Environmental and Resource Economics (2022) 82:601–624
https://doi.org/10.1007/s10640-022-00668-z

International Environmental Agreements and Black Technology

Gilbert Kollenbach[1]

Accepted: 28 February 2022 / Published online: 31 March 2022

Abstract
This paper analyzes the stability of international environmental agreements in a dynamic game when the generation of both renewables and fossil fuel based energy requires specialized capital stocks or technologies, respectively. Two contract types are considered. At an incomplete (a complete) contract, the coalition coordinates only (both) CO_2 emissions (and renewable energy investments) of its members. In contrast to the results of Battaglini and Harstad (J Polit Econ 124:160–204, 2016) who endorse incomplete contracts to increase the coalition size, only small coalitions are stable regardless of whether the contract is complete or incomplete. This result also holds if black technology is temporary not completely used or transfers are considered.

Keywords International environmental agreements · Black capacity · Complete contract · Incomplete contract

JEL Classification H87 · Q54 · Q55

1 Introduction

Global warming, which is mainly driven by CO_2 emissions, belongs to the most serious problems of our time.[1] The public good property of the atmosphere implies that countries burn too much fossil fuels and, therefore, emit too much CO_2 than socially optimal by non-cooperative behavior. International environmental agreements (IEA), such as the Kyoto-Protocol or the Paris Agreement, are a widely discussed option to coordinate the countries' climate policies and, therefore, reduce carbon emissions. However, there is no

[1] Cf. IPCC (2014).

Helpful comments from two anonymous referees are gratefully acknowledged. Remaining errors are the author's sole responsibility.

✉ Gilbert Kollenbach
gilbert.kollenbach@fernuni-hagen.de

[1] Department of Economics, University of Hagen, Universitätsstr. 41, 58097 Hagen, Germany

supranational authority that can force sovereign countries to join an IEA, so that IEAs must be self-enforcing. Another important aspect of fossil fuel use is its high capital intensity.[2] The extraction and utilization of fossil fuels requires a specialized capital stock, such as mines, oil platforms, or fossil fuel fired power plants, which constitutes a limited capacity to use fossil fuels. Limited extraction capacities are discussed, among others, by Campbell (1980), Cairns and Lasserre (1991), Holland (2003), and Kollenbach (2017).

This paper aims at analyzing the role of a limited fossil fuel utilization (*black*) capacity for the formation of IEAs. Both climate change and the accumulation of a black capacity are dynamic problems. Therefore, this paper rests on the seminal work of Battaglini and Harstad (2016), who consider a dynamic game where countries choose whether to join a climate coalition, the contract length of the IEA, emissions, and investments in green technologies, such as solar panels, wind turbines, hydroelectric facilities or nuclear power plants.[3] Thus, Battaglini and Harstad (2016) consider a green capacity necessary to produce renewable energy, but they neglect a black capacity by assuming that a practically unlimited amount of fossil fuels can be used. By introducing a black capacity, fossil fuel use becomes limited.

If the black capacity is completely used, that is if the black capacity constraint binds, every country uses less fossil fuel and more renewables than with a non-binding constraint. However, the promising result of Battaglini and Harstad (2016), that large climate coalition can be stable if countries only negotiate on emissions but determine green technology investments non-cooperatively (incomplete contract), disappears. Without a black capacity, every signatory realizes that its green technology investments will not affect the emissions of the other countries in the next period, but only weaken its bargaining position in the negotiations of a new climate agreement. Due to this hold-up problem, the signatories reduce their investments, which increases their emissions and, therefore, the climate damage of the following period. Thus, the signatories can credibly threaten every deflecting country to sign a short term contract and, therefore, antedating the higher climate damages. With a black capacity, this credible threat no longer exists. Because the black capacity determines emissions, the signatories know that their green capacity investments of the last contract period will payoff. Consequently, they do not reduce their investments eliminating the hold-up problem, so that only small climate coalitions are stable.

This result also holds if the black capacity constraint does not bind initially but will bind in the future. In this case, the hold-up problem appears but it is only possible to postpone it as long as the capacity constraint is non-binding. Consequently, only a small climate coalition is stable in the period directly before the constraint becomes binding. Because the countries anticipate that a large coalition is not stable in this period, the hold-problem cannot be postponed beyond it. Therefore, also only a small coalition is stable in the prior

[2] Cf. Cairns (1998, p. 234). According to the projections of Birol et al. (2019, p. 266), the capacity of gas power stations will increase from less than 1000GW in 2000 to more than 2000GW in 2040. With respect to coal, the capacity may stay constant at around 2000GW. In this case, 690GW capacity will be added mainly in developing economies, while 600GW will be retired. For renewables the numbers are as follows: photovoltaic increases from approx. 500GW in 2019 to more than 3000GW, wind increases from approx. 500GW to almost 2000GW or more, hydro increases from approx. 1250GW to approx. 2000GW. Thus, while Birol et al. (2019) predict that renewables make up the majority of capacity additions, there are still considerably investments into black technologies.

[3] Whether nuclear power is a renewable energy source is internationally disputed. On the one hand, nuclear power is carbon free. On the other hand, it produces atomic waste and requires uranium, which has a static range of 120 years. Cf. Toth (2014).

period, and so forth. In other words, the standard backward induction argument implies that a large coalition is never stable.

IEAs are studied in economic literature since the early 1990s. Pioneering works are Hoel (1992), Carraro and Siniscalco (1993) and Barrett (1994). Using the stability concept of d'Aspremont et al. (1983), this literature analyses the stable size of an environmental agreement or climate coalition, respectively. That is, it is analyzed at which size no country has an incentive to join or to leave the coalition. The core result is that stable IEAs are either small and deep or large and shallow.[4] There are several reasons that can explain the disparity to the key result of the literature. Using a static model, Breton and Sbragia (2019) show that larger coalitions can be stable if signatories not only coordinate their carbon emission abatement but also their adaption measures. The stabilizing effect of research investments is discussed by Barrett (2006), Hoel and de Zeeuw (2010), Rubio (2017) and Masoudi and Zaccour (2018). Another channel that can stabilize an IEA is foresightedness. According to Diamantoudi and Sartzetakis (2018), a large coalition can be stable if potential deviants take the policy changes of the remaining signatories into account. Carbon border adjustments may also stabilize an IEA by reducing free-riding incentives as shown by Al Khourdajie and Finus (2020). Due to the countries' asymmetry, the cooperation benefits vary among countries. Therefore, transfer mechanisms, which reallocate the benefits either among signatories, or between fringe countries and signatories may stabilize large coalitions.[5] Carraro et al. (2006) discuss both options and also differentiate between several transfer schemes. Kornek and Edenhofer (2020) consider international transfer funds and analyze which fund design performs best with respect to a reduction of free-riding. The role of the countries' asymmetry is highlighted by Fuentes-Albero and Rubio (2010). Although I consider asymmetric countries, I show that transfers among signatories are not able to stabilize large coalitions.

Because climate change is a dynamic problem, a growing number of studies make use of multi-period models.[6] The dismal key result that IEAs are either small or ineffective is confirmed by Barrett (1999). In contrast, Asheim et al. (2006) and Asheim and Holtsmark (2009) show that larger coalitions can be stable if the signatories have the option to punish deviators. A common feature of these studies is that they focus on a repeated game but assume carbon emissions to be a flow pollutant. A carbon emission stock is considered by Breton et al. (2010), Mason et al. (2017), Rubio and Casino (2005), and Rubio and Ulph (2007). While the results of the latter two are in line with Barrett (1999), Breton et al. (2010) and Mason et al. (2017) show that a punishment option stabilizes larger IEAs. Kováč and Schmidt (2021) show that large coalitions are stable if countries have the opportunity do delay negotiations. Closest to my study is the dynamic game considered by Battaglini and Harstad (2016). They differentiate between a complete contract and an incomplete contract. In the former case, countries coordinate both emissions and green capacity investments, and the stable climate coalition is small. In contrast, a large climate coalition can be stable if the signatories only coordinate emissions but leave the decision with respect to green investments to the national authorities. In this case, a hold-up problem emerges that stabilizes coalitions.

The remainder of the paper is organized as follows: The model and preliminary results are presented in Sect. 2. Sect. 3 analyzes the stability of climate coalition under both a complete and an incomplete contract given that the black capacity constraint binds. The

[4] See also Dixit and Olson (2000), Finus (2003), and Barrett (2005).

[5] Finus (2003) gives an overview of the related literature.

[6] A review of this literature is given by Calvo et al. (2012).

case of an initially non-binding capacity constraint is discussed in Sect. 4. Section 5 analyzes whether transfers can stabilize large coalitions. 6 concludes.

2 The Model

I consider an economy in discrete time that relies on building blocks of the model of Battaglini and Harstad (2016). That is, I follow their assumptions if not stated otherwise. The $n \in N$ countries of the model are divided into the groups M and F. The first group represents the members $\{1, ..., m\} \in M$ of an international environmental agreement. This environmental *coalition* coordinates (parts of) its policy to limit climate change. The remaining countries $\{m+1, ..., n\} \in F$ are non-signatories, also referred to as fringe. In every period τ these fringe countries set their policy non-cooperatively.

2.1 Utility, Pollution, Capacity, and Technology

The benefit of country i in period $\tau \in \{1, ..., \infty\}$ is given by

$$U_i(R_{i,\tau} + B_{i,\tau}) = -\frac{a}{2}\left(\bar{y}_i - R_{i,\tau} - B_{i,\tau}\right)^2. \tag{1}$$

Thus, the benefit depends on energy consumption, which is generated by means of renewable sources $R_{i,\tau}$ and fossil fuel $B_{i,\tau}$. The exogenously given satiation point of energy consumption is denoted by $\bar{y}_i$ and $a > 0$ measures the disutility of deviating from the satiation point.[7] Note that the benefit function concavely increases in both $R_{i,\tau}$ and $B_{i,\tau}$. For simplicity, the two energy types are perfect substitutes.

The generation of renewables requires specialized capital goods. Examples for this *green capacity* or green technology, respectively, are wind turbines, solar panels and hydropower stations. By appropriate unit choice, I assume that one capacity unit generates one energy unit, so that $R_{i,\tau}$ denotes both the green capacity necessary to generate renewable energy and the amount of renewable energy. To build up green capacity, the countries invest in the respective capital goods. Green capacity of country i at time $\tau + 1$ is given by

$$R_{i,\tau+1} = q_R R_{i,\tau} + r_{i,\tau}, \tag{2}$$

where $r_{i,\tau}$ denotes the capacity investments and $(1 - q_R) \in [0, 1]$ is the depreciation rate of green capacity. Following Battaglini and Harstad (2016), I assume a quadratic relationship between investments costs $\kappa(\cdot)$ and the new technology level $R_{i,\tau+1}$, which yields $\frac{\partial \kappa}{\partial R_{i,\tau+1}} = kR_{i,\tau+1}$, where k is a positive parameter. By requiring that investment costs are nil without investments, I get[8]

[7] To ensure the comparability with the model of Battaglini and Harstad (2016), I adopt their functional forms. In particular, I use linear-quadratic functions throughout the model. The benefit function (1) can be written as $U_i(R_{i,\tau} + B_{i,\tau}) = -\frac{a\bar{y}_i^2}{2} + a\bar{y}_i(R_{i,\tau} + B_{i,\tau}) - \frac{a}{2}(R_{i,\tau} + B_{i,\tau})^2$. Because $-\frac{a\bar{y}_i^2}{2}$ is exogenous, the term can be omitted, which gives a standard linear-quadratic function. By using (1), the parameter $\bar{y}_i$ can be interpreted as a satiation point.

[8] Solving $\frac{\partial \kappa}{\partial R_{i,\tau+1}} = kR_{i,\tau+1}$ gives $\kappa(\cdot) = \frac{k}{2}R_{i,\tau+1} + Q$, with Q as the constant of integration. Because of $\kappa(\cdot) = 0$ for $r_{i,\tau} = 0$ and (2), I get $Q = -\frac{k}{2}q_R^2 R_{i,\tau}^2$.

$$\kappa(R_{i,\tau+1}, R_{i,\tau}) = \frac{k}{2}\left(R^2_{i,\tau+1} - q_R^2 R^2_{i,\tau}\right). \tag{3}$$

While Battaglini and Harstad (2016) consider a limited green capacity, they assume that fossil fuel extraction is costless and that the fuels are directly usable. However, extracting and using fossil fuels require specialized capital goods, such as mines, pump stations and fossil fuel fired power plants. Therefore, I introduce a corresponding *black capacity*. Again, appropriate unit choice allows that $B_{i,\tau}$ represents both the black capacity necessary to extract and burn fossil fuels and the amount of fossil fuel based (black) energy.[9] With respect to capacity investments and corresponding costs, I adopt the modeling of the green capacity. That is, black capacity of country i at time $\tau + 1$ reads

$$B_{i,\tau+1} = q_B B_{i,\tau} + b_{i,\tau}, \tag{4}$$

where $b_{i,\tau}$ denotes the capacity investments and $(1 - q_B) \in [0, 1]$ is the depreciation rate of black capacity. The investments costs are given by

$$\lambda(B_{i,\tau+1}, B_{i,\tau}) = \frac{l}{2}\left(B^2_{i,\tau+1} - q_B^2 B^2_{i,\tau}\right), \tag{5}$$

with l as a positive investment cost parameter.[10] While renewable energy is produced by making use of a freely available flow of resources such as solar radiation, fossil fuels need to be extracted. To take account of this fact, I consider extraction costs. For country i, these are given by $hB_{i,\tau}$, with positive marginal costs h.

These assumptions imply that both capacities are binding for all points of time, i.e. that the capacities are completely used. With respect to black capacity, I relax this assumption for early periods in sect. 4, where I assume that the black capacity $B_{i,\tau}$ is not completely used initially. In case of the green capacity, I follow Battaglini and Harstad (2016) and assume that the capacity is always completely used. Because green energy is clean and not associated with further utilization costs, a country has no incentives to leave some green capacity unused as long as the capacity does not exceed the satiation point $\bar{y}_i$. By assuming $R_{i,1} + B_{i,1} < \bar{y}_i$ for all $i \in N$, this case is ruled out, because no country will invest into an excess capacity. In particular, $R_{i,1} + B_{i,1} < \bar{y}_i$ ensures that energy generation never exceeds the satiation point.

In contrasts to green energy, fossil fuel extraction is costly and burning fossil fuels contributes to global warming. The latter is caused by the accumulation of CO_2 in the atmosphere. The corresponding stock is denoted by G_τ and evolves in time according to

$$G_\tau = q_G G_{\tau-1} + \sum_{j\in N} B_{j,\tau}, \tag{6}$$

where $(1 - q_G) \in [0, 1]$ is the natural regeneration rate of the atmosphere. Using appropriate units, I assume that every black energy unit causes one carbon emission unit, so that $B_{i,\tau}$ denotes also the carbon emission of country $i \in N$ at time τ. The climate damages caused by the accumulated CO_2 stock are given by $D(G_\tau)$. Golosov et al. (2014, p.65,

[9] Following Battaglini and Harstad (2016), I consider a practically infinite fossil fuel stock. According to Andruleit et al. (2012), the static range of coal reserves and resources exceeds 5000 years.

[10] (5) can be derived in a similar way as (3) by using the assumptions of a quadratic relationship between the new capacity level $B_{i,\tau+1}$ and the investments costs $\lambda(\cdot)$, and of nil investments costs in the absence of investments.

78) argue that the CO_2 concentration to global temperature relation is concave, while the global temperature to damage relation is convex. Therefore, they consider a linear damage function a reasonable approximation. Following their argument and Battaglini and Harstad (2016), I assume

$$D(G_\tau) = cG_\tau, \tag{7}$$

with $c > 0$ as marginal climate damage.

Note that countries are asymmetric with respect to the satiation point $\bar{y}_i$ and the capacity endowments $R_{i,1}$ and $B_{i,1}$. In all other respects, they are identical.

2.2 Value Function

This paper uses discrete time, with each period τ lasting Ξ moments. At the beginning of each period, the countries simultaneously decide about their capacity investments. Subsequently, energy is produced and carbon emitted. Similar to Battaglini and Harstad (2016), I consider an investment time-lag. One period is needed to realize capacity investments. Therefore, energy generation and corresponding carbon emissions of period τ are determined by the decisions of period $\tau - 1$.[11] The utility of country i in period τ reads

$$u_{i,\tau} = -\frac{a}{2}\left[\bar{y}_i - R_{i,\tau} - B_{i,\tau}\right]^2 - hB_{i,\tau} - cG_\tau - \frac{k}{2}\left[R^2_{i,\tau+1} - q_R^2 R^2_{i,\tau}\right] - \frac{l}{2}\left[B^2_{i,\tau+1} - q_B^2 B^2_{i,\tau}\right]. \tag{8}$$

Following Battaglini and Harstad (2016), I solve the model for Markov-perfect-equilibria (MPE) in pure strategies, so that the current state of the economy determines the strategy of all countries.[12] In other words, the history that led to the state is irrelevant. Let $\rho > 0$ denote the time preference rate and define $\delta := e^{-\rho\Xi}$ as discount factor. Then, the value function of country i can be written as $v_{i,t} = \hat{v}_{i,t} - \zeta$, with[13]

$$\hat{v}_{i,t} = \sum_{\tau=t}^{\infty} \delta^{\tau-t}\left[-\frac{a\delta}{2}d^2_{i,\tau+1} - \delta C \sum_{j\in N}\left(\bar{y}_j - d_{j,\tau+1} - R_{j,\tau+1}\right) - \delta h\left(\bar{y}_i - d_{i,\tau+1} - R_{i,\tau+1}\right) - \frac{K}{2}R^2_{i,\tau+1} - \frac{L}{2}\left(\bar{y}_i - R_{i,\tau+1} - d_{i,\tau+1}\right)^2\right] + \zeta. \tag{9}$$

$d_{i,\tau} := \bar{y}_i - R_{i,\tau} - B_{i,\tau}$ denotes the difference between the satiation point $\bar{y}_i$ and energy generation at period τ given by $R_{i,\tau} + B_{i,\tau}$. The effective costs of green and black capacity investments are measured by $K := k\left(1 - q_R^2\delta\right)$ and $L := l\left(1 - q_B^2\delta\right)$, respectively. Finally, the social costs of carbon of one CO_2 unit emitted at period t read $C := \frac{c}{1-q_G\delta} = c\sum_{\tau=t}^{\infty}(\delta q_G)^{\tau-t}$. The parameter $\zeta := -\frac{a}{2}d^2_{i,t} - Cq_G G_{t-1} - C\sum_{j\in N} B_{j,t} - hB_{i,t} + \frac{k}{2}q_R^2 R^2_{i,t} + \frac{l}{2}q_B^2 B^2_{i,t}$ is exogenously given at period t, so that it can be omitted.

[11] Battaglini and Harstad (2016) assume that the pollution decision precedes the investment decision. I abstain from two decision points to simplify notation. As long as an investment time-lag exists and investments are realized at the time of the next pollution decision, the results are not affected by a second decision point.

[12] See Maskin and Tirole (2001), Harstad (2012), Battaglini and Harstad (2016) for more detailed discussions of MPE.

[13] See Online-Appendix A.1.

2.3 Business as Usual and First-Best

At first, I characterize the non-cooperative MPE also denoted as business as usual (BAU) and the first-best allocation. Both cases will serve as benchmarks. In case of BAU, the governments of all countries $i \in N$ maximize their value function $v_{i,t}$ with respect to $d_{i,\tau}$ and $R_{i,\tau},\ \tau \in \{t+1, ..., \infty\}$. Solving the first-order conditions yields

$$B^{BAU}_{i,\tau} = \Psi_i - \Delta(C+h) \quad \Leftrightarrow \; b_{i,\tau-1} = \Psi_i - \Delta(C+h) - q_B B_{i,\tau-1}, \tag{10}$$

$$R^{BAU}_{i,\tau} = \Phi_i + \Lambda(C+h) \quad \Leftrightarrow \; r_{i,\tau-1} = \Phi_i + \Lambda(C+h) - q_R R_{i,\tau-1}, \tag{11}$$

$$d^{BAU}_{i,\tau} = \Gamma_i + \Pi(C+h), \tag{12}$$

with $\Psi_i = \frac{a\delta K}{KL+a\delta(L+K)}\bar{y}_i$, $\Delta = \frac{\delta(K+a\delta)}{KL+a\delta(L+K)}$, $\Phi_i = \frac{a\delta L}{KL+a\delta(L+K)}\bar{y}_i$, $\Lambda = \frac{\delta^2 a}{KL+a\delta(L+K)}$, $\Gamma_i = \frac{KL}{KL+a\delta(L+K)}\bar{y}_i$, and $\Pi = \frac{K\delta}{KL+a\delta(K+L)}$. Note $\Psi_i, \Phi_i, \Gamma_i, \Delta, \Lambda, \Pi > 0$, which implies $d^{BAU}_{i,\tau} > 0$. Thus, the gap between the satiation point $\bar{y}_i$ and energy generation $R_{i,\tau} + B_{i,\tau}$ is positive for all countries ruling out all investment paths $(r_{i,\tau}, b_{i,\tau})$ which lead to an excess energy consumption characterized by $R_{i,\tau} + B_{i,\tau} > \bar{y}_i$.[14] According to (12), the energy gap $d_{i,\tau}$ is small [large] for countries with a low [high] satiation point. (10) and (11) show that a higher satiation point increases black energy consumption by the factor $\frac{a\delta K}{KL+a\delta(L+K)} \in (0,1)$, while green energy consumption is increased by the factor $\frac{a\delta L}{KL+a\delta(L+K)} \in (0,1)$. Thus, a higher satiation point boosts energy consumption but the increase is not sufficient to close the energy gap. Rather, the gap increases by the factor $\frac{KL}{KL+a\delta(L+K)} \in (0,1)$.[15]

The first-best allocation is determined by the maximization of $\sum_{j \in N} v_{j,t}$ with respect to $d_{i,\tau}$ and $R_{i,\tau},\ \tau \in \{t+1, ..., \infty\}$ and reads

$$B^{FB}_{i,\tau} = \Psi_i - \Delta(nC+h) \quad \Leftrightarrow \; b_{i,\tau-1} = \Psi_i - \Delta(nC+h) - q_B B_{i,\tau-1}, \tag{13}$$

$$R^{FB}_{i,\tau} = \Phi_i + \Lambda(nC+h) \quad \Leftrightarrow \; r_{i,\tau-1} = \Phi_i + \Lambda(nc+h) - q_R R_{i,\tau-1}, \tag{14}$$

$$d^{FB}_{i,\tau} = \Gamma_i + \Pi(nC+h). \tag{15}$$

According to (10–15), countries invest more in renewable energy and less into black capacity, if the social costs of carbon $C = \frac{c}{1-\delta q_G}$ are high and fossil fuel extraction costly. Because $\frac{\partial \Psi_i}{\partial K} > 0$, $\frac{\partial \Delta}{\partial K} < 0$, $\frac{\partial \Phi_i}{\partial K} < 0$ and $\frac{\partial \Lambda}{\partial K} < 0$, lower costs of green capacity accumulation reflected by a lower $K = k(1 - q_R^2\delta)$ lead to higher green capacity investments and lower black capacity investments. In contrast, a variation of the cost parameter of black capacity investments $L = l(1 - q_B^2\delta)$ causes two opposing effects. On the one hand, $\frac{\partial \Psi_i}{\partial L} < 0$ and $\frac{\partial \Phi_i}{\partial L} > 0$ imply that higher black capacity costs decrease investments into black capacity but increase green capacity investments. On the other hand, the corresponding reductions of climate

[14] This statement holds for all cases discussed in this paper.

[15] Similar remarks hold for all subsequently discussed cases characterized by a binding black capacity constraint. If the constraint does not bind, the energy gap and green energy consumption are independent from the satiation point implying that fossil fuel consumption increases with the satiation point.

damages and extraction costs imply a counter effect, which is reflected by $\frac{\partial \Delta}{\partial L} < 0$ and $\frac{\partial \Lambda}{\partial L} < 0$.

Compared with the first-best solution, countries invest too much into black capacity and too little into green capacity under the BAU regime. Carbon emissions are inefficiently high, because they are directly determined by black capacity. This reflects the non-internalized environmental damage externality.

3 IEA-Contracts

In this section, I turn to the formation of international environmental agreements or climate coalition, respectively. These agreements have to be self-enforcing, because a supranational authority able to enforce an agreement is missing.[16] At first, I introduce the timing of coalition formation. Subsequently, I present the stability conditions of climate coalitions with a complete and with an incomplete contract. In the first case, the coalition members coordinate their policies with respect to both green and black capacity investments. Note that the latter directly determine carbon emissions. In the second case, the coalition cooperatively sets carbon emission but delegates the decision about green capacity investments to the national authorities.

3.1 Timing

The timing of the model is illustrated in Fig. 1.

At first, suppose that no coalition exists at the beginning of a period. In this case, the countries independently and simultaneously decide whether to join the coalition or not (stage (1): *coalition formation stage*). Subsequently, the coalition countries negotiate the climate contract by determining the contract duration T and their policy $\left(r_{i,\tau}, b_{i,\tau}\right)$ for all $i \in M$ and all $\tau \in \{1, ..., T\}$ (stage (2): *negotiation stage*). Finally, at the *investment and emission stage* (stage (3)), the fringe countries of group $F = N \setminus M$ independently and simultaneously determine their policy $\left(r_{i,\tau}, b_{i,\tau}\right)$ for the current period and all countries invest and pollute according to their plans. The timing implies a Stackelberg game structure, where the coalition moves first in the first contract period and the fringe countries subsequently in every period $\tau \in \{1, ..., T\}$. If a climate coalition exists at the beginning of a period, the coalition formation stage and the negotiation stage are omitted.

3.2 Complete Contract

Consider an agreement where the m coalition members agree on $r_{i,\tau}$ and $b_{i,\tau}$ and, therefore, on $R_{i,\tau+1}$ and $B_{i,\tau+1}$ for all $\tau \leq T$ and $i \in M$. In other words, the coalition acts as a single player who plays non-cooperatively against the fringe countries by maximizing the aggregated value function $\sum_{j \in M} v_{j,t}$ with respect to $d_{i,\tau}$ and $R_{i,\tau}$, $\tau \in \{t+1, ..., T+1\}$. In contrast, the fringe countries act independently, so that their strategy is given by (10) and (11). Thus, the strategy of the fringe depends neither on the coalition's strategy nor on the

[16] Following Hoel (1992), Carraro and Siniscalco (1993), Barrett (1994) and Battaglini and Harstad (2016), I apply the stability concept of d'Aspremont et al. (1983).

size of the climate coalition. The reason is that the climate damage function (7) is linear and that there are no other externalities. Therefore, the coalition does not enjoy a first-mover advantage although it determines its policy during the first contract period for all $\tau \in \{1, .., T\}$, while fringe countries determine their strategy in every period.

Solving the first-order conditions of the coalition for $B_{i,\tau}$ and $R_{i,\tau}$ yields

$$B^M_{i,\tau} = \Psi_i - \Delta(mC + h) \quad \Leftrightarrow \quad b_{i,\tau} = \Psi_i - \Delta(mC + h) - q_B B_{i,\tau-1}, \tag{16}$$

$$R^M_{i,\tau} = \Phi_i + \Lambda(mC + h) \quad \Leftrightarrow \quad r_{i,\tau} = \Phi_i + \Lambda(mC + h) - q_R R_{i,\tau-1}, \tag{17}$$

$$d^M_{i,\tau} = \Gamma_i + \Pi(mC + h). \tag{18}$$

If the coalition does not encompass all countries, the investments into green (black) capacity are inefficiently low (high). However, the inefficiencies are smaller than in the BAU-case, because the coalition internalizes the environmental externality within the coalition but ignores the effect of its investment decisions on the fringe countries.

To analyze the optimal contract duration let m^* denote the size of the stable coalition. In Online-Appendix A.2, I prove Lemma 1, which determines the optimal contract length.

Lemma 1 *Suppose the coalition members coordinate their black capacity investments and their green capacity investments (complete contract). The optimal contract length is given by*

$$T^* \begin{cases} = 1, & \text{if } m < m^* \\ \in \{1, 2, ..., \infty\}, & \text{if } m = m^* \\ = \infty, & \text{if } m > m^*. \end{cases} \tag{19}$$

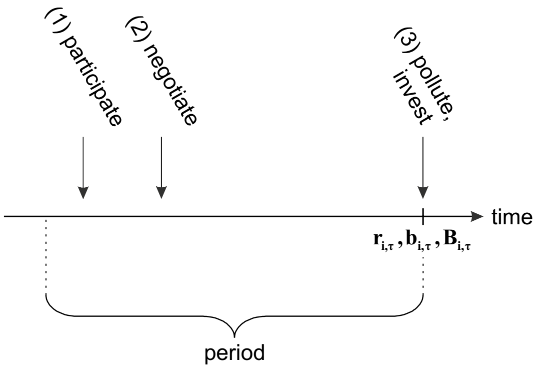
According to Lemma 1, the coalition countries sign an one-period contract if the coalition size falls short of the size of the stable coalition. If the coalition is larger than the stable one, the contract is signed for an unlimited time period. In case of the stable coalition, the members are indifferent between all possible contract lengths, because an identical contract will be signed if the current contract expires.

Finally, I turn to the stability of the coalition. An IEA is self-enforcing or stable if no fringe country has an incentive to sign the agreement (external stability) and no coalition country has an incentive to defect (internal stability).[17] Consider the stable coalition characterized by (m^*, T^*). According to Lemma 1, the coalition countries are indifferent with respect to the contract length. If one coalition country defects, the coalition size falls short of m^* and the remaining $m^* - 1$ countries will sign a short-term contract of only one period. By definition, the stable coalition is established subsequently. This implies that the defecting country will enjoy its free-riding benefits for only one period. Formally, the defection does not pay-off, if the internal stability condition

$$v^c_i(m^*, T^*) \geq u^f_{i,\tau}(m^* - 1) + \delta v^c_i(m^*, T^*) \tag{20}$$

[17] Cf. d'Aspremont et al. (1983).

Fig. 1 Timing



holds for the signatories $i \in M$.[18] If a fringe country joins the stable coalition (m^*, T^*), the coalition size exceeds m^*. In this case, Lemma 1 implies that the coalition countries will sign a contract for all eternity. Thus, the acceding country will be part of the coalition for all time. The accession of the fringe country does not payoff if the external stability condition

$$v_i^f(m^*, T^*) \geq v_i^c(m^* + 1, \infty) \tag{21}$$

is satisfied for all fringe countries $i \in F$. In Online-Appendix A.3, I prove

Proposition 1 *Suppose contracts are complete and the black capacity limits fossil fuel use for all points of time. The stable coalition consists of maximal three countries.*

Proposition 1 reproduces the result of Battaglini and Harstad (2016). The free-riding incentives with complete contracts are large, so that only a small coalition of $m^* = 3$ is stable. Battaglini and Harstad (2016) suggest incomplete contracts, where coalition countries only agree on emission but not on green capacity investments, to increase the size of the stable coalition. They find that in the last period of the contract a hold-up problem emerges, which reduces green capacity investments of signatories. By extending the contract length, these low capacity investments can be postponed to future periods. In case that one signatory defects, the remaining countries can credibly threat to sign a short-term contract, which antedates the underinvestment into green capacity. It is this credible threat that stabilizes coalitions with more than three members. Whether the mechanism also works in the presence of a black capacity is investigated in the following section.

3.3 Incomplete Contracts

In case of an incomplete contract, the signatories only coordinate their emission policies but leave the decision with respect to green capacity investments to the national authorities. Because emissions are determined by black capacity, the coordination of emissions implies the cooperative determination of black capacity investments.

[18] I mark the coalition members with the superscript c and fringe countries with f.

I adopt the timing of Battaglini and Harstad (2016), which is similar to the complete contract case. Thus, after the coalition formation stage, the coalition chooses its black capacity investments for all $i \in M$ and all $\tau \in \{t, .., T\}$ at the negotiation stage. At the investment and emission stage, all countries simultaneously and non-cooperatively determine their national policies and the strategies are realized. In case of the fringe, the strategy consists of both green and black capacity investments. In contrast, coalition countries only decide on their green capacity investments independently.[19]

Because the signatories determine their green capacity investments in every period, the timing implies a Stackelberg game structure with the coalition as the leader and the signatories as the follower. A similar structure exists with respect to the coalition and the fringe. However, the optimization problem of the fringe countries is neither affected by the degree of cooperation inside the coalition nor by the coalition size. Therefore, their policies are given by (10) and (11) implying that the coalition does not enjoy a first-mover advantage with respect to the fringe's strategy.[20] As shown subsequently, this is not true with respect to the national policies of the signatories.

To solve the optimization problem of the coalition and its members, I rewrite the value function of a coalition member i as

$$v_{i,1}(m,T) = \sum_{\tau=1}^{T} \delta^{\tau-1} \left[-\frac{a\delta}{2}\left(\bar{y}_i - R_{i,\tau+1} - B_{i,\tau+1}\right)^2 - \delta C \sum_{j\in N} B_{j,\tau+1} - \delta h B_{i,\tau+1} - \frac{K}{2}R_{i,\tau+1}^2 - \frac{L}{2}B_{i,\tau+1}^2 \right] + \delta^{T+1} v_{i,T+1}. \quad (22)$$

The coalition determines its strategy in the first contract period for all $\tau \in \{1, ..., T\}$, while the signatories' national strategies are determined in every period. In other words, the coalition moves first and the problem is solved by backward induction. Differentiating (22) with respect to $R_{i,\tau}$, $\tau \in \{2, ..., T+1\}$, yields the signatories' reaction function

$$R_{i,\tau} = \frac{a\delta}{K + a\delta}\left[\bar{y}_i - B_{i,\tau}\right]. \quad (23)$$

Thus, the coalition anticipates that it can increase the green capacity investments of the signatories by reducing its black capacity investments.[21] In particular, (23) is valid for period $T+1$, because the investment decision determining $R_{i,T+1}$ is made in T. Therefore, the coalition influences emissions in the period following the expiration of the contract. Substituting (23) into (22) and maximizing $\sum_{j\in M} v_j$ with respect to $B_{i,\tau}$, $\tau \in \{2, ..., T+1\}$ yields

$$B_{i,\tau} = \Psi_i - \Delta(mC + h). \quad (24)$$

The black capacity and, therefore, the emissions chosen by the coalition are identical to the case of a complete contract. Substituting (24) into (23) gives

$$R_{i,\tau} = \Phi_i + \Lambda(mC + h), \quad (25)$$

so that the green capacity investments chosen by the signatories are identical to the green capacity investments of the coalition under a complete contract. Thus, the coalition is able

---

[19] If a coalition already exists at the beginning of a period, the first two stages are omitted.

[20] As discussed in sect. 3.2, the coalition has no first-mover advantage with respect to the fringe's strategy because of the linearity of the damage function (7).

[21] Formally, the reaction function exists, because signatory i's objective function (22) is quadratic in $B_{i,\tau+1}$. In contrast, the fringe's objective function (9) is linear in both $B_{i,\tau+1}$ and $R_{i,\tau+1}$, $i \in M$, so that the coalition cannot influence the fringe's strategy.

to implement the same policy as in case a complete contract. Because this policy represents the welfare maximizing strategy for the coalition given its size m, the coalition enjoys a first-mover advantage.[22] In Online-Appendix A.4, I prove

Proposition 2 *Suppose the coalition only coordinates the emission policy (incomplete contracts) and the black capacity limits fossil fuel use for all points of time.*

(i) *The optimal contract length is given by*

$$T^* \begin{cases} = 1, & \text{if } m < m^* \\ \in \{1, 2, ..., \infty\}, & \text{if } m = m^* \\ = \infty, & \text{if } m > m^* \end{cases} \tag{26}$$

(ii) *The stable coalition consists of maximal three countries.*

Proposition 2 is identical with the complete contract result of sect. 3.2. This contrasts with the findings of Battaglini and Harstad (2016), who advocate incomplete contracts to increase the size of the stable coalition. The disparity is driven by the implicit assumption of an unlimited black capacity made by Battaglini and Harstad (2016), which causes a hold-up problem. With an unlimited black capacity, the coalition cannot control the carbon emissions of its members in period $T + 1$. Consequently, the countries realize that their green capacity investments in the last contract period will only allow them to easily reduce their emissions in the next period but will not affect the emissions of other countries. In other words, high green capacity investments will weaken the position of coalition members in the negotiation of a new climate contract. To avoid this weak position, the countries reduce their green capacity investments to the BAU-level. Because low investments imply more fossil fuel use and, therefore, higher climate damages, all countries are interested in postponing the hold-up problem. This stabilizes larger coalitions, because coalition countries can credibly threat to sign a short-term agreement if one country defects.

Consider now the case of a limited black capacity, such that fuel use at period $\tau + 1$ is determined in the previous period τ by the black capacity investments. Recall that the coalition has a first-mover advantage, which it uses to influence the green capacity investments of its members such that these equal the value of the complete contract case. In particular, this influence is present at the last period of the contract T, so that the green capacity investments of this period are not reduced to the BAU-level. Consequently, the hold-up problem and its stabilizing effect on larger coalitions is eliminated.

4 Initially Non-binding Capacity Constraint

In the previous section, I implicitly assumed that the installed black capacity constrains fossil fuel use. In other words, I assumed a binding capacity constraint for all periods. This assumption is relaxed in this section in the sense that the capacity constraint is non-binding

[22] Under a complete contract, the coalition maximizes the aggregated welfare of the signatories with respect to both green and black capacity.

initially.[23] Thus, I assume that $x_{i,1} < B_{i,1}$, with $x_{i,\tau}$ denoting fossil fuel use of country i in period τ. By using Online-Appendix A.1, the value function of country i can be written as $v_{i,1} = \hat{v}_{i,1} - \zeta_x$, with

$$\hat{v}_{i,1} = \zeta_x + \sum_{\tau=1}^{\infty} \delta^{\tau-1}\left[-\frac{a}{2}d_{i,\tau}^2 - C\sum_{j\in N}\left(\bar{y}_j - d_{j,\tau}\right) - h\left(\bar{y}_j - d_{i,\tau}\right) + \delta C\sum_{j\in N} R_{j,\tau+1} + \delta h R_{i,\tau+1} - \frac{K}{2}R_{i,\tau+1}^2 - \frac{L}{2}B_{i,\tau+1}^2\right], \tag{27}$$

where

$$d_{i,\tau} = \begin{cases} \bar{y}_i - R_{i,\tau} - x_{i,\tau}, & \text{if } x_{i,\tau} < B_{i,\tau} \\ \bar{y}_i - R_{i,\tau} - B_{i,\tau}, & \text{if } x_{i,\tau} = B_{i,\tau} \end{cases}$$

and $\zeta_x := -Cq_G G_0 + \frac{k}{2}q_R^2 R_{i,1}^2 + \frac{l}{2}q_B^2 B_{i,1}^2 + hR_{i,1} + C\sum_{j\in N} R_{j,1}$ is a constant. Differentiating with respect to $B_{i,\tau+1}$ yields

$$\frac{\partial v_{i,1}}{\partial B_{i,\tau+1}} < 0, \tag{28}$$

if the capacity constraint does not bind in $\tau + 1$. Thus, country i should reduce its black capacity as fast as possible. If arbitrary large disinvestments are possible, the capacity is immediately reduced to the value of a binding capacity constraint. In other words, the constraint binds in period 2. However, I assume that capacity can be only reduced by $(1 - q_B)B_{i,\tau} + \alpha$ per period, with $\alpha \geq 0$. Consequently, several periods are needed until the capacity constraint becomes binding in period $T_i > 2$. For simplicity, I assume $T_x = T_i = T_j$ for all $i \neq j \in N$. This assumption is relaxed in Sect. 4.4.

4.1 Business as Usual and First-Best

Before I turn to the stability of the IEA, I characterize the business as usual and the first-best solutions, which serve as benchmark cases. For this purpose, I rewrite the value function as

$$\begin{aligned} v_{i,1} = &\sum_{\tau=1}^{T_x-2} \delta^{\tau-1}\left[-\frac{a}{2}d_{i,\tau}^2 - C\sum_{j\in N}\left(\bar{y}_j - d_{j,\tau}\right) - h\left(\bar{y}_i - d_{i,\tau}\right) + \delta C\sum_{j\in N} R_{j,\tau+1} + \delta h R_{i,\tau+1} - \frac{K}{2}R_{i,\tau+1}^2 - \frac{L}{2}B_{i,\tau+1}^2\right] \\ &+ \delta^{T_x-1-t}\left[-\frac{a}{2}d_{i,T_x-1}^2 - C\sum_{j\in N}\left(\bar{y}_j - d_{j,T_x-1}\right) - h\left(\bar{y}_i - d_{i,T_x-1}\right) + \delta C\sum_{j\in N} R_{j,T_x} + \delta h R_{i,T_x} \right. \\ &\left. - \frac{K}{2}R_{i,T_x}^2 - \frac{L}{2}\left(\bar{y}_i - R_{i,T_x} - d_{i,T_x}\right)^2\right] \\ &+ \delta^{T_x-t}\sum_{\tau=T_x}^{\infty} \delta^{\tau-T_x}\left[-\frac{a}{2}d_{i,\tau}^2 - C\sum_{j\in N}\left(\bar{y}_j - d_{j,\tau}\right) - h\left(\bar{y}_i - d_{i,\tau}\right) + \delta C\sum_{j\in N} R_{j,\tau+1} + \delta h R_{i,\tau+1} \right. \\ &\left. - \frac{K}{2}R_{i,\tau+1}^2 - \frac{L}{2}\left(\bar{y}_i - d_{i,\tau+1} - R_{i,\tau+1}\right)^2\right]. \end{aligned} \tag{29}$$

[23] If the constraint is non-binding for all periods, the model is identical with the one of Battaglini and Harstad (2016). Therefore, this case is not discussed in the following.

The business as usual solution is given by the maximization of (29) with respect to $d_{i,\tau}$ and $R_{i,\tau+1}$, which yields[24]

$$d_{i,\tau}^{BAU} = \frac{C+h}{a} \Leftrightarrow \quad x_{i,\tau}^{BAU} = \bar{y}_i - \frac{C+h}{a} - R_{i,\tau}, \quad \tau \in \{1, ..., T_x - 1\}, \tag{30}$$

$$R_{i,\tau+1}^{BAU} = \delta \frac{C+h}{K} \Leftrightarrow \quad r_{i,\tau}^{BAU} = \delta \frac{C+h}{K} - q_R R_{i,\tau}, \quad \tau \in \{1, ..., T_x - 2\}, \tag{31}$$

$$R_{i,T_x}^{BAU} = \Phi_i + \Lambda(C+h) \Leftrightarrow \quad r_{i,T_x-1}^{BAU} = \Phi_i + \Lambda(C+h) - q_R \delta \frac{C}{K}, \tag{32}$$

$$B_{i,T_x}^{BAU} = \Psi_i - \Delta(C+h) \Leftrightarrow \quad b_{i,T_x-1}^{BAU} = \Psi_i - \Delta(C+h) - q_B B_{i,T_x-1}. \tag{33}$$

For $\tau \geq T_x + 1$ the first-order conditions replicate (10) and (11).

Maximizing $\sum_{j \in N} v_j$ with respect to $d_{i,\tau}$ and $R_{i,\tau+1}$ determines the first-best allocation, which reads

$$d_{i,\tau}^{FB} = \frac{nC+h}{a} \Leftrightarrow \quad x_{i,\tau}^{FB} = \bar{y}_i - \frac{nC+h}{a} - R_{i,\tau}, \quad \tau \in \{1, ..., T_x - 1\}, \tag{34}$$

$$R_{i,\tau+1}^{FB} = \delta \frac{nC+h}{K} \Leftrightarrow \quad r_{i,\tau}^{FB} = \delta \frac{nC+h}{K} - q_R R_{i,\tau}, \quad \tau \in \{1, ..., T_x - 2\}, \tag{35}$$

$$R_{i,T_x}^{FB} = \Phi_i + \Lambda(nC+h) \Leftrightarrow \quad r_{i,T_x-1}^{FB} = \Phi_i + \Lambda(nC+h) - q_R \delta \frac{C}{K}, \tag{36}$$

$$B_{i,T_x}^{FB} = \Psi_i - \Delta(nC+h) \Leftrightarrow \quad b_{i,T_x-1}^{FB} = \Psi_i - \Delta(nC+h) - q_B B_{i,T_x-1}, \tag{37}$$

and (13) and (14) for $\tau \geq T_x + 1$. Comparing (10) and (30) as well as (13) and (34) reveals that more fossil fuel is used without a binding capacity constraint under both regimes. Analogously, (11), (14), (31), and (35) show that green energy investments are lower in case of a non-binding black capacity constraint. In the last period of the non-binding black capacity constraint, the investments for green and black capacity are chosen such that the economy is adapted to the binding constraint in the next period.

4.2 Complete Contract

Consider an agreement that coordinates the green capacity investments and the carbon emissions of the signatories. Note that the latter implies a coordination of black capacity investments in period $T_x - 1$. According to proposition 1, the stable coalition has maximal three members for $\tau \geq T_x$. Subsequently, I analyze whether larger coalitions are possible for $\tau < T_x$. For that purpose, differentiate $\sum_{j \in M} v_j$ with respect to $d_{i,\tau}$ and $R_{i,\tau+1}$, which yields

[24] Recall that the black capacity investments are minimal until $t = T_x - 1$ for both cases business as usual and first-best.

$$d_{i,\tau}^{M} = \frac{mC+h}{a} \Leftrightarrow \quad x_{i,\tau}^{M} = \bar{y}_i - \frac{mC+h}{a} - R_{i,\tau}, \quad \tau \in \{1, ..., T_x - 1\}, \tag{38}$$

$$R_{i,\tau+1}^{M} = \delta \frac{mC+h}{K} \Leftrightarrow \quad r_{i,\tau}^{M} = \delta \frac{mC+h}{K} - q_R R_{i,\tau}, \qquad \tau \in \{1, ..., T_x - 2\}, \tag{39}$$

$$R_{i,T_x}^{M} = \Phi_i + \Lambda(mC+h) \Leftrightarrow \quad r_{i,T_x-1}^{M} = \Phi_i + \Lambda(mC+h) - q_R \delta \frac{C}{K}, \tag{40}$$

$$B_{i,T_x}^{M} = \Psi_i - \Delta(mC+h) \Leftrightarrow \quad b_{i,T_x-1}^{M} = \Psi_i - \Delta(mC+h) - q_B B_{i,T_x-1}, \tag{41}$$

and (16) and (17) for $\tau \geq T_x + 1$. Similar to the case of a binding black capacity constraint, coalition countries use more fossil fuel than socially optimal but less than fringe countries, while their green capacity investments are inefficiently low but higher than in the fringe.

To analyze the size of the stable coalition, I make use of backward induction. Suppose the coalition has more than three members in the last period with a non-binding constraint ($T_x - 1$). Then, the contract has to end in this period, because a binding constraint implies a coalition of maximal three countries. The coalition is stable in $T_x - 1$ if the internal stability condition

$$\begin{aligned} &u_i^c(m^*, T_x - 1) + \delta v_{i,T_x} \geq \\ &u_i^f(m^* - 1, T_x - 1) + \delta v_{i,T_x} \end{aligned} \tag{42}$$

holds for all $i \in M$ and the external stability condition

$$\begin{aligned} &u_i^f(m^*, T_x - 1) + \delta v_{i,T_x} \geq \\ &u_i^c(m^* + 1, T_x - 1) + \delta v_{i,T_x} \end{aligned} \tag{43}$$

holds for all $i \in F$, where v_{i,T_x} denotes the value of country i for $\tau \geq T_x$. In Online-Appendix A.5, I prove that the coalition at time $T_x - 1$ has maximally three members. Thus, I can check the stability for period $T_x - 2$ and, subsequently, for all periods $\tau < T_x - 2$. The corresponding internal and external stability conditions read

$$\begin{aligned} &u_i^c(m^*, T_x - \tau) + \delta v_{i,T_x-\tau+1} \geq \\ &u_i^f(m^* - 1, T_x - \tau) + \delta v_{i,T_x-\tau+1}, \end{aligned} \tag{44}$$

$$\begin{aligned} &u_i^f(m^*, T_x - \tau) + \delta v_{i,T_x-\tau+1} \geq \\ &u_i^c(m^* + 1, T_x - \tau) + \delta v_{i,T_x-\tau+1}. \end{aligned} \tag{45}$$

Online-Appendix A.5 together with proposition 1 prove proposition 3.

Proposition 3 *Suppose contracts are complete and that the black capacity constraint does not bind initially but becomes binding in all countries at period T_x. For all periods, the stable coalition has maximally three members.*

Without a black capacity constraint, Battaglini and Harstad (2016) show that the climate coalition has not more than three members under a complete contract. According to proposition 1, this result holds in case of a binding capacity constraint. Therefore, it is not surprising that in a combination of both settings the stable coalition size is not larger.

4.3 Incomplete Contract

In case of an incomplete contract, Battaglini and Harstad (2016) find that coalitions with more than three members can be stable, while proposition 2 shows that this result does not hold if the black capacity constraint binds. In the following, I am going to analyze whether there can be large coalitions before the capacity constraint binds under an incomplete contract. As in sect. 3.3, I adopt the timing of Battaglini and Harstad (2016), so that the coalition determines its emission policy on the second stage. Subsequently, the coalition members and the fringe countries choose their national policies in every period. Again, the timing implies a Stackelberg-structure between the coalition and the signatories and between the coalition and the fringe. However, the linearity of the damage function implies that the fringe's strategy is not affected by the coalition and, therefore, given by (30–33) for $\tau \le T_x - 1$. Following the established procedure, I use backward induction to determine the strategies of the coalition and its members.

By taking note of that $v_{i,T_x} = \hat{v}_{i,T_x} - \zeta$, with $\hat{v}_{i,T_x}$ given by (9) evaluated at T_x, is independent from the decisions made before T_x, the value function of country i can be rewritten as

$$v_{i,1} = \delta^{T_x-1} v_{i,T_x} + \sum_{\tau=1}^{T_x-1} \delta^{\tau-1} \Big[-\frac{a}{2}\left(\bar{y}_i - R_{i,\tau} - x_{i,\tau}\right)^2 - C\sum_{j\in N}\left(x_{j,\tau} + R_{j,\tau}\right) - h(x_{i,\tau} + R_{i,\tau}) + \delta C \sum_{j\in N} R_{j,\tau+1} + \delta h R_{i,\tau+1} - \frac{K}{2}R_{i,\tau+1}^2 - \frac{L}{2}B_{i,\tau+1}^2 \Big] + \delta^{T_x - t}\Big[-\frac{a}{2}\left(\bar{y}_i - R_{i,T_x} - B_{i,T_x}\right)^2 - C\sum_{j\in N}\left(R_{j,T_x} + B_{j,T_x}\right) - h\left(R_{i,T_x} + B_{i,T_x}\right)\Big]. \tag{46}$$

Suppose that a climate contract exists at period $\tau = \{1, ..., T_x - 1\}$.[25] Maximizing (46) with respect to $R_{i,\tau}$, $\tau \in \{2, ..., T_x\}$, yields[26]

$$R_{i,\tau} = \frac{a\delta}{K + a\delta}\left(\bar{y}_i - x_{i,t}\right), \qquad \tau \in \{2, ..., T_x - 1\}, \tag{47}$$

$$R_{i,T_x} = \frac{a\delta}{K + a\delta}\left(\bar{y}_i - B_{i,T_x}\right). \tag{48}$$

The green capacity investments of the coalition members are the higher the stricter the coordinated emission policy, which is anticipated by the coalition. Substituting (47) and (48) into (46) and differentiating $\sum_{j\in M} v_j$ with respect to $x_{i,\tau}$ with $\tau \in \{1, ..., T_x - 1\}$ and B_{i,T_x} yields

$$x_{i,\tau}^M = \bar{y}_i - R_{i,\tau} - \frac{mC + h}{a}, \quad \tau \in \{1, ..., T_x - 1\}, \tag{49}$$

[25] For $\tau \ge T_x$ the coalition has maximal three members.

[26] For $\tau < T_x$, the differentiation of (46) with respect to $B_{i,\tau}$ implies the fastest possible reduction of black capacity.

$$B^M_{i,T_x} = \Psi - \Delta(mC + h). \tag{50}$$

Finally, substituting into (47) and (48) gives

$$R^M_{i,\tau} = \delta \frac{mC + h}{K}, \qquad \tau \in \{2, ..., T_x - 1\}, \tag{51}$$

$$R^M_{i,T_x} = \Phi_i + \Lambda(mC + h) \tag{52}$$

and, therefore,

$$d^M_{i,\tau} = \frac{mC + h}{a}, \quad \tau \in \{1, ..., T_x - 1\}. \tag{53}$$

Due to its first mover advantage, the coalition can induce its members to choose their green investments level such that they equal the coordinated policy of the complete contract. In particular, this is true for the first period with a binding capacity constraint (T_x), because of the coalition's black capacity investment strategy of period $T_x - 1$. Therefore, the hold-up problem identified by Battaglini and Harstad (2016) does not occur. To analyze the stability of the coalition, I use backward induction. In Online-Appendix A.6, I prove lemma 2.

Lemma 2 *Suppose the climate coalition only coordinates the emission policy (incomplete contract) and the black capacity constraint does not bind initially but becomes binding in all countries in period T_x. In period $T_x - 1$, the stable coalition has maximally three members.*

According to Battaglini and Harstad (2016), a hold-up problems reduces the green capacity investments in the last period of the contract below the coalition level. In case of an unlimited black capacity, this allows the coalition countries to credibly threaten any deviating country to sign a short-term agreement. The corresponding antedating of the reduced green capacity investments stabilizes the coalition. In other words, the countries are willing to form coalitions with more than three members to postpone the hold-up problem. However, if black capacity is limited and becomes binding in the next period, the hold-up problem cannot be postponed to future periods. Rather, the problem is eliminated, because the decisions of period $T_x - 1$ determine the use of both renewables and fossil fuels in period T_x. Consequently, there is no stabilizing effect and the stable climate coalition has maximally three members.

An analogous argument can be repeated for all periods $\tau \leq T_x - 2$. That is, in period $T_x - 2$ the hold-up problem cannot be postponed by forming a large coalition, because any coalition with more than three members has to expire in period $T_x - 2$ to comply with lemma 2. At period $T = T_x - 2$, the coalition members choose[27]

$$R^M_{i,T+1} = \delta \frac{C + h}{K}. \tag{54}$$

[27] If the contract ends at period $T_x - 2$, the coalition countries maximize $v_{i,1} = \delta^{T_x-2} v_{i,T_x-1} + \sum_{\tau=1}^{T_x-2} \delta^{\tau-1} \Big[-\frac{a}{2}\left(\bar{y}_i - R_{i,\tau} - x_{i,\tau}\right)^2 - C\sum_{j\in N}\left(x_{j,\tau} + R_{j,\tau}\right) - h\left(x_{i,\tau} + R_{i,\tau}\right) + \delta C \sum_{j\in N} R_{j,\tau+1} + \delta h R_{i,\tau+1} - \frac{K}{2} R^2_{i,\tau+1} - \frac{L}{2} B^2_{i,\tau+1} \Big]$

At the last period of the agreement (T) the coalition members anticipate that high green capacity investments only reduce their emissions in the next period but do not affect the emissions of other countries. In other words, high green capacity investments weaken country i's bargaining power in the negotiation of a new climate contract creating the hold-up problem. Therefore, the coalition countries underinvest in T. For $\tau < T$, (47) holds such that the coalition's emission policy is not affected by the hold-up problem and (49) holds for all periods of the contract. The stability of the climate coalition is analyzed in Online-Appendix A.7, where I prove proposition 4.

Proposition 4 *Suppose the climate coalition only coordinates the emission policy (incomplete contract) and the black capacity constraint does not bind initially but becomes binding in all countries in period T_x. In all periods $\tau \leq T_x - 1$, the stable coalition has maximally three members.*

Although the hold-up problem appears, it cannot stabilize larger coalitions as in Battaglini and Harstad (2016). The reason is that the hold-up problem cannot be postponed beyond period $T_x - 1$ eliminating the credible threat to sign a short-term agreement. Therefore, the coalition size at this period cannot exceed three. This, in turn, implies that the hold-up problem cannot be postponed beyond period $T_x - 2$ giving a maximal coalition size of three. Repeating the argument for all periods $\tau < T_x - 2$ yields proposition 4. In other words, a black capacity constraint which becomes binding in the future renders the threat of the coalition countries to end the coalition if one country diverges incredible.

4.4 Asymmetry in T_i

In this section, I relax the assumption that the black capacity constraint of all countries will become binding at the same point in time by differentiating between the group $N_I \subset N$ and $N_{II} \equiv N \setminus N_I$, where n_i denotes the size of group N_i, $i = I, II$. The subset $M_i \subset N_i$ denotes the signatories from group N_i, $i = I, II$ with size m_i, so that the number of fringe countries is given by $\sum_{i=I,II}(n_i - m_i)$. The black capacity constraint is non-binding for both groups initially but becomes binding for group N_I at time T_I and for group N_{II} at time T_{II}, with $T_I < T_{II}$. Thus, I can distinguish between three time phases. In Phase 1, which lasts from the first period until $T_I - 1$, the capacity constraint is non-binding in all countries. During Phase 2, the capacity constraints binds in group N_I, while the countries of group N_{II} still face a non-binding constraint. This Phase begins with period T_I and ends at $T_{II} - 1$. Finally, the capacity constraint is binding in all countries in Phase 3, i.e. for all $\tau \geq T_{II}$. Because Phase 3 is identical with the setting discussed in Sect. 3, propositions 1 and 2 hold, so that the stable coalition has maximally three members for all $\tau \geq T_{II}$ irrespective whether the climate contract is complete or incomplete.

In Online-Appendix A.8 the problem is solved for the complete contract case by using backward induction and by following McGinty (2007). Proposition 5 summarizes the results.

Proposition 5 *Consider the case of a complete climate contract. During both Phase 1 and Phase 2 the stable climate coalition has maximally three members. Tab. 1 and 2 state the possible stable coalitions, which depend on the size of $\delta a \Delta$.*

Because Phase 1 *equals the setting of Sect.* 4.2, *the size of the stable coalition equals three. This is also true for Phase* 2, *although some countries face a binding capacity constraint, while some face a non-binding constraint. If* $\delta a\Delta$ *is small, only countries with a binding capacity constraint are members of the climate coalition at period* $\tau \in \{T_I - 1, ..., T_{II} - 1\}$. *If* $\delta a\Delta$ *is sufficiently large, which is the case if both green capacity investments and black capacity investments are cheap (low L and low K), energy consumption is important (high a) and the discount factor* δ *is high, also a coalition of three countries without a binding capacity constraint can be stable. However, in contrast to Phase* 1 *and Phase* 3, *a mixed coalition is not stable at period* $\tau \in \{T_I - 1, ..., T_{II} - 1\}$.

Finally, consider the case of an incomplete contract, which is discussed in detail in Online-Appendix A.9.

Proposition 6 *Consider the case of an incomplete climate contract. During both Phase 1 and Phase 2 the stable climate coalition has maximally three members. Tab.* 3 *and* 4 *state the possible stable coalitions, which depend on the size of* $\delta a\Delta$.

Similar to the complete contract case, the coalition in Phase 1 has maximally three members. At period period $\tau \in \{T_I - 1, ..., T_{II} - 2\}$, a coalition of countries which face a binding capacity constraint is only possible if $\delta a\Delta$ is low, while a large value of $\delta a\Delta$ implies that the coalition is formed by countries without a binding capacity constraint.

To understand the role of $\delta a\Delta$ during Phase 2, compare the reduction of fossil fuel use of a country without a binding black capacity constraint with the reduction of a country facing a binding constraint when joining a climate coalition. The difference can be written as $dx_{i,\tau} - dB_{i,\tau} = C(m-1)\left(\frac{1}{a} + \frac{\delta}{K} - \Delta\right) > 0$. Differentiating with respect to L, a, δ and K shows that the difference is the larger the larger L and the smaller a. While the difference increases in K, a small K ensures that the difference is the larger the smaller δ. Thus, if $\delta a\Delta$ is small, a country belonging to group N_{II} abates considerably more CO_2 emissions than a country of group N_I when joining a climate coalition. However, this large abatement implies strong free-riding incentives for other group N_{II} countries, so that the coalition $m^*_{II} = 3$ is not stable. Because the free-riding incentives are weaker the larger $\delta a\Delta$, a coalition of countries of group N_{II} is stable if $\delta a\Delta$ is sufficiently large.

5 Transfers

Transfers are well-established in the literature as a means to enhance the stable coalition size.[28] Because I consider asymmetric countries, I analyze in this sect. whether transfers within the coalition can increase the size of the stable coalition. For this purpose, I follow Caparrós and Péreau (2017), Caparrós and Finus (2020) and Fuentes-Albero and Rubio (2010) and check whether the cooperation surplus

[28] See Finus (2003) for an overview of related literature.

Table 1 Stable coalitions in Phase 1 under a complete contract

$\delta a\Delta \in$	$\tau \in \{t, ..., T_I - 2\}$	$\tau = T_I - 1$
$\left(\frac{3}{2}\frac{a\delta^2}{K} + \frac{1}{2}, \infty\right)$	$m^* = 3$	$m^*_{II} = 3$
$\left(\frac{a\delta^2}{K}, \frac{3}{2}\frac{a\delta^2}{K} + \frac{1}{2}\right]$	$m^* = 3$	$m^*_I = 3$ and $m^*_{II} = 3$
$\frac{a\delta^2}{K}$	$m^* = 3$	$m^* = 3$
$\left[\frac{2}{3}\frac{a\delta^2}{K} - \frac{1}{3}, \frac{a\delta^2}{K}\right)$	$m^* = 3$	$m^*_I = 3$ and $m^*_{II} = 3$
$\left(0, \frac{2}{3}\frac{a\delta^2}{K} - \frac{1}{3}\right)$	$m^* = 3$	$m^*_I = 3$

Table 2 Stable coalitions in Phase 2 under a complete contract

$\delta a\Delta \in$	$\tau \in \{T_I, ..., T_{II} - 2\}$	$\tau = T_{II} - 1$
$\left[\max\{2, 2\beta\}, \infty\right)$	$m^*_I = 3$ and $m^*_{II} = 3$	$m^*_I = 3$ and $m^*_{II} = 3$
$\left[2, 2\beta\right)$	$m^*_I = 3$	$m^*_I = 3$ and $m^*_{II} = 3$
$\left[2\beta, 2\right)$	$m^*_I = 3$ and $m^*_{II} = 3$	$m^*_I = 3$
$\left(0, \min\{2, 2\beta\}\right)$	$m^*_I = 3$	$m^*_I = 3$

Table 3 Stable coalitions in Phase 1 under an incomplete contract

$\delta a\Delta \in$	$\tau \in \{t, ..., T_I - 2\}$	$\tau = T_I - 1$
$\left(\frac{1}{2}, \infty\right)$	$m^* = 3$	$m^*_{II} = 3$
$\left(0, \frac{1}{2}\right]$	$m^* = 3$	$m^*_I = 3$ and $m^*_{II} = 3$

Table 4 Stable coalitions in Phase 2 under an incomplete contract

$\delta a\Delta \in$	$\tau \in \{T_I, ..., T_{II} - 2\}$	$\tau = T_{II} - 1$
$\left[2, \infty\right)$	$m^*_{II} = 3$	$m^*_I = 3$ and $m^*_{II} = 3$
$\left(\frac{3}{2}, 2\right)$	$m^*_{II} = 3$	$m^*_I = 3$
$\left(1, \frac{3}{2}\right]$	$m^*_I = 3$ and $m^*_{II} = 3$	$m^*_I = 3$
1	$m^* = 3$	$m^*_I = 3$
$\left[\frac{2}{3}, 1\right)$	$m^*_I = 3$ and $m^*_{II} = 3$	$m^*_I = 3$
$\left(0, \frac{2}{3}\right)$	$m^*_I = 3$	$m^*_I = 3$

$$S(m) = \sum_{j \in M} v^c_{j,t}(m) - \sum_{j \in M} v^f_{j,t}(m - 1) \tag{55}$$

is positive. The idea is that country i will only participate in a coalition if it receives at least the payoff $v^f_{i,t}(m-1)$ it would get as a fringe country. Obviously, a coalition is only stable if its aggregated welfare $\sum_{j \in M} v^c_{j,t}(m)$ is sufficiently large to grant each member its fringe payoff. In Online-Appendix A.10, I prove

Proposition 7 *If the capacity constraint is binding initially, transfers within the coalition are not able to stabilize a coalition with more than three members.*

The main asymmetry in my model are the different satiation points $\bar{y}_i$. While this asymmetry affects the countries' strategies (cf. (10–15) and (16–18)), it enters a signatory's value function only via a country specific constant (cf. e.g. (9)). Due to the assumed functional forms, all terms which interrelate the asymmetry with the coalition size cancel out.[29] Therefore, no coalition country is willing to pay a transfer to other countries to enlarge the coalition.[30]

In case of an initially non-binding black capacity constraint, an additional asymmetry arises due to different switching points T_i. However, Online-Appendix A.11 proves

Proposition 8 *If the black capacity constraint is non-binding initially, transfers within the coalition are not able to stabilize a coalition with more than three members.*

Similar to the case with an initially binding capacity constraint, transfers are not able to stabilize larger coalitions with an initially non-binding black capacity constraint. To further understand Proposition 8, consider $T_I < T_{II}$, a complete contract, period $T_{II} - 1$ and suppose that every signatory $i \in M_I$ pays a transfer μ to every signatory $i \in M_{II}$ to enlarge the coalition. The internal stability conditions of country $i \in M_I, M_{II}$ read

$$(m-3)(m-1) - \frac{2m_{II}}{\delta C^2 \Delta}\left(\frac{C^2}{a} - \mu\right) \leq 0, \quad \text{for } i \in M_I, \tag{56}$$

$$(m-3)(m-1) + \frac{2m_I}{\delta C^2 \Delta + \frac{C^2}{a}}\left(\frac{C^2}{a} - \mu\right) \leq 0, \quad \text{for } i \in M_{II}. \tag{57}$$

Without transfers, no country $i \in N_{II}$ will participate in a coalition with more than three members. If the country receives a transfer of $\mu > \frac{C^2}{a}$, its internal stability condition is relaxed, so that it may join. However, a transfer of $\mu > \frac{C^2}{a}$ implies that the internal stability condition for signatories $i \in M_I$ is violated for all $m > 3$. In other words, the transfer necessary to prompt a country of group N_{II} to join a coalition with more than three members will induce the signatories of group N_I to leave the coalition.

6 Conclusion

This paper analyzes the stability of international environmental agreements (IEA) or climate coalitions, respectively, in a dynamic game where the production of both renewables and fossil fuel based energy requires the accumulation of energy generation capacities. The results show that only small climate coalitions are stable, no matter whether the contract coordinates both CO_2 emissions and renewable energy investments (complete contract) or only CO_2 emissions (incomplete contract). This contrasts with the findings of Battaglini and Harstad (2016), who endorse incomplete contracts to increase the size of stable climate coalitions.

[29] A similar remark holds with respect to fringe countries.

[30] Clearly, other functional forms, such as asymmetric cost functions, may render transfers beneficial with respect to the coalition size. However, these considerations go beyond the scope of this paper and are, therefore, left for further research.

To understand the disparity, note that Battaglini and Harstad (2016) implicitly assume an unlimited black capacity. In the last period of an incomplete climate contract, the coalition members realize that high green capacity investments will only weaken their position in the negotiations of a new climate contract and will leave their investments at low levels. Because all countries want to postpone this underinvestment, the hold-up problem stabilizes larger climate coalition. The deviation of countries is prevented by the credible threat of the remaining signatories to sign a short-term agreement.

In case of a limited black capacity, fossil fuel use in one period is determined by the black capacity investments made in the previous period. Because the coalition sets its policy directly after the climate contract was signed, it uses its first-mover advantage such that the signatories set higher green capacity investments. In particular, this is true for the last period of the climate contract. Therefore, the signatories' green capacity investments are not reduced in the last contract period, so that the hold-up problem does not exist. Consequently, there is also no credible threat which could stabilize a large coalition.

Interestingly, this result also holds if the black capacity is limited but temporarily not completely used. In this case, the hold-up problem emerges if the climate contract expires at a period with a non-binding black capacity constraint. While the hold-up problem could be postponed for some time, the countries anticipate that this is only possible as long as the capacity constraint is non-binding. Consequently, there is no credible threat to sign a short term agreement in the last period with a non-binding constraint, so that only a small climate coalition is stable. By using the standard backward-induction argument, there cannot be a large stable coalition in the prior period or in any other previous period. Neither the introduction of asymmetry with respect to the point in time the countries face a binding black capacity constraint nor of transfers within the coalition leads to more optimistic results.

To ensure the analytical tractability of the model, I adopted the linear-quadratic framework with Markovian equilibria of Battaglini and Harstad (2016). Different solution concepts and functional forms my affect the results. In particular, this seems to be true for transfers. If countries differ with respect to climate damages or investments costs, transfers may very well stabilize larger coalitions.[31] However, these cases are left for further research, as are other ignored factors, such as renegotiation and trade.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10640-022-00668-z.

Funding Open Access funding enabled and organized by Projekt DEAL.

[31] Cf. Fuentes-Albero and Rubio (2010).

References

Al Khourdajie A, Finus M (2020) Measures to enhance the effectiveness of international climate agreements: the case of border carbon adjustments. Eur Econ Rev 124:103405

Andruleit H, Babies HG, Bahr A, Kus J, Meßner J, Schauer M (2012) Energiestudie 2012: Reserven, Ressourcen und Verfügbarkeit. DERA Rohstoffinformationen 15

Asheim GB, Froyn CB, Hovi J, Menz FC (2006) Regional versus global cooperation for climate control. J Environ Econ Manag 51(1):93–109

Asheim GB, Holtsmark B (2009) Renegotiation-proof climate agreements with full participation: Conditions for Pareto-efficiency. Environ Resource Econ 43(4):519–533

Barrett S (1994) Self-enforcing international environmental agreements. Oxf Econ Pap 46:878–894

Barrett S (1999) A theory of full international cooperation. J Theor Polit 11(4):519–541

Barrett S (2005) The theory of international environmental agreements. Handb Environ Econ 3:1457–1516

Barrett S (2006) Climate treaties and “breakthrough'' technologies. Am Econ Rev 96(2):22–25

Battaglini M, Harstad B (2016) Participation and duration of environmental agreements. J Polit Econ 124(1):160–204

Birol F et al (2019) World energy outlook 2019. International Energy Agency, Paris

Breton M, Sbragia L (2019) The impact of adaptation on the stability of international environmental agreements. Environ Resource Econ 74(2):697–725

Breton M, Sbragia L, Zaccour G (2010) A dynamic model for international environmental agreements. Environ Resource Econ 45(1):25–48

Cairns RD (1998) The microeconomics of mineral extraction under capacity constraints. Nonrenewable Res 7(3):233–244

Cairns RD, Lasserre P (1991) The role of investment in multiple-deposit extraction: Some results and remaining puzzles. J Environ Econ Manag 21(1):52–66

Calvo E, Rubio SJ (2012) Dynamic models of international environmental agreements: a differential game approach. Int Rev Environ Resour Econ 6(4):289–339

Campbell HF (1980) The effect of capital intensity on the optimal rate of extraction of a mineral deposit. Canadian J Econ 1:349–356

Caparrós A, Finus M (2020) Public good agreements under the weakest-link technology. J Public Econ Theory 22(3):555–582

Caparrós A, Péreau J-C (2017) Multilateral versus sequential negotiations over climate change. Oxf Econ Pap 69(2):365–387

Carraro C, Eyckmans J, Finus M (2006) Optimal transfers and participation decisions in international environmental agreements. Rev Int Organ 1(4):379–396

Carraro C, Siniscalco D (1993) Strategies for the international protection of the environment. J Public Econ 52(3):309–328

d'Aspremont C, Jacquemin A, Gabszewicz JJ, Weymark JA (1983) On the stability of collusive price leadership. Can J Econ 16:17–25

Diamantoudi E, Sartzetakis ES (2018) International environmental agreements - the role of foresight. Environ Resource Econ 71(1):241–257

Dixit A, Olson M (2000) Does voluntary participation undermine the Coase Theorem? J Public Econ 76(3):309–335

Finus M (2003) Stability and design of international environmental agreements: the case of transboundary pollution. International yearbook of environmental and resource economics 4 (4)

Fuentes-Albero C, Rubio SJ (2010) Can international environmental cooperation be bought? Eur J Oper Res 202(1):255–264

Golosov M, Hassler J, Krusell P, Tsyvinski A (2014) Optimal taxes on fossil fuel in general equilibrium. Econometrica 82(1):41–88

Harstad B (2012) Buy Coal! A Case for Supply-Side Environmental Policy. J Polit Econ 120(1):77–115

Hoel M (1992) International environment conventions: the case of uniform reductions of emissions. Environ Resource Econ 2(2):141–159

Hoel M, de Zeeuw A (2010) Can a focus on breakthrough technologies improve the performance of international environmental agreements? Environ Resource Econ 47(3):395–406

Holland SP (2003) Extraction capacity and the optimal order of extraction. J Environ Econ Manag 45(3):569–588

IPCC (2014) Climate change 2014: synthesis report. Online at: https://www.ipcc.ch/report/ar5/syr/

Kollenbach G (2017) Endogenous growth with a limited fossil fuel extraction capacity. Can J Econ 50(1):233–272

Kornek U, Edenhofer O (2020) The strategic dimension of financing global public goods. Eur Econ Rev 127:103423
Kováč E, Schmidt RC (2021) A simple dynamic climate cooperation model. J Public Econ 194:104329
Maskin E, Tirole J (2001) Markov perfect equilibrium. J Econ Theory 100(2):191–219
Mason CF, Polasky S, Tarui N (2017) Cooperation on climate-change mitigation. Eur Econ Rev 99:43–55
Masoudi N, Zaccour G (2018) Adaptation and international environmental agreements. Environ Resource Econ 71(1):1–21
McGinty M (2007) International environmental agreements among asymmetric nations. Oxf Econ Pap 59(1):45–62
Rubio SJ (2017) Sharing R&D investments in breakthrough technologies to control climate change. Oxf Econ Pap 69(2):496–521
Rubio SJ, Casino B (2005) Self-enforcing international environmental agreements with a stock pollutant. Span Econ Rev 7(2):89–109
Rubio SJ, Ulph A (2007) An infinite-horizon model of dynamic membership of international environmental agreements. J Environ Econ Manag 54(3):296–310
Toth FL (2014) Nuclear energy and sustainable development. Energy Policy 74:1–4