a special case. Since B&H's results are robust with respect to changes in the timing of investments and emissions, B&H's game coincides with the polar case of our game in which countries do not invest in the short-lag technology. This happens if the investment costs of the short-lag technology are prohibitively high (see Corollary 1(i)). Analogously, our game also entails the polar case in which countries do not invest in the long-lag technology. This happens if the investment costs of the long-lag technology are prohibitively high. Given our timing, the hold-up problem disappears and significant participation is not feasible (see Corollary 1(ii)). B&H's main result is sensitive to investment lags in combination with the timing of pollution and investments.

The extension with respect to technologies with different investment lags is motivated by time lags of emissions and real-world clean investments. In both B&H's and our model, emissions build up the pollution stock and cause climate damage in the current period. There is empirical evidence (Ricke and Caldeira 2014, Zickfeld and Herrington 2014) of lags between the point in time at which carbon dioxide emissions have ceased and the point at which the maximum warming and climate damages caused by these emissions set in. In view of this evidence, a period is a time span of five or ten years. However, Tierney and Bird (2020) advertises that it takes less than two years to build up solar and onshore wind capacity. The World Commission on Dams (2000, p. 10) reports that it takes 5 to 10 years on average to build a large hydroelectric power plant. The realization of offshore wind farms is between 4 and 13 years (Voormolen et al. 2016, p. 443). The aforementioned arguments confirm that the length of a period in a dynamic game is somewhat arbitrary, there is no correct timing of investments and emissions,[2] and there are clean technologies with different investment lags and some of them build their stock faster than emissions.[3]

[2]In reality, investments and pollution are made at (multiple) points in time within a period and would require a continuous game, which is beyond the scope of the present paper but an important task for future research.

[3]In addition, there is empirical evidence that short-lag technologies are cheaper than long-lag technologies. For example, the levelized cost of solar energy is lower than that of offshore wind energy (UNEP).

The remainder of the paper is organized as follows. Section 2 presents the building blocks of the model, the timing of the game and derives the countries' value function. Section 3 briefly characterizes the first-best outcome and the non-cooperative equilibrium which is also denoted as business as usual (BAU). Section 4 analyzes incomplete contracts and Section 5 briefly turns to complete contracts.[4] Section 6 concludes.

2 The model

2.1 Utility, pollution, capacity, and technology

We consider a modified version of B&H's model. At each period in time t, each country $i \in \{1, ..., N\}$ consumes energy $y_{i,t} = g_{i,t} + R_{i,t} + S_{i,t}$, which is generated by means of fossil fuels $g_{i,t}$ and two clean technologies $R_{i,t} + S_{i,t}$, which differ in investment lags and will be defined more carefully below. The benefit function

$$B_i(y_{i,t}) = -\frac{b}{2} \left(\bar{y}_i - y_{i,t}\right)^2 \tag{1}$$

of energy consumption satisfies $B_i' > 0$ and $B_i'' < 0$ for $y_{i,t} < \bar{y}_i$. The exogenous bliss point $\bar{y}_i$ represents country i's ideal energy consumption if there were no pollution concerns. The parameter b measures the disutility of reducing energy consumption relative to the bliss point.

Greenhouse gas emissions of country i are proportional to its fossil fuel consumption, such that $g_{i,t}$ denotes both fossil fuel use and greenhouse gas emissions of country i at period t. Emissions accumulate in the atmosphere, with G_t representing the stock of pollution at period t. According to

$$G_t = q_G G_{t-1} + \sum_{j \in N} g_{j,t}, \tag{2}$$

the stock increases with world-wide emissions, and decreases due to the decay rate $(1 - q_G) \in [0, 1]$. The stock of pollution causes the climate damage according to the function

$$D(G_t) = cG_t, \tag{3}$$

where c is a positive parameter. The damage function increases with the stock of accumulated CO_2 emissions.

[4] In the case of complete contracts, i.e. when both emissions and investments are contractible, the stable coalition consists of at most 3 countries. Complete contracts in our game are equivalent to those in B&H's game.

To substitute fossil fuels, countries can produce clean energy. For sake of simplicity, clean energy production in country i is proportional to the sum of accumulated stocks of clean technologies, i.e. $R_{i,t}$ and $S_{i,t}$ denote both the production of clean energy by means of technology R and S and the stock of clean technology R and S in country i at period t. Clean technologies differ with respect to investment lags. Following B&H, investments $r_{i,t-1}$ in the *long-lag* technology R_i have an *inter*-period lag and build up the stock in the next period t. The corresponding equation of motion is

$$R_{i,t} = q_R R_{i,t-1} + r_{i,t-1}, \tag{4}$$

where $(1 - q_R) \in [0, 1]$ denotes the depreciation rate of technology R. In addition, departing from B&H,[5] there are investments in clean technologies with an *intra*-period lag. Investments $s_{i,t}$ in the *short-lag* technology S_i are realized within the period of investment t, and the technology stock evolves over time according to

$$S_{i,t} = q_S S_{i,t-1} + s_{i,t}, \tag{5}$$

where $(1-q_S) \in [0, 1]$ denotes the depreciation rate of technology S. Obviously, the difference between the two types of clean investment is that investments in (4) build up technology stock in the next period, whereas investments in (5) build up technology stock in the same period.

The costs $\kappa_R(\cdot)$ and $\kappa_S(\cdot)$ of technology investments $r_{i,t-1}$ and $s_{i,t}$ depend on both the technology level and investments. We assume a quadratic relationship between costs $\kappa_L(\cdot)$ and the targeted technology level $L_{i,t}$ for $L = R, S$ such that $\frac{\partial \kappa_L}{\partial L_{i,t}} = k_L L_{i,t}$, where $k_L > 0$ is a cost parameter. In addition, we assume that investment costs are nil if no investments are made, formally $\kappa_R(\cdot) = 0$ for $r_{i,t-1} = 0$ and $\kappa_S(\cdot) = 0$ for $s_{i,t} = 0$. These assumptions imply the cost functions[6,7]

$$\kappa_R(R_{i,t}, R_{i,t-1}) = \frac{k_R}{2} \left(R_{i,t}^2 - q_R^2 R_{i,t-1}^2 \right), \tag{6}$$

$$\kappa_S(S_{i,t}, S_{i,t-1}) = \frac{k_S}{2} \left(S_{i,t}^2 - q_S^2 S_{i,t-1}^2 \right). \tag{7}$$

2.2 Timing

The timing of the sequential four-stage game is illustrated in Figure 1. The timing for complete and incomplete contracts is identical in stage 1 and 4 and different in stage 2

[5] In B&H there is only one type of technology which is identical to our technology R.

[6] Solving $\frac{\partial \kappa_L}{\partial L_{i,t}} = k_L L_{i,t}$ yields $\kappa_L(\cdot) = \frac{k_L}{2} L_{i,t}^2 + Q_L$ for $L = R, S$. $\kappa_R(\cdot) = 0$ for $r_{i,t-1} = 0$ and (4) imply $Q_R = -\frac{k_R}{2} q_R^2 R_{i,t-1}^2$. $\kappa_S(\cdot) = 0$ for $s_{i,t} = 0$ and (5) imply $Q_S = -\frac{k_S}{2} q_S^2 S_{i,t-1}^2$.

[7] Both cost functions κ_R and κ_S are identical to B&H's cost function κ. With regard to cost functions, our model does not deviate from B&H.

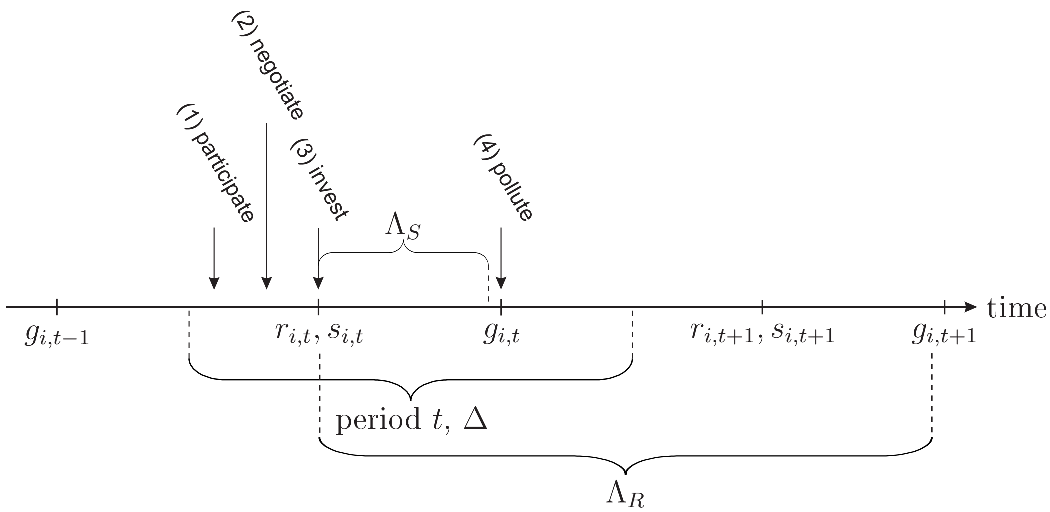

Figure 1: Timing in the game

and 3. If no coalition exists at the beginning of a period, at stage 1, country $i \in N$ decides whether to participate in a new climate coalition M. In case of *incomplete* contracts, at stage 2, the coalition countries first negotiate[8] the contract duration T, and then emissions $(g_{i,t})$ for all $i \in M$ and $t \in \{1, \ldots, T\}$. At stage 3, every country $i \in N$ non-cooperatively chooses investments $(r_{i,t}, s_{i,t})$. In case of *complete* contracts, at stage 2 the coalition countries first negotiate the contract duration T, and then both emissions and investments $(g_{i,t}, r_{i,t}, s_{i,t})$ for all $i \in M$ and $t \in \{1, \ldots, T\}$. At stage 3, every non-participant $i \in N \setminus M$ non-cooperatively chooses investments $(r_{i,t}, s_{i,t})$ and every coalition country invests as agreed. Both in case of the complete and the incomplete contract, at stage 4 every non-participant $i \in N \setminus M$ non-cooperatively chooses emissions $(g_{i,t})$ and every coalition country pollutes as agreed.

In Figure 1, $\Delta > 0$ denotes both the time from one investment/pollution decision to the next and the length of one period. $\Lambda_S > 0$ and $\Lambda_R > 0$ are the investment lags of investments $s_{i,t}$ and $r_{i,t}$, respectively, i.e. the time the investment needs to build up new technology. Investments $s_{i,t}$ develop new technology after the time Λ_S in period t, and before emissions $g_{i,t}$ are released. Investments $r_{i,t}$ develop new technology after time Λ_R in period $t + 1$, and before $g_{i,t+1}$ are emitted. The investment lag Λ_S is *intra*-periodic, whereas the investment lag Λ_R is *inter*-periodic.

Comparing Figure 1 and B&H's Figure 1 shows the differences in timing and in the investment lags between our game and B&H's. With regard to the timing in B&H's game, countries first pollute at stage 3 and then invest at stage 4, whereas in our game the timing

[8]Following the literature on IEAs, we assume that the outcome of the negotiation stage 2 is the cooperative solution which maximizes the utilitarian welfare of the coalition without any side transfers.

is in the reverse order, i.e. countries first invest at stage 3 and then pollute at stage 4. With regard to investment lags, in B&H there is one type of clean technology which is identical to our technology R, and the corresponding investments $r_{i,t}$ increase the technology stock in period $t+1$ according to (4). The investment lag overlaps two periods and is inter-periodic. In contrast, in our paper, there are two types of clean technology. In addition to the long-lag technology R, there is also the short-lag technology S. The corresponding investments $s_{i,t}$ build up technology in the same period according to (5). The investment lag of technology S is intra-periodic, i.e. it is within a period and does not overlap two periods. We augment B&H's long-lag technology R by means of the short-lag technology S.

B&H (2016, p. 169) have mentioned that reversing the order of stages 3 and 4 does not alter their results. Therefore, B&H's game is included in ours as a special case in which countries do not invest in technology S. Below we point out that introducing technology S and corresponding investments $s_{i,t}$ *and* modifying B&H's timing of emissions and investments, do not change the size of the stable coalition of complete contracts, but have far-reaching consequences for the size of the stable coalition of incomplete contracts.

2.3 Value function

Throughout the paper we restrict our attention to Markov-perfect equilibria (MPE) in pure strategies. Let $\rho > 0$ denote the time preference rate and define $\delta = e^{-\rho\Delta} \in (0,1)$ as the discount factor. The utility of country i of period t is given by[9]

$$\begin{aligned} u_{i,t} = & -\frac{b}{2}\left(\bar{y}_i - g_{i,t} - R_{i,t} - S_{i,t}\right)^2 - cG_t - \frac{k_R}{2}\left(R_{i,t+1}^2 - q_R^2 R_{i,t}^2\right)e^{\rho(\Lambda_S+\epsilon)} \\ & -\frac{k_S}{2}\left(S_{i,t}^2 - q_S^2 S_{i,t-1}^2\right)e^{\rho(\Lambda_S+\epsilon)}, \end{aligned} \tag{8}$$

where ϵ denotes the time from the realization of investments $s_{i,t}$ until emissions $g_{i,t}$ are released.

At Markov-perfect equilibria in pure strategies, the decisions of each country depend only on the current state of the economy but not its history. The *value function* of country i can be written as[10]

$$\begin{aligned} v_i = \sum_{\tau=t}^{\infty}\delta^{\tau-t}\Bigg[& -\frac{b}{2}d_{i,\tau}^2 - C\sum_{j\in N}(\bar{y}_j - d_{j,\tau}) + C\sum_{j\in N}S_{j,\tau} + \delta C\sum_{j\in N}R_{j,\tau+1} \\ & -\frac{K_S}{2}S_{i,\tau}^2 - \frac{K_R}{2}R_{i,\tau+1}^2\Bigg], \end{aligned} \tag{9}$$

[9]We measure utility at the pollution stage.

[10]Cf. Appendix A.1.

with $d_{i,t} := \bar{y}_i - g_{i,t} - R_{i,t} - S_{i,t}$ as the difference between the bliss point and instantaneous energy consumption, $K_R := k_R(1 - q_R^2\delta)e^{\rho(\Lambda_S+\epsilon)} > 0$ and $K_S := k_S(1 - q_S^2\delta)e^{\rho(\Lambda_S+\epsilon)} > 0$ as the effective technology investment cost parameter, and $C := \frac{c}{1-\delta q_G} = \sum_{\tau=t}^{\infty} c(\delta q_G)^{\tau-t}$ as the social costs of carbon of one emission unit released in period t.

3 First-best outcome and BAU

In this section, we briefly characterize the first-best outcome and the non-cooperative equilibrium (BAU). Maximizing (9) and $\sum_{j\in N} v_j$ with respect to $d_{i,t}$, $S_{i,t}$ and $R_{i,t+1}$, for $t \geq 1$, yield the emission quotas and technology levels

$$(a)\ g_{i,t}^{BAU} = \bar{y}_i - \frac{C}{K_S} - \frac{C}{b} - R_{i,t}, \qquad (b)\ g_{i,t}^{FB} = \bar{y}_i - n\frac{C}{K_S} - n\frac{C}{b} - R_{i,t}, \tag{10}$$

$$(a)\ S_{i,t}^{BAU} = \frac{C}{K_S}, \qquad (b)\ S_{i,t}^{FB} = n\frac{C}{K_S}, \tag{11}$$

$$(a)\ R_{i,t+1}^{BAU} = \frac{\delta C}{K_R}, \qquad (b)\ R_{i,t+1}^{FB} = n\frac{\delta C}{K_R}, \tag{12}$$

in the cases of non-cooperation (BAU) and first best (FB). Comparing BAU and FB reveals the underlying dynamic common pool problem. Investments are a public good reducing climate damage, and emissions are a public bad increasing the climate damage. In the case of non-cooperation, countries choose their emissions and investments without taking into account that their action influences the climate damage of other countries. Therefore, BAU emissions are inefficiently high and BAU investments are inefficiently low.

4 Incomplete contract

Next, we consider incomplete climate contracts. Whereas non-participants always choose BAU emissions and BAU investments, the members of the coalition only coordinate their emissions, but not their investments. That is, emissions $g_{i,t}$ for all $i \in M$ and all $t \in \{1, ..., T\}$ are cooperatively set at the negotiation stage 2 (in period $t = 1$), while the investments $s_{i,t}$ and $r_{i,t}$ are non-cooperatively set by the coalition countries at investment stage 3 in each period $t \in \{1, ..., T\}$. The timing of the game implies that a Stackelberg game arises in which the coalition countries non-cooperatively choose investments as Stackelberg followers, and the coalition cooperatively sets pollution as a Stackelberg leader. Solving the game by backward induction, at stage 3 of period $t \in \{1, ..., T\}$, coalition members maximize v_i from (9) with respect to $S_{i,t}$ and $R_{i,t+1}$ for a given negotiated $g_{i,t}$ (from stage 2). In Appendix

A.2 we show that the first-order conditions yield the reaction functions

$$S_{i,1} = \frac{b}{b+K_S}\left(\bar{y}_i - g_{i,t} - R_{i,1}\right) \qquad \forall i \in M, \tag{13}$$

$$S_{i,t} = \frac{bK_R}{K_R K_S + bK_R + \delta b K_S}\left(\bar{y}_i - g_{i,t}\right) \qquad \forall i \in M,\ t \in \{2,...,T\}, \tag{14}$$

$$R_{i,t} = \frac{\delta b K_S}{K_R K_S + bK_R + \delta b K_S}\left(\bar{y}_i - g_{i,t}\right) \qquad \forall i \in M,\ t \in \{2,...,T\}. \tag{15}$$

Recall that investments $r_{i,t}$ increase the technology stock $R_{i,t+1}$, whereas investments $s_{i,t}$ increase the technology stock $S_{i,t}$. According to (14) - (15), the coalition can induce coalition countries to raise their investments $s_{i,t}$ and $r_{i,t-1}$ and hence their technology stock $S_{i,t}$ and $R_{i,t}$ by reducing its pollution $g_{i,t}$ for $t \in \{2,...,T\}$. In the first period ($t = 1$), the technology stock $R_{i,1}$ is exogenous and the coalition can only affect the stock $S_{i,1}$ with its pollution $g_{i,1}$ according to (13).

Because a coalition country $i \in M$ does not know whether the coalition will persist in period $T+1$, it chooses BAU investments $r_{i,T}^{BAU}$ in period T such that

$$R_{i,T+1} = \frac{\delta C}{K_R} \qquad \forall i \in M. \tag{16}$$

The underinvestment in technology R in the last contract period leads to a hold-up problem.

At the negotiation stage 2, the coalition takes into account the reaction functions (13) - (15) and maximizes $\sum_{j\in M} v_j$ with respect to $g_{i,t}$. Appendix A.2 proves that the resulting MPE is characterized by

$$g_{i,t} = \bar{y}_i - m\frac{C}{K_S} - m\frac{C}{b} - R_{i,t} \qquad \forall i \in M,\ t \in \{1,...,T\}, \tag{17}$$

$$S_{i,t} = m\frac{C}{K_S} \qquad \forall i \in M,\ t \in \{1,...,T\}, \tag{18}$$

$$R_{i,t} = m\frac{\delta C}{K_R} \qquad \forall i \in M,\ t \in \{2,...,T\}. \tag{19}$$

Comparing (16) - (19) for $m = n$ with (10)(b) - (12)(b) shows that the incomplete contract implements the first-best outcome for all $t \in \{1,...,T\}$ with the exception of excessively low investments $r_{i,T}$ in the last contract period. For $1 < m < n$, a coalition country's investments and emissions internalize the climate externalities inside the coalition, but not outside it. This holds for emissions and investments in all periods $t \in \{1,...,T\}$ of the contract except for $r_{i,T}$.

The technology stocks $R_{i,t}$ in (16) and (19) are as in B&H, whereas the technology stocks $S_{i,t}$ in (18) are different from those of B&H. In the case of investments $r_{i,t}$, which are realized with an inter-period lag, the coalition can influence the coalition members' investments for all periods $t \in \{1, \ldots, T-1\}$, apart from the last contract period T. As a consequence, coalition countries underinvest in period T such that $R_{i,T+1} = \frac{\delta C}{K_R}$ and a hold-up problem arises. In the case of investments $s_{i,t}$, which are realized with an intra-period lag, the coalition can influence the coalition members' investments for all contract periods such that $S_{i,t} = m\frac{C}{K_S}$ for all $t \in \{1, \ldots, T\}$. With respect to technology S, there is no underinvestment in the last contract period and no hold-up problem.

To make our results comparable with B&H, we introduce the relative cost $x_L = \frac{K_L}{b\delta}$ of technology $L = R, S$. Lemma 1, which is proven in Appendix A.2, characterizes the optimal contract length.

Lemma 1. *Suppose coalition members negotiate only their emissions, but not their investments (incomplete contract). Let m^* denote the stable coalition size.*

(i) The optimal contract length for the coalition of size m is

- $T^* = 1$, *if* $m < \hat{m}$,
- $T^* \in \{1, ..., \infty\}$, *if* $m = \hat{m}$,
- $T^* = \infty$, *if* $m > \hat{m}$,

where

$$\hat{m} := m^* - (m^* - 1)\left(1 - \sqrt{\frac{x_R + \frac{x_R}{\delta x_S} + \delta}{x_R + \frac{x_R}{\delta x_S} + 1}}\right) < m^*.$$

(ii) The optimal contract length for the stable coalition of size m^* is $T^* = \infty$.

In view of Lemma 1(i), there is a threshold $\hat{m}$ at which a coalition of size m is indifferent between all contract lengths. Because of $\hat{m} < m^*$, it is optimal for the participants of a stable coalition to sign a long-term agreement ($T = \infty$, see Lemma 1(ii)). If a country abandons the stable coalition m^*, the remaining countries sign a short-term agreement ($T = 1$) only if $m^* - 1 \leq \hat{m}$, or equivalently only if

$$m^* \leq m_M(x_R, x_S) := 1 + \frac{1}{1 - \sqrt{\frac{x_R + \frac{x_R}{\delta x_S} + \delta}{x_R + \frac{x_R}{\delta x_S} + 1}}}. \tag{20}$$

(20) is referred to as *discipline constraint*. If (20) is satisfied, the defection of a country induces the remaining coalition to sign an one-period contract and coalition countries underinvest in technology R. If the discipline constraint is violated, the remaining coalition signs a long-term contract $T = \infty$ and there is no underinvestment. Proposition 1, which is proven in Appendix A.2, characterizes the size of the stable coalition.

Proposition 1. *Suppose coalition members negotiate only their emissions, but not their investments (incomplete contract). The stable coalition size is then given by*

(i) If $m^* \geq m_M$*, then* $m^* \in \{2,3\}$*.*

(ii) If $m^* < m_M$ *and*

(a) and $x_R \leq \frac{\delta}{1+\frac{1}{\delta x_S}}$*, then* $m^* \leq \min\{m_M, n\}$*,*

(b) and $x_R > \frac{\delta}{1+\frac{1}{\delta x_S}}$*, then* $m^* \leq \min\{m_M, m_I, n\}$*,*

where

$$m_I(x_R, x_S) := 3 + \frac{2\delta}{x_R + \frac{x_R}{\delta x_S} - \delta} \geq 3. \tag{21}$$

(21) is the *internal stability condition.*[11] In view of Proposition 1(i), a significant participation is not feasible if the discipline constraint (20) is violated, because only small climate coalitions of maximal 3 countries are stable. Otherwise, larger coalitions can be stable. If the relative cost of technologies R and S satisfy $x_R \leq \frac{\delta}{1+\frac{1}{\delta x_S}}$, either the grand coalition is stable or the stable coalition conforms to the discipline constraint m_M. If $x_R > \frac{\delta}{1+\frac{1}{\delta x_S}}$, the stable coalition is given by the minimum of n, m_M or m_I. Proposition 1 is illustrated in Figure 2 for $\delta = 0.95$ and $n = 60$. Ignore for the moment the $m_M(x_R, 0.5)$-line and the $m_I(x_R, 0.5)$-curve. For $x_S = 10$, the blue line reflects the discipline constraint $m_M(x_R, 10)$, and the violet curve reflects the internal stability condition $m_I(x_R, 10)$. Observe that in the example of Figure 2, the discipline constraint is satisfied and it holds that $\frac{\delta}{1+\frac{1}{\delta x_S}} = 0.86$. For $x_R < 0.44$, we have $\min\{m_M, n\} = m_M$, and the stable coalition lies on the line AB. For $0.44 < x_R < 0.89$, we have $\min\{m_M, m_I, n\} = n$ and the stable coalition lies on the $n = 60$-line. For $x_R > 0.89$, the stable coalition is characterized by the internal stability condition $m_I(x_R, 10)$ and lies on the segment CD. To sum up, for $x_R = 10$, the stable coalition is on the polyline $ABCD$.

Proposition 1 resembles B&H's Proposition 8. The driving force behind larger climate coalitions is the hold-up problem associated with technology R. If one coalition country defects and the discipline constraint holds, the remaining coalition countries sign a short-term agreement, in order to wait for the deviant to return in the next period. Due to the hold-up problem, the coalition countries reduce their investments in technology R to the

[11] Both the discipline constraint (20) and the internal stability condition (21) are written in such a way that they are easily comparable with B&H's discipline constraint in the Corollary to Proposition 7 and with B&H's internal stability condition in Proposition 8. Setting $\frac{x_R}{\delta x_S} \equiv 0$ in (20) and (21) yields B&H's discipline constraint and internal stability condition, respectively.

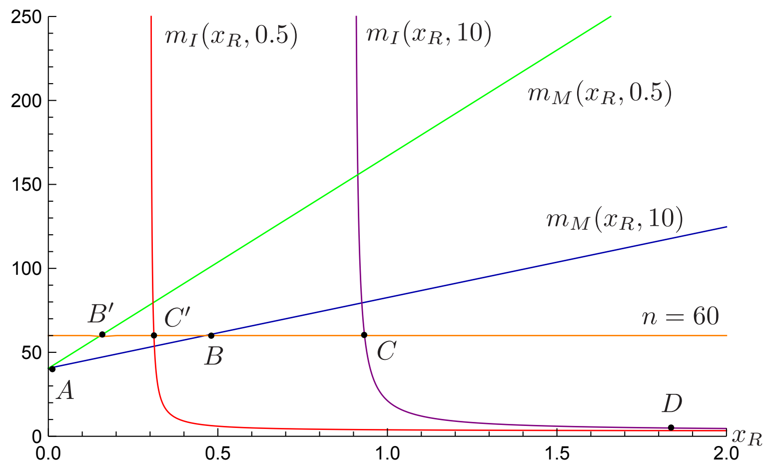

Figure 2: The size of the stable coalition m^* for $\delta = 0.95$ and $n = 60$

BAU-level, implying higher climate damages in the next period. Because all countries suffer from these additional damages, they counter the free-riding incentives of the deviant. In other words, the hold-up problem leads to a credible threat which stabilizes larger climate coalitions. However, in contrast to B&H, the size of the stable coalition not only depends on the relative costs of long-lag technology x_R, but also on the relative costs of short-lag technology x_S.

To explain the driving forces most clearly, consider the extreme cases of prohibitively costly technology S ($x_S \to \infty$) and prohibitively costly technology R ($x_R \to \infty$). If $x_S \to \infty$, countries invest only in technology R and not in technology S, with the consequence that Proposition 1 is identical to B&H's Proposition 8.[12] The hold-up problem reduces free-riding incentives. When a country leaves the coalition, the remaining countries sign a short-term agreement while they wait for the deviant to return to the coalition in the next period. With that short-term agreement, investments in technology R are as in BAU, due to the hold-up problem. The underinvestment of the short-term agreement is a credible threat that reduces free-riding incentives, and significant participation is feasible.

If $x_R \to \infty$, countries invest only in technology S and not in technology R. For $x_R \to \infty$ it holds $\lim_{x_R \to \infty} m_M(x_R, x_S) = \infty$ and $\lim_{x_R \to \infty} m_I(x_R, x_S) = 3$. Hence, Proposition 1(ii) applies with $m^* \leq \min\{m_M, m_I, n\} = m_I = 3$. Since investments in S build up technology in the same period, there is no underinvestment in the last contract period. Consequently, the credible threat of underinvestment in short-term agreements that stabilizes larger coalitions is lacking. Therefore, significant participation is not feasible and only small coalitions with

[12] For $x_S \to \infty$ it holds $\lim_{x_S \to \infty} \frac{x_R}{\delta x_S} = 0$, and the discipline constraint (20) and the internal stability condition (21) are equal to those of B&H, see footnote 11.

maximal 3 members are stable. We summarize these results in

Corollary 1. *Suppose coalition members negotiate only their emissions, but not their investments (incomplete contract).*

(i) If $x_S \to \infty$, the stable coalition size is characterized by B&H's Proposition 8.
(ii) If $x_R \to \infty$, the stable coalition size is given by $m^* \in \{2, 3\}$.

Next, consider the simultaneous use of technologies R and S. Differentiating m_M and m_I with respect to x_S yields

$$\frac{\mathrm{d}m_M}{\mathrm{d}x_S} < 0, \quad \frac{\mathrm{d}m_I}{\mathrm{d}x_S} > 0. \tag{22}$$

A cheaper technology S relaxes the discipline constraint, while it tightens the internal stability condition. Figure 2 shows the movement of curves in the transition from $x_S = 10$ to $x_S = 0.5$. The curve of the discipline constraint m_M rotates upwards, whereas the curve of the internal stability condition shifts to the left. As a consequence, the stable coalition moves from the polyline $ABCD$ to the polyline $AB'C'D$. If the technology S becomes cheaper, the technology stock $S_{i,t}$ and, therefore, total energy consumption $g_{i,t} + R_{i,t} + S_{i,t}$ increase. In contrast, the technology stock $R_{i,t}$ is not affected, implying that the absolute strength of the hold-up problem, and therefore of the additional climate damages that arise from signing a short-term contract, remain constant. However, the greater benefit from total energy consumption reduces the relative strengths of both the hold-up problem and of (additional) climate damages. With regard to coalition countries, the reduced relative strength of (additional) climate damages relaxes the discipline constraint. With regard to the defecting country, the reduced relative strength of (additional) climate damages tightens the internal stability condition. In the transition from $x_S = 10$ to $x_S = 0.5$, the parameter set of large coalitions becomes smaller.[13] This result can be generalized beyond the numerical example of Figure 2. Denoting the set of economies in which the cost parameter x_S is given and the size of the stable coalition is $m^* = h$, by $\mathcal{E}(x_S, h) := \left\{(\delta, n, x_R)\middle| m^* = h\right\}$ we prove in Appendix A.4 for[14] $h \in]3, n]$

$$\frac{\mathrm{d}\left|\mathcal{E}(x_S, h)\right|}{\mathrm{d}x_S} \geq 0. \tag{23}$$

According to (23), the set of economies in which the stable coalition is $m^* = h$ becomes smaller when the relative cost x_S is reduced. The effect of tightening the internal stability condition overcompensates for the effect of relaxing the discipline constraint.

[13] Increasing x_S from $x_S = 10$ to $x_S \to \infty$ shifts the $m_I(x_R, 10)$-curve of Figure 2 slightly to the right and rotates the $m_M(x_R, 10)$-line slightly downwards. As a consequence, Figure 2 turns into B&H's Figure 2.

[14] $|A|$ denotes the cardinality of the set A.

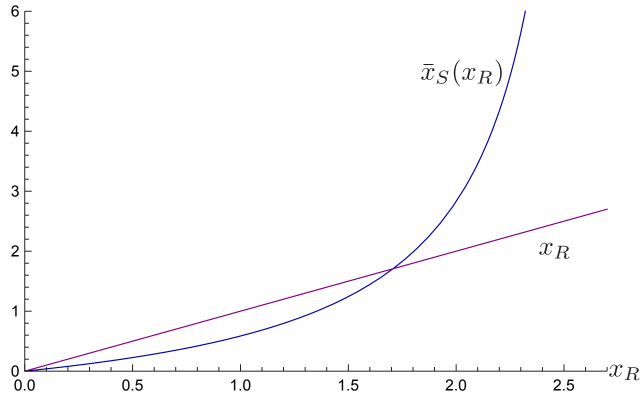

Figure 3: The threshold $\bar{x}_S(x_R)$ for $\delta = 0.95$

Finally, suppose that technology S is sufficiently cheaper than technology R ($x_S \ll x_R$).[15] For any given cost parameter x_R, there is a threshold $\bar{x}_S(x_R) := \frac{x_R}{(3\delta - x_R)\delta}$ such that the size of the stable coalition is at most 3 if $x_S < \bar{x}_S$. In that case, the share of clean energy S is so large relative to the clean energy R that the hold-up problem becomes irrelevant. The threshold $\bar{x}_S(x_R)$ is illustrated in Figure 3. We summarize these results in

Corollary 2 . *Suppose coalition members negotiate only their emissions, but not their investments (incomplete contract).*

(i) The set of economies, in which the stable coalition is $m^* = h$ with $h \in]3, n]$, becomes smaller when x_S reduces.

(ii) If $x_S < \bar{x}_S(x_R)$, the stable coalition size is $m^* \in \{2, 3\}$.

5 Complete contract

Finally, we briefly turn to complete contracts. Whereas non-participants act non-cooperatively and always set their BAU emissions and BAU investments, coalition countries choose emissions and investments cooperatively. Formally, they maximize $\sum_{j \in M} v_j$ with respect to $d_{i,t}$,

[15] According to UNEP (2020), in 2019 the levelized cost of energy are: solar 57 \$/MWh, onshore wind energy 50 \$/MWh and offshore wind energy 89 \$/MWh. Solar and onshore wind energies belong to technology S and offshore wind to technology R.

$S_{i,t}$ and $R_{i,t+1}$, for all $t \in \{1, ..., T\}$. The first-order conditions yield

$$g_{i,t} = \bar{y}_i - m\frac{C}{K_S} - m\frac{C}{b} - R_{i,t} \qquad \forall i \in M, \ t \in \{1, ..., T\}, \tag{24}$$

$$S_{i,t} = m\frac{C}{K_S} \qquad \forall i \in M, \ t \in \{1, ..., T\}, \tag{25}$$

$$R_{i,t+1} = m\frac{\delta C}{K_R} \qquad \forall i \in M, \ t \in \{1, ..., T\} \tag{26}$$

(24) - (26) are identical to emissions and technologies of incomplete contracts with the exception of the last contract period T, in which coalition countries build up the technology $R_{i,T+1} = m\frac{\delta C}{K_R} > R^{BAU}_{i,t+1} = \frac{\delta C}{K_R}$. In the case of complete contracts there is no underinvestment and no hold-up problem. In Appendix A.5 we prove

Proposition 2. *Suppose that coalitions members negotiate both emissions and investments. The stable coalition size is* $m^* \in \{2, 3\}$

Because the hold-up problem is absent in the case of complete contracts, significant participation is not feasible. The complete contract of our game has the same properties as B&H's complete contract.

6 Concluding remark

Dynamic games map real-world international environmental agreements at a high level of abstraction. In this paper, we extend B&H's game by a short-lag technology and reverse the order of pollution and investments. The aim of our paper is to point out how robust the size of the stable coalition is with respect to the modeling of timing and investment lag. If the short-lag technology is cheaper than the long-lag technology, the potential of incomplete contracts to yield large and effective stable climate coalitions may be more limited than expected.

References

Battaglini, Marco and Bård Harstad. 2016. Participation and duration of environmental agreements. *Journal of Political Economy* 124: 160-204.

Benchekroun, Hassan and Ngo van Long. 2012. Collaborative environmental management: A review of the literature. *International Game Theory Review* 20: 119-133.

Calvo, Emilio and Santiago J. Rubio. 2012. Dynamic models of International Environmental Agreements. *International Review of Environmental and Resource Economics* 6: 289-339.

Harstad, Bård. 2016. The dynamics of climate agreements. *Journal of European Economic Association* 14: 719-752.

Harstad, Bård. 2012. Climate contracts: A game of emissions, investments, negotiations, and renegotiations. *Review of Economic Studies* 79: 1527-1557.

Karp, Larry and Hiroaki Sakamoto. 2021. Sober optimism and the formation of international environmental agreements. *Journal of Economic Theory* 197: 105321.

Kováč, Eugen and Robert C. Schmidt. 2021. A simple dynamic climate cooperation model. *Journal of Public Economics* 194: 104329.

Ricke, Katharine L. and Ken Caldeira. 2014. Maximum warming occurs about one decade after a carbon dioxide emission. *Environmental Research Letters* 9: 124002.

Voormolen J.A., Junginger, H.M. and W.G.J.H.M. van Sark. 2016. Unravelling historical cost developments of offshore wind energy in Europe. *Energy Policy* 88: 435-444.

World Commission on Dams. 2000. Cross-Check Survey: Final Report, Cape Town, South Africa.

Tierney, Susan and Lori Bird. 2020. Setting the Record Straight about Renewable Energy. World Resources Institute. https://www.wri.org/blog/2020/05/setting-record-straight-about-renewable-energy.

UNEP. 2020. Global trends in renewable energy investment 2020, Frankfurt School of Finance & Management GmbH.

Zeeuw de, Aart. 2018. Dynamic games of international pollution control: A selective review. In *Handbook of Dynamic Games Theory*, ed. T. Basar and G. Zaccour. Springer Nature, Switzerland.

Zickfeld, Kirsten and Tyler Herrington. 2015. The time lag between a carbon dioxide emission and maximum warming increases with the size of the emission. *Environmental Research Letters* 10: 03001.

A Appendix

A.1 Value Function

The value of country i at time t reads $\hat{v}_i = \sum_{\tau=t}^{\infty} \delta^{\tau-t} u_{i,t}$. We use the definition of $d_{i,t}$,

$$-c\sum_{\tau=t}^{\infty}\delta^{\tau-t}G_\tau = -\sum_{\tau=t}^{\infty}\delta^{\tau-t}\left[C\sum_{j\in N}(\bar{y}_j - d_{j,\tau}) - C\sum_{j\in N}S_{j,\tau} - \delta C\sum_{j\in N}R_{j,\tau+1}\right] - Cq_G G_{t-1} + C\sum_{j\in N}R_{j,t},$$

$$-\sum_{\tau=t}^{\infty}\delta^{\tau-t}\frac{k_S}{2}\left(S_{i,\tau}^2 - q_S^2 S_{i,\tau-1}^2\right)e^{\rho(\Lambda_S+\epsilon)} = -\sum_{\tau=t}^{\infty}\delta^{\tau-t}\frac{K_S}{2}S_{i,\tau}^2 + \frac{k_S}{2}q_S^2 S_{i,t-1}^2 e^{\rho(\Lambda_S+\epsilon)},$$

$$-\sum_{\tau=t}^{\infty}\delta^{\tau-t}\frac{k_R}{2}\left(R_{i,\tau+1}^2 - q_R^2 R_{i,\tau}^2\right)e^{\rho(\Lambda_S+\epsilon)} = -\sum_{\tau=t}^{\infty}\delta^{\tau-t}\frac{K_R}{2}R_{i,\tau+1}^2 + \frac{k_R}{2}q_R^2 R_{i,t}^2 e^{\rho(\Lambda_S+\epsilon)}$$

to get

$$\hat{v}_i = \sum_{\tau=t}^{\infty}\delta^{\tau-t}\left[-\frac{b}{2}d_{i,\tau}^2 - C\sum_{j\in N}(\bar{y}_j - d_{j,\tau}) + C\sum_{j\in N}S_{j,\tau} + \delta C\sum_{j\in N}R_{j,\tau+1} - \frac{K_S}{2}S_{i,\tau}^2 - \frac{K_R}{2}R_{i,\tau+1}^2\right]$$
$$- Cq_G G_{t-1} + C\sum_{j\in N}R_{j,t} + \frac{k_S}{2}q_S^2 S_{i,t-1}^2 e^{\rho(\Lambda_S+\epsilon)} + \frac{k_R}{2}q_R^2 R_{i,t}^2 e^{\rho(\Lambda_S+\epsilon)}.$$

Because $S_{j,t-1}$, $R_{j,t}$ and G_{t-1} are given values at period t, they are pay-off irrelevant. Therefore, the last four terms can be omitted, which gives (9). ■

A.2 Incomplete contract

A.2.1 Derivation of (13) - (16)

The value function (9) can be rewritten as

$$v_i = \frac{1}{1-\delta^T}\left\{\sum_{t=1}^{T}\delta^{t-1}\left[-\frac{b}{2}(\bar{y}_i - g_{i,t} - S_{i,t} - R_{i,t})^2 - C\sum_{j\in N}g_{j,t} - \frac{K_S}{2}S_{i,t}^2 - \frac{K_R}{2}R_{i,t+1}^2\right] + \delta^{T-1}\delta C\sum_{j\in N}R_{j,T+1} - C\sum_{j\in N}R_{j,1}\right\} \tag{27}$$

At every point in time $t \in \{1, ..., T\}$ a coalition member maximizes (27) with respect to $S_{i,t}$ and $R_{i,t+1}$. The first-order conditions give

$$S_{i,t} = \frac{b}{b+K_S}(\bar{y}_i - g_{i,t} - R_{i,t}), \qquad t \in \{1, ..., T\}, \tag{28}$$

$$R_{i,t+1} = \frac{\delta b}{\delta b + K_R}(\bar{y}_i - g_{i,t+1} - S_{i,t+1}), \quad t \in \{1, ..., T-1\}, \tag{29}$$

$$R_{i,T+1} = \frac{\delta C}{K_R}. \tag{30}$$

Using (28) and (29) yields (13) - (15), while (30) is identical with (16). ■

A.2.2 Derivation of (17) - (19)

The coalition takes account of (13) - (16) when maximizing $\sum_{j \in M} v_j$ for all $t \in \{1, ..., T\}$. By substituting into (27) we can write the value function of a coalition member as

$$\begin{aligned} v_i = \frac{1}{1-\delta^T} \Bigg\{ & -\frac{b}{2} \frac{K_S^2 (\bar{y}_i - g_{i,1} - R_{i,1})^2}{(K_S + b)^2} - C \sum_{j \in N} g_{j,1} - \frac{K_S}{2} \frac{b^2 (\bar{y}_i - g_{i,1} - R_{i,1})^2}{(K_S + b)^2} \\ & + \sum_{t=2}^{T} \delta^{t-1} \Bigg[-\frac{b}{2} \frac{K_S^2 K_R^2 (\bar{y}_i - g_{i,t})^2}{(K_S K_R + b K_R + \delta b K_S)^2} - C \sum_{j \in N} g_{j,t} \\ & - \frac{K_S}{2} \frac{b^2 K_R^2 (\bar{y}_i - g_{i,t})^2}{(K_S K_R + b K_R + \delta b K_S)^2} - \frac{K_R}{2} \frac{\delta b^2 K_S^2 (\bar{y}_i - g_{i,t})^2}{(K_S K_R + b K_R + \delta b K_S)^2} \Bigg] \\ & + \delta^{T-1} \sum_{j \in N} R_{j,T+1} - \delta^{T-1} \frac{1}{2} \frac{\delta^2 C^2}{K_R} - C \sum_{j \in N} R_{j,1} \Bigg\} \end{aligned} \tag{31}$$

The first-order conditions of the coalition's maximization give

$$g_{i,1} = \bar{y}_i - m \frac{C}{K_S} - m \frac{C}{b} - R_{i,1}, \tag{32}$$

$$g_{i,t} = \bar{y}_i - m \frac{C}{K_S} - m \frac{C}{b} - m \frac{\delta C}{K_R}. \tag{33}$$

Substituting into (28) and (29) yields (18) and (19). ■

A.2.3 Proof of lemma 1

Suppose that the stable coalition has the size m^* and lasts for T^* periods. When the contract expires, an identical contract is implemented. By substituting (10)(a) - (12)(a) and (16) - (19) into (27), the value function of a coalition country reads

$$\begin{aligned} v_c(m^*, T^*) = & \frac{1}{1-\delta} \left[\left(\frac{m^{*2}}{2} + (n - m^*) \right) \left(\frac{C^2}{K_S} + \frac{\delta^2 C^2}{K_R} + \frac{C^2}{b} \right) - C \sum_{j \in N} \bar{y}_j \right] \\ & - \frac{\delta^{T^*-1}}{1-\delta^{T^*}} \frac{\delta^2 C^2}{K_R} \frac{(m^* - 1)^2}{2}. \end{aligned} \tag{34}$$

For $T^* \to \infty$ the last term vanishes, so that $v_c(m^*, \infty) - v_c(m^*, T) > 0$ implying $T^* = \infty$.

Consider an arbitrary coalition (m, T). When the contract expires, the stable coalition is established. The value function of a coalition country reads

$$\begin{aligned} v_c(m, T) = & \frac{1-\delta^T}{1-\delta} \left[\left(\frac{m^2}{2} + (n - m) \right) \left(\frac{C^2}{K_S} + \frac{\delta^2 C^2}{K_R} + \frac{C^2}{b} \right) - C \sum_{j \in N} \bar{y}_j \right] \\ & - \delta^{T-1} \frac{\delta^2 C^2}{K_R} \frac{(m-1)^2}{2} + \delta^T v_c(m^*, T^*). \end{aligned} \tag{35}$$

Differentiating with respect to T yields

$$\frac{\partial v_c}{\partial T} = \frac{|ln(\delta)|\delta^T}{2(1-\delta)}\left[(m-1)^2-(m^*-1)^2\right]\left(\frac{C^2}{K_S}+\frac{\delta^2 C^2}{K_R}+\frac{C^2}{b}\right)+|ln(\delta)|\delta^{T-1}\frac{\delta^2 C^2}{K_R}\frac{(m-1)^2}{2}.$$

The optimal contract length equals 1 if $\frac{\partial v_c}{\partial T}<0$, which yields $m<\hat{m}$. The optimal contract is $T\in\{1,...,\infty\}$ if $\frac{\partial v_c}{\partial T}=0$, which yields $m=\hat{m}$. The contract is signed forever if $\frac{\partial v_c}{\partial T}>0$, which yields $m>\hat{m}$. ■

A.3 Proof of proposition 1

For the determination of the stable coalition size consider the value function of a non-participant for one period and for T periods, respectively. By using (10)(a) - (12)(a) and (17) - (19) we get

$$u_f(m) = \left(m^2+n-m-\frac{1}{2}\right)\left(\frac{C^2}{K_S}+\frac{\delta^2 C^2}{K_R}+\frac{C^2}{b}\right)-C\sum_{j\in N}\bar{y}_j-\frac{\delta^2C^2}{K_R}m(m-1), \tag{36}$$

$$\begin{aligned} v_f(m,T) = {} & \frac{1-\delta^T}{1-\delta}\left[\left(m^2+n-m-\frac{1}{2}\right)\left(\frac{C^2}{K_S}+\frac{\delta^2 C^2}{K_R}+\frac{C^2}{b}\right)-C\sum_{j\in N}\bar{y}_j\right] \\ & -\delta^{T-1}\frac{\delta^2C^2}{K_R}m(m-1)+\delta^T v_f(T+1). \end{aligned} \tag{37}$$

Suppose that one coalition member defects from the coalition m^* and that the disciplinary constrain (20) does not hold. Then, the defection does not pay-off if $v_c(m^*,\infty)\geq v_f(m^*-1,\infty)$, which yields

$$m^*\leq 3. \tag{38}$$

The accession of one fringe country to the coalition does not pay-off if $v_c(m^*+1,\infty)\leq v_f(m^*,\infty)$, which yields

$$m^*\geq 2. \tag{39}$$

(38) and (39) prove proposition 1(i).

Suppose that one coalition member defects from the coalition m^* and that the disciplinary constrain (20) holds. Then, the defection does not pay-off if $v_c(m^*,\infty)\geq u_f(m^*-1)+\delta v_c(m^*,\infty)$, which yields

$$m^*\left(1+\frac{1}{\delta x_S}-\frac{\delta}{x_R}\right)\leq 2\left(1+\frac{1}{\delta x_S}\right)+\left(1+\frac{1}{\delta x_S}-\frac{\delta}{x_R}\right). \tag{40}$$

If $1+\frac{1}{\delta x_S}>\frac{\delta}{x_R}\Leftrightarrow x_R>\frac{\delta}{1+\frac{1}{\delta x_S}}$, the inequality (40) yields $m^*\leq m_I$. If $1+\frac{1}{\delta x_S}<\frac{\delta}{x_R}\Leftrightarrow x_R<\frac{\delta}{1+\frac{1}{\delta x_S}}$, the inequality (40) yields

$$m^*\geq 1+2\frac{1+\frac{1}{\delta x_S}}{1+\frac{1}{\delta x_S}-\frac{\delta}{x_R}}<1, \tag{41}$$

which holds for all m^*. If $1+\frac{1}{\delta x_S}=\frac{\delta}{x_R}$, the inequality (40) holds for all m^*. The accession of one fringe country to the coalition does not pay-off if $v_c(m^*+1,\infty)\leq v_f(m^*,\infty)$, which yields (39). The results are summarized by proposition 1. ■

A.4 Proof of (23)

Solving $m_M = n$ and $m_I = n$ with respect to x_R we obtain

$$x_R^{Mn} = \frac{\left[n^2(1-\delta)-2n(2-\delta)+4-\delta\right]\delta x_S}{(2n-3)(1+\delta x_S)}, \tag{42}$$

$$x_R^{In} = \frac{(n-1)\delta^2 x_S}{(n-3)(1+\delta x_S)} > 0. \tag{43}$$

$x_R^{Mn}>0$ if $n\geq\frac{2-\delta+\sqrt{\delta}}{1-\delta}=:\underline{n}$. Subtracting x_R^{Mn} from x_R^{In} yields

$$x_R^{In}-x_R^{Mn}=\frac{\left[-(1-\delta)n^2+n(5-\delta)-6\right](n-2)\delta x_S}{(n-3)(2n-3)(1+\delta x_S)}. \tag{44}$$

$x_R^{In}-x_R^{Mn}>0$ if $n>\frac{5-\delta-\sqrt{1+14\delta+\delta^2}}{2(1-\delta)}=:\overline{n}$ and $n<\frac{5-\delta+\sqrt{1+14\delta+\delta^2}}{2(1-\delta)}$. Differentiating $x_R^{In}-x_R^{Mn}$ with respect to x_S leads to

$$\frac{\mathrm{d}(x_R^{In}-x_R^{Mn})}{\mathrm{d}x_S}=\frac{\left[-(1-\delta)n^2+n(5-\delta)-6\right](n-2)\delta}{(n-3)(2n-3)(1+\delta x_S)^2}. \tag{45}$$

Presupposed $x_R^{In}-x_R^{Mn}>0$ we get $\frac{\mathrm{d}(x_R^{In}-x_R^{Mn})}{\mathrm{d}x_S}>0$.

Solving $m_M = m_I$ with respect to x_R we obtain

$$x_R^{MI}=\frac{\left(1+\delta+\sqrt{1+14\delta+\delta^2}\right)\delta x_S}{6(1+\delta x_S)}. \tag{46}$$

Inserting x_R^{MI} in turn in m_I we get

$$m_I\left(x_R^{MI},x_S\right)=\frac{3\left(1-\delta+\sqrt{1+14\delta+\delta^2}\right)}{1-5\delta+\sqrt{1+14\delta+\delta^2}}. \tag{47}$$

In the following we make a case distinction. Since we argue with the movement of curves, it may be helpful to consider Figure 2.

If $n\leq\underline{n}$, then $m_M>n$ for all x_R and x_S. In that case the m_M-line lies above the n-line. It applies Proposition 1(ii) with

$$\begin{aligned} m^* &\leq n, && \text{for } x_R\leq\frac{\delta}{1+\frac{1}{\delta x_S}},\\ m^* &\leq \min\left\{m_I,n\right\}, && \text{for } x_R>\frac{\delta}{1+\frac{1}{\delta x_S}}. \end{aligned} \tag{48}$$

$\frac{\mathrm{d}m_I}{\mathrm{d}x_S} > 0$ establishes (23).

If $\underline{n} < n < \overline{n}$, there exists an intersection point of the m_M-line with the n-line and an intersection point of the n-line with the m_I-curve. The grand coalition is stable for all $x_R \in [x_R^{Mn}, x_R^{In}]$. Due to (45) the set of economies in which the grand coalition is stable becomes larger and we get $\frac{\mathrm{d}\left|\mathcal{E}(x_S,h)\right|}{\mathrm{d}x_S} \geq 0$ for $3 < h < n$.

If $n \geq \overline{n}$, there exists an intersection point of the m_M-line with the m_I-curve. This intersection point is determined by $\left(x_R^{MI}, m_I\left(x_R^{MI}, x_S\right)\right)$ in (46) and (47) and lies below the n-line implying that the grand coalition is never stable. Increasing x_S enhances x_R^{MI} but leaves $m_I\left(x_R^{MI}, x_S\right)$. Hence, we conclude $\frac{\mathrm{d}\left|\mathcal{E}(x_S,h)\right|}{\mathrm{d}x_S} \geq 0$ for $3 < h \leq n$.

A.5 Complete contract

Suppose that the stable coalition has the size m^* and lasts for T^* periods. When the contract expires, an identical contract is implemented. By substituting (10)(a) - (12)(a), (24) - (26) into (9) we get

$$v_c(m^*, T^*) = \frac{1}{1-\delta}\left[\left(\frac{m^{*2}}{2} + (n - m^*)\right)\left(\frac{C^2}{K_S} + \frac{\delta^2 C^2}{K_R} + \frac{C^2}{b}\right) - C\sum_{j\in N}\bar{y}_j\right]. \tag{49}$$

Consider an arbitrary coalition (m, T). When the contract expires, the stable coalition is established. The value function of a coalition country reads

$$v_c(m, T) = \frac{1-\delta^T}{1-\delta}\left[\left(\frac{m^2}{2} + (n - m)\right)\left(\frac{C^2}{K_S} + \frac{\delta^2 C^2}{K_R} + \frac{C^2}{b}\right) - C\sum_{j\in N}\bar{y}_j\right] + \delta^T v_c(m^*, T^*).$$

Differentiating with respect to T yields

$$\frac{\partial v_c}{\partial T} = \frac{|ln(\delta)|\delta^T}{2(1-\delta)}\left[(m-1)^2 - (m^*-1)^2\right]\left(\frac{C^2}{K_S} + \frac{\delta^2 C^2}{K_R} + \frac{C^2}{b}\right).$$

We get $\frac{\partial v_c}{\partial T} < 0 \Leftrightarrow m < m^*$, $\frac{\partial v_c}{\partial T} = 0 \Leftrightarrow m = m^*$, and $\frac{\partial v_c}{\partial T} > 0 \Leftrightarrow m > m^*$.

For the determination of the stable coalition consider the value function of a non-participant for one period and for T^* periods, respectively. By using (10)(a) - (12)(a) and (24) - (26) we get

$$u_f(m) = \left(m^2 + n - m - \frac{1}{2}\right)\left(\frac{C^2}{K_S} + \frac{\delta^2 C^2}{K_R} + \frac{C^2}{b}\right) - C\sum_{j\in N}\bar{y}_j, \tag{50}$$

$$\begin{aligned} v_f(m, T) = {} & \frac{1-\delta^T}{1-\delta}\left[\left(m^2 + n - m - \frac{1}{2}\right)\left(\frac{C^2}{K_S} + \frac{\delta^2 C^2}{K_R} + \frac{C^2}{b}\right) - C\sum_{j\in N}\bar{y}_j\right] \\ & + \delta^T v_f(T+1). \end{aligned} \tag{51}$$

Suppose that one coalition member defects from the stable coalition. The remaining members will sign an one-period contract. After that period the coalition (m^*, T^*) is established. The defection does not pay-off if the internal stability condition $v_c(m^*, T^*) \geq u_f(m^*-1)+\delta v_c(m^*, T^*)$ holds, which is equivalent to (38). Suppose that one non-participant accedes the coalition, such that the contract lasts forever. The accession does not pay-off if the external stability condition $v_f(m^*, T^*) \geq v_c(m^* + 1, \infty)$ holds, which is equivalent to (39). Consequently, $m^* \in \{2, 3\}$. ■