Free-riding, Participation, Coalitions

Bård Harstad

Lecture Notes 3

Abstract

This lecture note builds on the previous one (the theoretical framework is quite similar), but here we explicitly discuss the possibility to free ride vs. participate in an environmental coalition. This choice endogenizes the coalition size and shows what it depends on.

I. Introduction

International environmental agreements (henceforth, IEAs) typically include many countries. The average number of countries in an IEA is 25 and more than three quarters of them include more than 5 countries. Some agreements include well over a hundred countries. In the first commitment period of the Kyoto Protocol, 35 countries committed to an average emission reduction target of 5% compared to 1990-levels.

It is quite natural that countries may desire agreements in order to limit free riding, since a more healthy environment is a global public good. Participation in an IEA, however, is itself comparable to a public good contribution: besides the cost of the negotiation, it ultimately involves voluntary restrictions on economic activity that also benefit countries that do not participate. Participation, therefore, should be hindered by free-rider problems. Indeed, a recent influential literature has shown there is no simple theoretical reason to expect that many countries will voluntarily accept to participate in an IEA, casting serious doubts on the efficiency of IEAs characterized by the three features mentioned above. Summarizing this literature, Kolstad and Toman (2005) describe its findings as the "Paradox of International Agreements:" while IEAs seem to be ubiquitous, economic theory suggest that they should not exist, or at least they should not be effective in the form in which they are observed. Why, then, do we nevertheless observe a large number of countries that participate in IEAs? What are the consequences of the fact that such agreements are often "incomplete contracts" which specify emissions but not investments in green technology? How should environmental agreements be designed to be more effective?

This note presents a new dynamic theory to answer these questions. As before, we will allow countries to choose both emission levels and the amount of resources to invest in "green technologies," which are strategic substitutes for polluting activity. Differently from the other lecture notes, here countries also decide whether to free ride or participate in an IEA. The length and depth of the cooperative agreement are endogenous: the coalition members negotiate the number of years for which the agreement holds and the abatement level for each participant. We consider both a "complete contracting" environment, where the agreement can also specify the investments; and we consider an "incomplete contracting" environment where such investments are not contractible. Confirming the previous literature, we show that very few countries find it optimal to cooperate in an environment with complete contracts - regardless of the discount factor and other parameters of the model. Surprisingly, the coalition may be much larger if contracts are incomplete. Under some conditions, even the first-best outcome may be feasible. Thus, our analysis shows that incomplete contracts can be beneficial and explains why environmental coalitions are often quite large.

An important part of this mechanism is the hold-up problem discussed already. If a country has a large stock of green technology, it will be required to abate more in any efficient agreement or reasonable bargaining game. Anticipating this, countries have few incentives to invest in green technologies during a short-lasting agreement when the next bargaining round is just around the corner.

To understand our results, we need to clarify how the duration of the contract depends on the size of the coalition. Suppose a country that is expected to participate instead chooses to deviate by not participating in a particular period. This generates two effects: First, it makes the agreement less ambitious since the policies are chosen to minimize only the externalities generated by the participating countries (it therefore reduces the "depth" of the agreement). Second, and more importantly, the deviation may reduce the duration of the agreement. Indeed, the remaining participants expect the deviator to return to the equilibrium strategy and thus the bargaining table next period, so they find it optimal to "wait" a period, by signing a short-term agreement, rather than to lock in an inefficient long-term agreement. With complete contracts, the duration of the contract is not very important: the IEA will exploit the complete nature of contracts to ensure that countries invest. This is not possible when contracts are incomplete, and short-term agreements will then discourage investments thanks to the hold-up problem. The hold-up problem generated by a short-term agreement is thus a credible "threat" which reduces the incentives to free ride.

We should also think about the consequences of endogenizing the contractual environment. Allowing the countries to choose whether to make investments contractible may or may not influence the details of the equilibrium - this depends on the exact timing of the decision process - but in any case, only incomplete contracts are signed in equilibrium. The model can thus explain why existing climate negotiations do not attempt to contract on investments.

Our positive analysis has important normative implications as well. First, the fact that the Kyoto Protocol is "incomplete" should not necessarily be seen as an accidental design flaw: an effort to closely monitor and control green investments may be counterproductive. Second, it is important to let the final coalition negotiate the duration of the agreement, rather than announcing a length before countries have fully committed on whether or not to join. Third, there are multiple equilibria regarding the coalition size. If one could coordinate on the equilibrium with the largest coalition size, then the coalition members would benefit and welfare would increase. Perhaps likely participants can influence the equilibrium selection by announcing an appropriate target for the coalition size.

The next section presents the model, the equilibrium concept, and two benchmark cases: the first-best solution and the noncooperative "business as usual." Section III solves the game in an environment where contracts can be complete and confirms the classic result that few countries are willing to participate. Incomplete contracts are considered and proven to be more efficient in Section IV. Section V endogenizes the contractual environment and derives the optimal degree of incompleteness. Various extensions are presented in Section VI, while Section VII discusses further readings. As in the other lecture notes, most proofs are in the appendix "Proofs".

II. Model and Preliminaries

A. Consumption, Pollution, and Technology

The model here is more similar to the one in Lecture Note 1, Section 5.

We consider an economy with many countries and an infinite number of periods. In every period $t \geq 1$, each country $i \in N = \{1, ..., n\}$ benefits from consuming $y_{i,t}$, perhaps best interpreted as country i's level of energy. As in much of the literature, we assume the benefit of consumption, $B_i(y_{i,t})$, is represented by a quadratic and concave function:

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

The variable $\overline{y}_i$ is an exogenous satiation point that should be assumed to be large: it represents the consumption or energy level country i would choose if there were no concern for climate change. The parameter $b > 0$ measures the disutility of reducing consumption relative to the satiation point.

While consumption is privately beneficial, it contributes to a public bad. We will say that the emission level of country i at time t is:

$$g_{i,t} = y_{i,t} - R_{i,t}, \tag{2}$$

where $R_{i,t}$ represents the level of green technology. The stock $R_{i,t}$ may therefore measure the quantity of potential emissions ($y_{i,t}$) that country i can clean thanks to the accumulated abatement technology. Or, as in our favored interpretation, $R_{i,t}$ can measure the quantity of energy generated by country i's renewable energy sources. When $g_{i,t}$ is the quantity of fossil fuel consumption, i's total energy consumption is $y_{i,t} = g_{i,t} + R_{i,t}$, implying (2). We allow $\overline{y}_i$ and the initial stock $R_{i,1}$ to vary across the i's, but countries are otherwise assumed to be identical.

The stock of pollution is $G_t = q_G G_{t-1} + \sum_{i \in N} g_{i,t}$, where $1 - q_G \in [0,1]$ measures the natural depreciation rate of greenhouse gases. At each point in time, country i's environmental harm is cG_t where $c > 0$ is assumed to be a constant.

The technology stock depreciates at the rate $1 - q_R \in [0,1]$, and if country i invests $r_{i,t}$ units today, the technology available tomorrow is:[1]

$$R_{i,t+1} = q_R R_{i,t} + r_{i,t}. \tag{3}$$

In general, the investment cost, $\kappa_t(\cdot)$, may depend on both the investment level and the level of existing technology. Because of this, we assume that the cost is convex and the marginal cost increases proportionally with the stock of capital. This reflects the fact that existing technological solutions can be ranked according to costs and that the cheapest technology options are developed and installed first. Specifically, we assume that the marginal cost of a unit of technology is:

$$\partial\kappa(\cdot)/\partial R_{i,t+1} = kR_{i,t+1}. \tag{4}$$

It follows that $\kappa(\cdot)$ takes the form $\kappa(R_{i,t+1}, R_{i,t}) = \frac{k}{2}\left(R_{i,t+1}^2 - q_R^2 R_{i,t}^2\right)$ when the investment is durable ($q_R > 0$), and $kr_{i,t}^2/2$ with full depreciation ($q_R = 0$).[2] Although there may be uncertainty, learning by doing, and increasing returns to scale in reality, cost functions are normally assumed to be both increasing and convex in the literature in order to ensure interior solutions. Assumption (4) is also standard in the literature and so make our work more comparable with existing findings.[3]

In section VI.A we extend the analysis to other functional forms, showing that the quadratic forms of $\kappa(\cdot)$ and $B_i(\cdot)$ are not driving the results.

B. Timing

Time can be continuous or discrete. However, we assume that countries invest simultaneously at discrete points in time; they consume simultaneously at discrete points in time; and the consumption stages and the investment stages alternate. In a continuous-time setting, let $\rho > 0$ be the discount rate, $\Delta > 0$ be the time from one emission/consumption decision to the next, and $\Lambda \in (0, \Delta]$

[1] We do not assume that $r_{i,t}$ is necessarily positive.

[2] To see this, just solve the differential equation $\partial\kappa(\cdot)/\partial R_{i,t+1} = kR_{i,t+1}$ to get $\kappa(\cdot) = kR_{i,t+1}^2/2$ plus a constant or variable which must be independent of $R_{i,t+1}$. Requiring $\kappa = 0$ when $r_{i,t} = 0 \Rightarrow R_{i,t+1} = q_R R_{i,t}$ pins down this constant and thus $\kappa(\cdot)$.

[3] For example, the same assumption of a convex cost of investments in abatement technology with marginal costs that increase linearly in the stock of capital is made by Dutta and Radner (2004) who empirically calibrate their theoretical model to study the dynamic effect of environmental agreements. Dutta and Radner also assume as we do that the marginal benefit of investments is linear (see (3)).

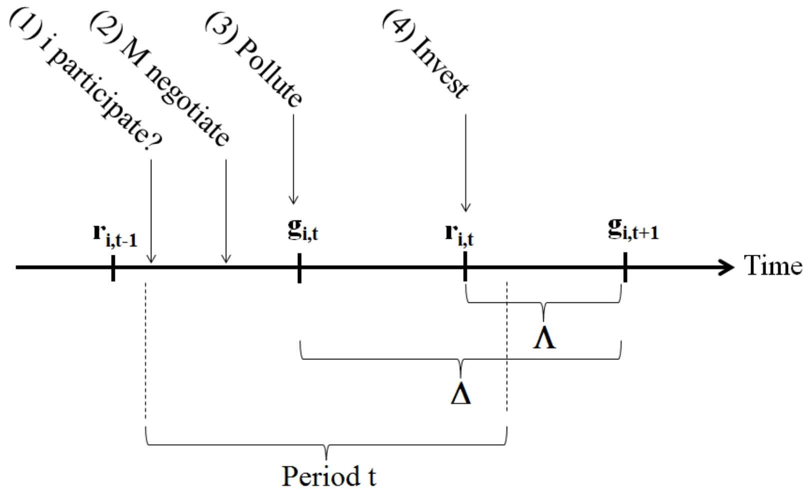

Figure 1: *The timing of the game*

be the time required to develop new technology. The optimal and equilibrium time between the investment stage and the next emission stage is then Λ, thus the time between the emission stage and the next investment stage is $\Delta - \Lambda$. We define a period to start with the emission stage and end with the investment stage. Given this, the utility of country i in period t is:

$$u_{i,t} = -\frac{b}{2}(\overline{y}_i - g_{i,t} - R_{i,t})^2 - cG_t - \frac{k}{2}\left(R_{i,t+1}^2 - q_R^2 R_{i,t}^2\right) e^{-\rho(\Delta-\Lambda)},$$

for every $i \in N$. Country i at time t seeks to maximize $\sum_{\tau \geq t} \delta^{\tau - t} u_{i,\tau}$, where next period's utility is discounted by the factor $\delta \equiv e^{-\rho\Delta} \in (0,1)$.[4]

We do not take a stand on what the contractual environment actually is. Instead, we analyze and compare all scenarios we believe are of interest. At the end of this section we derive two benchmark cases—the first-best outcome and the noncooperative, business-as-usual environment, in which nothing is contractible.

Section III analyzes the complete contracting environment. In this case, the stage game is as follows (see the figure for an illustration). (1) Coalition formation stage: if there exists no coalition, every $i \in N$ independently and simultaneously decides whether to become a member of a new coalition, M. The remaining countries, $L \equiv N \backslash M$, remain independent. (2) Negotiation stage: the coalition members first negotiate the duration of the agreement T, and then

[4] If $\Lambda = \Delta$, emissions and investments are decided simultaneously. If $\rho \to \infty$ or, equivalently, $\delta \to 0$, then there will be no investment and the next period becomes irrelevant. The model is then as in Barrett (2005, Section 6.4).

every $g_{i,t}$ and $r_{i,t}$ for $i \in M$ and $t \in \{1, ..., T\}$.[5] (3) Emission stage: every nonparticipant $i \in L$ simultaneously and independently chooses $g_{i,t}$, while the coalition members pollute as agreed. (4) Investment stage: every nonparticipant $i \in L$ simultaneously and independently chooses $r_{i,t}$, while the coalition members invest as agreed. Since $R_{i,t}$ is given by the investment stage in the previous period, deciding on $r_{i,t}$ is equivalent to directly choosing $R_{i,t+1}$. If an agreement already existed at the start of the period, the first two stages are skipped.

Section IV considers an incomplete contracting environment in which emissions, but not investments, are contractible. In this case, the coalition members negotiate the $g_{i,t}$s while the $r_{i,t}$s are chosen noncooperatively at stage 4.

As will be clarified in Section VI.C, we do not need to impose strong assumptions on the outcome of the bargaining stage 2. As a start, however, it is convenient to simply assume that any coalition M cooperatively chooses a policy vector (T, $g_{i,t}$, and, if contracts are complete, $r_{i,t}$) that maximizes the utilitarian welfare of the coalition without any accompanying side transfers. This is the standard assumption in the literature (see the survey by Barrett, 2005).

Our results are also quite robust with respect to timing. For example, stage 3 and stage 4 can occur simultaneously or their timing can be reversed (requiring $\Delta < \Lambda$) without affecting any of the conclusions. Stage 2 and stage 3 may also occur simultaneously or in reversed order: to us it is irrelevant whether or not the coalition acts as a Stackelberg leader since the environmental harm is linear in the stock.

C. The Equilibrium Concept and Preliminaries

There is typically a large number of subgame-perfect equilibria in dynamic games. We focus on Markov-perfect equilibria in pure strategies since these are simple, robust, and the strategies depend on the payoff-relevant variables only. These equilibria are also empirically plausible.[6]

Because of the linearity of the payoffs and technology, the game has a simple structure that allows a practical characterization of all equilibria. To see this, note that the players' preferences can be restated as follows.

Lemma 1. *At any time t, the utility of $i \in N$ is independent of all past stocks and can be represented by the continuation value function $v_{i,t} = \sum_{\tau=t}^{\infty} \delta^{\tau-t} \widehat{u}_{i,\tau}$,*

[5] Whether the choices of the policies and the duration are simultaneous or sequential is irrelevant for the results. In the following it will prove convenient for expositional reasons to separate these decisions as if they were sequential.

[6] There is an emerging experimental literature showing that Markov perfect equilibria provide a good description of behavior in dynamic free rider problems, see Battaglini et al. (2012), (2013) and Vespa (2012) for recent contributions. Dixit and Olson (2000) and Hong and Karp (2012) analyze equilibria in mixed strategies.

where

$$\widehat{u}_{i,t} \equiv -\frac{b}{2}{d_{i,t}}^2 - C\sum_{j\in N}\left(\overline{y}_i - d_{j,t}\right) - \delta\frac{K}{2}R_{i,t+1}^2 + \delta C\sum_{j\in N}R_{j,t+1}, \tag{5}$$

$$d_{i,t} \equiv \overline{y}_i - \left(g_{i,t} + R_{i,t}\right),$$

$$K \equiv k\left(1 - e^{-\rho\Delta}q_R^2\right)e^{\rho\Lambda}, \text{ and}$$

$$C \equiv \frac{c}{1-\delta q_G}.$$

Proof. Note that $\sum_{\tau=t}^{\infty}\delta^{\tau-t}\widehat{u}_{i,\tau} = \sum_{\tau=t}^{\infty}\delta^{\tau-t}u_{i,\tau} + e^{-\rho(\Delta-\Lambda)}q_R^2R_t^2k/2$, where the latter term is a constant not affecting the ranking of any vectors of future actions. QED

The present-discounted cost of emission is represented by C, while K is the net cost of technology given that some of it survives to later periods. The variable $d_{i,t}$ measures how much i decreases consumption relative to the bliss level. Since $g_{i,t} = \overline{y}_i - d_{i,t} - R_{i,t}$, country i can reduce $g_{i,t}$ either by decreasing consumption or by investing in technology.

The representation in (5) makes clear that the accumulated stocks of greenhouse gases and green technologies enter linearly in the players' objective functions. Because of this, these stocks do not affect the marginal cost or benefit of the policies, nor the players' reaction functions. This fact is key for a simple characterization of the Markov-perfect equilibria and their associated strategies. Since the stocks are "payoff-irrelevant", the Markov-perfect strategies are conditioned on neither G_t nor the $R_{i,t}$s. The only relevant state variables are whether an IEA is in force or not and, if so, the prescription of that contract. In particular, all nodes at which there is no contract in place are equivalent.[7]

D. The First-Best Outcome

Consider a welfare function $W\left(v_{1,t}, v_{2,t}, ..., v_{n,t}\right)$ which is symmetric, concave, and increasing in each of its arguments. A special case is the utilitarian welfare function $W(\cdot) = \sum_{i\in N} v_{i,t}$. Since $W(\cdot)$ is symmetric and every function $v_{i,t}$ is symmetric and concave in the vectors of $d_{i,t}$s and $R_{i,t}$s, the first-best requires that the $d_{i,t}$s and the $R_{i,t}$s are identical across the countries. So, even if countries have different ideal points $\overline{y}_i$, it is efficient that they all decrease their consumption level, relative to their ideal point, by the same amount $d_{i,t}$. Furthermore, these uniform policies must be such that each $v_{i,t}$ is maximized. The first-order conditions are then straightforward to derive from (5) (the second-order conditions hold trivially).

Proposition 1. *(i) The first-best investments ensure that:*

$$R_{i,t+1} = n\frac{C}{K} \Leftrightarrow r_{i,t} = n\frac{C}{K} - q_R R_{i,t},\ \forall t \geq 1.$$

[7] A detailed description of the players' strategies will be presented in Sections 3 and 4 before we analyze the games with contractual completeness and incompleteness.

(ii) The first-best emission levels are given by:

$$d_{i,t} = n\frac{C}{b} \Leftrightarrow g_{i,t} = \overline{y}_i - R_{i,t} \ - n\frac{C}{b},\ \forall t \geq 1.$$

Intuitively, if the cost of emission and the number of countries are both large, then it is optimal that each country consumes less as well as invest more in green technology. The two means of reducing emissions should be combined in a sensible way: the technological solution ought to dominate the total abatement effort if K is small, while consumption-reduction is cheaper if b is small. The first-best ratio between the two instruments is as follows:

$$\frac{d_{i,t}}{R_{i,t}} = x \equiv \frac{K}{b},\ \forall t > 1. \tag{6}$$

By definition, x measures how the present-value of the marginal cost of investing (taking future cost-savings into account) increases in $R_{i,t}$ relative to how the marginal cost of reducing consumption from the bliss level increases in the level of this reduction. At the first-best, this ratio dictates by how much it is optimal to reduce consumption relative to the optimal green technology stock. Since both $d_{i,t}$ and $R_{i,t}$ are proportional to C, the ratio x is independent of C.

E. No Cooperation (Business as Usual)

Suppose instead that each country decides $g_{i,t}$ and $r_{i,t}$ noncooperatively. In a Markov-perfect equilibrium, i anticipates that its choices of $d_{i,t}$ and $R_{i,t}$ do not affect the future choice of $d_{j,\tau}$ and $R_{j,\tau}$ for any player j or time τ. Thus, when each country is simply maximizing $v_{i,t}$, or equivalently $\widehat{u}_{i,t}$ in (5), we get the following outcome.

Proposition 2. *There is a unique Markov-perfect equilibrium.*
(i) The noncooperative investments ensure that:

$$R_{i,t+1} = \frac{C}{K} \Leftrightarrow r_{i,t} = \frac{C}{K} - q_R R_{i,t}\ \forall t \geq 1. \tag{7}$$

(ii) The noncooperative emission levels are given by:

$$d_{i,t} = \frac{C}{b} \Leftrightarrow g_{i,t} = \overline{y}_i - \frac{C}{b} - R_{i,t}\ \forall t \geq 1. \tag{8}$$

The noncooperative equilibrium coincides with the first-best only if $n = 1$. With multiple countries, each country invests too little while it pollutes and consumes too much. Note, however, that the *ratio* of consumption-reduction to technology is exactly as in the first-best:

$$\frac{d_{i,t}}{R_{i,t}} = x \equiv \frac{K}{b},\ \forall t > 1.$$

III. Contractible Investments

This section analyzes the model in Section 2 assuming that the coalition can contract on investment as well as emission levels. A pure-strategy equilibrium will specify a coalition M^*, a duration strategy $T^*(M)$ and a policy $(g_{i,t}(T,M), R_{i,t+1}(T,M))_{t=1}^{T}$. Here M^* is the set of countries whose strategy is to join the coalition when there is an opportunity to do so (i.e., in period one and in a period following the expiration of an agreement).[8] The function $T^*(M)$ specifies, for any coalition of countries that has chosen to join the IEA, the length of the agreement.[9] The functions $(g_{i,t}(T,M), R_{i,t+1}(T,M))_{t=1}^{T}$ specify the levels of emissions and investments for all periods following the formation of the IEA. The participants collectively choose $g_{i,t}$ and $r_{i,t}$ for every $i \in M$ and $t \in \{1, 2, ..., T\}$ at the start of period 1.[10] The nonparticipating countries choose $g_{i,t}$ and $r_{i,t}$ independently in every period.

We first present the equilibrium $g_{i,t}$s and $r_{i,t}$s, assuming a duration T and coalition M, before we derive T and, finally, M. Because the model is symmetric, the identity of the countries in M is irrelevant, i.e., if we have an equilibrium with coalition M, then we have an equilibrium with any other coalition $M' \neq M$ with $|M'| = |M|$. In the remainder we will ignore the identity of countries in the equilibrium coalition and simply focus on the characterization of the number of countries $m^* = |M^*|$ that join the IEA.

A. Emissions and Investments

For the reasons described in the business-as-usual case above, every nonparticipant acts according to (7)-(8). The coalition ensures that the externalities of the m coalition members are taken into account, but it does not internalize the environmental harm on the nonparticipants. Negotiating the $r_{i,t}$s is equivalent to negotiating the $R_{i,t+1}$. Furthermore, agreeing on $g_{i,t}$ is equivalent to agreeing on $d_{i,t} = \overline{y}_i - R_{i,t} - g_{i,t}$.

Proposition 3. *(i) For every coalition member, equilibrium investment levels ensure that:*

$$R_{i,t+1} = m\frac{C}{K} \Leftrightarrow r_{i,t} = m\frac{C}{K} - q_R R_{i,t}, \forall i \in M, t \in \{1, .., T\}.$$

[8] Because we study pure strategy Markov-perfect equilibria, if a country's strategy prescribes to join with probability one at $t = 0$, then the same country will choose to join with probability one at any period following the expiration of an agreement.

[9] Naturally, in equilibrium we will observe only $T^*(M^*)$, since only countries in M^* join the IEA in equilibrium. However, we still need to specify the reaction function for all the possible coalitions M that can be reached by a unilateral deviation.

[10] Because we focus on Markov-perfect equilibria, the period τ in which the IEA is formed is irrelevant, so these function are independent of τ. If the coalition is formed in period τ, then pollution and investments in the following T periods will be $g_{i,\tau+t} = g_{i,t}(T,M)$ and $R_{i,\tau+t+1} = R_{i,t+1}(T,M)$ for $t = 1, ..., T$.In this and the following sections we normalize the period when the coalition is formed to "period one." Thus, a T-period agreement expires at the end of period T.

(ii) Equilibrium consumption and emission are given by:

$$d_{i,t} = m\frac{C}{b}, t \in \{1, ..., T\} \Rightarrow$$
$$g_{i,1} = \overline{y}_i - R_{i,1} - m\frac{C}{b} \text{ and } g_{i,t} = \overline{y}_i - m\frac{C}{K} - m\frac{C}{b}, \; t \in \{2, ..., T\}.$$

Proof. Since every country has the identical preference $\widehat{u}_{i,t}$, the negotiated $d_{i,t}$s and the $R_{i,t}$s will be identical for every $i \in M$ and these maximize $\sum_{j \in M} \widehat{u}_{j,t}$. The first-order conditions in Proposition 3 follow and the second-order conditions are trivially fulfilled. QED

Every coalition member invests more and consumes less if the coalition size is large. The investment and abatement levels are first-best if $m = n$, but they are otherwise too low. It is interesting to note that independent of m, and even if $m < n$, the ratio of consumption-reduction to technology stock is efficient: the coalition chooses the right mixture of investments relative to general abatement.

Corollary to Proposition 3. *(i) We have $d_{i,t}/R_{i,t} = x$ for every $t \in \{2, ..., T\}$.*
(ii) *If $m = n$, the outcome would be first-best for every $t \in \{1, ..., T\}$ regardless of T.*

Finally, note that the coalition's optimal $d_{i,t}$ and $R_{i,t+1}$ are independent of any past stocks, the duration of the agreement, and what the countries expect will replace it.

B. Duration of the Agreement

While Proposition 3 holds for any contract length, no matter where it comes from, we can also ask for the equilibrium T when the countries can freely negotiate it. The choice of T will depend on the composition of the current coalition, M, as well as on what the countries believe will replace the agreement. As noted already, the equilibrium coalition, M^*, will be independent of any stock, history, or time in a Markov-perfect equilibrium. Thus, no matter the actual composition of the current coalition, M, everyone expects that, once the current agreement expires, the next coalition will be M^*. The next proposition characterizes the equilibrium duration as a function of the coalition's actual size.[11]

Proposition 4. *Let M^* denote an equilibrium coalition of size $m^* \equiv |M^*|$ and assume that $M \subseteq M^*$ or $M^* \subseteq M$. Then, a coalition of size $m = |M|$ finds it optimal to contract for $T(m)$ periods, where:*

$$T(m) = \begin{cases} 1 & \text{if } m < m^* \\ \{1, ..., \infty\} & \text{if } m = m^* \\ \infty & \text{if } m > m^* \end{cases}.$$

[11] Proposition 4 does not specify the players' reaction function when neither $M \subseteq M^*$ nor $M^* \subseteq M$. The reaction function after these out of equilibrium histories is irrelevant for the equilibrium conditions since a coalition reached after a unilateral deviation must be such that either $M \subseteq M^*$ or $M^* \subseteq M$.

From Proposition 4 we learn that if the coalition happens to be smaller than the equilibrium coalition, the coalition strictly prefers a one-period agreement, since a larger coalition is to be expected next period. If the current coalition equals the equilibrium coalition, then any length is a best choice. If the length is $T < \infty$, for example, the identical coalition (comprising of the same m countries) will form and negotiate the identical terms in period $T+1$, generating the same payoffs to everyone, irrespective of the choice of T.

C. Participation

We can now analyze the first stage of the game. For M^* to be an equilibrium coalition it must be both externally stable and internally stable. External stability requires that every $j \in N \backslash M^*$ should be unwilling to join. It can be shown that this condition is satisfied whenever $|M^*| > 1$. Internal stability requires that every $i \in M^*$ does not strictly prefer to free ride.

When a country contemplates whether or not to join the coalition, it anticipates the reaction function described in Proposition 4. In particular, if a country which is supposed to participate in equilibrium considers a deviation, then it understands that the consequence will be a one-period contract and that the country will be expected, and find optimal, to join the coalition next period. The country must then balance the gains from its own lower investment-cost and higher consumption today, with the fact that the coalition members will not take the externality on i into account (i.e., they will consume more and invest less when the coalition is smaller). This trade-off determines whether a country would like to join the coalition.

Proposition 5. *M^* is an equilibrium coalition if and only if $m^* = |M^*| \leq 3$.*

The result is dismal. Even with extremely patient players and large externalities, the equilibrium coalition size will be very small. The gain from participating is the fact that the other coalition members will take the entrant's externality into account and thus further reduce consumption and raise investment. Proposition 5 shows that these gains cannot motivate more than three countries to join.[12]

Recall that a special case of our model is the workhorse model with one period and no investments (achieved by letting $\delta = 0$ and x and K approach infinity). A well-known result from that literature is that at most three countries will join the coalition (Barrett, 2005). This result is quite robust in that it is independent of any parameters of the model. Proposition 5 shows that this discouraging result continues to hold even if we have multiple periods, investment in green technologies, and if countries can contract on all these choices for any length of time.

IV. Incomplete Contracts

[12] The reason that the discount factor doesn't help in obtaining a larger coalition is intuitive: as δ increases, the benefit of joining a coalition increases, but so does the benefit of staying out and free ride. The result is that the size remains small even as $\delta \to 1$.

As discussed in the Introduction, real climate negotiations have mainly focused on emission levels, leaving the investment decisions to individual countries. To also capture this situation, we now relax the assumption that the policy is fully contractible and assume that countries can commit to emission levels but not to specified levels of investments. We investigate how investments are influenced by the negotiated emission quotas, how the quotas are decided taking into account the effect on investments, and how the contractual incompleteness influences the equilibrium duration as well as coalition size.

As in the previous section, a pure-strategy equilibrium specifies a coalition M^*, a duration strategy $T^*(M)$ and a policy $(g_{i,t}(T,M), R_{i,t+1}(T,M))_{t=1}^{T}$. The coalition chooses the duration T and commits to $g_{i,t}$ for every $i \in M$ and $t \in \{1, 2, ..., T\}$ at the start of period 1. The level of investments, however, are independently chosen by the individual members in every period. Nonparticipating countries choose both $g_{i,t}$ and $r_{i,t}$ independently in every period.

A. Emissions and Investments

Just as in the previous sections, nonparticipants find it optimal to consume and invest according to (7)-(8). For coalition members, however, the optimal investment levels will depend on the negotiated quotas. If $g_{i,t}$ is small, then the marginal utility of energy consumption is very large unless $R_{i,t}$ is large. Thus, the smaller the quota, the larger the incentives to invest.

Proposition 6. *(i) For every $i \in M$, equilibrium investment ensures that the technology stock decreases in the emission quota:*

$$R_{i,t} = \frac{b\left(\overline{y}_i - g_{i,t}\right)}{b+K},\ t \in \{2, ..., T\}\ ,\ \textit{but}\ R_{i,T+1} = \frac{C}{K}.$$

(ii) The equilibrium emission quotas satisfy:

$$\begin{aligned} & g_{i,1} = \overline{y}_i - R_{i,1} - m\frac{C}{b}\ \textit{and}\ g_{i,t} = \overline{y}_i - m\frac{C}{K} - m\frac{C}{b},\ t \in \{2, ..., T\} \\ \Rightarrow\ & R_{i,t} = m\frac{C}{K},\ \textit{but}\ R_{i,T+1} = \frac{C}{K},\ t \in \{2, ..., T\} \\ \Rightarrow\ & d_{i,t} = m\frac{C}{b},\ t \in \{1, ..., T\}\,. \end{aligned}$$

Part (i) states that country i, in general, invests more if $g_{i,t}$ is small, as is intuitive. In the last period of the agreement, however, the countries realize that the impact of a higher $R_{i,T+1}$ is simply to reduce total emissions (and i's quota) one by one: their investment choices are "sunk" and not payoff-relevant in the following period when the countries will choose the $d_{i,T+1}$s and $R_{i,T+2}$s. Thus, the marginal benefit to country i of increasing the technological stock is just C: this explains why the equilibrium level of $R_{i,T+1}$ is only C/K. This under-investment can be interpreted as a consequence of the traditional hold-up problem, where parties invest too little when they fear being "held up" in future negotiations.

Part (ii) describes the equilibrium negotiated quotas. For every period and country, quotas ensure that the marginal benefit of another unit of consumption equals the coalition's cost of more emissions. Since the latter is constant over time, the implication is that $d_{i,t}$ is the same for every $i \in M$ and $t \in \{1, ..., T\}$. The countries will then invest the ideal amount for the coalition as a whole, except for the last period, in which every country invests too little. So, except for the last period, emission and investment levels are identical to the complete contracting outcome, if we take T and M as given.

Corollary to Proposition 6. *(i) We have $d_{i,t}/R_{i,t} = x$ for every $t \in \{2, ..., T\}$.*
(ii) If $m = n$, the outcome would be first-best for every $t \in \{1, ..., T\}$ if and only if $T = \infty$.

Note that the above corollary is similar to the corollary to Proposition 3. The only difference is that if we had $m = n$ for every agreement, then complete contracts would implement the first-best for any T, while incomplete contracts would implement the first-best only if $T = \infty$. When T is finite, every country invests too little in the last period if investments are noncontractible. If we had $m = n$ and T finite, complete contracts would lead to the first-best while incomplete contracts would not.[13]

B. Duration of the Agreement

Proceeding as in the previous section, we next determine equilibrium contract length T, given an arbitrary coalition, M.

Proposition 7. *Let M^* denote the equilibrium coalition of size $m^* \equiv |M^*|$ and assume that $M \subseteq M^*$ or $M^* \subseteq M$. Then, a coalition of size $m = |M|$ finds it optimal to contract for $T(m)$ periods, where:*

$$T(m) = \begin{cases} 1 & \text{if } m < \widehat{m}(x, m^*) \\ \{1, ..., \infty\} & \text{if } m = \widehat{m}(x, m^*) \\ \infty & \text{if } m > \widehat{m}(x, m^*) \end{cases}, \text{ with}$$

$$\widehat{m}(x, m^*) \equiv m^* - (m^* - 1)\left(1 - \sqrt{\frac{x+\delta}{x+1}}\right) < m^*.$$

In Proposition 4, assuming complete contracts, the coalition was indifferent to T if $M = M^*$, and any smaller coalition made them strictly prefer a one-period contract. This is no longer the case. With incomplete contracts, the small investments generated by the hold-up problem create a cost of signing short-term agreements. This cost must be weighed against the benefit of waiting for a larger coalition in the future. If the current coalition size, m, is smaller but close to the equilibrium size, m^*, then a long-term agreement with a smaller coalition is none-the-less preferred. The threshold making the coalition indifferent, $\widehat{m}(x, m^*)$, is thus strictly smaller than m^*.

[13] A similar result is derived in the literature on international trade (see Bagwell and Staiger 2001, where $T = 1$ and $n = 2$).

Proposition 7 allows us to predict the duration if a country deviates from the equilibrium by not participating. In particular, for a unilateral deviation to trigger $T = 1$ it must be the case that $m^* - 1 \leq \widehat{m}(x, m^*)$. This inequality implies that m^* cannot be too large.

Corollary to Proposition 7. *If a single country deviates by not participating, the remaining coalition sets $T = 1$ only if $m^* \leq m_M(x)$, where*

$$m_M(x) \equiv 1 + \frac{1}{1 - \sqrt{(x+\delta)/(x+1)}}.$$

We will refer to the inequality $m^* \leq m_M(x)$ as the *discipline constraint.* If it is violated, then even if a country $i \in M^*$ deviates by not participating, the remaining participants will proceed by signing a long-term agreement ($T = \infty$). If the discipline constraint is instead satisfied, then whenever some $i \in M^*$ deviates by not participating, the remaining coalition signs a one-period agreement only while it waits for i to return to the equilibrium strategy in the next period.

C. Participation

Just as before, any equilibrium coalition M^* must ensure that every $i \in M^*$ prefers to participate. Larger coalitions require larger reductions in pollution from their members (in line with Proposition 6), and this makes it more tempting to free ride. The individual participation constraint thus requires that $m^* = |M^*|$ cannot be too large.

If $m^* > m_M(x)$, such that the discipline constraint is violated, then the coalition signs a long-lasting agreement ($T = \infty$) whether i participates or deviates. Proposition 6 then fully characterizes the impact of the smaller m, and investments are exactly as with complete contracts. Compared to the situation with complete contracts, the only differences are that now free-riding gives a benefit in *every* period rather than just for one period; but the cost (i.e., the coalition pollutes more) is *also* suffered in every period rather than just one. By comparison, the participation constraint still requires that the (one-period) cost is larger than the (one-period) benefit.[14] As shown in the previous section, this participation constraint is satisfied if and only if $m^* \leq 3$.

If the discipline constraint holds, so that $m^* \leq m_M(x)$, then $i \in M^*$ anticipates that free-riding would lead to a one-period agreement, triggering the hold-up problem. That is, free-riding implies that even participants will reduce technology stocks from $m^* C/K$ to simply C/K (rather than to $(m^* - 1)\, C/K$ as with complete contracts). Thus, the punishment for free-riding is now higher and so the individual participation constraint can be satisfied for a larger m^*. The next result determines this threshold, $m_I(x)$, and allows us to characterize all the Markov equilibria.

[14] When a country free-rides in every period rather than in just one period, the benefit as well as the cost must be multiplied by $1/(1-\delta)$.

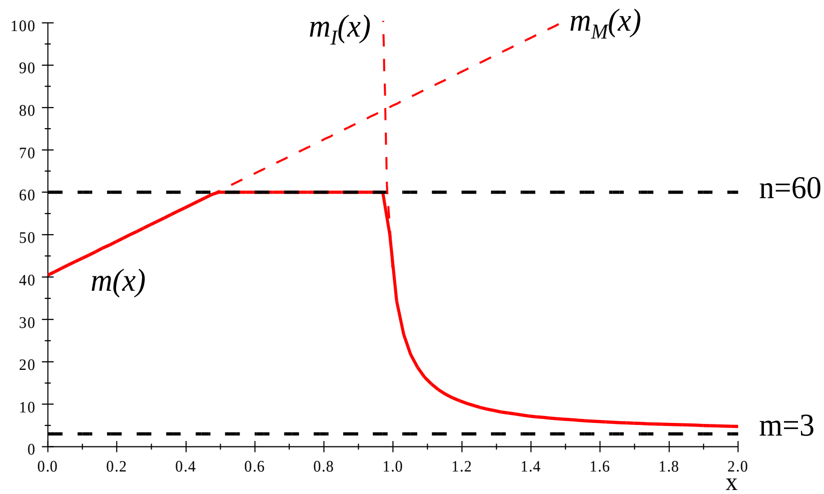

Figure 2: *The coalition size m^* must be below all three curves*

Proposition 8. *M^* is an equilibrium coalition if and only if either $m^* = |M^*| \leq 3$ or $3 < m^* \leq \min\{n, m(x)\}$, where*

$$m(x) = \min\{m_I(x), m_M(x)\} = \begin{cases} m_M(x) & \text{if } x < \widehat{x} \\ m_I(x) & \text{if } x \geq \widehat{x} \end{cases}, \tag{9}$$

with

$$m_I(x) \equiv 3 + \frac{2\delta}{x - \delta}$$

and

$$\widehat{x} = \frac{1}{6}\left((1+\delta) + \sqrt{(1+\delta)^2 + 12\delta}\right) \in \left(\frac{1}{3}, 1\right).$$

Just as before, we do have equilibria where the coalition size is just two or three. In addition, the equilibrium coalition size m^* can now be much larger, as long as it satisfies $m^* \leq m(x)$. In fact, if $n \leq m(x)$, the grand coalition is an equilibrium outcome and the first-best outcome would be implemented. Figure 2 illustrates $m(x)$ as a function of x. The figure shows that, even for very small discount factors, equilibrium participation can be significantly larger than 3 countries, which is the upper bound with complete contracts. In the example in Figure 2 there is an interval for x in which all countries choose to join the IEA and thus the outcome is efficient.

As the above furmulae make clear a key variable is the relative cost of technology, x. This variable has interesting but ambiguous effects on the coalition size. Intuitively, a larger x means that technological investment becomes both more expensive and less important as a policy relative to simply reducing consumption. Thus, when x is large, the under-investment problem following a short-term agreement is less important. This has two consequences. On the one hand, this makes the coalition more willing to sign a short-term agreement and wait for a larger coalition in the future: the discipline constraint is thus relaxed and $m_M(x)$ increases. On the other hand, it becomes more tempting for $i \in M^*$ to deviate, since the subsequent hold-up problem is, in any case, less important: the participation constraint is thus strengthened and $m_I(x)$ declines. When $x < \widehat{x}$, the binding constraint is $m_M(x)$. If $x > \widehat{x}$, the binding constraint is $m_I(x)$. To satisfy both constraints, x must be moderate.

D. Comparing Contractual Environments

In both contractual environments, the equilibrium coalition will be formed and an ever-lasting agreement will be signed. Every $g_{i,t}$ and $r_{i,t}$ would thus be exactly the same in the two cases if the coalitions were the same. The comparison between the two environments thus boils down to the coalition sizes that can be sustained. The coalition size is is important, since utilitarian welfare increases monotonically in the equilibrium coalition size, m^*.

Proposition 8 makes clear that in an incomplete contracting environment we can always sustain a coalition with three countries. For a precise comparison of the equilibrium outcomes in a complete and an incomplete contractual environment, it is useful to recast the result of Proposition 8 to characterize the conditions under which a given coalition size can be supported in equilibrium. To this end, note that for every potential equilibrium coalition size m^*, the discipline constraint $m^* \leq m_M(x)$ requires:

$$x \geq \underline{x}(\delta, m^*) \equiv \frac{(m^*-2)^2 - \delta(m^*-1)^2}{(m^*-1)^2 - (m^*-2)^2}. \tag{10}$$

Similarly, the participation constraint $m^* \leq m_I(x)$ requires:

$$x \leq \overline{x}(\delta, m) \equiv \delta + \frac{2\delta}{m-3}. \tag{11}$$

It follows that a coalition size $m^* \in (3, n]$ is feasible in equilibrium if and only if x is moderate in the following sense:

$$\underline{x}(\delta, m^*) \leq x \leq \overline{x}(\delta, m^*). \tag{12}$$

Since utilitarian welfare is increasing in m^*, (12) allows us to characterize when a coalition of size $m > 3$ is feasible and, therefore, when the best MPE with incomplete contracts is strictly superior to the best MPE with complete contracts. Expression (12) also allows us to characterize when a coalition of size $m = n$ is feasible and, thus, when the best MPE with incomplete contracts achieves the first-best outcome.

Proposition 9. *(i) The maximal coalition size is always weakly larger with incomplete contracts than with complete contracts.*

(ii) It is strictly larger if and only if:

$$x \in [\underline{x}(\delta, 4), \overline{x}(\delta, 4)] = \left[\frac{1}{5}(4 - 9\delta), 3\delta\right],$$

a set that is non-empty if $\delta \geq 1/6$.
(iii) Moreover, for any n*, the best equilibrium with incomplete contracts implements the first-best outcome if and only if:*

$$x \in [\underline{x}(\delta, n), \overline{x}(\delta, n)] = \left[\frac{(n-2)^2 - \delta(n-1)^2}{(n-1)^2 - (n-2)^2}, \delta + \frac{2\delta}{n-3}\right],$$

a set that is non-empty if $\delta \geq (n-2)(n-3)/n(n-1) < 1$.

It is interesting to note how these constraints and regions depend on the discount factor. As expected, if a coalition of size m is feasible at x with some δ, then it remains feasible for any $\delta' > \delta$: the more patient the agents are, the larger is the set of parameters that support a given coalition size. However, an efficient outcome is not always possible, even if δ is arbitrarily large. From (iii) we can see that if $x > \overline{x}(1, n) = (n-1)/(n-3)$, then there is no $\delta \leq 1$ such that all countries find it optimal to join.[15]

A crucial assumption in the above analysis is that the game is dynamic and that the contract length can be endogenously negotiated among participating countries. It is easy to show that if the duration T were exogenous, the equilibrium coalition size would be $m^* \leq 3$ regardless of the contractual environment. Since incomplete contracts generate under-investments in period T, the complete contracting environment would strictly Pareto dominate the incomplete contracting environment for any fixed $T < \infty$.[16] In the robustness section, we discuss how our results survive if the duration is endogenous but limited by a finite upper threshold, $T \leq \overline{T}$.

[15] We have chosen to emphasize the effects of x rather than the impact of δ since the discount factor has multiple interpretations (as patience or period-length, for example). The alternative interpretations would have conflicting implications for how one should change the model's other parameters when δ changes.

[16] The fact that, if T is exogenous, we obtain an inefficient outcome highlights the differences of our theory with Bagwell and Staiger's (2001) theory of trade agreement mentioned in the introduction. First, in our model efficiency only requires control of pollution limits; in Bagwell and Staiger (2001) if only tariffs are controlled, the equilibrium is inefficient. As the analysis in this section however headlights, our efficiency result is true only if the length and the participation in the agreement are endogenous, features that characterize our model. Second, in Bagwell and Staiger (2001) an efficient allocation is possible even without explicit regulation of domestic standards (and so with an "incomplete contract" in our terminology) if countries can commit to a given level of market access and despite the model is static. Efficiency however is possible because the are no non-pecuniary externalities and because participation to the agreement is exogenous. In our model, efficiency is possible because the model is dynamic and both participation and the length of the agreement are endogenous. Finally, in our model complete contracts make efficiency impossible; in Bagwell and Staiger (2001) complete contracts are always good. This occurs because participation is exogenous in their paper; one of the main results of our paper is to show that contractual completeness is harmful to participation.

V. Endogenous Incomplete Contracts

In the preceding sections we have analyzed several alternative situations: complete contracts, no contracts, and incomplete contracts.[17] This way, we did not need to take a definitive stand on what the appropriate contractual environmental is likely to be. Traditionally, the literature on incomplete contracts assumes the nature of the contractual environment is exogenous: we should expect incomplete contracts when investments are ex-post observable by the negotiating partners but not verifiable by a third party, such as an international court.[18] Contractual incompleteness, therefore, seems appropriate when it is hard or costly to describe the exact nature of the investment and its expected payoffs in all conceivable contingencies. For our specific application, there are at least two reasons for assuming contractual incompleteness. First, part of the investment in green technology is in basic research and this may be difficult to describe ex-ante. Second, establishing unanimous criteria to evaluate the effectiveness of new technologies is often more controversial in environmental matters because of its political nature: there are well-funded lobbyists that can produce countervailing evidence on the feasibility of new green technologies and make the issue controversial (even if the science is not).[19]

There are situations in which the contracting environment is best viewed as endogenously determined. To consider this case, we first introduce the possibility of acquiring a "contracting technology" such that investments can be contracted on (for example by establishing standards of measurements and monitoring facilities). We prove that contracts will always remain incomplete, in equilibrium, but the possibility to switch to complete contracts might nevertheless influence the outcome. In Section V.B we allow a variety of technologies where some are contractible while others are not. Measuring the degree of contractual incompleteness we can extend the results from Section IV as well as derive the optimal (and equilibrium) degree of incompleteness.

A. Endogenizing the Contractual Environment

We now let countries decide on whether to make the contractual environment complete: for instance, they can write detailed rules regarding how investments should be measured and establish regulatory agencies that verify and measure each country's investment.[20] Establishing such a monitoring technology on a

[17] For the sake of brevity, we have withheld the analysis of the (counter-factual) case in which investments are contractible but emissions are not.

[18] See Hart (1995) and Tirole (1999) for authoritative discussions of the conditions under which it is plausible to assume incomplete contracts.

[19] For example, in September 2013, after the adminstration of the *Environmental Proptection Agency* Gina McCarthy proposed to limit new coal power plants to 1100 pounds of carbon dioxide per Megawatt hour, opponents as the *Electric Reliability Coordinating Council* have responded that the technology to reduce emissions is not yet available (see *New York Times*, 19 September 2013).

[20] On the importance of establishing technological standards, note that the *Environmental Protection Agency* has established a pilot program of cooperation with international authorities to establish mutual recognition of environmental technology verification programs. See for example http://ec.europa.eu/environment/etv/international.htm.

country may potentially require a cost $h \geq 0$ and some time $\Delta_m \geq 0$. It may be reasonable to assume that (a) the monitoring technology is durable and so the cost h is paid only the first time investments are measured, but we will also consider (b) the nondurable case where the cost must be paid in every period in which the monitoring technology is used.

As in the preceding analysis, let $m(x) \geq 3$ refer to the largest possible coalition size under incomplete contracts. If $m(x) = 3$, the equilibrium outcomes under incomplete and complete contracts coincide, so consider the case with $m(x) \geq 4$ and an equilibrium $m^* \in [4, m(x)]$.

Suppose first that $\Delta_m \geq \Delta$, so the decision to measure investments must be made before the coalition-formation stage. In this case it is clear that, regardless of h and whether monitoring is (a) durable or (b) reversible, it is optimal for the coalition to leave contracts incomplete. The coalition is larger under incomplete contracts, and the duration will, in both cases, be infinite in equilibrium.[21]

Suppose next that $\Delta_m < \Delta$, so the coalition can decide whether to contract on investments even after the coalition-formation stage. A complete contract is unnecessary if the actual number of coalition-members turned out to be $m \geq m^*$ since then a long-term agreement will be chosen.[22] Thus, suppose $m < m^*$. This out-of equilibrium possibility is (only) of interest to a country that is contemplating to free ride, so it is sufficient to consider the case $m = m^* - 1 \geq 3$. Two cases are relevant:

(a) Assume first that the technology is durable so that the decision of moving to complete contracts is essentially irreversible.[23] In this situation, the countries anticipate that after installing the measurement technology the equilibrium coalition-size is forever 3 (at best). The sum of payoffs (for the coalition-members) is then smaller than if the current coalition of size $m \geq 3$ commits to a long-term agreement; an option that is available without the measuring technology. Consequently, the coalition will never want to make an irreversible switch to a complete contracting environment no matter the levels of $h \geq 0$ or $m \geq 3$.

(b) Assume next that the technology is not durable and measurement cost h must be paid in every period. Switching to a complete contract in this period will then not affect any future MPE. As before, if only $m = m^* - 1$ countries participate at time t, the coalition will prefer to negotiate a short-term agreement (since $m^* \leq m_M(x)$). If investments are not part of the contract, Proposition 6 states that $R_{i,t+1} = C/K$. If investments are part of the con-

[21] As explained in footnote 26, it is possible that a fourth country would strictly prefer a complete contracting environment, but Coasian bargaining would predict that the surplus-maximizing incomplete contracting environment would prevail.

[22] Of course a coalition can simultaneously make the contractual environment complete and choose $T < \infty$. This, however, would not be optimal since after the end of the agreement no coalition larger than 3 is formed.

[23] This assumption is reasonable if setting up a measurement technology requires a fixed cost, while the cost of subsequently applying and maintaining the technology is negligible (for the result, it is sufficient to assume that the subsequent cost is strictly lower than the initial set-up cost).

tract, Proposition 3 states that instead $R_{i,t+1} = mC/K$. The one-period gain from contracting on investments is $\delta (m-1)^2 C^2/2K$ and this is less than the cost if:

$$\delta (m-1)^2 \frac{C^2}{2K} \leq h \Rightarrow m+1 \leq m_h \equiv 2 + \frac{\sqrt{2Kh/\delta}}{C}. \tag{13}$$

Consequently, any $m^* \leq m_h$ can be an equilibrium coalition size without violating the constraint that the coalition will stick to the incomplete contracts even when $m = m^* - 1$. Combined with Proposition 8, we can conclude that any $m^* > 3$ is an equilibrium coalition size if just $m^* \leq \min \{m_I(x), m_M(x), m_h, n\}$. While both m_M and m_h increases in K, m_I decreases in K. Thus, a simple sufficient condition for the threshold m_h to be nonbinding is $K > \delta C^2/2h \Leftrightarrow m_h > m_I$.

Proposition 10. *Suppose the countries can choose to sign complete contracts.*
(i) In equilibrium, contracts are always incomplete.
(ii) If the decision is irreversible or $\Delta_m \geq \Delta$, the equilibrium is as described by Propositions 6-8.
(iii) If the decision is reversible and $\Delta_m < \Delta$, Propositions 6-8 hold if (9) is replaced by: $m(x) = \min \{m_I(x), m_M(x), m_h\}$.

It is also possible to endogenize the measurement cost. Suppose that the recurring cost $h \in [\underline{h}, \overline{h}]$ can be reduced if the countries take appropriate action in advance (before the coalition-formation stage). For example, countries might be able to exert some effort (or up-front payments) in order to reduce the future cost h. What, then, is the equilibrium effort and h? The simple answer is that the countries will never exert *any* effort in reducing the future h, so $h = \overline{h}$. The (only) consequence of exerting effort would be that m_h and thus m^* may be reduced according to (13). In fact, countries may instead prefer to *raise* h (and thus m_h). This way, our model can explain why contracting on investments is costly (and perhaps artificially costly).

B. The Optimal Degree of Incompleteness

So far, there has been a stark distinction between complete and incomplete contracting environments. Since the reality may be somewhere in between, consider now a situation in which there is a large set of green technology investments (a continuum of mass one). The technologies are identical and characterized by the same investment cost, depreciation rate, perfect substitutability and effectiveness. The only difference is that a fraction $\alpha \in [0,1]$ of these technologies (and the associated investments) are contractible while the others are not. In an agreement with duration T, this implies that in the last period, investments ensure that $R_{i,T+1} = mC/K$ for the mass α of contractible investments while $R_{i,T+1} = C/K$ for the noncontractible ones. The results from Section IV continue to hold as long as $\alpha < 1$, but now we obtain more nuanced results in which the feasibility set depends on α.

Proposition 11. *Suppose a fraction $\alpha \in [0,1]$ of investments are contractible. M^* is an equilibrium coalition if and only if either $m^* = |M^*| \leq 3$ or $3 < m^* \leq$*

$\min\{n, m(x;\alpha)\}$, *where*

$$m(x;\alpha) = \min\{m_I(x;\alpha), m_M(x;\alpha)\}, \text{ with}$$
$$m_I(x;\alpha) \equiv 3 + \frac{2\delta(1-\alpha)}{x + 2\delta\alpha - \delta};$$
$$m_M(x;\alpha) \equiv 1 + \mu(\alpha) + \sqrt{\mu(\alpha)[\mu(\alpha)-1]};$$
$$\mu(\alpha) \equiv \frac{1 + x - \alpha(1-\delta)}{(1-\alpha)(1-\delta)} > 1.$$

The proof is available in the appendix. The threshold for the participation constraint, $m_I(x;\alpha)$, is decreasing in α for the same reason that $m_I(\cdot)$ was and still is decreasing in x: when the hold-up problem becomes less important (because either x or α increases), then a country fears the consequences of a one-period agreement less and free-riding becomes tempting unless the coalition is sufficiently small. At the same time, it becomes less costly for the coalition to sign a one-period agreement if a deviator free rides. Thus, the threshold for the discipline constraint, $m_M(x;\alpha)$, is increasing in α as well as in x. Combined, $m(x;\alpha)$ increases in α when $m_M(x;\alpha) < m_I(x;\alpha)$ but decreases in α otherwise.

To complement the previous subsection, we can endogenize the contractual environment by deriving the preferred level of incompleteness—if the countries could decide on α. For the sake of brevity, we consider only the case where the countries cooperatively decide on α before the coalition formation stage (i.e., $\Delta_m \geq \Delta$, using the notation in the previous subsection). They would then prefer to set α such that the coalition size would would be as large as possible. In the appendix we prove:

Proposition 12. *Let* $\alpha^*(x) = \arg\max_\alpha m(x;\alpha)$: *(i) if* $x \geq \widehat{x}$, *then* $\alpha^*(x) = 0$*; (ii) if* $x < \widehat{x}$, *then* $\alpha^*(x) \in (0,1)$, $\alpha^*(x)$ *decreases in* x *and it is such that* $m_I(x;\alpha^*(x)) = m_M(x;\alpha^*(x))$.

Note that it is *always* efficient to have some degree of contractual incompleteness: $\alpha^* < 1$. The reason is that in the limit when $\alpha \uparrow 1$, then $m_I(x;\alpha) \downarrow 3$ at the same as $m_M(x;\alpha) \uparrow \infty$. So, for (almost) complete contracts, the binding constraint is always $m_I(x;\alpha)$, which is decreasing in α.

On the other hand, it is possible that $\alpha^* = 0$. If $x \geq \widehat{x}$, defined in Proposition 8, then $m_I(x;\alpha) \leq m_M(x;\alpha)$ even when $\alpha^* = 0$, and thus the binding constraint is $m_I(x;\alpha) < m_M(x;\alpha)$ for every $\alpha \in (0,1]$. In this case, $m(x;\alpha) = m_I(x;\alpha)$ is always decreasing in α and thus we have the corner solution $\alpha^* = 0$.

The importance of the treshold $\widehat{x}$ is therefore intuitive: if $x < \widehat{x}$, then $m_I(x;\alpha) > m_M(x;\alpha)$ when $\alpha^* = 0$. Since we also know that $m_I(x;1) < m_M(x;1)$ and because both thresholds are continuous in α, there exists an $\alpha^* \in (0,1)$ such that the two thresholds cross, $m_I(x;\alpha^*) = m_M(x;\alpha^*)$. The best degree of contractual incompleteness is then ensuring that both constraints are binding and equalized. Since $\partial m_I(x;\alpha)/\partial x < 0$ while $\partial m_M(x;\alpha)/\partial x > 0$, we have that α^* must decrease in x to ensure $m_I(x;\alpha^*) = m_M(x;\alpha^*)$. Plainly, if the green investments are relatively expensive, then a larger fraction of them should remain noncontractible.[24]

VI. Robustness

In this section we discuss a few extensions of the basic model to show that the results are robust with respect to a number of modeling choices we made for convenience. In particular, we (A) generalize the quadratic formulae above, (B) permit limits on the possibility to commit, (C) show that the bargaining outcome we have assumed can be derived in a noncooperative bargaining game, and (D) discuss how to relax our equilibrium refinement. All the extensions build on the model above (rather than building on each other), so they can be read isolated and in any order.

A. Relaxing the Functional Forms

The adoption of a model with quadratic preferences and cost functions is a convenient choice in the preceding analysis: first, it allows us to directly compare our result with the previous literature (that has made the same assumption); second, it permits simple, closed form solutions and thus keeps the analysis clean. The intuition for why incomplete contracts are helpful, however, does not hinge on the quadratic formulation. To see how this result generalizes, suppose the disutility of consumption-reduction, $B(d_{i,t})$, is a general increasing concave function while the investment-cost is $\delta K(R_{i,t})$, a general increasing and convex function. Suppose $q_R = 0$ for simplicity. If, in addition, we continue to let the marginal disutility of pollution be the constant C, then a complete contract implies: $d_{i,t} = B^{'-1}(Cm)$ and $R_{i,t} = K'^{-1}(Cm)$ for the members and $d_{i,t} = B^{'-1}(C)$ and $R_{i,t} = K'^{-1}(C)$ for nonparticipants. The same is true for the case where an incomplete contract lasts forever.

Just as before, we can show that the largest possible coalition size under incomplete contracts is larger than the largest possible coalition size under complete contracts.

To see this result, note that each member of an m-sized coalition receives the following payoff from this period's choices (analoguous to $\widehat{u}_{i,t}$ in Lemma 1):

$$\widehat{u}_m^M = -B\left(B^{'-1}(Cm)\right) - \delta K\left(K'^{-1}(Cm)\right) - C\sum_{i\in N}\overline{y}_i$$
$$+ mC\left[B^{'-1}(Cm) + \delta K'^{-1}(Cm)\right] + (n-m)\,C\left[B^{'-1}(C) + \delta K'^{-1}(C)\right].$$

[24] If $n < m(x;\alpha)$ for some $\alpha \in [0,1]$, neither constraint is binding. There is then an interval $[\underline{\alpha}(x), \overline{\alpha}(x)] \subset \Re$ such that for every $\alpha \in [\underline{\alpha}(x), \overline{\alpha}(x)]$, $m(x;\alpha) = n$ and the first-best is possible. The lower threshold is defined by $m_M(x;\underline{\alpha}(x)) = n$ while the upper threshold is defined by $m_I(x;\overline{\alpha}(x)) = n$. If $\overline{\alpha}(x) < \underline{\alpha}(x)$, the interval is empty and the coalition size is maximized by $\alpha^* = \arg\max_\alpha m(x;\alpha) < n$, as described by Proposition 12.

If a country deviates from this equilibrium, this nonparticipant's one-period payoff is given by the following if contracts are complete:

$$\begin{aligned}\widehat{u}_{m-1}^{CC} &= -B\left(B^{'-1}\left(C\right)\right) - \delta K\left(K'^{-1}\left(C\right)\right) - C\sum_{i\in N}\overline{y}_i \\ &+ \left(m-1\right)C\left[B^{'-1}\left(C\left(m-1\right)\right) + \delta K'^{-1}\left(C\left(m-1\right)\right)\right] \\ &+ \left(n-m+1\right)C\left[B^{'-1}\left(C\right) + \delta K'^{-1}\left(C\right)\right].\end{aligned}$$

If the contract is instead incomplete, investments will be lower in this period so the deviator's payoff becomes:

$$\begin{aligned}\widehat{u}_{m-1}^{IC} &= -B\left(B^{'-1}\left(C\right)\right) - \delta K\left(K'^{-1}\left(C\right)\right) - C\sum_{i\in N}\overline{y}_i \\ &+ \left(m-1\right)CB^{'-1}\left(C\left(m-1\right)\right) + \left(n-m+1\right)CB^{'-1}\left(C\right) + \delta nCK'^{-1}\left(C\right).\end{aligned}$$

Clearly, we must have $\widehat{u}_{m-1}^{CC} > \widehat{u}_{m-1}^{IC}$, since the participants invest more under complete contracts and this is beneficial for a nonparticipant.

The participation constraint for a coalition of size m requires that each member must find participation better than free-riding one period. For complete contracts, this implies $\widehat{u}_m^M \geq \widehat{u}_{m-1}^{CC}$ but for incomplete contracts, the condition is $\widehat{u}_m^M \geq \widehat{u}_{m-1}^{IC}$, which is weaker since $\widehat{u}_{m-1}^{CC} > \widehat{u}_{m-1}^{IC}$. It follows that every potential member finds free-riding less attractive if the contract is incomplete than if it is complete. Thus, the upper boundary m_I (i.e., the largest m satisfying $\widehat{u}_m^M \geq \widehat{u}_{m-1}^{IC}$) must be larger for incomplete contracts.

The necessary condition $\widehat{u}_m^M \geq \widehat{u}_{m-1}^{IC}$ is sufficient if it is indeed optimal for the coalition to sign a one-period agreement when one of the countries deviates by not participating. This requires:

$$\frac{\widehat{u}_{m-1}^M}{1-\delta} \leq \widehat{u}_{m-1}^{M,1} + \frac{\delta \widehat{u}_{m-1}^M}{1-\delta}, \tag{14}$$

where $\widehat{u}_{m-1}^{M,1}$ is the first-period payoff for one of the $m-1$ coalition-members. Clearly, this condition is nonbinding if δ is sufficiently close to one since $\widehat{u}_m^M > \widehat{u}_{m-1}^M$. If condition (14) fails, then the deviator's payoff is $\widehat{u}_{m-1}^{CC}$ as with complete contracts. In other words, the equilibrium coalition size is larger for incomplete than for complete contracts even if $B(\cdot)$ and $K(\cdot)$ are non-quadratic.[25]

B. Commitment and Time Horizon

In the preceding analysis we have assumed that countries in the IEA can commit to a policy for the entire length of an agreement and we have therefore

[25] Of course, even when contracts are complete the coalition size might be much larger than three (this point has been made by Karp and Simon (2013). Furthermore, we cannot conclude that the coalition is always *strictly* larger with incomplete contracts since m must be an integer and the largest integer satisfying $\widehat{u}_m^M \geq \widehat{u}_{m-1}^{IC}$ may equal the largest integer satisfying $\widehat{u}_m^M \geq \widehat{u}_{m-1}^{CC}$ even if $\widehat{u}_{m-1}^{CC} > \widehat{u}_{m-1}^{IC}$. It is for such reasons it is helpful with specific functional forms, such as those we have above.

focused on the issue of participation. In doing this we are following a typical approach in the literature on environmental international agreements (see Barrett 2005). The question of whether such commitment is actually possible is ultimately empirical and it has been closely scrutinized in the literature. In one of the most comprehensive empirical studies, Breitmeier et al. (2006) conclude that although compliance problems are frequently encountered, "the majority of member states comply with the majority of international environmental rules most of the time." (see Chapter 3, p.66). These significant levels of compliance are explained as the result of explicit enforcement mechanisms in the agreements,[26] but more often by other factors that are often ignored in game theoretic models: for example establishing mechanisms of compliance monitoring or performance assessment that increase media scrutiny and peer pressure (See Young 2011, Peterson 1997).[27]

Still, it is clear that problems with incentive compatibility and compliance in international agreements may limit their effectiveness. To explore this issue we study how the analysis changes when the countries can only commit for the entire length of the agreement. An interesting benchmark is the case in which countries cannot commit for more that $\overline{T}$ periods. It can be shown that the presence of an upper bound does not change the equilibrium characterization when there is contractual completeness as in Section III. In an incomplete contracting environment we can show following the exact same steps as in Section IV that an equilibrium exists if and only if:

$$\underline{x}(\delta, m^*) \leq x \leq \overline{x}(\delta, m^*, \overline{T}),$$

where $\underline{x}(\delta, m^*)$ is defined as in Corollary 1 to Proposition 8 while

$$\overline{x}(\delta, m^*, \overline{T}) \equiv \delta \left(1 - \frac{(1-\delta)\,\delta^{\overline{T}-1}}{1-\delta^{\overline{T}}}\right) \frac{m^*-1}{m^*-3}. \tag{15}$$

The analysis is therefore as in the previous sections, except that the upper bound of the feasibility set, $\overline{x}(\delta, m^*, \overline{T})$, is now an increasing function of $\overline{T}$: the smaller $\overline{T}$ is, the smaller the region of parameters that sustains an IEA of size m^* is. The intuition is that if $i \in M^*$ deviates by not participating, the hold-up problem is moved forward from $\overline{T}$ to the current period. If $\overline{T}$ is small, this "penalty" is small so the participation constraint strengthens and, to satisfy it, x must be smaller. However, as can be easily verified from (15), quite large coalitions are feasible in an incomplete contracting environment even

[26] Explicit enforcement procedures are contemplated, for example, in the Montreal Protocol, the Protocols of the Geneva Convention, the Basel Protocol, the Aarhus Convention on the Access to Justice in Environmental Matters, the Carthagena Protocol on Biosafety, the International Treaty on Plant Genetic Resources for Food and Agriculture, the Stockholm Convention on Persistent Organic Pollutant (see Breitmeier et al. 2006).

[27] Examples are mandatory reporting systems of routine information (for example, in the International Maritime Organization), mechanisms to publicly report deviant behavior to the central organization (for example, in the Montreal Protocol, Madrid Protocol on Antarctic Environmental Protection), mechanisms of performance assesment (for example the Convention on North Pacific Anadromous Stock (Annex II)); see Peterson (1997) for details.

when the expected length of the agreement is short. Naturally, the upper bound converges to $\overline{x}(\delta, m^*)$ as $\overline{T} \to \infty$.

C. Noncooperative Bargaining

In the analysis presented above we have assumed that the policies in the IEA are chosen cooperatively. In this section we present a simple microfoundation of the cooperative decision rule used in the previous sections. To achieve this, we adopt a bargaining model introduced by Baron and Ferejohn (1989), now a standard workhorse model in the political economy literature. Bargaining, in this model, follows a simple dynamic protocol. First, one of the signatory countries is randomly selected to make a proposal. The proposal consists of a time horizon, pollution limits $g_{i,t}$ and (if possible) investments $r_{i,t}$ for each country and each period of the agreement, and a vector of monetary transfers z_i for each country that satisfy budget balance ($\sum_N z_i = 0$).[28] Each country has the same probability of being selected to make a proposal. Countries observe the proposal and unanimity is required. If the proposal is accepted, then it is implemented and bargaining ends; if the proposal is rejected, then another country is selected to be the proposer and the process is repeated. The process stops when a policy is chosen. The time between subsequent offers is close to zero, so we ignore discounting between offers.

It is relatively straightforward to prove that if an IEA is an equilibrium of the games studied in the previous sections, then it is an equilibrium of the corresponding game in which policies in the IEA chosen with the non-cooperative bargaining protocol described above. The intuition behind this result is as follows.[29] Take the problem faced by a country selected to propose an IEA. For simplicity, consider only the case with incomplete contracts (the case with complete contracts is almost identical). Let $u_l(g_{j,l})$ be the indirect utility of country j at time l given the equilibrium investment in green technology $R_{j,l}(g_{j,l})$ from Proposition 6.[30] The proposing country desires to maximize its expected utility, but will be forced to make a proposal sufficiently appealing to be approved by all other participants. Formally, the proposer's problem at time t can be

[28] In Baron and Ferejohn's bargaining model countries are allowed to make monetary transfers among each other. As we have said in the introduction, monetary transfers are not typically observed in IEAs. Since in the equilibrium described below transfers are zero, however, this evidence is not necessarily in contrast with the bargaining model with transfers of this section.

[29] Proofs for this result, and the other results in this section, are available from the authors.

[30] Formally $u_l(g_{j,l})$ is equal to $-\frac{b}{2}\left(Y_{i,l} - g_{i,l} - R_{i,l}\right)^2$ for $l = t$, where $R_{i,t}$ is taken as given from the previous period; to $-\frac{b}{2}\left(Y_{i,l} - g_{i,l} - R_{i,t}(g_{i,l})\right)^2 - \frac{K}{2} R_{i,l}\left(g_{i,l}\right)^2$ for $l = t+1, ...T-1$ where $R_{i,l}\left(g_{i,l}\right)$ is given by Proposition 4.1; and to $-\frac{b}{2}\left(Y_{i,l} - g_{i,l} - R_{i,l}\right)^2 - \frac{K}{2} R_{i,l}^2 - \delta \frac{K}{2}\left(\frac{C}{K}\right)^2 + \delta^T C \sum_{j \in N} \frac{C}{K}$ for $l = T$.

stated as:

$$\max_{g_{j,l},t_j,T}\left\{\begin{array}{c}\sum_{l=t}^{t+T}\delta^{l-t}\left(u_l(g_{j,l})-c\sum_{j\in M}g_{j,l}\right)-\sum z_j \\ s.t.\ z_j+\sum_{l=t}^{t+T}\delta^{l-t}\left(u_l(g_{j,l})-c\sum_{j\in M}g_{j,l}\right)+\delta^T v_j\geq V_j(M)\end{array}\right\},\quad (16)$$

where $V_j(M)$ is the outside option for a country that refuses the proposal: that is, the expected utility of entering a new round of bargaining before knowing who the proposer will be. The inequality in (16) is the individual rationality constraint: each agent j must be better off accepting the proposer's offer (the left-hand side of the inequality) than by rejecting it (the right-hand side). It can be shown that the inequality holds as an equality, so we have:

$$z_j=V_j(M)-\sum_{l=t}^{t+T}\delta^{l-t}\left(u_l(g_{j,l})-c\sum_{j\in M}g_{j,l}\right)-\delta^T v_j. \quad (17)$$

It is important to note that although endogenous in the model, from the point of view of the proposer, $V_j(M)$ is a constant independent of his or her proposal. Given this, it is easy to see that, modulo a constant that is irrelevant for the solution, we can rewrite (16) as:

$$\max_{g_{j,l},T}\left\{\sum_{j\in M}\sum_{l=t}^{t+T}\delta^{l-t}\left(u_l(g_{j,l})-c\sum_{j\in M}g_{j,l}\right)+\delta^T v_j\right\}, \quad (18)$$

which is the utilitarian problem we have been assuming. Note, moreover, that the proposer does not need to make a transfer to have the policy accepted (and will not be able to extract any surplus). If the other countries are expecting a utilitarian solution with no transfer, their expected continuation is $V_j(M)=\frac{1-\delta^{T^*-1}}{1-\delta}\left(u_l(g^*_{j,l})-c\sum_{j\in M}g^*_{j,l}\right)+\delta^{T^*}v_j$, where $g^*_{j,l},T^*$ is the solution of (18). Condition (17) then implies that $z_j=0$. Therefore, the cooperative solution assumed in Sections 3 and 4 is an equilibrium of this non-cooperative bargaining.

D. The Equilibrium Concept

Up to this point we have focused the analysis to the study of Markov-perfect equilibria (MPEs). These equilibria are appealing because they are simple and they do not rely on complex punishment strategies that may seem unrealistic in many environments, including the coalition formation problem studied above. Although it is hard to test empirically what type of equilibrium is actually played in real world strategic interactions, recent experimental work has provided evidence in support of MPEs as the appropriate equilibrium concept in dynamic environment with state variables.[31] Because of this, MPEs are widely

[31] See for example, Battaglini, Palfrey Nunnari (2012) and (2013), Vespa (2013).

adopted to study dynamic strategic interactions.[32]

It is however important to recognize that more efficient equilibria are possible using history dependent strategies if the discount factor is sufficiently high. In an MPE, every time that the countries can choose an agreement, the coalition that is formed is history independent. A natural extension is to consider equilibria in which the coalitions that are formed after a deviation may depend on the history of coalitions before the deviation h^t (even though this history is payoff irrelevant). In this case, not only we can show that large agreements are possible using history dependent strategies, but that they can also be constructed with relatively simple strategies. We say that a Subgame-perfect equilibrium (SPE) is *simple* if after any history h^t, a coalition $M(h^t)$ is formed for all remaining periods. Differently from MPEs, the coalition may be history dependent; but the SPE is "simple" since it is unnecessary to construct a complex sequence of changing coalitions to discourage free-riding. In the appendix we formally prove that, if δ is sufficiently large, there exists a simple equilibrium in which any number $m \leq n$ of countries join an agreement, even in environments with complete contracts.

In these equilibria, a deviation is punished by the formation of a particular coalition designed to penalize the deviator for the remaining periods: this is done by forming a smaller and less efficient coalition in which the deviating country has to participate. Of course, these punishing coalitions must be equilibrium coalitions in the subgame following the deviation. This implies that the equilibria in these subgames must be supported by even worse threats, and even smaller coalitions. Although these equilibria are substantially more complicated than our MPEs because they require a nested chain of punishment phases, they may appear plausible in some environments. In these cases the differences between environments with and without complete contracts that we have highlighted in the previous pages may be less marked, since efficiency can be achieved in both cases, at least for high discount factors.

In an influential contribution, however, Barrett (2005, 1994) has argued that equilibria in coalition formation games should be at least *weakly renegotiation proof* as defined in Farrell and Maskin (1989).[33] The MPEs derived in the previous sections are all robust to this refinement, since MPEs are weakly renegotiation proof by construction.[34] In the appendix, however, we formally prove that *if* weakly renegotiation proof equilibria with $m > 4$ exist, then they can not be *simple* as defined above. This result, therefore, suggests that *if* larger coalitions can be sustained as renegotiation-proof equilibria with complete contracts, then these equilibria must rely on quite complex punishment strategies. So complex strategies may be unrealistic in the context of international envi-

[32] See, among others, Levhari and Mirman (1980), Dutta and Radner (2004), Battaglini and Coate (2007, 2008), Besley and Persson (2011), Harstad (2012), Battaglini et al. (2014) and many others.

[33] In our game an equilibrium is weakly renegotiation proof if there are no two histories h^t and $\widetilde{h}^t$ where an agreement is formed in which the continuation equilibrium strategies σ_{h^t} and $\sigma_{\widetilde{h}^t}$ are such that σ_{h^t} strictly Pareto dominates $\sigma_{\widetilde{h}^t}$.

[34] In each equilibrium, after any history, the continuation strategies and value function are uniquely defined, so the renegotiation proofness condition is automatically satisfied.

ronmental agreements.

VII. Further Readings

This lecture note follows Battaglini and Harstad (2016) which, in turn, builds on a large literature. There is a substantial literature on the hold-up problems associated with noncontractible investments (going back to Grossman and Hart, 1986, and surveyed by Segal and Whinston, 2010), but contractual incompleteness is generally either harmful or, at best, irrelevant, if the externalities are small and the contract is sufficiently long-lasting (Guriev and Kvasov, 2005). An important exception is Bernheim and Whinston (1998) who construct simple two-player, two-stage games in which, if some aspects of performance is not contractible, the optimal contract may also leave other aspects of performance unspecified when the players actions' are strategic complements. In this note we do not require pre-existing contractual incompleteness and because we focus on coalition formation, we study games with many players and an infinite horizon.

In environmental economics, there is an emerging literature that uses insights from the hold-up problem to study the relationship between investments in green technologies and international cooperation (see Buchholz and Konrad 1994, Harstad 2012, 2016, Becherle and Tirole 2011 and Helm and Schmidt 2013). These papers develop the idea that individual countries fear that investments in green technology today will weaken their bargaining position in the future, when new commitments are to be negotiated. However, these papers take participation as exogenously given and focus on the harmful effects of the hold-up problem. We integrate the hold-up problem with an endogenous model of coalition formation and agreement length to show how the hold-up problem can be beneficial and lead to a larger equilibrium coalition.[35]

A second strand of related literature in environmental economics focuses on the size of coalitions or IEAs. Building on the work by Palfrey and Rosenthal (1984) and D'Aspermont et al. (1983), this research has highlighted the fact that cooperative agreements are a form of public good, so countries should be expected to free ride on any form of negotiation.[36] The main result of this literature is that international agreements are incentive compatible only if they involve a very small number of countries (Hoel 1992, Carraro and Siniscalco 1983, Carraro et al. 2006, Barrett 1994, Dixit and Olson 2000). This is related to the "Paradox of International Agreements," mentioned above. The timing in these models, is, as here, that countries first decide whether or not to participate in a coalition, and, second, the coalition-members negotiate an agreement that maximizes the sum of the members' payoffs.[37] The prediction of small coalitions has been found to be robust by a large subsequent literature which concludes

[35] While relatively few papers focus on the hold-up problem, several permit both technological investments and emissions (Dutta and Radner 2004; van der Ploeg and de Zeeuw 1992). Hoel and de Zeeuw (2010) and Barrett (2006) even include a coalition formation stage.

[36] See the surveys by Barrett (2005) and Aldy and Stavins (2007, 2009) among others. A more general survey of the field of climate change economics can be found in Kolstad and Toman (2005).

[37] With the two stages, Coasian bargaining is prevented since a party can commit to not

that significant international cooperation is possible only if monetary transfers between countries are feasible (Carraro and Siniscalco 1993, Hoel and Schneider 1997, Bosello et al. 2003), or if the environmental technology is characterized by increasing returns or similar technical conditions (Barrett 2005, 2006, Heal and Kunreuther 2011, Karp and Simon 2013). Although this literature is primarily static, dynamic extensions have been presented by Barrett (1994), Rubio and Casino (2005), Rubio and Ulph (2007) with similar conclusions (see Calvo and Rubio 2012 for a survey).

This note has built on this literature and extended it in two directions. First, in the preceding literature negotiations of IEAs are confined to pollution limits lasting for an exogenous length, typically one period. In our dynamic model, the duration of the agreement is endogenously negotiated, so the length becomes a function of the coalition size. Second, we allow for investments in technology and consider environments in which complete contracts are admissible and environments in which only emission levels are contractible. We find that the small-coalition prediction is robust to each of these realistic extensions in isolation, but not when they are combined.[38]

Finally, this note is also related to the literature on international trade agreements.[39] Particularly related is Bagwell and Staiger (2001) who study an economy in which countries choose tariffs and other domestic policies to manipulate their terms of trade. As in our model, with no international agreement countries achieve an inefficient equilibrium (inefficiently low market access to foreign competitors). A trade agreement can be signed to achieve an efficient outcome if it allows countries to commit to a given level of market access to foreign competitors. Incomplete trade agreements that only set limits on tariffs, however, are inefficient. In contrast to us, Bagwell and Staiger (2001) study a static model in which participation is given and not voluntary, and they explicitly rule out nonpecuniary externalities.

negotiate later (Dixit and Olsson 2000, Ellingsen and Paltseva 2012). Alternative coalition-formation models are presented by, among others, Chwe (1994), Rey and Vohra (2001) and, applied to a dynamic model of climate treaties, de Zeeuw (2008).

[38] We also find that there is a positive relationship between the coalition size and depth, which contrasts with the typical observations in the literature (Barrett 2002, Finus and Maus 2008).

[39] See Bagwell and Staiger (2010) for an extensive review of this literature.

References

[1] Aldy, Joseph and Stavins, Robert (2007): *Architectures for Agreement,* Cambridge U. Press.

[2] Aldy, Joseph and Stavins, Robert (2009): *Post-Kyoto International Climate Policy,* Cambridge U. Press.

[3] Bagwell Kyle and R. Staiger (2001): "Domestic Policies, National Sovereignty and International Economic Institutions," *Quarterly Journal of Economics*, 116(2): 519-562.

[4] Bagwell Kyle and R. Staiger (2001): "The Worlds Trade Organization: Theory and Practice," *Annual Reviews of Economics*, 2: 223-256.

[5] Barrett, Scott (1994): "Self-enforcing international environmental agreements," *Oxford Economic Papers*, 46, p. 878-94.

[6] Barrett, Scott (1997): "The strategy of trade sanctions in international environmental agreements," *Resource and Energy Economics* 19, p. 345-61

[7] Barrett, Scott (2002): "Consensus Treaties," *Journal of Institutional and Theoretical Politics*, 158: 519–41."

[8] Barrett, Scott (2003): *Environment and Statecraft*, Oxford University Press.

[9] Barrett, Scott (2005): "The Theory of International Environmental Agreements," *Handbook of Environmental Economics 3*, edited by K.-G. Mäler and J.R. Vincent.

[10] Barrett, Scott (2006): "Climate Treaties and "Breakthrough" Technologies,"*American Economic Review, papers and proceedings* 96 (2): 22-5.

[11] Battaglini, Marco and Coate, Stephen (2007): "Inefficiency in Legislative Policy-Making: A Dynamic Analysis," *American Economic Review*, 97(1), 118-149.

[12] Battaglini, Marco, and Coate, Stephen (2008): "A Dynamic Theory of Public Spending, Taxation and Debt." *American Economic Review*, vol. 98, pp. 201-36.

[13] Battaglini, Marco and Harstad, Bård (2016): "Participation and Duration of Environmental Agreements." *Journal of Political Economy*, 124, 160-204.

[14] Battaglini, Marco, Nunnari, Salvatore and Palfrey, Thomas (2012): "Legislative Bargaining and the Dynamics of Public Investment," *American Political Science Review*, May 2012, 106 (2), pp. 407-429.

[15] Battaglini, Marco, Nunnari, Salvatore and Palfrey, Thomas (2013): "The Dynamic Free-rider Problem: A Laboratory Study," mimeo.

[16] Battaglini, Marco, Salvatore Nunnari, and Thomas R. Palfrey (2014). "The Free Rider Problem: a Dynamic Analysis," *American Economic Review*, forthcoming.

[17] Beccherle, Julien and Tirole, Jean (2011): "Regional Initiatives and the Cost of Delaying Binding Climate Change Agreements," *Journal of Public Economics* 95(11-12): 1339-48.

[18] Bernheim Douglas and Michael Whinston (1998): "Incomplete Contracts and Strategic Ambiguity," *American Economic Review*, 88(4): 902-932.

[19] Bosello, Francesco, Buchner, Barbara and Carraro, Carlo (2003): "Equity, Development, and Climate Change Control," *Journal of the European Economic Association* 1: 601–11.

[20] Breitmeier Helmut, Oran Young and Michael Zurn (2006): *Analyzing International Environmental Regimes: From Case Studies to Database*. MIT Press, Cambridge, USA.

[21] Buchholz, Wolfgang and Konrad, Kai (1994): "Global Environmental Problems and the Strategic Choice of Technology," *Journal of Economics* 60 (3): 299-321.

[22] Calvo, Emilio and Rubio, Santiago (2012): "Dynamic Models of International Environmental Agreements: A Differential Game Approach", forthcoming, *International Review of Environmental and Resource Economics.*

[23] Carraro, Carlo, Eyckmans, Johan, and Finus, Michael (2006): "Optimal Transfers and Participation Decisions in International Environmental Agreements," *The Review of International Organizations* 1(4): 379-96.

[24] Carraro, Carlo and Siniscalco, Domenico (1993): "Strategies for the international protection of the environment," *Journal of Public Economics* 52 (3): 309-28.

[25] Chwe, Michael (1994): "Farsighted Coalitional Stability," *Journal of Economic Theory* 63: 299-325.

[26] d'Aspremont, Claude., Jacquemin, Alexis, Gabszewicz, Jean jaskold and Weymark, John A. (1983): "On the stability of collusive price leadership," *The Canadian Journal of Economics* 16(1): 17–25.

[27] Dixit, Avinash and Olson, Mancur (2000): "Does voluntary participation undermine the Coase Theorem?" *Journal of Public Economics* 76 (3): 309-35.

[28] Dutta, Prajit K. and Radner, Roy (2004): "Self-enforcing climate-change treaties," *Proceedings of the National Academy of Science*, 101: 4746-51.

[29] Ellingsen, Tore and Paltseva, Elena (2012): "The Private Provision of Excludable Public Goods: An Inefficiency Result," *Journal of Public Economics* 96 (9–10): 658–69.

[30] Farrell, Joseph and Maskin, Eric (1989): "Renegotiation in Repeated Games," *Games and Economic Behavior*, 1, 327-360.

[31] Finus, Michael and Maus, Stefan (2008): "Modesty may pay!" *Journal of Public Economic Theory* 10: 801–26.

[32] Guriev, Sergei and Kvasov, Dmitriy (2005): "Contracting on Time," *The American Economic Review* 95 (5): 1269-1385.

[33] Harstad, Bård (2012): "Climate Contracts: A Game of Emissions, Investments, Negotiations, and Renegotiations," *Review of Economic Studies* 79(4): 1527-57.

[34] Harstad, Bård (2016): "The Dynamics of Climate Agreements," *Journal of the European Economic Association*, 14, 719-52.

[35] Hart Oliver (1995): *Firms Contracts and Financial Structure*, Oxford University Press, Oxford, UK.

[36] Heal, Geoffrey, Kunreuther, H. (2011): "Tipping Climate Negotiations," NBER WP 16954.

[37] Helm, Carsten and Schmidt, Robert C. (2013): "Climate cooperation with technology investments and border carbon adjustment," mimeo.

[38] Hoel, Michael (1992): "International environmental conventions: the case of uniform reductions of emissions," *Environmental and Resource Economics* 2(2): 141-159.

[39] Hoel, Michael and Schneider, Kerstin (1997): "Incentives to Participate in an International Environmental Agreement," *Environmental and Resource Economics*, 9: 153-70.

[40] Hoel, Michael and de Zeeuw, Aart (2010): "Can a Focus on Breakthrough Technologies Improve the Performance of International Environmental Agreements?" *Environmental and Resource Economics* 47 (3): 395-406.

[41] Hong, Fuhai and Karp, Larry (2012): "International Environmental Agreements with Mixed Strategies and Investment," *Journal of Public Economics* 96 (9–10): 685–97.

[42] Karp, Larry and Simon, Leo (2013): "Participation games and international environmental agreements: A nonparametric model," *Journal of Environmental Economics and Management* 65 (2): 326–344.

[43] Kolstad, Charles D. and Toman, Michael (2005): "The Economics of Climate Policy," *Handbook of Environmental Economics* 3: 1562-93.

[44] Levhari, D. and L. J. Mirman (1980): "The Great Fish War: An Example Using Nash–Cournot Solution," *Bell Journal of Economics*, 11, 322–334.

[45] Nordhaus, William D. (2006): "After Kyoto: Alternative Mechanisms to Control Global Warming," *American Economic Review* 96(2): 31–4.

[46] Palfrey, Thomas R. and Rosenthal, Howard (1984): "Participation and the provision of discrete public goods: a strategic analysis," *Journal of Public Economics* 24(2): 171-93.

[47] Peterson M.J. (1997): "International Organizations and the Implementation of Environmental Regime," in *Global Governance: Drawing Insights from the Environmental Experience* (Oran Young, Ed.), MIT Press, Cambridge, USA.

[48] Ploeg, Frederick van der, and de Zeeuw, Aart (1992): "International aspects of pollution control," *Environmental and Resource Economics* 2 (2): 117-39.

[49] Rey, Debraj and Vohra, Rajiv (2001): "Coalitional Power and Public Goods," *Journal of Political Economy* 109 (6): 1355-84.

[50] Rubio, Santiago J., and Casino, Begona (2005): "Self-enforcing international environmental agreements with a stock pollutant," *Spanish Economic Review* 7 (2): 89-109.

[51] Rubio, Santiago J., and Ulph, Alistar (2007): "An infinite-horizon model of dynamic membership of international environmental agreements," *Journal of Environmental Economics and Management* 54 (3): 296-310.

[52] Segal, Ilya and Whinston, Michael D. (2010): "Property Rights," prepared for *Handbook of Organizational Economics* (Gibbons and Roberts, Eds), Elsevier.

[53] Tirole Jean (1999): "Incomplete Contracts: Where do we Stand," *Econometrica*, 67(4):741-781.

[54] UNFCCC (2011): "Report of the Conference of the Parties on its sixteenth session, held in Cancun from 29 November to 10 December 2010, Addendum, Part Two: Action taken by the Conference of the Parties at its sixteenth session," United Nations FCCC/CP/2010/7/Add.1.

[55] UNFCCC (2012): "Report of the Conference of the Parties on its seventeenth session, held in Durban from 28 November to 11 December 2011, Addendum, Part Two: Action taken by the Conference of the Parties at its seventeenth session," United Nations FCCC/CP/2011/9/Add.1.

[56] Vespa, E. (2012), "Cooperation in Dynamic Games: An Experimental Investigation," mimeo, UCSB.

[57] Young Oran (2011): "Effectiveness of international environmental regimes: Existing knowledge, cutting-edge themes, and research strategies," *Proceedings of the National Accademy of Science*, 108(50): 19854-60.

[58] de Zeeuw, Aart (2008): "Dynamic effects on the stability of international environmental agreements," *Journal of Environmental Economics and Management* 55(2): 163–74.

Proofs

Proof of Proposition 4

Let $m^* \equiv |M^*|$ while T^* is the equilibrium agreement length. If m countries participate in a T-period contract, every i's continuation value can be written as (when substituting from Proposition 3):

$$\begin{aligned} v(m,T) &= \sum_{t=1}^{T} \delta^{t-1} \left[\begin{array}{c} -\frac{b}{2}\left(\frac{mC}{b}\right)^2 \\ -C\left(\overline{y}_i - \left(m^2+n-m\right)\left(\frac{C}{b}+\frac{\delta C}{K}\right)\right) \\ -\delta \frac{K}{2}\left(m\frac{C}{K}\right)^2 \end{array} \right] + \delta^T v(m^*,T^*) \\ &= -\frac{1-\delta^T}{1-\delta} C\left[\overline{y}_i - C\left(\frac{m^2}{2}+n-m\right)\left(\frac{1}{b}+\frac{\delta}{K}\right)\right] + \delta^T v(m^*,T^*). \end{aligned}$$

This implies:

$$v(m^*,T^*) = -\frac{1}{1-\delta} C\left[\overline{y}_i - C\left(\frac{m^{*2}}{2}+n-m^*\right)\left(\frac{1}{b}+\frac{\delta}{K}\right)\right], \qquad (19)$$

and therefore:

$$\begin{aligned} v(m,T) = &-\frac{1-\delta^T}{1-\delta} C\left[\overline{y}_i - C\left(\frac{m^2}{2}+n-m\right)\left(\frac{1}{b}+\frac{\delta}{K}\right)\right] \\ &- \frac{\delta^T}{1-\delta} C\left[\overline{y}_i - C\left(\frac{m^{*2}}{2}+n-m^*\right)\left(\frac{1}{b}+\frac{\delta}{K}\right)\right] \end{aligned} \qquad (20)$$

Note that the derivate of $v(m,T)$ w.r.t. T, or equivalently w.r.t. $-\delta^T$, is always negative if and only if:

$$\frac{C^2}{1-\delta}\left(\frac{m^2}{2}+n-m\right)\left(\frac{1}{b}+\frac{\delta}{K}\right) \leq \frac{C^2}{1-\delta}\left(\frac{m^{*2}}{2}+n-m^*\right)\left(\frac{1}{b}+\frac{\delta}{K}\right),$$

requiring $m \leq m^*$. QED

Proof of Proposition 5

First, note that a trivial equilibrium is where no country joins the coalition (if no-one does, it is irrelevant if i does). In the formulae, this would correspond to the situation where $m^* = 1$.

Following Proposition 4, if a participant deviates, then $m = m^* - 1 < m^*$, so $T = 1$ and the participant is expected to join the coalition next period. Such a one-period deviation is not beneficial to i if:

$$v(m^*,T^*) \geq -\frac{b}{2}\left(\frac{C}{b}\right)^2 - \left[\begin{array}{c} C\left(\overline{y}_i - \left(m^2+n-m\right)\frac{C}{b}\right) + \delta\frac{K}{2}\left(\frac{C}{K}\right)^2 \\ -\delta C\left(m^2+n-m\right)\frac{C}{K} - \delta v(m^*,T^*) \end{array} \right].$$

Substituting expression (19) for $v(m^*, T^*)$, this condition can be written as:

$$-C\left[\begin{array}{c}\overline{y}_i\\ -C\left(\frac{m^{*2}}{2}+n-m^*\right)\left(\frac{1}{b}+\frac{\delta}{K}\right)\end{array}\right]\geq-\left[\begin{array}{c}\frac{b}{2}\left(\frac{C}{b}\right)^2+C\left(\overline{y}_i-\left(m^2+n-m\right)\frac{C}{b}\right)\\ +\delta\frac{K}{2}\left(\frac{C}{K}\right)^2-\delta C\left(m^2+n-m\right)\frac{C}{K}\end{array}\right].$$

Simplified, this becomes:

$$\left(\frac{m^{*2}}{2}-m^*\right)\left(\frac{1}{b}+\frac{\delta}{K}\right)\geq\left(m^{*2}-3m^*+\frac{3}{2}\right)\left(\frac{1}{b}+\frac{\delta}{K}\right), \tag{21}$$

requiring $(m^*-1)(m^*-3)\leq 0$. It follows that if $m^*=3$, each participant is indifferent whether to join; if $m^*=2$, each participant strictly prefers to join. If $m^*>3$, no participant would be willing to join. QED

Proof of Proposition 6

(i) Once the quotas $g_{i,t}$ for $i\in M$ and $t\in\{1,..,T\}$ are negotiated in period 1, country i's continuation payoff can be written recursively as follows (where we drop the subscripts for period t):

$$v_i=\sum_{t=1}^{T}\delta^{t-1}\left[-\frac{b}{2}\left(\overline{y}_i-g_{i,t}-R_{i,t}\right)^2-C\left(\sum_{j\in N}g_{j,t}\right)-\delta\frac{K}{2}R_{i,t+1}^2\right]+\delta^T v_i+\delta^T C\sum_{j\in N}R_{j,T+1}. \tag{22}$$

This recursive formulation recognizes that the game starting at time $T+1$ is identical to the game starting in period 1 (as before, the stocks are payoff irrelevant at the start of period $T+1$ as well as period 1, since the stocks do not change the ranking of any vector of future actions).[40]

It follows that the first-order conditions for the $R_{i,t}$s are:

$$R_{i,t}=\frac{b}{K}\left(\overline{y}_i-g_{i,t}-R_{i,t}\right)\text{ for }t\in\{2,...,T\},\quad R_{i,T+1}=\frac{C}{K}.$$

This implies:

$$R_{i,t}=\frac{\overline{y}_i-g_{i,t}}{K/b+1}\Rightarrow\overline{y}_i-g_{i,t}-R_{i,t}=\frac{K}{b}\frac{\overline{y}_i-g_{i,t}}{K/b+1},\ t\in\{2,...,T\}. \tag{23}$$

(ii) Substituting (23) into (22) and defining $a_{i,t}\equiv\overline{y}_i-g_{i,t}$, we see that every i is identical with respect to the $a_{i,t}$s. Negotiating the $g_{i,t}$s is equivalent to negotiating the $a_{i,t}$s, so, in equilibrium, the $a_{i,t}$s will be identical and such as to maximize a participant's continuation value. The first-order condition w.r.t. $a_{i,t}=a_t$, $t\in\{2,...,T\}$ gives:

$$-b\left(\frac{K/b}{K/b+1}\right)^2a_t+mC-K\left(\frac{1}{K/b+1}\right)^2a_t=0\Rightarrow\overline{y}_i-m\frac{C}{K}-m\frac{C}{b}=g_{i,t}.$$

[40] Also, note v_i does not account for the fact that a larger technology stock at the outset reduces emission in the first period (this benefit has already been accounted for): this is why the term $\delta^T C\sum_{j\in N}R_{j,T+1}$ must be added at the end of (22).

For $t = 1$, the countries are, in effect, negotiating the $d_{i,1}$s directly (since $R_{i,1}$ is given), and all countries have symmetric preferences over the $d_{i,1}$s and the preferred $d_{i,1} = d_1$ is $d_1 = mC/b \Rightarrow g_{i,1} = \overline{y}_i - R_{i,1} - mC/b$. QED

Proof of Proposition 7

It is first useful to prove the following lemma.

Lemma A1. *On the equilibrium path of a Markov equilibrium, $T^* = \infty$.*

Proof. Assume not, so that $T^* < \infty$. First note that in a Markov equilibrium the decision to join a coalition is stationary, so the continuation value for a participant can be written recursively as:

$$\begin{aligned} v\left(m^*, T^*\right) = &-\frac{1-\delta^{T^*}}{1-\delta} C\left[\overline{y}_i - C\left(\frac{m^{*2}}{2} + n - m^*\right)\left(\frac{1}{b} + \frac{\delta}{K}\right)\right] \\ &- \delta^T \frac{C^2}{2K}\left(m^* - 1\right)^2 + \delta^T v(m^*, T^*) \end{aligned} \tag{24}$$

where the second term follows from the fact that, in an incomplete contracting environment, each coalition member receives the additional "benefit" that in the last period, it can invest less, although that, in turn, generates more pollution in period $T+1$. Compared to the complete contracting situation, the net additional benefit is:

$$\delta^{T-1}\left(\delta\frac{K}{2}\left(m\frac{C}{K}\right)^2 - \delta\frac{K}{2}\left(\frac{C}{K}\right)^2\right) - \delta^T C\left(m\left(m\frac{C}{K}\right) - m\left(\frac{C}{K}\right)\right) = -\delta^T\frac{C^2}{2K}\left(m-1\right)^2 < 0.$$

Equation (24) implies that:

$$\begin{aligned} v\left(m^*, T^*\right) &= -\frac{1-\delta^{T^*}}{1-\delta} C\left[\overline{y}_i - C\left(\frac{m^{*2}}{2} + n - m^*\right)\left(\frac{1}{b} + \frac{\delta}{K}\right)\right] - \frac{\delta^{T^*}}{1-\delta^{T^*}}\frac{C^2}{2K}\left(m^* - 1\right)^2 \\ &< -\frac{1-\delta^{T^*}}{1-\delta} C\left[\overline{y}_i - C\left(\frac{m^{*2}}{2} + n - m^*\right)\left(\frac{1}{b} + \frac{\delta}{K}\right)\right] = v\left(m^*, \infty\right) \end{aligned}$$

where the last term is the utility that the coalition would achieve if it committed to an infinite agreement. It follows that $T^* < \infty$ cannot be optimal. ■

We can now prove Proposition 7. Given Lemma A.1, the value of a T-period agreement for each member of a coalition of size m is:

$$\begin{aligned} v\left(m, T\right) = &-\frac{1-\delta^T}{1-\delta} C\left[\overline{y}_i - C\left(\frac{m^2}{2} + n - m\right)\left(\frac{1}{b} + \frac{\delta}{K}\right)\right] \\ &- \frac{\delta^T}{1-\delta} C\left[\overline{y}_i - C\left(\frac{m^{*2}}{2} + n - m^*\right)\left(\frac{1}{b} + \frac{\delta}{K}\right)\right] - \delta^T\frac{C^2}{2K}\left(m-1\right)^2. \end{aligned}$$

Note that the derivate of $v(m,T)$ w.r.t. T, or equivalently w.r.t. $-\delta^T$, is always negative if and only if:

$$\left(\frac{m^2}{2}-m\right)+\frac{1-\delta}{2K}\left(\frac{bK}{K+\delta b}\right)(m-1)^2 \leq \left(\frac{m^{*2}}{2}-m^*\right)$$

That is, after some algebraic manipulations, if and only if $m \leq \widehat{m}(x)$, as defined in Proposition 7. QED

Proof of Proposition 8

Suppose $m^* \leq \overline{m}_M$. If a country which joins the coalition in equilibrium deviates, then the coalition size will be $m = m^*-1$ and the coalition will form a one-period contract rather than a long-term contract. The participant is expected to join the coalition next period. Such a one-period deviation is not strictly beneficial to i if:

$$v(m^*,T^*) \geq -\left[\begin{array}{c}\frac{b}{2}\left(\frac{C}{b}\right)^2+C\left(\overline{y}_i-\left(m^2+n-m\right)\frac{C}{b}\right)\\ +\delta\frac{K}{2}\left(\frac{C}{K}\right)^2-\delta Cn\frac{C}{K}\end{array}\right]+\delta v(m^*,T^*),$$

where $m = m^*-1$. Simplifying, we obtain:

$$\left(\frac{m^{*2}}{2}-m^*\right)\left(\frac{1}{b}+\frac{\delta}{K}\right) \geq \left(m^{*2}-3m^*+\frac{3}{2}\right)\frac{1}{b}-\frac{\delta}{2K}.$$

Summing and subtracting $\left(m^{*2}-3m^*+\frac{3}{2}\right)\frac{\delta}{K}$, we obtain:

$$\left(\frac{m^{*2}}{2}-m^*\right)\left(\frac{1}{b}+\frac{\delta}{K}\right) \geq \left(m^{*2}-3m^*+\frac{3}{2}\right)\left(\frac{1}{b}+\frac{\delta}{K}\right)-\left(m^{*2}-3m^*+2\right)\frac{\delta}{K}.$$

After some algebra, this inequality reduces to:

$$2\frac{\delta}{x} \geq (m^*-3)\left(1-\frac{\delta}{x}\right).$$

To prevent a deviation from a nonparticipating country, we also need to satisfy the condition that a nonparticipant does not find it profitable to join the coalition:

$$\begin{aligned}&-\frac{C}{1-\delta}\left[\overline{y}_i-C\left(\frac{(m^*+1)^2}{2}+n-m^*-1\right)\left(\frac{1}{b}+\frac{\delta}{K}\right)\right]\\ &\leq -\frac{C}{1-\delta}\left[\overline{y}_i-C\left(m^{*2}+n-m^*-\frac{1}{2}\right)\left(\frac{1}{b}+\frac{\delta}{K}\right)\right],\end{aligned}$$

which is implied by $m^*(m^*-2) \geq 0$, or $m^* \geq 2$, which is always satisfied. From Proposition 7 we can conclude that an equilibrium of size $m^* \in [2,n]$ exists if $2\frac{\delta}{x} \geq (m^*-3)\left(1-\frac{\delta}{x}\right)$ and $m^* \leq 1+\frac{1}{1-\sqrt{\frac{x+\delta}{x+1}}}$ or, rewriting these two conditions, if $m^* \leq \min\{m_I(x), m_M(x)\}$. It is easy to verify that $m_I(x) \geq m_M(x)$ if and only if $x \leq \widehat{x} = \frac{1}{6}\left((1+\delta)+\sqrt{(1+\delta)^2+12\delta}\right)$, which proves the sufficiency of $m^* \leq m(x)$.

The fact that $T^* = \infty$ follows from Proposition 7. For the remaining results, we proceed in 2 steps.

Step 1. Assume $m^* = 2$. In this case $m^* \leq 1 + \frac{1}{1-\sqrt{\frac{x+\delta}{x+1}}}$ is always satisfied. Condition $2\frac{\delta}{x} \geq (m^*-3)\left(1-\frac{\delta}{x}\right)$ is satisfied if $x \geq \delta$ or, in case $x < \delta$, if $m^* \geq 3 + \frac{2\delta}{x-\delta}$, that is if $x \geq \delta + \frac{2\delta}{m^*-3} = -\delta$, which is always true. If $m^* = 3$, condition $2\frac{\delta}{x} \geq (m^*-3)\left(1-\frac{\delta}{x}\right)$ is always true. Condition $m^* \leq 1 + \frac{1}{1-\sqrt{\frac{x+\delta}{x+1}}}$, is true if $x > \frac{\left(\frac{m^*-2}{m^*-1}\right)^2-\delta}{1-\left(\frac{m^*-2}{m^*-1}\right)^2} = \frac{1/4-\delta}{3/4} = \frac{1}{3}-\frac{4}{3}\delta$. Assume $x < \frac{1}{3}-\frac{4}{3}\delta$. In this case a unilateral deviation is not optimal if $m^* \leq 3$. To see this note that if a country does not join the coalition, the other countries in the coalition will still find it optimal to commit to an agreement that lasts for an infinite number of periods. In this case, staying out of the coalition is not profitable if:

$$\begin{aligned}
v(m^*, T^*) &= -\frac{C}{1-\delta}\left[\overline{y}_i - C\left(\frac{m^{*2}}{2}+n-m^*\right)\left(\frac{1}{b}+\frac{\delta}{K}\right)\right] \\
&\geq -\frac{1}{1-\delta}\left[\begin{array}{c} \frac{b}{2}\left(\frac{C}{b}\right)^2 + C\left(\overline{y}_i - \left((m^*-1)^2+n-m^*+1\right)\frac{C}{b}\right) \\ +\delta\frac{K}{2}\left(\frac{C}{K}\right)^2 - \delta\left((m^*-1)^2+n-m^*+1\right)\frac{C^2}{K} \end{array}\right] \\
&= -\frac{C}{1-\delta}\left[\overline{y}_i - C\left(m^{*2}+n-3m^*+\frac{3}{2}\right)\left(\frac{1}{b}+\frac{\delta}{K}\right)\right]
\end{aligned}$$

Note that this inequality is the same as (21) studied in Proposition 5: it is satisfied if $m^* \leq 3$.

Step 2. We now prove that the conditions of Proposition 8 are necessary. To this end, it will be suffice to show that $m^* > 3$ cannot be an equilibrium if it is not the case that $m^* < m_I(x)$ and $m < m_M(x)$. These two inequalities can be written as (10) and (11). We therefore need to consider only 3 cases:

a. $x > \overline{x}(m^*, \delta)$, $x > \underline{x}(m^*, \delta)$. By the definition of $\overline{x}(m^*, \delta)$, we have that at least one agent has an incentive to free ride by not participating.

b. $x < \underline{x}(m^*, \delta)$. In this case if a country deviates and does not participate, the remaining coalition members commit to a contract that lasts for an infinite number of periods. In this case, the argument presented in Step 1 above shows that it is optimal to deviate if $m^* > 3$.

c. $x > \overline{x}(m^*, \delta)$, $x = \underline{x}(m^*, \delta)$. In this case if there are $m^* - 1$ countries in the coalition, then the coalition members are indifferent between choosing any T'. Assume that if there are $m^* - 1$ participants, then they choose to commit to an agreement for T' periods, where T' can be anything from 1 to infinity.

The deviation of agent i is profitable if:

$$
\begin{aligned}
v\left(m^*, T^*\right) < &-\frac{1-\delta^{T'}}{1-\delta} C\left[\overline{y}_i - C\left(\left(m^*-1\right)^2 + n - \left(m^*-1\right) - \frac{1}{2}\right)\left(\frac{1}{b} + \frac{\delta}{K}\right)\right] \\
&- \delta^{T'} \frac{C^2}{K}(m-2)(m-1) + \delta^{T'} v\left(m^*, T^*\right) = v'\left(m^*-1, T'\right).
\end{aligned}
$$

Note that:

$$
\begin{aligned}
v'\left(m^*-1, T'\right) = &-\frac{1}{1-\delta} C\left[\begin{array}{c} -\frac{C}{2b} + \overline{y}_i - C\left(\begin{array}{c} \left(m^*-1\right)^2 + n \\ -\left(m^*-1\right) \end{array}\right)\frac{1}{b} \\ +\frac{C\delta}{2K} - n\frac{C}{K} \end{array}\right] \\
&+ \frac{\delta}{1-\delta}\frac{C^2}{K}\left(\left(m^*-1\right)^2 - \left(m^*-1\right)\right)\frac{C^2\delta}{K} \\
&- \frac{\delta^{T'}}{1-\delta^{T'}}\frac{C^2}{K}(m-2)(m-1) \\
= &-\frac{1}{1-\delta} C\left[\begin{array}{c} -\frac{C}{2b} + \overline{y}_i \\ -C\left(\left(m^*-1\right)^2 + n - \left(m^*-1\right)\right)\frac{1}{b} + \frac{C\delta}{2K} - n\frac{C}{K} \end{array}\right] \\
&+ \left(\frac{\delta}{1-\delta} - \frac{\delta^{T'}}{1-\delta^{T'}}\right)\frac{C^2}{K}(m-2)(m-1).
\end{aligned} \tag{25}
$$

The right-hand side of (25) is increasing in T', so the condition is satisfied if it is satisfied for $T' = 1$. By the definition of $\overline{x}(m^*, \delta)$, we have that (25) is satisfied for $T' = 1$ if $x > \overline{x}(m^*, \delta)$. So when $x > \overline{x}(m^*, \delta)$ and $x = \underline{x}(m^*, \delta)$, agent i has a profitable deviation. QED

Appendix

Proof of Proposition 11

Suppose a coalition of size m agrees on a long-term agreement (of duration $T > 1$). Then, the equilibrium contract and policy ensures that for each participant $\widehat{u}_{i,t}$, as defined by Lemma 1, is the following for every $t < T$:

$$\widehat{u}^m = -\frac{b}{2}\left(m\frac{C}{b}\right)^2 - C\left[\overline{y}_i - \left(m^2+n-m\right)\frac{C}{b}\right] - \delta\frac{K}{2}\left(\frac{mC}{K}\right)^2 + \delta C\left(m^2+n-m\right)\frac{C}{K}.$$

Intuitively, this payoff does not depend on the degree of contractual completeness, α, since investments are the same whether they are contractible or not (before T). At $t = T$, however, each member's $\widehat{u}_{i,t}$ is exactly as it would have been following a one-period agreement:

$$\begin{aligned}\widehat{u}^{m,1} &= -\frac{b}{2}\left(m\frac{C}{b}\right)^2 - \delta\alpha\frac{K}{2}\left(\frac{mC}{K}\right)^2 - \delta\frac{K}{2}(1-\alpha)\left(\frac{C}{K}\right)^2 \\ &\quad - C\left[\overline{y}_i - \left(m^2+n-m\right)\frac{C}{b}\right] + \delta C\alpha\left(m^2+n-m\right)\frac{C}{K} + \delta C(1-\alpha)n\frac{C}{K}.\end{aligned}$$

Note that $\widehat{u}^{m,1} = \widehat{u}^m$ if $\alpha = 1$. If $\alpha < 1$, it is easy to see that $\widehat{u}^{m,1} < \widehat{u}^m$. This implies that the equilibrium coalition prefers $T = \infty$.

Next, consider a country that has deviated by not participating. If the remaining coalition, of size $m-1$, signs a one-period agreement, then the free-rider's $\widehat{u}_{i,t}$ becomes:

$$\begin{aligned}\widehat{u}^{n,1} &= -\frac{b}{2}\left(\frac{C}{b}\right)^2 - \delta\frac{K}{2}\left(\frac{C}{K}\right)^2 - C\left(\overline{y}_i - \left[(m-1)^2+n-(m-1)\right]\frac{C}{b}\right) \\ &\quad + \delta C\alpha\left[(m-1)^2+n-(m-1)\right]\frac{C}{K} + \delta C(1-\alpha)n\frac{C}{K}.\end{aligned}$$

We will first derive the individual participation constraint and thus $m_I(x;\alpha)$. If it is anticipated that a coalition of size $m-1$ will prefer a one-period agreement, then a supposed-to-be member will not strictly prefer a one-period deviation if $\widehat{u}^m - \widehat{u}^{n,1} \geq 0$. Using the expressions above we get:

$$\begin{aligned}\widehat{u}^m - \widehat{u}^{n,1} &= -\frac{C^2}{2b}\left(m^2-1\right) + \frac{C^2}{b}(2m-1) - \frac{C^2}{b} \\ &\quad - \delta\alpha\frac{C}{2K}^2\left(m^2-1\right) + \delta\alpha\frac{C}{K}^2(2m-1) - \delta\alpha\frac{C}{K}^2 \\ &\quad - \delta(1-\alpha)\frac{C}{2K}^2\left(m^2-1\right) + \delta(1-\alpha)\frac{C}{K}^2\left(m^2-m\right)\end{aligned}$$

$$= -\frac{C^2}{2b}\left(m^2-4m+3\right) - \delta\alpha\frac{C}{2K}^2\left(m^2-4m+3\right) + \delta(1-\alpha)\frac{C}{2K}^2(m-1)^2$$

$$= \delta(1-\alpha)\frac{C}{2K}^2(m-1)^2 - \frac{C^2}{2K}(x+\delta\alpha)(m-1)(m-3),$$

which is positive if and only if

$$\delta(1-\alpha)(m-1) > (x+\delta\alpha)(m-3) \Leftrightarrow$$
$$m \leq m_I(x;\alpha) \equiv \frac{3x+4\delta\alpha-\delta}{x+2\delta\alpha-\delta} = 3+\frac{2\delta(1-\alpha)}{x+2\delta\alpha-\delta} = 2+\frac{x+\delta}{x+2\delta\alpha-\delta}.$$

Note that $m_I(x;\alpha)$ decreases in α and, if $\alpha < 1$, $m_I(x;\alpha)$ decreases in x. Note also that $m_I(x;0) = 3+\frac{2\delta}{x-\delta}$, as in Section 4, and $m_I(x;1) = 3$ for all $x \geq 0$.

We next derive the discipline constraint and $m_M(x;\alpha)$. Suppose the current coalition size is $m-1$, but there is one deviator who is expected to return to equilibrium behavior in the next period (giving the a coalition of size m). As before, it is easy to see that payoffs are linear in δ^{T+1} so only $T \in \{1,\infty\}$ can be optimal. Further, $T=1$ is indeed better for the coalition than $T=\infty$ if $\widehat{u}^{m-1,1}+\widehat{u}^m\delta/(1-\delta) \geq \widehat{u}^{m-1}/(1-\delta)$, implying $(\widehat{u}^m-\widehat{u}^{m-1,1}) \geq (\widehat{u}^{m-1}-\widehat{u}^{m-1,1})/\delta$. Note that the left-hand side is:

$$\begin{aligned}
&\widehat{u}^m-\widehat{u}^{m-1,1} = -\frac{C}{2b}^2(2m-1)+\frac{C^2}{b}(2m-1)-\frac{C^2}{b}+\delta\alpha\frac{C}{2K}^2(2m-1)\\
&-\delta\alpha\frac{C}{K}^2-\delta(1-\alpha)\frac{C}{2K}^2\left(m^2-1\right)+\delta(1-\alpha)\frac{C}{K}^2\left(m^2-m\right)\\
&=\frac{C^2}{2b}(2m-3)+\delta\alpha\frac{C^2}{2K}(2m-3)+\delta(1-\alpha)\frac{C^2}{2K}(m-1)^2\\
&=\frac{C^2}{2}\left(\frac{1}{b}+\frac{\delta\alpha}{K}\right)(2m-3)+\delta(1-\alpha)\frac{C^2}{2K}(m-1)^2.
\end{aligned}$$

Next, use the expressions above to derive $\widehat{u}^{m-1}-\widehat{u}^{m-1,1}$:

$$\begin{aligned}
&-\delta(1-\alpha)\frac{C}{2K}^2\left((m-1)^2-1\right)+\delta(1-\alpha)\frac{C}{K}^2\left((m-1)^2-(m-1)\right)\\
&=\delta(1-\alpha)\frac{C}{2K}^2\left[2\left(m^2-3m+2\right)-\left(m^2-2m\right)\right]=\delta(1-\alpha)\frac{C}{2K}^2(m-2)^2.
\end{aligned}$$

Combined, we get $(\widehat{u}^m-\widehat{u}^{m-1,1}) \geq (\widehat{u}^{m-1}-\widehat{u}^{m-1,1})/\delta$ if and only if:

$$\frac{C^2}{2}\left(\frac{1}{b}+\frac{\delta\alpha}{K}\right)(2m-3)+\delta(1-\alpha)\frac{C^2}{2K}(m-1)^2 \geq (1-\alpha)\frac{C}{2K}^2(m-2)^2 \Leftrightarrow$$
$$(x+\delta\alpha)(2m-3)+\delta(1-\alpha)(m-1)^2 \geq (1-\alpha)(m-2)^2,$$

which always holds when $\alpha = 1$ and $m \geq 2$. If $\alpha < 1$, the inequality becomes:

$$\left(\frac{x+\delta\alpha}{1-\alpha}\right)(2m-3) + \delta(m-1)^2 \geq (m-2)^2 \Leftrightarrow$$

$$m^2(1-\delta) - 2m\left[(1-\delta) + \left(1 + \frac{x+\delta\alpha}{1-\alpha}\right)\right] + (1-\delta) + 3\left(1 + \frac{x+\delta\alpha}{1-\alpha}\right) \leq 0 \Leftrightarrow$$

$$m^2 - 2m[1+\mu] + 1 + 3\mu \leq 0; \tag{26}$$

$$\mu \equiv \frac{1+x-\alpha(1-\delta)}{(1-\alpha)(1-\delta)} > 1.$$

It is easy to check that the lowest m satisfying (26) is smaller than 2. Thus, for every $m \geq 2$, the discipline constraint (26) requires that $m \leq m_M(x;\alpha)$, where $m_M(x;\alpha)$ is the largest m satisfying (26):

$$m_M(x;\alpha) = 1 + \mu + \sqrt{(1+\mu)^2 - 1 - 3\mu} = 1 + \mu + \sqrt{\mu^2 - \mu}$$

$$= 1 + \frac{1+x-\alpha(1-\delta)}{(1-\alpha)(1-\delta)} + \sqrt{\left(\frac{1+x-\alpha(1-\delta)}{(1-\alpha)(1-\delta)}\right)^2 - \left(\frac{1+x-\alpha(1-\delta)}{(1-\alpha)(1-\delta)}\right)}$$

The upper treshold $m_M(x;\alpha)$ is increasing in μ and, since μ increases in both x and α, $m_M(x;\alpha)$ is increasing in both α and x. QED

Proof of Proposition 12

Since efficiency increases in $m(x;\alpha) = \min\{m_I(x;\alpha), m_M(x;\alpha)\}$, it follows that efficiency (and $m(x;\alpha)$) increases in α when $m_I(x;\alpha) < m_M(x;\alpha)$ but decreases in α when $m_I(x;\alpha) > m_M(x;\alpha)$.

(i) When $\alpha \downarrow 0$, Proposition 8 shows that $m_I(x;\alpha) > m_M(x;\alpha)$ if and only if $x < \widehat{x}$. Therefore, if $x \geq \widehat{x}$, $\min\{m_I(x;\alpha), m_M(x;\alpha)\} = m_I(x;\alpha)$ for all $\alpha \in [0,1]$ and utilitarian welfare (and $m(x;\alpha) = m_I(x;\alpha)$) is maximized when $\alpha = 0$.

(ii) On the other hand, $\lim_{\alpha\uparrow 1} m_M(x;\alpha) = \infty$ for every $x \geq 0$. We have already noted that $m_I(x;1) = 3$, so when $\alpha \uparrow 1$, $\min\{m_I(x;\alpha), m_M(x;\alpha)\} = m_I(x;\alpha)$, which is increasing in α.

Suppose $x < \widehat{x}$. Efficiency increases when α increases from zero since $m_M(x;\alpha)$ increases. For some $\alpha^* \in (0,1)$, $m_M(x;\alpha^*) = m_I(x;\alpha^*)$ and a further increase in α will reduce $m_I(x;\alpha)$ and thus m^*. Efficiency is thus maximized when $\alpha = \alpha^*$. Since $m_M(\cdot)$ increases in both arguments while $m_I(\cdot)$ decreases in both arguments, it follows that α^* decreases in x. By solving $m_I(x;\alpha^*) = m_M(x;\alpha^*)$ we can also show that this requires:

$$\alpha^* = \sqrt{\left[\frac{7\delta x - \delta^2 + x + \delta}{8\delta^2}\right]^2 + \frac{x + \delta + \delta x - 3x^2}{4\delta^2}} - \frac{7\delta x - \delta^2 + x + \delta}{8\delta^2}.$$

This completes the proof. QED

Proof of the results presented in Section 6.D

In this section we first prove that for any $m \leq n$ there is a $\delta_m < 1$ such that, even in the complete contracting environment, a simple equilibrium in which an agreement of size m is force in all periods exists for $\delta > \delta_m$ exists. We then prove that no simple SPE with $m > 4$ is weakly renegotiation proof.

Existence of simple SPE with $m > 3$**.** Let $h^t = \{(M_1, T_1), ...(M_k, T_k)\}$ be the history of coalitions formed up to period t, where k is the number of coalition formed up to period t, M_i is the ith coalition formed starting from period 0, and T_i is its length. We prove by induction that for any $m \leq n$ there is a $\delta_m < 1$ such that a simple equilibrium in which m countries join an agreement in all periods is an equilibrium for $\delta > \delta_m$. The proof is by induction. From Proposition 5 we know that for any $\delta < 1$ and any coalition M' with $|M'| = 3$ there is a Markov equilibrium in which M' is formed in all periods. A Markov equilibrium is a simple equilibrium, so for any $\delta < 1$ and M' with $|M'| = 3$ there is a simple equilibrium in which M' is formed in all periods.

Assume it is true that there is a $\delta_{m-1} < 1$ such that for any $\delta > \delta_{m-1}$ and any M' with $|M'| = m - 1$ there is a simple equilibrium in which M' is formed in all periods. We now show that there is a $\delta_m < 1$ such that for any $\delta > \delta_m$ and M with $|M| = m$ there is a simple equilibrium in which M is formed in all periods.

Consider the following strategies:

Phase 1. A coalition M with $|M| = m$ is formed for $T = \infty$ periods. If a country deviates and does not join the agreement, then a coalition $M \backslash i$ with the remaining $m - 1$ players in M is formed for $T = 1$ periods and we move to Phase 2.

Phase 2. If country $i \in M$ has stayed out of the agreement in Phase 1, then after the agreement of the remaining $M \backslash i$ is dissolved the countries play a continuation equilibrium in which a set M_i with $|M_i| = m - 1$ and $i \in M_i$ form a new agreement and choose a length $T = \infty$. If a set M^* of countries with $|M^*| > 1$ stays out in Phase 1, then the remaining countries $M \backslash M^*$ form an agreement with $T = 1$; at the end of this agreement, we return to Phase 1.

From the induction hypothesis we have that the subgame when a new coalition is formed after a unilateral deviation at t is an equilibrium for $\delta > \delta_{m-1}$. It is moreover easy to see that it is optimal for the $M \backslash i$ coalition to choose $T = 1$, since the expected continuation utility with the new coalition M_i is not lower than the utility obtained keeping the coalition for $T > 1$ for all $j \in M \backslash i$. To see that no deviation is strictly optimal at $t = 0$, consider a unilateral deviation by i: it yields a net payoff of $(1-\delta)\left(u_i(M \backslash i) - u_i(M)\right) + \delta\left(u_i(M_i) - u_i(M)\right)/(1-\delta)$. Since $u_i(M_i) - u_i(M) < 0$, it is clear that there must be a $\delta_m \geq \delta_{m-1}$ with $\delta_m < 1$ such that this deviation is not profitable for $\delta \geq \delta_m$. Since multilateral deviations are irrelevant when considering a Nash equilibrium, it follows that the strategies described above are an equilibrium for $\delta \geq \delta_m$. QED

No simple SPE with $m > 4$ is weakly renegotiation proof. We now prove that there is no simple equilibrium that supports an agreement M in which $|M| = m > 4$ countries participate in all periods and that it is weakly renegotiation proof. Assume by contradiction that such an equilibrium exists. In this equilibrium, after any history h^t we must have a coalition $M(h^t)$ of size $m(h^t) = |M(h^t)|$ that, if there is no deviation in the subgame starting at h^t, remains in force for all following periods. Let M_* be the smallest coaltion to which coaltions can converge after some history: i.e. M_* is such that $|M_*| \leq |M(h^t)|\ \forall h^t$. Since the number of players and their possible permutations is finite, M_* is well defined. Without loss of generality, let i be such that $i \in M_*$ and h_*^t be a history at which $M(h_*^t) = M_*$. One of two conditions must be satisfied at h_*^t. If after a deviation in which i stays out of the agreement, the coalition $M(h_*^t)\backslash i$ chooses to remain in force for all remaining periods, then i finds the deviation unprofitable only if $u_i(M(h_*^t)\backslash i) \leq u_i(M(h_*^t))$: this condition is possible only if $|M(h_*^t)| \leq 3$. If after the deviation the coalition $M(h_*^t)\backslash i$ chooses to remain in force for $T_i(h_*^t) < \infty$ periods, then i finds the deviation unprofitable only if:

$$\left(1-\delta^{T_i(h_*^t)}\right)\left(u_i(M(h_*^t)\backslash i) - u_i(M(h_*^t))\right) + \delta^{T_i(h_*^t)}\left(u_i(M(D_i(h_*^t)) - u_i(M(h_*^t))\right)$$

is not positive, where $u_i(M')$ is i's utility when a generic set M' is the equilibrium coalition, $D_i(h_*^t)$ is the history following h_*^t in which i does not join the coalition $M(h_*^t)$ and $M(D_i(h_*^t)$ is the coalition formed after i's deviation. Assume, by contradiction, that $|M(h_*^t)| > 3$. In this case $u_i(M(h_*^t)\backslash i) - u_i(M(h_*^t)) > 0$, so $u_i(M(D_i(h_*^t)) < u_i(M(h_*^t))$. Since by construction $|M(h_*^t)| \leq |M(D_i(h_*^t))|$ and $i \in M(h_*^t)$, however, it must be that $u_i(M(h_*^t)) \leq u_i(M(D_i(h_*^t)))$, which is a contradiction. It must therefore be that $|M(h_*^t)| \leq 3$.

In equilibrium, the minimal utility obtained by a country when coalition M is in force must be smaller than the maximal utility obtained at history h_*^t, otherwise the continuation equilibrium at h_*^t would be Pareto dominated by the equilibrium utility and so it would violate the condition for weak renegotiation proofness. We must have $u_j(M(h_*^t)) > u_k(M)$ for $j \notin M(h_*^t)$ and $k \in M$, else we would have $u_j(M(h_*^t)) \leq u_j(M)\ \forall j$, strict for some j, contradicting weak renegotiation proofness. This condition implies that:

$$\begin{aligned} & -Cn\overline{y}_i + C^2\left(\frac{1}{b}+\frac{\delta}{K}\right)\left[\begin{array}{c} \frac{\left(m(h_*^t)\right)^2}{2} + n - m(h_*^t) \\ + \frac{\left(m(h_*^t)\right)^2 - 1}{2} \end{array}\right] \\ & \geq -Cn\overline{y}_i + C^2\left(\frac{1}{b}+\frac{\delta}{K}\right)\left[\left(\frac{(m)^2}{2} + n - m\right)\right] \end{aligned} \tag{27}$$

where $m(h_*^t) = |M(h_*^t)|$. Condition (27) can be rewritten as $m(h_*^t)^2 - 2m(h_*^t) + \frac{1}{2} \geq \frac{m^2}{2} - m$. The left hand side of (27) is the utility of a country outside the coalition when the size of the coalition is $m(h_*^t)$; the right hand side of (27) is the per period utility of a country in a coalition if the country is in the coalition and the size of the coalition is m. Since the left hand side is increasing in $m(h_*^t)$ and $m(h_*^t) \leq 3$, condition (27) implies $(m)^2 - 2m - 11 \leq 0$. This inequality is satisfied only for $m \leq 2\sqrt{3} + 1 = 4.4641$. The maximal integer consistent with this condition is 4, as claimed in Section 6.D. QED