Self-enforcing climate coalitions for farsighted countries: integrated analysis of heterogeneous countries[*†]

Sareh Vosooghi[‡] Maria Arvaniti[§] Frederick van der Ploeg[¶]

February 2022

Abstract

This paper studies the formation of international climate coalitions by heterogeneous countries. Countries rationally predict the consequences of their membership decisions in climate negotiations. We offer an approach to characterise the equilibrium number of coalitions and their number of signatories independent of their heterogeneity, and we suggest a tractable algorithm to fully characterise the equilibrium. In a dynamic game analysis of a general equilibrium model of the economy integrated with climate dynamics, a grand climate coalition or multiple climate coalitions may form in equilibrium, but if the policymakers are patient, the number of signatories in all climate treaties is a Tribonacci number. Our results are robust to the possibility of renegotiation and investment in green technologies besides fossil fuels.

Key words: climate economics; international environmental agreements; coalition formation; heterogeneous countries; integrated assessment models

JEL Classification: Q54; D70; D50

[*]We are grateful to Bard Harstad, Tim Worrall, Margaret Meyer, Marek Pycia, Piotr Dworczak, Michal Kobielarz, Paula Onuchic, Ludvig Sinander, Aart de Zeeuw, Simon Cowan and conference, seminar and workshop participants at the European Winter Meeting of the Econometric Society, Oxford, Transatlantic Theory Workshop and Nuffield Economic Theory Workshop for their helpful comments and conversations.

[†]For the latest version of the paper, please check here.

[‡]sareh.vosooghi@economics.ox.ac.uk, University of Oxford, Department of Economics, Manor Road, Oxford, United Kingdom, OX1 3UQ

[§]maria.arvaniti@unibo.it, University of Bologna, Department of Economics, Italy

[¶]rick.vanderploeg@economics.ox.ac.uk, University of Oxford, Department of Economics, Manor Road, Oxford, United Kingdom, OX1 3UQ. Also affiliated with University of Amsterdam, CEPR, and CESifo.

1 Introduction

The biggest planetary tragedy has been the failure of countries to work together to curb anthropogenic greenhouse gases (GHG) emissions and combat global warming effectively. Unfortunately, after three decades of climate negotiations, there is still no effective and self-enforcing internationally cooperative climate policy with a real chance of being implemented. Still, the only way to meet the targets agreed upon in the Paris COP21 Agreement is for countries to cooperate and implement an ambitious climate policy. Unfortunately, there is an enormous gap between the current pledges agreed at the Glasgow COP26 Agreement and the targets of the Paris Agreement. In the international arena there has been too much focus on the formation of the unlikely grand climate coalition of all countries concerned. Instead, it might be more worthwhile to search under the umbrella of these international agreements for multiple climate coalitions among smaller group of countries that are stable and overall more ambitious than what we observe today.

In this spirit, we model individual countries' decision-making in joining climate treaties and suggest a more pragmatic approach. We develop an integrated assessment model (IAM) that considers the interactions of different aspects of climate problems: the ecosystem, the asymmetric countries and their long-run incentives, and the possibility of cooperation on climate agreements among countries. Signatories of climate coalitions commit to choose climate policies jointly such that they maximise the benefits of their block. There is a consensus in the literature that the larger the number of signatories of a climate coalition, the more ambitious their climate policy is. The problem is, however, that not all large coalitions are self-enforceable. We therefore address three questions regarding self-fulfilling international climate treaties. First, how do we model the problem of coalition formation among heterogeneous countries? Second, do multiple climate coalitions of potentially different sizes form in equilibrium? Third, how many signatories commit to each of these climate treaties?

Our main contribution is the analysis of coalition formation by heterogeneous countries. For any number of heterogeneous countries we characterise the equilibrium number of coalitions and their number of signatories. Furthermore, in a complex environment of an IAM with farsighted countries, we show that in equilibrium, the number of signatories to a climate treaty (corresponding to the grand coalition or a smaller coalition) is always a Tribonacci number formed from a sequence of numbers where each element is the sum of the preceding three elements. Tribonacci numbers belong to the family of Fibonacci sequences, and although they have not been used in the economics literature before, they occur in the natural world whenever an efficient way of packing elements together is called for.[1]

[1]For example, the number of petals of flowers, bracts of pinecones, trees branching tend to be from a Fibonacci sequence (Campbell, 2020; Minarova, 2014; and Sinha, 2017). Tribonacci numbers were first

The literature on coalition formation with farsighted countries has focused only on the case of symmetric countries. In that case, equilibria of coalitions need to be characterised only in terms of the number of coalitions and the number of signatories, since the identity of symmetric countries is indeterminate in equilibrium. However, in the more realistic case of heterogeneous countries, there can be multiple coalitions which have the same number of signatories. Our analysis allows for heterogeneity across countries with respect to the initial stocks of capital, total factor productivities, the initial levels of fossil fuel reserves (and the associated scarcity rents) and the cost of investment in green technologies. With asymmetric countries it is not possible to directly use the conventional coalition formation methodologies as these have been developed for symmetric countries. However, we show that the problem can be *decoupled* as follows: first we find the equilibrium number of signatories, and then we check the allocation of countries across climate coalitions.

More specifically, we study four sources of heterogeneity and in each case, our decoupling result holds for a different reason. In particular, since in the setting of our IAM heterogeneity with respect to capital stock or total factor productivity affect the countries' payoffs linearly, it turns out that the equilibrium number of coalitions and their number of signatories can be characterised independent of the heterogeneity. Heterogeneity still affect the countries equilibrium payoffs, and indeed after characterising the equilibrium number of coalitions and number of their signatories, we check whether the countries choose an equilibrium coalitional composition which reduces the global emission externality (i.e. improves constrained efficiency). Importantly, our decoupling approach can be used in any setting where the reduced-form payoffs of countries (for which we offer micro foundations) is affected by heterogeneity linearly.

Most of the literature on international climate coalition formation abstracts from the details of macroeconomic outcomes and their underlying determinants, and works with very stylised models without micro foundations and restrictive assumptions. Our main objective is to capture broader incentives for the policymakers in climate negotiations. Climate economists have developed *global* and *multi-country* IAMs. These are typically macroeconomic growth models which allow for the effects of the economy on global warming and vice versa, and which can analyse a wide range of climate policies. Even though such IAMs have been used to analyse and contrast fully internationally cooperative and fully non-cooperative outcomes, these IAMs have not often be used to analyse the strategic interactions of countries seeking climate agreements.

The two strands of literature have developed almost independently. We build a bridge between the literature on climate coalition formation and the literature on optimal climate policies from IAMs. However, there are few macroeconomic growth models and IAMs with analytical solutions, and our analysis needs to accommodate both climate

found by Charles Darwin in the *Origin of Species*.

dynamics and international coalition formation. Golosov et al. (2014) have put forward a tractable IAM, which adds simple climate dynamics to Brock and Mirman's (1973) tractable model of economic growth. They find a closed-form solution for the optimal social cost of carbon (SCC) which is proportional to current output and is independent of future values of output or consumption. They show numerically that their characterisation replicates the properties of general IAMs such as the Dynamic Integrated Climate-Economy (DICE) model of Nordhaus (1993) reasonably well. Hassler and Krusell (2012) extend the closed economy IAM of Golosov et al. (2014) to multiple countries, which is able to match the outcomes of the Regional Integrated Climate-Economy (RICE) model of Nordhaus and Yang (1996) also relatively well. Here, in capturing the countries' incentives in climate negotiations, we use payoff specifications from a multi-country model in the spirit of Hassler and Krusell (2012). We follow van der Ploeg and Rezai (2021) and modify, however, the climate dynamics and temperature structure based on recent advances in climate science, and we integrate it with an analysis of the decision of individual countries to participate in international climate agreements.

For the strategic side of international coalition formation, we assume that the planner of each country is farsighted. We thus allow negotiating countries to rationally consider all self-enforceable unilateral and multilateral deviations from their membership decisions, and consequently predict the entire structure of conceivable coalitions. This is in sharp contrast to the most commonly used solution concept of cartel stability, which corresponds to a Nash equilibrium, and makes the myopic assumption that countries in coalition formation are only concerned about the immediate gain or losses of their unilateral deviations, ignoring the reactions of other countries. Ignoring the possibility of retaliation or generally optimal reactions by other countries after breaking off climate negotiations increases incentives for free riding. Thus, as is well known, the use of the cartel stability solution concept results in the formation of very small coalitions. Since it is unrealistic to assume that countries are myopic, we thus investigate how robust the finding of small coalitions is when countries are farsighted.

To find the equilibrium number of signatories, the immunity of the equilibrium to deviations must be checked. Farsightedness implies that deviations from equilibrium strategies which are not themselves farsighted must be excluded. Hence, in conventional studies using the concept of farsightedness, the characterisation of the equilibrium structure of coalitions relies on algorithms which recursively identify the set of the total number of countries for which the equilibrium is the formation of a grand coalition. The recursion in such algorithms starts from the smallest total number of countries that can form a self-fulfilling coalition and continues to a finite number N. This set determines the possible farsighted deviations. Furthermore, the number of members of equilibrium coalition(s) is a subset of this set. In this strand of literature, the countries have a one-off payoff. This implies that the comparison of payoffs and the characterisation in each step

of the algorithm is not too demanding. In an infinite-horizon IAM, the recursion process can be onerous and typically requires one to resort to numerical simulations. However, due to the special structure of our IAM based on the assumptions made in Golosov et al. (2014), we are able to obtain intuitive and analytical results.

We show that for our IAM, the set of total number of countries for which the equilibrium is the formation of a grand coalition is the set of Tribonacci numbers (i.e. 1, 1, 2, 4, 7, 13, 24, 44, 81, 149, ...). This is a known set, so that there is no need to check the payoffs of the countries recursively. Thus we suggest a tractable and intuitive algorithm to characterise analytically the equilibrium number of climate coalitions and their number of signatories in an IAM. Importantly, this algorithm does not require any recursions.

Furthermore, we show that the equilibrium number of signatories of any coalition, either grand or non-grand, must be a Tribonacci number. An interesting property of these numbers is that the elements of this sequence increase rapidly. Thus, depending on the total number of countries, the result of Tribonacci numbers of signatories implies that equilibrium climate coalitions can be large. This is due to our more realistic assumption of farsightedness of countries, which reduces their incentive to free ride.

Fixing the equilibrium number of signatories, we show that in all equilibrium coalition structures the heterogeneous participating countries in climate negotiations prefer coalitions which are more efficient in terms of emission mitigation.

We show that our results generalise to situations where countries can walk away from agreed climate treaties and renegotiate them in future. Furthermore, our results are robust to cases where the energy sector of countries includes investment in a green technology as a perfect substitute to fossil fuels (such as solar).

Given the analytical characterisation of climate coalitions, we can back out the macroeconomic policies, global temperature, growth rate, energy consumption, and the optimal SCCs for the various countries associated with self-enforceable climate treaties. This analysis takes account of interactions between anthropogenic emissions, the ecosystem and the incentives of heterogeneous countries, and enriches the usual economic approaches that have been used for this purpose in the literature on international climate coalition formation.

The remainder is organised as follows. Related literature is reviewed in the next section. Our multi-country IAM with climate coalition formation is presented in section 3. We analyse how heterogeneous countries arrive at climate treaty memberships within the context of our IAM in Section 4. Section 5 generalises our results to the case of reversible agreements, where the countries can renegotiate any existing agreement. In section 6, we present an extension of our model where the energy sector includes both fossil fuels and green technologies. Section 7 concludes. All proofs are provided in the Appendix.

2 Related Literature

In the theoretical literature on environmental economics, there are two main approaches to analyse climate problems: International Environmental Agreements (IEAs) or climate governance, and macroeconomic analysis of climate policy using IAMs. Our paper bridges these two strands of literature. Research on IEAs and international cooperation by forming climate coalitions has led to an extensive literature. It has provided some valuable inputs into the design of international climate treaties, including the Paris Climate Accords. Seminal papers include Carraro and Siniscalco (1993) and Barrett (1994). Benchekroun and Long (2012) and Battaglini and Harstad (2016) review the literature on IEAs.

Most of this literature employs the solution concept of cartel stability which requires internal and external stability. The former requires that no country inside the coalition has an incentive to leave the coalition and the latter means that no country outside the coalition has an incentive to join the coalition. Cartel stability implies that only unilateral deviations are checked while taking the membership decision of the complementary set of players as given. This assumption leads to the result of the formation of small coalitions (of maximum size three). This result is known as the *small coalition paradox* and is remarkably robust.[2]

To achieve the formation of larger coalitions using the concept of cartel stability, some *remedies* have been suggested. These include international transfers (Carraro and Siniscalco, 1993; and Hoel and Schneider, 1997; Carraro et al., 2006); the adoptation of a "breakthrough" green technology that exhibits increasing returns in a critical number of countries (Barrett, 2006); "modest" agreements (Finus and Maus, 2008); use of a refunding club where signatories of the treaty pay an upfront fee which is invested and the return on the fund is redistributed according to how successful countries have been in reducing emissions (Gersbach et al. 2021); gaining from the trade-off between R&D costs and the costs of adopting the breakthrough technology (Hoel and de Zeeuw, 2010); markets for fuel and tradable rights to extract fossil fuel in other countries (Harstad, 2012); non-quadratic functional forms (Karp and Simon, 2013); linkage to trade clubs (Nordhaus, 2015); effects of incomplete contracts of the green technologies on emission coalitions (Battaglini and Harstad, 2016).

The d'Aspremont et al. (1983) notion of cartel stability focuses on the formation of one single coalition beside the fringe. In the theory of IEAs and practice, there has been much emphasis on forming a single coalition of countries. Despite being widely quoted and used, this result of insisting on only one climate coalition is unnecessarily restrictive both from a theoretical and a policy perspective. Our paper contributes to the literature which allows the formation of multiple coalitions. In the literature on

[2]See Battaglini and Harstad (2016) for the literature on the robustness of this result.

IEAs with Nash equilibrium and open-membership this was first suggested by Yi and Shin (2000). Asheim et al. (2006), Finus and Rundshagen (2003 and 2009), Finus et al. (2009) also relax the assumption of a single coalition and allow for multiple coalitions under the cartel stability.

As mentioned earlier, we use a different solution concept and assume that all the countries are farsighted. The internal and external stability conditions are necessary, but not sufficient for farsightedness. The early literature on coalition formation and farsightedness is due to Aumann and Myerson (1988), Dutta et al. (1989), Chwe (1994), Bloch (1996), and Ray and Vohra (1997), Ray and Vohra (1999), Chatterjee et al. (1993) among others. This literature on coalition formation abstracts from any externalities across the coalitions such as the global warming externality we are concerned with. However, Ray and Vohra (2001) generalise the farsighted coalition formation of Ray and Vohra (1999) to the case of public goods. Analysing IEAs, Vosooghi (2017) uses the assumption of farsighted stability in a stochastic setting while Diamantoudi and Sartezetakis (2018) and de Zeeuw (2008) analyse it in deterministic settings. De Zeeuw (2008) studies the effect of a gradual adjustment of emission reduction in a simplified IEA, and shows numerically that the stable number of signatories under farsightedness depends on the relative cost of emission adjustment and climate damages.[3]

We examine a *dynamic* game extension of Ray and Vohra (2001) and do this within the context of integrated assessment models of the economy and the climate. Our analysis allows for heterogeneous countries and reversible agreements, both of which the above studies abstract from.

Our paper also relates to the literature on IAMs, which relative to the models used in the literature on IEAs, are more general and have a different focus. IAMs use macroeconomic growth models with a combination of economic and geophysical assumptions in order to understand the interactions between anthropogenic greenhouse gas (GHG) emissions and the ecosystem. While abstracting from international climate agreements, these models try to capture the global economy and some of them include different economic regions too. These IAMs are often very large, detailed and too complicated to be solved analytically, so that numerical methods are used to analyse them. These IAMs address a wide range of analyses of climate policy. The main ones are the DICE model developed by Nordhaus (2014), the FUND model with effects of uncertainties and different climate regions put forward by Anthoff and Tol (2013), and the PAGE model with regional temperatures leading to global average temperature developed by Hope (2011). Analytical expressions for the optimal SCC and climate policies have been obtained from IAMs by Golosov et al. (2014), Hassler and Krusell (2012) and van der Ploeg and Rezai

[3]There has been much work on farsighted sets recently, e.g. Ray and Vohra (2019) and Dutta and Vohra (2017). However, since these solution concepts use the cooperative approach and rely on the characteristic function, they do not accommodate study of externalities.

(2021).[4]

Only a small subset of the literature combines the two fields of literature on IEAs and IAMs. An early paper is Tol (2001), which considers coalition formation among climate regions of an IAM. Eyckmans and Tulkens (2006), Yang (2008), Buchner and Carraro (2009) have put forward similar models, but in contrast to our approach to modelling coalition formation, these papers use a cooperative game-theoretic approach. Due to the external effects of emissions on climate change, the use of cooperative game theory in modelling such games and climate treaties has been criticised.[5]

Some authors have combined these two strands using a non-cooperative game-theoretic approach to coalition formations. They add a coalition-formation stage to numerical IAMs such as RICE (a multi-country version of DICE), STAC-3, CWS, or WITCH, and then use numerical simulations to examine the stability of an international climate coalition. Specifically, Lessmann et al. (2009) examine the effect of trade sanctions and tariffs on the size of stable coalitions; Bosetti et al. (2013) use the IAM referred to as WITCH and show that an ambitious grand climate coalition is not internally stable; and finally Lessmann et al. (2015) investigate the effect of international transfers on the coalition sizes in five different types of IAMs. In modelling the details of coalition formation, all these papers insist on cartel stability, which implies that in the absence of any remedies, their analysis results also in *small* climate coalitions. Furthermore, their analysis is entirely based on numerical results. Relative to these studies, our paper generalises the stability concept to allow for farsightedness. Here the size of stable climate coalitions can be large without relying on any of the above mentioned remedies. We offer a full characterisation of the equilibrium number of coalitions and signatories for any number of countries or regions.

3 The Model

Our IAM framework is an adaptation of the multi-country version of Golosov et al. (2014) where we have modified its climate dynamic modelling to take account of recent atmospheric science insights as in van der Ploeg and Rezai (2021) and Dietz et al. (2021). There are N countries; each country is indicated by the subscript $i \in I$, where $I \equiv \{1, 2, ..., N\}$. Furthermore, time is discrete and infinite, indexed by $t = 0, 1, 2, ...$ In climate negotiations, each country is represented by a planner, who can implement any desired policy in a competitive market economy.

[4]Van den Bremer and van der Ploeg (2021) use perturbation theory to obtain a tractable expression for the optimal risk-adjusted SCC in a macroeconomic growth model with a wide range of economic and climatic uncertainties.

[5]For more discussions see Rosenthal, (1971) and Ray and Vohra (2001).

3.1 The Economy

In each country, there is a representative household with lifetime utility from consumption of a final good, C_{it}, given by

$$\sum_{\tau=0}^{\infty} \beta^{\tau} U(C_{it+\tau}) \tag{3.1}$$

where $\beta \in (0,1)$ is a constant discount factor and her instantaneous utility function is given by $U(C_{it}) = \ln(C_{it})$. Thus, we assume that the intertemporal elasticity of substitution is constant and equal to one. Golosov et al. (2014) argue this is a reasonable assumption in long-run economic growth models. Barrage (2014) explores the sensitivity of the optimal SCC in the IAM of Golosov et al. (2014) to elasticities of intertemporal substitution different from one and shows numerically that their results are robust.[6]

The production process in each country i has two sectors: an energy sector, E_{it}, and the final goods sector, Y_{it}. Energy is produced using fossil fuels. We assume that the (marginal) cost of generating fossil fuel is zero, so that its production is constrained only by the given finite stock of the fossil fuel resource in each country, i.e.

$$E_{it} = R_{it} - R_{it+1} \tag{3.2}$$

where R_{it} is the stock of reserves of fossil fuel of country i at the beginning of period t. A finite stock of fossil fuel is particularly relevant for oil and gas resources. Thus using equation (3.2), we have

$$R_{it+1} = R_{i0} - \sum_{s=0}^{t} E_{it-s} \tag{3.3}$$

where R_{i0} is the exogenous stock of reserves of fossil fuel of country i in $t = 0$, and which can differ across countries.

Production of the aggregate output of final goods uses capital and energy which are endogenously determined. Following the DICE model of Nordhaus (1993) and the RICE model of Nordhaus and Yang (1996), global temperature negatively affects the aggregate production of final output. We let global warming damages be proportional to aggregate output. Golosov et al. (2014) show that an exponential functional form for damages related to the stock of atmospheric carbon approximates the ratio of global warming damages to aggregate output of the DICE and RICE models reasonably well. Production follows a Cobb-Douglas technology so that the aggregate output of country i at time t is

[6] Among the studies which support the use of logarithmic utility in macro models, Chetty (2006) suggests a method of estimating the coefficient of relative risk aversion and shows that the mean estimate is bounded and equals about one. Furthermore, Gandelman and Rubén Hernández-Murillo (2015) using a mega database of 75 countries show this coefficient varies closely around one, which corresponds to a logarithmic utility function.

$$Y_{it} = \exp(-\gamma T_t) A_i K_{it}^{1-\nu} E_{it}^{\nu} \tag{3.4}$$

where K_{it} is the aggregate capital stock at the beginning of period t, that is used in the production of final output; ν is the output elasticity of energy; and E_{it} is the energy use in the production of the final good. The current capital stock, K_{it}, and the initial capital stock, K_{i0}, can vary across countries.[7] Total factor productivity (TFP) has two multiplicative terms, a constant, A_i, which can vary across countries, and a negative exponential function of global temperature, T_t, where γ is the damage coefficient.[8] Jones (2005) provides a micro-founded justification for the assumption that the aggregate production function is Cobb-Douglas at the macroeconomic level if the parameters of the production technology are drawn from a Pareto distribution.[9] Moreover, Hassler et al. (2021) using historical data to calibrate an IAM, estimate an aggregate production function and show that the long-run input shares tend to be stationary which also suggests a Cobb-Douglas production technology. Finally, Miller (2008) surveys the literature on macroeconomic production functions and concludes that Cobb-Douglas production functions provide a good empirical fit across many data sets.

As a market-clearing condition, the fossil fuel depletion constraints (3.2) must be satisfied for each country. Furthermore, the feasibility constraint for the final good requires that aggregate consumption plus investment equals aggregate production in each country, so that

$$C_{it} + K_{it+1} = Y_{it} \tag{3.5}$$

Note that capital and energy are used only in the final goods sector. We assume zero adjustment cost of capital. In a decentralised economy the market for the final good clears at the national level. Hence, our IAM assumes that there is no international trade in fossil fuel. The only factor which links countries in our IAM is thus the externality resulting from global warming damages. We abstract from any other international interactions. To get tractable analytical solutions, we assume full capital depreciation in each period. Barrage (2014) shows that the characterisation of the optimal SCC in Golosov et al.

[7]The model can be interpreted as an AK growth model in the spirit of Romer (1986): by assuming $Y_{it} = \exp(-\gamma T_t) A_i K_{it}^{1-\alpha-\nu} E_{it}^{\nu} (\bar{K_{it}} L_{it})^{\alpha}$ where $\bar{K_{it}} L_{it}$ is effective labour input and $\bar{K_{it}}$ is the economy wide capital stock (e.g. human capital, research and development, infrastructure) which corresponds to the efficiency of labour. Since the efficiency of labour is proportional to the economy-wide capital stock, it is an endogenous AK growth. Without loss of generality we assume labour is supplied inelastically, and fixed at unity. In equilibrium, the economy wide capital is equal to the firm level capital, i.e. $\bar{K_{it}} = K_{it}$ and hence the expression in (3.4) results.

[8]The damage coefficient can be assumed to be an uncertain parameter. E.g., Golosov et al. (2014) replace it with the expectation of a fixed and common distribution of γ. We ignore that the damage coefficient can differ across countries (e.g., developing countries typically have higher damage coefficients than developed countries).

[9]Kortum (1997) shows within the context of a search-based model that a production technology only leads to steady-state growth if its parameters are drawn from Pareto distributions.

(2014) in the long run is numerically robust with respect to depreciation rates that are less than 100%.

3.2 Climate dynamics

For our explanation of temperature, we depart from Golosov et al. (2014). Based on recent insights in climate science (e.g., Allen et al. (2009) and Matthews et al. (2009)), we assume that temperature is a linear function of cumulative emissions of CO_2 (rather than relating temperature to the stock of atmospheric carbon).[10] Global temperature is thus given by

$$T_t = T_0 + \xi S_t \tag{3.6}$$

where S_t denotes the stock of cumulative emissions of CO_2; T_0 is pre-industrial temperature; and ξ is the transient climate response to cumulative emissions. The stock of cumulative emissions is the sum of past cumulative and current emissions, i.e.

$$S_t = S_{t-1} + E_t \tag{3.7}$$

where E_t is the flow of emissions produced by all countries at time t, i.e. $E_t \equiv \sum_{i=1}^{N} E_{it}$. Equation (3.7) is equivalent to

$$S_t = S_0 + \sum_{i=1}^{N} \sum_{s=0}^{t} E_{it-s} \tag{3.8}$$

where S_0 is the pre-industrial level of cumulative emissions (set to zero if temperature is measured from pre-industrial levels).

3.3 Climate Coalition Formation

We allow the countries to form climate coalitions to collectively reduce their emissions and cut their damages from global warming. At the beginning of period t, they have the choice of participating in climate negotiations. We focus on *no-delay* equilibria, thus we assume that if the negotiations do not come to a conclusion, all countries suffer an infinite loss. This assumption is also made in the IEA model of Ray and Vohra (2001). It ensures that the negotiations will lead to the formation of a coalition structure in period t. A coalition structure is a partition of the set of countries, I, into coalitions, $\mathbb{M} \equiv \{M_1, M_2, ..., M_k\}$. Let $m \leq N$ be a positive integer showing the cardinality (i.e., the number of members) of coalition M. A *numerical coalition structure*, $\mathcal{M} \equiv \{m_1, m_2, ..., m_k\}$, is a partition of

[10] To capture the carbon stock dynamics, one needs to distinguish a permanent component with no decay and at least one transient component of the stock of atmospheric carbon with a strictly positive rate of decay (e.g., due to absorption of CO_2 by the oceans). Cumulative emissions and stock of atmospheric carbon are mathematically only equivalent if there is no decay of atmospheric carbon.

N into the sizes of coalitions. If countries were identical, the identity of any particular country is indeterminate in equilibrium and characterising the equilibrium membership strategies thus involves only the sizes of the coalitions and the number of coalitions, i.e. the equilibrium numerical coalition structure. However, with heterogeneous countries both the identity of members and the equilibrium numerical coalition structure matter as there can be multiple coalitions with the same number of members.

We assume that formation of coalitions is costless and open, so that no country can be excluded from joining and no country can be forced to join. But joining a climate coalition requires signing a binding agreement with the other signatories of the coalition. This implies that upon signing an agreement, the signatories act cooperatively as a block in deciding on their common climate policy summarised by the SCC implemented by this coalition, for all $t \in \{0, 1, ..., \infty\}$ and all $i \in M$. Thus, implementing the decisions of a climate treaty is costless.

Assumption 1 . *Membership decisions are irreversible.*

In other words, countries do not have any chance of renegotiation.[11] After the *membership* stage in period t, all countries enter the compliance and *action* stage in that period, where the signatories set their climate policy as agreed at the membership stage. Then, each country $i \in M$, determines its equilibrium strategy for emissions, consumption, the next period capital stock (or saving) and resource extraction, i.e. $\{E_{it+\tau}(M, \mathbb{M}), C_{it+\tau}(M, \mathbb{M}), K_{it+\tau+1}(M, \mathbb{M}), R_{it+\tau+1}(M, \mathbb{M})\}_{\tau=0}^{\infty}$.

If at the beginning of period t a full coalition structure is already in place, the membership stage is skipped. At the end of each period, the countries observe emissions E_{it} of each country and payoffs for each country are realised.

The climate negotiation stage is modeled as a bargaining process with proposals and responses. In each sub-period of the membership stage (in period t), one country is chosen as the initial *proposer*. This captures that usually, climate negotiations take a couple of months. We assume that the length of time of sub-periods corresponding to the length of period t is fixed (e.g. 1/365), and that there is a cost of delay in sub-periods which is captured by the discount factor σ.[12]

The proposer makes a *proposal* to form a coalition to a group of *respondent* countries which are in the so-called negotiation room, i.e. to those who have not joined any other binding coalition yet.

The proposal consists of the identity of the members (thus of the size m too) and the optimal SCC of the coalition signatories, along with the corresponding emission plans and payoffs for the members of the treaty. We allow for any arbitrary split of payoffs, which in a climate game requires that we allow for transfers between the various countries in coalitions. This implies that we allow for transferable utilities. In principle, the

[11] We relax this assumption in section 5.

[12] The farsighted methodology that we use holds for $\sigma \to 1$ so that in the limit bargaining is frictionless. We assume $\sigma \neq 1$ to avoid multiplicity of equilibria.

proposal should be conditioned on the complementary formed coalition structure i.e the proposed emission plan could be conditioned on the emission plans of other coalitions in the coalition structure, $\mathbb{M}$. However, as will become clear in the next section, the coalitions have dominant strategies as both the marginal cost and benefit of the countries are proportional to the cumulative stock of emissions, when determining their optimal emissions. If $m = 1$, the proposer exits the negotiations as a singleton coalition,[13] and if $m > 1$, the proposer must at least include itself in the proposal.

After a proposal is made, it is the turn of the respondents. The strategy of the respondents is either to accept or reject the proposal. If the proposal is rejected by at least one country, no coalition forms in that sub-period. The next proposer may or may not include the initial proposer in its proposal.

The order by which the countries take action in the negotiation stage is determined by the *protocol*.

Definition 1 . *The protocol determines the rules of bargaining and the order of the initial proposers and all chosen respondents.*

The protocol is exogenous and is set at the very beginning of the negotiation stage. We use a special class of *rejector-friendly* protocols where the first rejector is the next proposer of a coalition M. Excluding countries in a public good game is never beneficial and the assumption of rejector-friendly protocol is in line with what is observed in climate negotiations. Furthermore, we assume that the order determined by the protocol is deterministic. Finally, we focus on protocols which require unanimity of members for a coalition to form. Hence, if a proposal is unanimously accepted, a binding coalition of size m forms and irreversibly leaves the negotiation room. Negotiation then continues among the remaining countries (set of active countries in the negotiation room). Once all treaties are concluded, the coalition structure $\mathbb{M}$ which corresponds to a numerical coalition structure, $\mathcal{M}$, is established.

4 Integrated analysis of IEAs

Dynamic games are typically characterised by a large number of subgame-perfect equilibria. To refine these equilibria, we focus on pure strategy Markov Perfect equilibria (MPE).[14] A MPE is a subgame-perfect equilibrium in which all countries use Markovian strategies. Markovian strategies depend only on payoff-relevant variables summarised in the current state, and history matters only through its effect on the current state.

[13] Committing to staying alone is a reasonable assumption in a public good game. We will show that a singleton coalition, if formed in equilibrium, has the highest payoff, while if it joins any other coalition, it has to set the same SCC as the rest of the signatories of that coalition, and will thus have a larger SCC.

[14] Focusing on pure strategies is a mainstream assumption in coalition formation theory. To the best of our knowledge, Dixit and Olson (2000) and Hong and Karp (2012) are the only papers which focus on mixed-strategy equilibria in coalition games with public goods.

In contrast to repeated games with no state or stocks, investigating MPEs in dynamic games is common.[15] Maskin and Tirole (2001) argue that MPEs are simple, robust and consistent with rationality.

In our framework, the current state includes the formed coalitions (if any); the number of countries that are negotiating (if any); the proposal (if ongoing or signed) and thus the identity of the proposing country; the capital stocks of the countries K_{it}; the global stock of cumulative emissions S_t; and the (per-unit) scarcity rent associated with their fossil fuel reserves, μ_{it}. In the next section, we show how these variables determine the payoffs. Finally, we recall that we focus on the farsighted stability concept of "equilibrium binding agreements" of Ray and Vohra (1999, 2001). To ensure sequential rationality, we solve the model backwards in time.

4.1 Climate policy decisions in a coalition

When choosing their optimal climate policy, the members of coalition M internalise the emissions externality they impose on other coalition members, while there is a non-cooperative behaviour among the coalitions. The members of each coalition maximise their joint discounted infinite-horizon payoff, which we call the total worth of coalition M and is given by

$$\sum_{i\in M}\sum_{\tau=0}^{\infty}\beta^{\tau}\{\ln(C_{it+\tau})\} \tag{4.1}$$

subject to the constraints for the depletion of fossil fuel reserves (3.2) and the feasibility conditions for the final goods in (3.5) for each $i \in M$. Optimal energy use for the membership in coalition M requires that

$$\frac{\nu Y_{it}}{E_{it}(m)} = \mu_{it}C_{it} + \hat{\Lambda}(m)Y_{it} \tag{4.2}$$

which implies that the marginal productivity of fossil fuel is set equal to its marginal cost which equals the scarcity rent $\mu_{it}C_{it}$ (as mentioned earlier, μ_{it} is the per-unit scarcity rent for country i at time t) plus the SCC, $\hat{\Lambda}(m)Y_{it}$, where the per-unit SCC is

$$\hat{\Lambda}_{it}(m) = \hat{\Lambda}(m) \equiv \frac{\gamma\xi m}{1-\beta} \tag{4.3}$$

The per-unit SCC is the SCC per unit of output of each signatory $i \in M$ for any period t [16] and corresponds to the present value of the sum of discounted climate damages for all members of coalition M from emitting one unit of carbon today.[17]

[15] Our dynamic game presented in the last section has $2N + 1$ state variables.

[16] The SCC can be implemented in a decentralised economy using for example a Pigouvian carbon tax where the revenues from this tax are rebated in a lump-sum fashion.

[17] The term *social* here is from the point of view of coalition M. Furthermore, the definition of the optimal SCC corresponds to the conventional definition in the field which is defined in *marginal* terms.

Hence, as $\hat{\Lambda}(m)$ increases linearly in the coalition size, the larger the coalition, the larger is the share of the damages associated with emissions that is internalised. Note that in equilibrium all members of coalition M are bound to set the same SCC. This rules out the possibility of the formation of a grand agreement in equilibrium which would include all countries but with different levels of SCC. The per-unit SCC also increases in the damage coefficient γ; the transient climate response of temperature to cumulative emissions ξ; and the discount factor, β. For example, more patient policymakers have a higher SCC and thus tax carbon more vigorously and reduce emissions more.

Equation (4.2) implies that the scarcity rent and optimal SCC are both proportional to aggregate economic activity. In Appendix A.1 we show that equation (4.2) in conjunction with the constant saving rate result from our general equilibrium analysis gives rise to our first result.

Proposition 1 . *A coalition M of m members sets the SCC per unit of output equal to $\hat{\Lambda}(m) \equiv \frac{\gamma\xi m}{1-\beta}$ for all $i \in M$ at any time t. The optimal emission level for each country of such a coalition is*

$$E_{it}(m) = \nu/[\mu_{it}(1-\beta(1-\nu)) + \hat{\Lambda}(m)] \tag{4.4}$$

where the scarcity rent per unit of consumption is

$$\mu_{it} = \beta^{-t}\mu_{i0} \tag{4.5}$$

and μ_{i0} is that value of the initial per-unit scarcity rent that exactly satisfies equations (3.3), (4.4) and (4.5) for a given stock of initial fossil fuel reserves, R_{i0}.

As explained earlier, larger coalitions agree on a proportionally larger SCC. This in turn leads to lower energy consumption and emissions for the coalition members of such coalitions. Furthermore, countries with large fossil fuel reserves have low scarcity rents and thus consume more energy and emit more.

An important consequence of our functional assumptions is that the per-unit SCC is independent of all stocks and independent of future values of output, consumption and cumulative emissions. The emission strategies are dominant in the sense that the emissions of complementary coalitions does not affect the emission strategies of any coalition. To see this, notice that having a Cobb-Douglas production function in the final good sector implies that the marginal products of capital and energy are proportional to output, that marginal damages, i.e. $\hat{\Lambda}(m)Y_{it}$ are proportional to output, and that a logarithmic utility function implies that the marginal utility of consumption is inversely proportional to output. This results in a decoupling of the economy and climate dynamics. The emissions plans of the members of a coalition depend only on two variables: the size of coalition itself (through its effect on the per-unit SCC) and the per-unit scarcity rent

of fossil fuel reserves. Therefore, in equilibrium a proposer does not need to condition its proposal on the SCC (or carbon price) of other coalitions. Although the emission strategies are dominant, their payoffs depend on the global cumulative emissions and thus on the entire coalition structure, and the associated energy use of all countries.

In Appendix A.2 we present two benchmarks: the non-cooperative outcome where the planner of each country chooses its energy consumption non-cooperatively (corresponding to a singleton coalition structure) and the fully globally cooperative outcome (i.e. the grand coalition, where $m = N$) which corresponds to the social optimum for the global economy. From the analysis, it follows that the per-unit social costs of carbon under the various outcomes satisfy $\hat{\Lambda}(N) \geq \hat{\Lambda}(m) \geq \hat{\Lambda}(1)$.

Equation (4.5) corresponds to the first-order optimality condition for R_{it+1}. As the natural resource is exhaustible, its per-unit shadow price increases over time (at the rate $1/\beta$) and the demand for fossil fuel decreases over time.[18] Therefore, emission levels become non-stationary. Note that the actual scarcity rent (i.e. multiplied by the level of aggregate consumption) grows at a rate equal to the marginal product of capital i.e. $(1-\nu)Y_{it}/K_{it}$ (equal to the rate of interest plus the depreciation rate, 1, in the market economy). This rule for the actual scarcity rent is known as the Hotelling rule. The initial scarcity rent, μ_{i0}, is such that cumulative fossil fuel use exhausts all of the initial fossil fuel reserves for each country either in finite time or asymptotically, i.e. $\lim_{t\to\infty} \sum_{s=0}^{t} E_{it-s} = R_{i0}$. Hence, at time t, after joining a (non-singleton) coalition, and by committing to a new per-unit SCC, the per-unit scarcity rent in each country in coalition M is adjusted. This leads to the following insight.

Corollary 1 . *The larger the size of the coalition, m, the smaller the per-unit scarcity rent of its signatories after the membership stage.*

So, the per-unit scarcity rent in countries which are signatories to larger coalitions is lower after the membership stage. The reason is that internalising the global warming externality implies that such countries will deplete their given reserves at a later time. We allow μ_{it} to be heterogeneous across countries and assume that the participating countries in climate negotiations have a finite scarcity rent. In other words, the total number of countries, N, consists of those countries which are contributing to the environmental externality.[19]

The other first-order conditions give rise to the following results.

[18] If fossil fuel reserves were abundant, as in the case of coal, the per-unit and actual scarcity rents would be zero.

[19] It may seem that by joining a non-singleton coalition, the emission level of signatories is affected by two counteracting factors: the decrease in the scarcity rent and the increase in coalition size and hence the per-unit SCC. However, note that the former is a result of the latter and it is plausible to assume that the effect of decrease in the scarcity rent should not dominate the effect of increase in the SCC on emissions.

Proposition 2 . *Aggregate output, consumption, the capital stock and growth rate of each member of coalition of size m at time t, are given by*

$$Y_{it}(\mathbb{M}) = exp(-\gamma T_t) A_y K_{it}^{1-\nu} (\nu/(\mu_{it}(1-s) + \hat{\Lambda}(m)))^{\nu}$$

$$C_{it}(\mathbb{M}) = (1-s) Y_{it}(\mathbb{M})$$

$$K_{it+1}(\mathbb{M}) = s Y_{it}(\mathbb{M})$$

$$Y_{it}(\mathbb{M})/Y_{it-1}(\mathbb{M}) = exp(-\tfrac{\gamma\xi}{\nu} E_t(\mathbb{M})) s^{1-\nu} \left(\tfrac{r_{it-1}}{1-\nu}\right)^{1-\nu} \left(\tfrac{\beta\mu_{it}(1-s)+\hat{\Lambda}(m)}{\mu_{it}(1-s)+\hat{\Lambda}(m)}\right)^{\nu}$$

respectively, where $s_{it} = s = \beta(1-\nu)$ is the countries' common and constant saving rate and $r_{it-1} \equiv (1-\nu)\frac{Y_{it-1}}{K_{it-1}}$ is the marginal product of capital.

Aggregate output increases in the saving rate s and total factor productivity, but decreases in global warming (cumulative emissions) and the price of fossil fuel (i.e. in the sum of the per-unit scarcity rent and the per-unit SCC), and thus decreases in the number of signatories of the coalition.

Consumption and investment are constant shares of output. From the reduced-form expression for aggregate output given in Proposition 2, we see that both of them depend negatively on the stock of cumulative emissions, S_t, through global temperature. Moreover, consumption and capital choices are non-stationary as they depend through aggregate output on the time-varying paths of μ_{it} and S_t.[20]

As in an endogenous growth model without global warming externalities and scarcity rents, the growth rate of aggregate output depends on the growth rate of technological progress (which we abstract from). But by introducing global warming and the scarcity rent, the growth rate in our modified AK model decreases in the per-unit scarcity rent, μ_{it}, for $i \in M$. Hence, during the decarbonisation period the rate of economic growth decreases. As the stock of fossil fuel exhausts and thus $\mu_{it} \to \infty$ and $E_{it}(m) \to 0$, the level and growth of aggregate output converge to zero, i.e. $Y_{it}(M, \mathbb{M})/Y_{it-1}(M, \mathbb{M}) \to 1$.[21]

4.2 Climate membership decisions

Within the above framework, countries decide about their membership in a climate coalition. The incentives of countries in the climate negotiation stage are determined by

[20] There is full deprecation of capital at the end of each period, but heterogeneity with respect to K_{i0} in this model implies that due to the constant and common saving rate all subsequent capital stocks K_{it} are the same fraction of output at the beginning of each subsequent period. Hence, the heterogeneity, at least weakly, spreads through future periods.

[21] By introducing technological change in combination with a green technology as a perfect substitute to fossil fuels (see Section 6) or using an inexhaustible fossil fuel, our model will exhibit a positive growth rate like in Golosov et al. (2014). However, our analysis of coalition formation of countries is not sensitive to their growth rate as long as the participating countries have a finite initial per-unit scarcity rent.

the optimum value function of a country in a coalition M. Before presenting our results under the farsightedness assumption, let us first briefly discuss the outcomes of our model under the more commonly used cartel (or internal and external) stability conditions. As shown in the Appendix A.4, under the assumption of cartel stability, the largest coalition size is $m^* = 3$ for any total number of countries N. This is the well-known small-coalition paradox. The cartel stability concept considers only unilateral deviations and upon a deviation, either totally disregards the possibility of updating the *membership* strategies by the remaining countries, or disregards an optimal update of strategies (as it is for the case we have analysed in the Appendix A.4). Hence, under such assumptions the countries act naively and, having a higher incentive to free ride, can only form a small coalition with a maximum size of three countries. This is true for any number of countries N.

Relative to the internal and external stability concept, where only *unilateral* deviations are considered, the *coalition-proofness* stability concept generalises the Nash equilibrium in that respect and includes the examination of *multilateral* deviations too. However, upon a deviation by a potential coalition member, the membership decisions of the complementary set are assumed to be fixed. Instead, we will use the *farsightedness* concept which relaxes this restrictive assumption.

Definition 2 . *A coalition structure is farsighted stable if it is immune to unilateral and multilateral deviations by the deviating group, and to deviations by the active players in the negotiation room or members of other coalitions (before signing binding agreements).*

From the point of view of a farsighted country, a potential group of deviating countries thus has to consider further possible deviations by the deviating group (similar to the coalition-proofness concepts, the deviating group can split further before signing their agreement), in addition to the consequence of their deviation on the active players in the negotiation room, i.e. they may disband too. Therefore, while fixing their membership decisions, the countries are required to rationally predict the entire $\mathbb{M}$. Let us denote the equilibrium coalition structure by $\mathbb{M}^*$ and the equilibrium numerical coalition structure by $\mathcal{M}^*$.

Farsightedness implies that potential deviations from a treaty must be constrained to be farsighted. This means that the farsighted set of possible equilibrium coalitions should be defined recursively. In each step of the recursion, we need to identify for which group of countries, a grand coalition forms in equilibrium. Starting from the smallest set of countries, i.e. $N = 2$, we should find $\mathbb{M}^*$ for each group of two countries. Then, knowing which group of two countries can strike a deal, all possible $\mathbb{M}^*$ have to be found for $N = 3$ and the process continues for $N = 4$, etc. With heterogeneous countries, there will be path-dependency in this analysis as the equilibrium outcome would depend on which countries are chosen in earlier stages of the recursive process. Clearly multiplicity of equilibria is expected and the analysis can be tedious.

In our model, the countries are heterogeneous in various ways: different initial capital levels, K_{i0}; different total factor productivity constant, A_i; different scarcity rents, μ_{it}; or different initial stocks of fossil fuel resource, R_{i0}, across countries (and in section 6, with respect to cost of investment in green technologies).[22] All of these sources of asymmetry have been important topics in climate negotiations. We begin with investigating the impact of heterogeneity with respect to capital level and TFP, and then we move to heterogeneity with respect to initial stocks of fossil fuel, and thus the scarcity rent.

Let us denote the optimum value function of country i in coalition M, when country j is the initial proposer as a function of M and the underlying $\mathbb{M}$ with $V_i^j(S_t, K_{it}, \mu_{it}, M, \mathbb{M})$, and the optimum value function of the country in a grand coalition $\{I\}$ with $V_i^j(S_t, K_{it}, \mu_{it}, I)$.

In contrast to the symmetric case, with heterogeneous countries, the concept of "average worth", i.e. payoff of one country in the coalition, in determining the equilibrium coalition structure is inadequate. Suppose j is the initial proposer. For any N, the farsighted country j needs to identify the most profitable deviation from the grand coalition and it is sufficient to compare the total payoff of the best profitable deviation by forming coalition $M \in \{M_1, M_2, ..., M_k\}$ versus the total payoff of the corresponding m members from staying in the grand coalition $\{I\}$. In other words, j needs to determine the sign of

$$\sum_{i=1}^{m} V_i^j(S_t, K_{it}, \mu_{it}, M, \mathbb{M}) - \sum_{i=1}^{m} V_i^j(S_t, K_{it}, \mu_{it}, I) \tag{4.6}$$

In the Appendix A.5 we show that this difference is independent of A_i and the capital stock for any discount factor β. Hence, the membership decisions are independent of heterogeneity with respect to K_{i0} and A_i.

Furthermore, the difference of payoffs in (4.6) is a linear function of emissions only. As discussed in the previous section, because of their dominant strategies, the emission plans of signatories of coalition M depend only on its *own* size, m, and importantly they do not need to be conditioned on the entire coalition structure. Furthermore, all members of a coalition of size m have the same per-unit SCC. Although V_i^j depends on the equilibrium coalition structure, the emission plans in the proposal depend only on m. This implies that we can direct our attention to characterising the numerical equilibrium coalition structure. In other words, we can use the conventional farsighted methodologies which are developed for symmetric countries in our asymmetric case here. Thus, the problem in (4.6) reduces to determining sign of

[22] There are not many studies in non-cooperative game theory on coalition formation of heterogeneous farsighted countries. To the best of our knowledge, only Ray (2007, p.130) studies games of heterogeneous agents with externalities. He derives the sufficient conditions for existence of equilibria without delay. These sufficient conditions are (i) coalitions which form subsequently have a lower average worth (ii) the larger the set of active countries in the negotiation room, the larger the equilibrium payoff of countries, (iii) the equilibrium payoff of being a proposer is greater than the equilibrium payoff of being proposed to (without any lapse of time or discounting). In our international climate game with free-riding incentives of the countries, all these conditions are satisfied.

$$V_i^j(S_t, K_{it}, \mu_{it}, m, \mathcal{M}) - V_i^j(S_t, K_{it}, \mu_{it}, N) \tag{4.7}$$

for each $i \in M$. As shown in the literature, although transfers and unequal split of payoffs are allowed, they play no role in characterising the equilibrium numerical coalition structure. Likewise for this purpose, the identity of the initial proposer is irrelevant too.

Heterogeneity with respect to the initial stock of fossil fuel R_{i0} affects the optimum value function through the associated scarcity rent, μ_{it}. From equation (4.4), the emission of country $i \in M$ depends negatively on its scarcity rent, i.e. a country with a higher scarcity rent, emits less, and vice versa. Note that, heterogeneity with respect to the stock of fossil fuel and μ_{it} causes heterogeneity with respect to emissions, even within a coalition M with m members.[23] This source of heterogeneity does not affect the countries' payoffs linearly. However, in the limit that $\beta \to 1$, the difference of payoffs in (4.6) becomes independent of μ_{it}. Therefore, the decision-making of farsighted countries in joining climate coalitions is also independent of heterogeneity with respect to the scarcity rent and the identity of the proposer. Thus independently of the source of heterogeneity, we can again characterise the numerical equilibrium coalition structure.[24]

The following proposition summarises this discussion. The proof is in Appendix A.5.

Proposition 3 . *The equilibrium numerical coalition structure $\mathcal{M}^*$ can be characterised independent of the heterogeneity with respect to K_{i0} and A_i. It can also be characterised independent of the heterogeneity with respect to R_{i0} and μ_{it}, if $\beta \to 1$.*

We call this the decoupling result, as we can decouple the problem of the cardinality of coalitions in equilibrium from the actual composition of countries in each $M^* \subseteq \mathbb{M}^*$. Hence, the numerical coalition structure can be characterised, while we can keep our focus off their heterogeneity. Then after finding $\mathcal{M}^*$, we direct our focus to the question of which m^* countries an initial proposer should propose to, and we answer the latter question from the point of view of efficiency improvement.

Our decoupling result for the case of heterogeneity with respect to K_{i0} and A_i is stronger, as it does not rely on assumptions regarding β. This result can be used with any reduced-form payoffs where the heterogeneity affects the countries' payoffs in an affine way. Instead, the decoupling result associated with R_{i0} and μ_{it} depends on the functional form and assumptions related to the discount factor.

As explained, the equilibrium payoffs and emissions, and therefore equilibrium global temperature depend on the identity of the initial proposer and the composition of the

[23] In a framework with heterogeneous countries, emission levels can differ for two reasons: firstly, countries are ex-ante asymmetric with respect to their initial fossil fuel and thus their scarcity rent, and secondly, by joining coalitions with different sizes, their emission path affects the trajectory path of their scarcity rent. In the membership stage of a reversible coalition formation, our focus is on the effect of ex-ante asymmetry on membership decisions.

[24] We show in the next section, that in our IAM, the characterisation of the equilibrium numerical coalition structure relies on the assumption of $\beta \to 1$ even for symmetric countries.

heterogeneous countries in each coalition. Thus the equilibrium payoffs may differ across countries. The importance of our result is that no matter what the protocol ordering of initial proposers is, every proposer selects the number of members which $\mathcal{M}^*$ prescribes. In the next section, we first answer the question of how many countries a proposer should include in its climate coalition proposal, and we show that the game with heterogeneous countries has a unique equilibrium numerical coalition structure.

4.2.1 Equilibrium numerical coalition structure

As described earlier, the optimum value function of a signatory of a coalition M with size m in a numerical coalition structure $\mathcal{M}$ is $V_i(S_t, K_{it}, \mu_{it}, m, \mathcal{M})$. Henceforth, we suppress all arguments not directly relevant for the analysis of characterising the equilibrium numerical coalition structure. Let us denote the optimum value function of a country as a function of m and the underlying $\mathcal{M}$ by $V_i(m, \mathcal{M})$, and the optimum value function of a country in a grand coalition $\{N\}$ by $V_i(N)$.

The equilibrium numerical coalition structure again needs to be identified recursively. For completeness, note that if $N = 1$, a singleton coalition forms. Then, we need to find $\mathcal{M}^*$ if $N = 2$. Given that, we then find $\mathcal{M}^*$ if $N = 3$, and continue the recursion until we have reached the total number of countries N that are in the global economy.

For the case $N = 2$, the problem reduces to whether $\{1, 1\}$ forms or $\{2\}$. It can be shown that this depends on the sign of

$$\begin{aligned} &V_i(1, \{1, 1\}) - V_i(\{2\}) = \\ &\frac{1}{1 - \beta(1 - \nu)} \Bigg\{ \nu \{ ln \left(\frac{E_{it}(1)}{E_{it}(2)} \right) + \beta ln \left(\frac{E_{it+1}(1)}{E_{it+1}(2)} \right) + ... \} \\ &- \frac{2\gamma\xi}{1 - \beta} \{ [E_{it}(1) - E_{it}(2)] + \beta [E_{it+1}(1) - E_{it+1}(2)] + ... \} \Bigg\} \end{aligned} \tag{4.8}$$

So, $V_i(1, \{1, 1\}) - V_i(\{2\})$ is independent of the capital stocks and the stock of cumulative emissions, and only depends on the emission paths under the two scenarios. The second line in equation (4.8) is the discounted infinite sum of a ratio of the benefit of emitting in a singleton coalition relative to the benefit of emitting in a grand coalition, and is clearly positive. The third line captures the discounted infinite sum of the losses resulting from the damages of emitting in a coalition structure of singleton relative to the damages of emitting in a grand coalition, and is negative.

In general, determining the sign of this equation requires a numerical analysis for a specific set of parameter values. However, if we focus on the cases $\sigma \to 1$ and $\beta \to 1$, it is easy to show that with two countries $\lim_{\beta \to 1}(V_i(1, \{1, 1\}) - V_i(\{2\})) < 0$ so that the grand coalition forms in equilibrium, i.e. the equilibrium coalition structure is $\mathcal{M}^* = \{2\}$.[25]

[25] Note that it is possible to keep the ratio of length of sub-periods to periods fixed, and analyse the model under $\sigma \to 1$ and $\beta \to 1$.

Continuing to the case $N = 3$, there are only three possible numerical coalition structures with symmetric countries: $\{3\}$, $\{2,1\}$, or $\{1,1,1\}$. From the last stage of the recursion we already know that (if one of the three countries leaves) a group of two countries would not unravel. Hence, due to the farsightedness of the countries, there is no need to check $\{1,1,1\}$, because $\{1,1\}$ is not a farsighted-stable deviation.

In a public good game, considering farsighted deviations implies splitting N (or any active number of players in the negotiation room) into coalitions where their sizes result from breaking up N into the largest possible integers at which a grand coalition was stable in previous stages of the recursion. Ray and Vohra (2001) show that in a public-good game with symmetric countries, it is sufficient to check the deviation by the smallest element of in $\mathcal{M}$ when upon this deviation, N countries have to split into the largest possible farsighted coalitions, i.e. those that result from *decomposition* of N. Since here the heterogeneity does not affect the equilibrium numerical coalition structure, checking such a deviation from the grand coalition is a sufficient condition for every country to prefer the grand coalition to any other coalition structure.

Definition 3 . *$\mathcal{T}^*$ is defined as the set of the total number of countries, N, for which a grand coalition forms in equilibrium.*

Definition 4 . *For any integer N, the decomposition $D(N)$ is defined as $\{m_1, m_2, ..., m_k\}$, such that m_k is the largest integer in $\mathcal{T}^*$ that is strictly smaller than N. Then any other m_i in $D(N)$, is the largest integer in $\mathcal{T}^*$ that is no greater than $N - \sum_{j=i+1}^{k} m_j$.*

For example, for the case $N = 3$, we know from previous stages of the recursion that $\mathcal{T}^* = \{1,2\}$, and thus the decomposition of N is $\{2,1\}$.

At each stage of the recursion, the optimum value of such a deviation should be compared with the optimum value of the grand coalition. This result reduces the number of checks. Ray and Vohra (2001) show that under low bargaining frictions ($\sigma \to 1$), the resulting numerical coalition structure or decomposition of N coincides with the equilibrium numerical coalition structure of the bargaining game described in section 3.3, as in Ray and Vohra (1999). Therefore, as the negotiations start, if the grand coalition is not stable, first a proposer makes an acceptable offer to the smallest equilibrium coalition (which can be to itself only, if it is a singleton), and without any delay the offer is accepted and the coalition forms. And a similar process continues among the remaining countries.

For example for the case $N = 3$, it is sufficient to check the sign of $V_i(1,\{2,1\}) - V_i(\{3\})$ only. This time, $\lim_{\beta\to 1}(V_i(1,\{2,1\}) - V_i(\{3\})) > 0$, thus in contrast to the case of $N = 2$, the grand coalition is not stable and in equilibrium there will be one free rider and a coalition of size two, i.e. the equilibrium numerical coalition structure is $\mathcal{M}^* = \{2,1\}$. Similarly, for $N = 4$, the only conceivable farsighted decomposition is $\{2,2\}$ which has to be considered against $\{4\}$, and it can be shown that here, a grand coalition forms again, i.e. $\mathcal{M}^* = \{4\}$. Hence, $\mathcal{T}^*$ expands to $\{1,2,4\}$. In fact, the equilibrium numerical coalition structure never consists of two coalitions with the same

size.[26] Therefore, the numerical equilibrium coalition structure is always unique.

Comparing the optimum value function of the smallest coalition in the decomposition with the value function of the grand coalition at each stage of the recursive procedure (i.e. for each N) can be demanding. In Appendix A.6, we show that in our IAM, the recursion process can be simplified and there is a general rule for inequalities like (4.8).

Lemma 1 . *Let $D(N) = \{m_1, m_2, ..., m_k\}$ be the decomposition of N, such that $m_1 < m_2 < ... < m_k$. For $\beta \to 1$, in our IAM, a grand coalition forms in equilibrium if*

$$ln(\frac{N}{m_1}) < (k-1) \tag{4.9}$$

This Lemma is proved in Appendix A.6 and provides a simple sufficient condition for the full characterisation of the set $\mathcal{T}^*$ for our IAM in the limit that $\beta \to 1$.[27] The LHS of (4.9) is the gain from emitting in the small coalition versus emitting in the grand coalition. The RHS of (4.9) is the externality damage resulting from forming $D(N)$ versus the grand coalition. Since in the limit as $\beta \to 1$, emissions are almost stationary, it is sufficient to compare the gains and losses of one period only: if damages are higher than the gains from emitting, a grand coalition forms in equilibrium.[28]

Using this lemma and continuing the recursion to find more elements of $\mathcal{T}^*$, we can show that the set $\mathcal{T}^*$ is a Tribonacci sequence.[29] The proof is in Appendix A.7.

Proposition 4 . *In our IAM with farsighted and heterogeneous countries, if $\beta \to 1$, a grand coalition occurs in equilibrium if the total number of countries N is an element of*

$$\mathcal{T}^* = \{1, 2, 4, 7, 13, 24, 44, 81, 149, 274, ...\} \tag{4.10}$$

which is the Tribonacci sequence with predetermined elements $\{0, 0, 1\}$. If $N \in \mathcal{T}^*$, then $\mathcal{M}^* = \{N\}$; if $N \notin \mathcal{T}^*$, then $\mathcal{M}^* = D(N)$, given $\mathcal{T}^*$. The unique numerical coalition structure is independent of the heterogeneity of the countries and thus independent of the identity of initial proposers. Furthermore, the equilibrum number of signatories in any climate coalitions, m^*, is a Tribonacci number.

Obtaining Tribonacci numbers is a novel result in economics.[30] As explained earlier,

[26] This holds for continuum action spaces, and not for binary action spaces.

[27] Note that although k and m_1 are endogenous and are to be determined, it is always true that $m_1 \leq N/k$.

[28] From equation (4.5) the per-unit scarcity rent rises at an infinitesimally small rate and thus from equation (4.4) emissions decline at an infinitesimally small rate.

[29] Tribonacci numbers are elements of a Fibonacci sequence of order 3.

[30] While the full proof can be found in Appendix A.7, the main idea is the following : given that N is at least 1, and starting from 1, 2 and 4, then 7 is the first Tribonacci element ($T_{n=1} = 1+2+4$) and they all satisfy Lemma 1. Then using strong induction, we show that for any positive integer n, an element $T_n \in \mathcal{T}^*$ is also the sum of the last three elements of the set given that this is the case for the preceding elements.

the equilibrium numerical coalition structure is constructed using the set $\mathcal{T}^*$. Thus our algorithm to characterise $\mathcal{M}^*$ in our IAM with farsighted and patient countries has two simple steps: first generate the Tribonacci set, and if the total number of countries belongs to the set of Tribonacci numbers, then a grand coalition forms in equilibrium. If not, then the equilibrium coalition structure is the decomposition of N using the elements of the Tribonacci set. For example, going back to the case of $N = 3$, since $3 \notin \mathcal{T}^*$, the grand coalition does not form and its decomposition, using the elements of $\mathcal{T}^*$ that are smaller than 3, determines the equilibrium coalition structure, thus $\mathcal{M}^* = \{2, 1\}$. Hence, in equilibrium there is one country on its own (a singleton) and one coalition of two countries.

The result in Proposition 4 simplifies the characterisation of the equilibrium coalition structure for any number of heterogeneous countries N. No matter how heterogeneous the countries are, the equilibrium numerical coalition structure is unique.

For our IAM, there is no need to use any recursions to find the set $\mathcal{T}^*$. Instead, Ray and Vohra's methodology can lead to any set of total number of countries for which a grand forms in equilibrium. They assume the countries are symmetric and have a one-off payoff. In other words, after bargaining, the countries receive their agreed payoffs and the game ends. The recursion thus is not demanding and it leads to an analytical solution. We use and extend their methodology to answer questions in an infinite-horizon IAM with time-varying payoffs. In general, in such a model numerical simulations would be called for at each stage of the recursion. De Zeeuw (2008) is the only infinite-horizon study with a public good game of farsighted countries, but this study obtains its results only numerically. However, in our model, Lemma 1 gives a sufficient condition to analytically characterise under which conditions a grand coalition forms. Given Proposition 4 there is no need to check the payoffs at each stage of the recursion to find the integers at which a grand coalition forms in equilibrium, because the Tribonacci set is an already known set.

In case the number of countries (or regions or cities, etc.) is large and using Proposition 4 is laborious, the formalisation of the Tribonacci set by mathematician Plouffe (1993) can be helpful:

$$T_n = \left\lfloor 3b \frac{(\frac{1}{3}(a_+ + a_- + 1))^n}{b^2 - 2b + 4} \right\rceil$$

where $a_\pm = \sqrt[3]{19 \pm 3\sqrt{33}}$ and $b = \sqrt[3]{586 + 102\sqrt{33}}$, and $\lfloor . \rceil$ is the nearest integer function.

The relative simplicity of Lemma 1 and Proposition 4 is the result of the solution concept and the special features of our IAM. In particular the special structure of our IAM yields a per-unit SCC which is independent of the aggregate economic outcomes, a constant saving rate and dominant emission strategies. Importantly, our assumption of

$\beta \to 1$ leads to unambiguous tractable outcomes and needs to be justified by a normative approach. There is a large literature on the fact that the social discount rate is smaller than the private or market-based discount rate. Arrow et al. (2003) argue that because of market imperfections, especially in long-run, using market observables such as the interest rate to identify the social discount rate can be misleading. Following Ramsey (1928) who argued that it is unethical to discount the welfare of future generations, climate economists have often used a near-zero rate of time preference (e.g., Stern Review, 2007; Dietz and Stern, 2015).

The fact that the equilibrium number of signatories is a Tribonacci number implies that the size of climate coalitions can be large. The small-coalition paradox that Lessmann et al. (2009, 2015) and Bosetti et al. (2013) find in their IAMs results from the Nash equilibrium solution concept and the single-coalition assumption. Adopting the farsighted-stability concept and without any of the known remedies to increase cooperation, we show that the Tribonacci-number of signatories depends on the number of countries N and the countries' payoffs and it can be significantly larger. If a grand coalition does not form, the largest stable climate coalition can still be large. It is the largest integer in the set of Tribonacci numbers, $\mathcal{T}^*$, that is smaller than N. Moreover, multiple (non-singleton) climate coalitions can form, which have more ambitious climate policies compared to the singleton coalitions (like the fringe countries that occur under the assumption of cartel stability).

The farsighted stability concept is a more realistic assumption than the cartel stability concept as it does not assume that if a country breaks off the negotiations, other countries will not react by changing their membership strategies as assumed under cartel stability. Hence, in the farsighted set of coalitions $\mathcal{M}^*$ that we have characterised in Proposition 4, the countries have to rationally predict the entire reactions prior to their membership decisions. This, in turn, reduces their free-riding incentives and leads to the formation of larger coalitions.

4.2.2 Example: equilibrium numerical coalition structure with 195 countries

Let us assume $N = 195$. Based on the Tribonacci sequence in Proposition 4, because $195 \notin \mathcal{T}^*$, we can verify that $\mathcal{M}^* = \{149, 44, 2\}$, where $149 + 44 + 2 = 195$. In other words, three coalitions can form and the coalition of 2 signatories forms first, then the coalition with 44 signatories leaves the negotiation room and lastly the largest coalition with 149 members forms. Clearly, there is no small-coalition paradox and multiple coalitions emerge. The equilibrium numerical coalition structure is not too complicated, since only three coalitions form. There is a relatively large coalition of $m^* = 149$ which will have more ambitious climate policies than the coalition of 44 countries. The coalition of 44 countries has more ambitious policies than the coalition of 2 countries. Furthermore, the set $\mathcal{M}^*$ does not have a lot of small coalitions of a small

number of countries as it has only three coalitions with only one small coalition.

In contrast to the fully cooperative outcome among all countries in the world, which is what most IAMs examine, our results imply that a group of 149 countries forms a stable coalition and sets the per-unit SCC at $\hat{\Lambda}(m^*) = \frac{149\gamma\xi}{1-\beta}$ for all future periods. The other two smaller coalitions accordingly set their climate policy in their binding agreement leading to smaller per-unit SCC corresponding to fractions 49/149 and 4/149 of the per-unit SCC of the coalition of 149 countries. Under cartel stability the maximum coalition size is 3. This leads to an average SCC which is much smaller than the average SCC under farsightedness, i.e. $1.03 = (192\times1+3\times3)\div195 < (149\times149+44\times44+2\times2)\div195 = 123.8$. In words, the average SCC under farsightedness is about 120 times larger than this average under cartel stability.

Using Proposition 1, it is possible to find the equilibrium emission plans of all countries, and back out the cumulative stock of emissions of GHGs from equation (3.8). Furthermore, we can use Proposition 2 to track variables which were missing in conventional analyses of IEAs, namely the general equilibrium allocations of aggregate consumption, capital and the economic growth rate, carbon emissions of each signatory, and the global temperature associated with the equilibrium coalition structure.

4.2.3 Can countries be found to curb emissions globally further?

So far, we have shown that with heterogeneous countries, the equilibrium numerical coalition structure follows from the set of Tribonacci numbers. Next, we turn to the question of the composition of countries in the equilibrium coalition structure. With heterogeneous countries the identity of the countries matters for the equilibrium coalition structure, $\mathbb{M}^*$, which naturally leads to multiplicity of equilibria. Therefore, we try to address the narrower question of whether the countries, knowing the equilibrium number of signatories, have any incentive to approach a particular country in order to achieve a greater reduction in global emissions while they match to climate coalitions.

In our IAM, the fully efficient outcome is achieved when the grand coalition forms. This is the internationally cooperative outcome where global environmental damages are fully internalised. Any other equilibrium coalition structure will naturally result to some level of inefficiency as not all damages are internalised and emissions will be higher. In our heterogeneous setup, when the grand coalition is not stable, constrained efficiency is associated with the level of global emissions internalised, which will depend not only on the numerical coalition structure but also on the identity of the coalition members depending on their stock of reserves of fossil fuel, say. We will refer to this as "emission efficiency".

Let us retain the assumption of heterogeneity with respect to initial fossil fuel reserves and thus with respect to scarcity rents in the remainder of this section. If the countries are concerned with reducing emissions any further, it is important to ensure the the

countries with the largest stocks of fossil fuel join the largest possible coalitions.[31] This policy suggestion was missed in the literature, as the focus of the literature on IEAs is on a single coalition and symmetric countries. From a normative perspective, this *ought to* happen. But do policymakers support it?

Emission efficiency can be exogenously affected by the protocol. Although the (ordering by the) protocol is irrelevant in characterising $\mathcal{M}^*$, it is important for the composition of countries in $\mathbb{M}^*$, and thus for the equilibrium global mean temperature and equilibrium payoffs. Furthermore, given the exogenous protocol, an individual country chosen by the protocol may do its best in matching with other countries in coalitions to internalise the global emissions further, and thus improve constrained efficiency. Here we investigate the latter problem. We thus examine the problem of an initial proposer (and later its respondents) to a non-ultimate coalition i.e. the proposer knows that the next coalition is going to be larger and examines whether it is worthwhile to include less or more emitting countries in the smaller coalition of its own. In this analysis all other factors, e.g. the protocol or any element that can affect the transfers are held fixed.[32] [33]

Based on the proposition above, any initial proposer knows the equilibrium number of members in each coalition, while the identity (initial fossil fuel reserves and thus scarcity rent) of the initial proposer and of the other potential signatories still affect the efficiency of different equilibria. If the grand coalition is not stable, every initial proposer faces two challenges. On the one hand, a proposing country prefers countries that emit a lot (i.e. those with low scarcity rent) to be in the larger *subsequent* coalitions to ensure that a larger part of emissions are internalised over an infinite horizon. Thus, if it has a choice, the proposer of a smaller coalition prefers to approach the low-emitting countries rather than the high-emitting countries. On the other hand, a proposer knows that every low-emitting country faces the same dilemma, and (if possible) a low-emitting country may reject the offer and next sub-period propose to a country which has a lower emission path relative to that of the initial proposer. Especially, with low bargaining frictions this is a serious concern for any initial proposer.

To examine the equilibrium composition of countries in coalitions and the implications for global emission reductions explained above, consider the example of $I = \{1, 2, 3, 4, 5, 6\}$, where for all $i \in I$ we have $\mu_{it} > \mu_{i+1t}$. This assumption implies a strict order for the scarcity rent of the countries such that country 1 is the country that emits

[31] Likewise, in the model with green technologies as discussed in section 6, the countries with the least cost of developing green technologies can reduce their emission levels more ambitiously and thus for a given numerical coalition structure, in order to reduce the global temperature further, such countries should join the largest possible coalitions.

[32] Although in characterisation of equilibrium numerical coalition structure, transfers play no role, in coalition formation with heterogeneous countries transfers are important factors. The international transfers are determined by what the recipients can get in another coalition structure if they reject the offer.

[33] We can allow investment in green technology as a perfect substitute to the fossil fuel. So, no low-emitting country with a high scarcity rent fears collapse of its economy as its stock of fossil fuel reserves vanishes. See section 6 that in steady-state the countries can have a finite scarcity rent.

least and country 6 is the country that emits most, and no two countries have equal scarcity rents. From Proposition 4 we know that the equilibrium numerical coalition structure is $\mathcal{M}^* = \{2, 4\}$. So, the first initial proposer needs to make an acceptable offer to one other country and they would leave with a binding agreement. But which country should be selected? Does it depend on the scarcity rent of the initial proposer? If country 2 is the initial proposer at the beginning of the climate negotiations, it knows that country 1, although it emits less, will not reject its offer. If country 1 is concerned with the global temperature, then by rejecting the offer of country 2, the most emission-efficient candidate to propose to is country 2 again. Therefore, if either of them are the initial proposers and are concerned with efficiency, the most efficient coalition structure $\{\{1, 2\}, \{3, 4, 5, 6\}\}$ forms in equilibrium.

But in a no-delay equilibrium, do proposing countries try to find countries with similar emission levels in their equilibrium coalitions? Let us consider a case where at the beginning of the game, country 4 is the initial proposer. To ensure the formation of a coalition with no delay which includes country 4, it seems the best option is to make offers to either country 5 or country 3. Clearly country 5 would not reject. But can country 4 do any better in reducing the global temperature by offering to country 3? Equivalently, this leads to the question of whether country 3 prefers the formation of $\{\{3, 4\}, \{1, 2, 5, 6\}\}$ to rejecting the offer and proposing to the country that emits least to ensure that the maximum possible efficiency (i.e. the minimum possible global temperature) is achieved. This would lead to the formation of $\{\{1, 3\}, \{2, 4, 5, 6\}\}$ after lapse of a sub-period. It can be shown that for any β, and keeping everything else fixed, country 4 prefers the formation of most emission-efficient coalitions (likewise, does country 3). Furthermore, country 4 is not ready to make any drastic sacrifice by making an unacceptable offer to either country 1 or 2 to ensure the formation of the most efficient coalition structure of $\{\{1, 2\}, \{3, 4, 5, 6\}\}$ next period. Therefore, country 4 prefers approaching country 1, and if country 4 adequately compensates, country 1 accepts the proposal. We can show that this is a general property. In other words, keeping everything fixed, permuting μ_{it} can lead to an emission-efficiency gain.

Proposition 5 . *Assume that the grand coalition is not stable. For any β, all proposers and respondents prefer the most emission-efficient coalitions among all equilibrium coalition structures, $\mathbb{M}^*$, which only differ in initial fossil fuel reserves.*

The preferences of countries for emission efficiency must be interpreted with care. This proposition does not state that countries prefer the most emission-efficient coalitions among all different coalitions which correspond to the same numerical coalition structure. But the efficiency preference here is a weaker statement, as we have kept all other factors, e.g. capital level, transfers, etc. fixed. Thus, our efficiency result helps to break ties when countries are indifferent, as we kept all other factors fixed.

The formal proof is in Appendix A.8. Here, fixing the equilibrium numerical coalition structure, $\mathcal{M}^*$, characterised by Proposition 4, we check whether a proposer choosing a more efficient coalition of the same size can improve its payoff. It is sufficient to check the deviation to the coalition that is potentially most efficient, i.e. by choosing the countries with the highest scarcity rent and lowest initial fossil fuel reserves from the active players in the negotiation room. Given that changing the number of signatories would not be an equilibrium strategy, the proposer would compare the (direct) gains and externality damages of such a reshuffling of players across two coalitions with similar number of signatories. Since the emission of the proposer itself depends only on m^* and its own scarcity rent, there is no direct gain of switching to another coalition with the same number of members. The question is whether the resulted emission damages would be different. We show that any such proposer would prefer the most efficient coalition. The same result can be obtained by checking the problem of a respondent, who can reject and in the next sub-period proposes to the most possible efficient coalition of the same size.

Another factor that affects the formation of a coalition is the transfer. As mentioned before, transfers do not affect the numerical equilibrium coalition structure but they exist in equilibrium. In a no-delay equilibrium, every initial proposer of a coalition M^*, at every history that it proposes, makes an acceptable offer to m^* number of countries. In fact, in a public good game, no proposer should lose the opportunity of being a proposer (of a non-grand coalition), because if the offer is rejected, the initial proposer may or may not be included in the next proposal, and after this coalition, the subsequent coalitions are larger and set a higher SCC. Thus any initial proposer if needed offers adequate transfers to the respondents, and the offers are accepted with no delay.

5 Reversible agreements

In this section, we allow the countries to renegotiate any existing agreement at no cost, and we relax Assumption 1. We continue to focus on Markovian strategies, so all that matters in any period t is the current state. Hence, the solution concept is again a MPE. We maintain our definition of the current state which includes the formed coalitions (if any), the number of countries that are negotiating (if any), and the proposal (if ongoing or signed), the identity of the proposing country, the stock of global cumulative emissions S_t and the individual capital stocks K_{it}, and individual scarcity rents μ_{it} for all $i \in I$. We assume that the negotiations start while coalition structure $\mathbb{M}$ is in place at the beginning of period t (this can be the coalition structure of singleton). We keep to focus on no-delay equilibria, where at the end of each period of negotiations a coalition structure is concluded. Change of coalition structure over time under reversible coalition formation can be thought of as moving from one Markov state to another, i.e. from one coalition

structure to another. Note that if at the beginning of period t, the countries are in a coalition structure other than the grand or the coalition structure of a singleton, then the countries are heterogeneous with respect to K_{it}, R_{it} and μ_{it} (even if ex-ante they were symmetric). Thus in our IAM, the analysis of reversible agreements builds on the analysis of coalition formation of heterogeneous countries in Section 4.

Let us also assume there is a fixed protocol for all periods.[34] As before, it is a deterministic protocol. The assumption of binding agreements and reversibility can coexist. In fact, we assume that the agreements are binding. As suggested by Hyndman and Ray (2007) in a reversible setting, the assumption of binding agreements is justified if we assume that an *approval committee* which includes all parties of an existing binding coalition can approve the move to another state. Some members of this coalitional agreement will be affected by the new state: either their membership will be affected, and/or their payoffs will be directly affected.[35] The inclusion of the approval committee is to ensure that the rights of those in any existing binding agreements are protected.

Thus, we adjust the proposal to include an *approval committee*. We generalise the coalition formation model stated in Section 3.3 to allow for climate negotiations in every period t. Hence, at the beginning of every period t, in each sub-period, a proposer, selected by the protocol, makes a proposal to a group of countries. If the approval committee approves it, subsequently extra respondents (if any) can respond. If accepted by all, the negotiation game moves to a new state in the next sub-period in period t.

There is no literature about the role of this approval committee in a public good game with farsighted countries. It turns out that such a committee has an important role here. If the equilibrium coalition structure is not a grand coalition, then the approval committee of any non-ultimate coalition, will reject inclusion of any new members in the coalition because it would mean that they should internalise more of the global warming externality and should agree to a higher SCC per unit. They would reject being excluded from the existing coalition too, because the next coalition that forms in equilibrium will be larger.

Proposition 6 . *With reversible coalition formation of farsighted and heterogeneous countries, if $\beta \to 1$, a grand coalition forms in equilibrium if the total number of countries is an element of the Tribonacchi set, i.e. $\mathcal{T}^* = \{1, 2, 4, 7, 13, 24, ...\}$, at any time $t \in \{0, 1, 2, ...\}$. If $N \in \mathcal{T}^*$, then $\mathcal{M}^* = \{N\}$; if $N \notin \mathcal{T}^*$, then $\mathcal{M}^* = D(N)$, given $\mathcal{T}^*$. Furthermore, the MPE has an absorbing membership state with the same equilibrium*

[34] Relaxing this assumption would require imposing other assumptions to reconcile the rights of signatories of existing binding agreements confronting a new protocol. Nonetheless, it can be shown that the equilibrium numerical coalition structure is renegotiation-proof if the protocol is not fixed, but $\mathbb{M}^*$ is not.

[35] By the direct effect on payoffs we do not refer to the indirect effect through the channel of externally of global temperature on the signatories of other coalitions, but we refer to the effect of a change of their per-unit SCC and emissions.

coalition structure $\mathbb{M}^*$. The distribution of international transfers (if needed to support the coalition structure $\mathbb{M}^*$) are renegotiation-proof too.

By absorbing membership state, we mean that from any initial coalition structure, the equilibrium moves to the same membership state. The proof is in Appendix A.9. Proposition 6 states that even if countries have the option to renegotiate every period, the equilibrium numerical coalition structure is the same as in the case of irreversible agreements. This is an extension of Proposition 4. In addition the approval committee ensures that not only the same numerical coalition structure, $\mathcal{M}^*$, forms, but also the same coalition structure $\mathbb{M}^*$ forms upon every renegotiation. Thus, if coalition structure $\mathbb{M}^*$ is formed in period t, moving to period $t+1$, we expect the same equilibrium coalition structure if they were to renegotiate.

A final point here is that the total number of countries, N, includes those countries which are contributing to the externality and want to internalise it. Thus, the membership strategies are renegotiation-proof as long as countries have a finite scarcity rent.

There can be a history in which the respondents of the initial proposal can reject and make a proposal to the same members but offering different transfer distributions. But the international transfers corresponding to the equilibrium coalition structure $\mathbb{M}^*$ are renegotiation proof, because these transfers in any coalition sum up to zero, i.e. they are budget balanced, and so, there is always at least one member of the approval committee which would reject the renegotiation offer.

6 Allowing for green energy as well as fossil fuel

In this section, we generalise the IAM described in section 3 to include a green technology. The results in this section also generalise to the case of reversible agreements. We assume that the total energy that is used in the production of the final good, E_{iyt}, can be sourced from both fossil fuel and green energy. By green energy we mean energy that is not sourced from fossil fuels and does not produce any emissions, for example wind or solar energy. We assume that these two sources of energy are perfect substitutes, as for example in Harstad (2012) and Battaglini and Harstad (2016). We therefore replace equation (3.4) by

$$Y_{it} = \exp(-\gamma T_t) A_i K_{it}^{1-\nu} E_{iyt}^{\nu} \tag{6.1}$$

and

$$E_{iyt} = E_{it} + g_{it} \tag{6.2}$$

where g_{it} is green energy use in the production of the final good. As this is a dynamic game with $3N + 1$ stocks, it can be difficult to solve analytically. We therefore assume that, in line with the assumption regarding the depreciation of the physical capital stocks used in the final goods sector, the stocks of green technology depreciate fully by the end of each period so that the stock of green technology is equal to the investment in green technology, g_{it} in each country. We assume that the cost of investment in green energy is given by a quadratic cost function

$$B(g_{it}) = \frac{d_i}{2} g_{it}^2 \tag{6.3}$$

with constant $d_i > 0$, and we allow for heterogeneity with respect to d_i across countries, as well as heterogeneity with respect to K_{i0}, R_{i0}, and A_i. The feasibility constraint for the final good thus becomes

$$Y_{it} = C_{it} + K_{it+1} + B(g_{it}) \tag{6.4}$$

so that output of the final good must match consumption plus investments in capital used in the final goods and green energy sectors of each country i at t.

As before, at the beginning of period t countries may join climate coalitions. Signatories of coalition M decide cooperatively about the profile of their per-unit SCC in all periods $\tau \geq t$. Subsequently, in period t and all future periods, they decide about their investment in green technology, g_{it}, and their emissions, E_{it}, independently and simultaneously. We assume that in period t, after the negotiations, there is sufficient time for investment in green technology before the emission compliance time at the end of the period. If the agreements were reversible, then given our timing assumptions, full depreciation of the stock of green technology and without any technological spillover across the countries, there are no hold-up problems for green technology investments.

When the countries within coalition M agree on a per-unit SCC, they compare the marginal productivity of total energy use, E_{iyt}, to the marginal cost of emissions (i.e. the scarcity rent plus the SCC). Thus, if $E_{it} > 0$, by negotiating a SCC, the countries in effect negotiate the marginal productivity of total energy use, E_{iyt}. Then, each country $i \in M$ finds its optimal level of investment in green technologies by equating the marginal productivity of E_{iyt} to the marginal cost of g_{it}. Hence, by choosing its optimal green technology investment, it also pins down its optimal emission level because the two types of energy are perfect substitutes.

Note that there is no bang-bang equilibrium in which the countries only use fossil fuel and then after its stock is exhausted they switch to the green technology as the only source of energy. This is due to the non-linearity of the cost function for investment in green technologies. If the scarcity rent is sufficiently small, then the countries start with a phase of decarbonisation in which both $E_{it}(m)$ and $g_{it}(m)$ are positive. In other words, there is an interior solution for both sources of energy and the countries simultaneously use the

two types of energy. This is because they optimally set the marginal productivity of total energy use equal to the marginal cost of fossil fuel energy, i.e. $[\mu_{it}(1-s_{it})+\hat{\Lambda}(m)]Y_{it}$, which is equal to the marginal cost of green technology, i.e. $d_i g_{it}(m)$ too. This implies that in such a phase, the growth rate of output is equal to the difference of growth rate of investment in green technology and growth rate of scarcity rent. Thus, if growth rate of investment in green technology is larger than growth rate of scarcity rent per unit, i.e. $\frac{1}{\beta}-1$, then growth rate of output is positive.

The second phase of decarbonisation is when the scarcity rent per-unit is large, such that marginal productivity of total energy use is less than marginal cost of E_{it}. Thus the complementary slackness condition of first-order condition of E_{it} leads to setting $E_{it}=0$. The important difference here relative to the model without green technologies is that, in this phase, the scarcity rent per-unit remains finite. In other words, the countries do not extract the fossil fuel to a point that scarcity rent explodes, and they find it optimal to keep some of the oil under ground. Still according to Hotelling ruel, the total scarcity rent of their fossil fuel reserve, i.e. $\mu_{it}C_{it}$ grows with rate of marginal product of capital, $r_{it}=\frac{(1-\nu)Y_{it}}{K_{it}}$, but if green technology grows enough fast, the growth rate of output and thus the growth rate of consumption can remain positive, while μ_{it} remains finite.

See the analytical counterpart of this discussion in Appendix A.10, and the resulting solutions for $E_{it}>0$ and $g_{it}>0$ in the first phase of decarbonisation, given in the following proposition of which the proof is given in Appendix A.10.

Proposition 7 . *In the IAM with green and fossil fuel energy, emissions and investment in green technology in each country $i\in M$ are, respectively,*

$$\begin{aligned} E_{it}(m) &= \frac{\nu}{\mu_{it}(1-s_{it})+\hat{\Lambda}(m)} - g_{it} \\ &= \frac{\nu - Y_{it}(m)/d_i[\mu_{it}(1-s_{it})+\hat{\Lambda}(m)]^2}{\mu_{it}(1-s_{it})+\hat{\Lambda}(m)} \end{aligned} \tag{6.5}$$

$$g_{it}(m) = \frac{\mu_{it}(1-s_{it})+\hat{\Lambda}(m)}{d_i} \tag{6.6}$$

where $\hat{\Lambda}_{it}(m)\equiv(1-s_{it})\sum_{i\in M}\sum_{\tau=0}^{\infty}\frac{\beta^\tau\gamma\xi}{1-s_{it+\tau+1}}$.

The emission strategies and investments in green technology are dominant against what other coalitions choose. The intuition about this is as before. Furthermore, emissions decrease with investment in green technology, and also decrease in the per-unit SCC and the scarcity rent of fossil fuel reserves. It is easy to see that the derivative of E_{it} with respect to d_i is positive, and with respect to the output of final good, y_{it}, is $-g_{it}$. In other words, for any increase in the production of final output, total consumption of fossil fuel energy falls exactly by the amount of investment in green technology. Investments in green technology also increase in the negotiated per-unit SCC, and the

scarcity rent of fossil fuel reserves, and are lower for a higher investment cost parameter, d_i.

Here again $(1-s_{it})\sum_{i\in M}\sum_{\tau=0}^{\infty}\beta^{\tau}\gamma\xi\frac{1}{s_{it+\tau}}$ is the per-unit SCC, but it is not possible to find an analytical solution for the saving rate, because this now includes a quadratic term of the per-unit SCC. Let us assume for the sake of argument that the saving rate is constant, say at $\bar{s}$, which is what is assumed in the DICE model of Nordhaus (1993) and the RICE model of Nordhaus and Yang (1996) and is the case in equilibrium for the IAM of Golosov et al (2014). This assumption also resonates with Peters et al. (2009), which show that in the long run the saving rate is relatively constant over time. Given this assumption, the per-unit SCC is $\hat{\Lambda}_{it}(m)=\hat{\Lambda}(m)=\frac{\gamma\xi m}{1-\beta}$ for all $i\in M$ at any time t.[36]

The countries which participate in climate negotiation are assumed to have a finite μ_{it}. Here, the nature of heterogeneity with respect to d_i is similar to K_{i0} and A_i, as the independence of (4.6) from d_i does not require the countries to be patient in the limit. However, here it is not the linearity of the optimum value function in d_i, but because mathematically, the sum of fossil fuel energy and green technology investment in equilibrium is equal to the total energy consumption in the model without green technology as described in section 4. The only difference is that here the per-unit scarcity rent remains finite. Hence, the reduced form of E_{iyt} is independent of d_i, and thus the equilibrium numerical coalition structure can be characterised independent of this sources of heterogeneity. Thus, in addition to Proposition 3, we get the following proposition.

Proposition 8 . *If we assume that the saving rate is constant with green energy as perfect substitute for the fossil fuel, the membership decision of countries leads to a coalition of maximum three members under cartel stability. Under farsighted stability, the equilibrium numerical coalition structure $\mathcal{M}^*$ can be characterised independent of the heterogeneity with respect to d_i, K_{i0}, A_i, R_{i0} and μ_{it}. Furthermore, the grand coalition occurs in equilibrium if the total number of countries, N, is a member of the Tribonacchi set, i.e. $\mathcal{T}^*=\{1,2,4,7,13,24,...\}$. If $N\in\mathcal{T}^*$, then $\mathcal{M}^*=\{N\}$; if $N\notin\mathcal{T}^*$, then $\mathcal{M}^*=D(N)$, given $\mathcal{T}^*$.*

The proof is given in Appendix A.11. Intuitively, as mentioned earlier, the countries at the negotiation stage ensure that total energy consumption of all coalition members is set optimally, regardless of how they split total energy between green energy and the fossil fuel. This indeed leads to our analysis of coalition formation in section 4 and hence our previous results go through. Proposition 8 implies that the equilibrium numerical coalition structure can be characterised in the same way described in the model without the green technologies. The number of signatories of climate coalitions in equilibrium is

[36]The exact value of saving rate and its dependence on the parameters of the model, does not have any impact on membership decisions of the countries, as long as $0<\bar{s}<1$.

a Tribonacci number again.

7 Conclusion

We have examined the formation of climate coalitions with heterogeneous countries and have put forward an approach to characterise the equilibrium numerical coalition structure of countries independent of their heterogeneity. We have fully characterised the unique equilibrium number of coalitions and their number of signatories. No matter how much countries differ in their development (initial capital stocks or total factor productivities), costs of green technology, or in their stocks of initial fossil fuel reserves, we have obtained unique predictions for these numbers.

We have shown that farsighted countries, which foresee the consequences of their climate membership decisions, form international climate treaties where the number of participating countries to a climate treaty is a Tribonacci number in equilibrium. The alignment of our results with phenomena in nature which follow numbers from Fibonacci sequences is not a matter of accident. Our result follows from adopting the solution concept of farsightedness rather than the more commonly used cartel or Nash stability concept. Furthermore, relative to the literature on international environmental agreements (IEAs), we capture various aspects of the incentives of countries that participate in international climate negotiations, by integrating an IEA with an integrated assessment model of the economy and global warming (IAM). Our analysis takes account of the general equilibrium features of the economy of each country and the resulting saving decisions and management of their exhaustible fossil fuel resources, as well as the climate dynamics of their emissions.

Given our results on the number of signatories being a Tribonacci, we offer a tractable algorithm to characterise the equilibrium coalition structure for any number of heterogeneous countries participating in international climate negotiations. We have shown that our results are robust if agreements are renegotiable, i.e. if countries can walk away from an agreement and renegotiate it. Furthermore, our results are robust if countries can invest in green technologies too.

In both the prevailing literature on IEAs and in practice, there is much focus on forming a single climate coalition despite that it is a fragile coalition and not ambitious enough. We have therefore departed from the cartel or Nash equilibrium solution concept of climate coalition formation and allow for the formation of multiple coalitions. We have shown that if the grand coalition does not form in equilibrium, multiple climate coalitions can form with different levels of ambitions regarding their emission mitigation strategies. Furthermore, as the number of signatories to the treaties must be a Tribonacci number, this number can be large and in particular much larger than three, than the more commonly used cartel stability solution concept predicts. For a world of 195 countries

our results imply that the average optimal social cost of carbon is 120 times larger under the farsightedness than what the cartel stability solution concept would result.

The only link among the countries in our IAM is the climate damages in their production function which depend on emissions in all countries. Thus, in future research other factors connecting the countries can be investigated. These might include international trade for fossil fuel, an international capital market, or international labour migration. Another line of research is to see how international climate treaties are hampered by political economy constraints on the size of international transfers between countries. Finally, our results rely on the observability of actions at the compliance stage in each period. This rules out any scope for strategic uncertainty about emissions. In practice, this assumption resembles the increasing emphasis of countries participating in climate negotiations on transparency of emissions and abatement actions. Accordingly, an important achievement of the Paris agreement has been to create a framework to improve transparency of emission levels of each country. The Task Force, a working group of the Intergovernmental Panel on Climate Change, is responsible for developing and implementing a unified methodology in measuring and reporting emissions and abatement of each country.

References

Acemoglu, D. 2009. Introduction to Modern Economic Growth (New Jersey, Woodstock, Oxfordshire, Princeton University Press).

Allen, Myles R, David J Frame, Chris Huntingford, Chris D Jones, Jason A Lowe, Malte Meinshausen, and Nicolai Meinshausen. 2009. Warming caused by cumulative carbon emissions towards the trillionth tonne. Nature 458 (7242): 1163–1166.

Anthoff, David, and Richard SJ Tol. 2013. The uncertainty about the social cost of carbon: A decomposition analysis using fund. Climatic Change 117 (3): 515–530.

Arrow, Kenneth J, Partha Dasgupta, and Karl-Göran Mäler. 2003. Evaluating projects and assessing sustainable development in imperfect economies. Environmental and Resource Economics 26 (4): 647–685.

Asheim, Geir B, Camilla Bretteville Froyn, Jon Hovi, and Fredric C Menz. 2006. Regional versus global cooperation for climate control. Journal of Environmental Economics and Management 51 (1): 93–109.

Aumann, RJ, and RB Myerson. 1988. Endogenous Formation of Links Between Coalitions and Players: An Application of the Shapley ValueB. The Shapley Value, Cambridge University Press, Cambridge.

Barrage, Lint. 2014. Sensitivity Analysis for Golosov, Hassler, Krusell, and Tsyvinski (2013): Optimal Taxes on Fossil Fuel in General Equilibrium. Econometrica Supplementary Material.

Barrett, Scott. 1994. Self-enforcing international environmental agreements. Oxford economic papers, pp. 878–894.

Barrett, Scott. 2006. Climate treaties and "breakthrough" technologies. American Economic Review 96 (2): 22–25.

Battaglini, Marco, and Bård Harstad. 2016. Participation and duration of environmental agreements. Journal of Political Economy 124 (1): 160–204.

Benchekroun, Hassan, and Ngo Van Long. 2012. Collaborative Environmental Management: A Review Of The Literature. International Game Theory Review 14, no. 04.

Bloch, Francis. 1996. Sequential formation of coalitions with fixed payoff division and externalities. Games and Economic Behavior 14 (0043): 90–123.

Bosetti, Valentina, Carlo Carraro, Enrica De Cian, Emanuele Massetti, and Massimo Tavoni. 2013. Incentives and stability of international climate coalitions: An integrated assessment. Energy Policy 55:44–56.

Bremer, Ton van den, and Frederick van der Ploeg. 2021. The Risk-Adjusted Carbon Price. American Economic Review 111:2782–2810.

Brock, William A, and Leonard J Mirman. 1973. Optimal economic growth and uncertainty: the no discounting case. International Economic Review, pp. 560–573.

Buchner, Barbara, and Carlo Carraro. 2009. Parallel climate blocs, incentives to cooperation in international climate negotiations. The Design of Climate Policy, pp. 137–163.

Campbell, Sarah C. 2020. Growing patterns: Fibonacci numbers in nature. Astra Publishing House.

Carraro, Carlo, Johan Eyckmans, and Michael Finus. 2006. Optimal transfers and participation decisions in international environmental agreements. The Review of International Organizations 1 (4): 379–396.

Carraro, Carlo, and Domenico Siniscalco. 1993. Strategies for the international protection of the environment. Journal of Public Economics 52 (3): 309–328.

Chatterjee, Kalyan, Bhaskar Dutta, Debraj Ray, and Kunal Sengupta. 1993. A noncooperative theory of coalitional bargaining. The Review of Economic Studies 60 (2): 463–477.

Chetty, Raj. 2006. A new method of estimating risk aversion. American Economic Review 96 (5): 1821–1834.

Chwe, Michael Suk-Young. 1994. Farsighted coalitional stability. Journal of Economic theory 63 (2): 299–325.

Darwin, Charles. 1859. On the Origin of Species by Means of Natural Selection. London: Murray. or the Preservation of Favored Races in the Struggle for Life.

d'Aspremont, Claude, Alexis Jacquemin, Jean Jaskold Gabszewicz, and John A Weymark. 1983. On the stability of collusive price leadership. Canadian Journal of economics, pp. 17–25.

Davis, Harry F. 1989. Fourier series and orthogonal functions. Courier Corporation.

Diamantoudi, Effrosyni, and Eftichios S Sartzetakis. 2006. Stable international environmental agreements: An analytical approach. Journal of public economic theory 8 (2): 247–263.

Dietz, Simon, and Nicholas Stern. 2015. Endogenous growth, convexity of damage and climate risk: how Nordhaus' framework supports deep cuts in carbon emissions. The Economic Journal 125 (583): 574–620.

Dietz, Simon, Frederick van der Ploeg, Armon Rezai, and Frank Venmans. 2021. Are economists getting climate dynamics right and does it matter? Journal of the Association of Environmental and Resource Economists 8 (5): 895–921.

Dixit, Avinash, and Mancur Olson. 2000. Does voluntary participation undermine the Coase Theorem? Journal of public economics 76 (3): 309–335.

Dutta, Bhaskar, Debraj Ray, Kunal Sengupta, and Rajiv Vohra. 1989. A consistent bargaining set. Journal of economic theory 49 (1): 93–112.

Dutta, Bhaskar, and Rajiv Vohra. 2017. Rational expectations and farsighted stability. Theoretical Economics 12 (3): 1191–1227.

Eyckmans, Johan, and Henry Tulkens. 2006. Simulating coalitionally stable burden sharing agreements for the climate change problem. In Public goods, environmental externalities and fiscal competition, 218–249. Springer.

Farrell, Joseph, and Eric Maskin. 1989. Renegotiation-proof equilibrium: Reply. Journal of Economic Theory 49 (2): 376–378.

Finus, Michael, and Stefan Maus. 2008. Modesty may pay! Journal of Public Economic Theory 10 (5): 801–826.

Finus, Michael, and Bianca Rundshagen. 2003. Endogenous coalition formation in global pollution control: a partition function approach. Endogenous Formation of Economic Coalitions, Edward Elgar, Cheltenham, UK, pp. 199–243.

Finus, Michael, and Bianca Rundshagen. 2009. Membership rules and stability of coalition structures in positive externality games. Social Choice and Welfare 32 (3): 389–406.

Finus, Michael, M Elena Saiz, and Eligius MT Hendrix. 2009. An empirical test of new developments in coalition theory for the design of international environmental agreements. Environment and Development Economics 14 (1): 117–137.

Gandelman, Nestor, and Rubén Hernández-Murillo. 2015. Risk aversion at the country level.

Gersbach, Hans, Noemi Hummel, and Ralph Winkler. 2021. Long-Term Climate Treaties with a Refunding Club. Environmental and Resource Economics, pp. 1–42.

Golosov, Mikhail, John Hassler, Per Krusell, and Aleh Tsyvinski. 2014. Optimal taxes on fossil fuel in general equilibrium. Econometrica 82 (1): 41–88.

Harstad, Bård. 2012a. Buy coal! A case for supply-side environmental policy. Journal of Political Economy 120 (1): 77–115.

Harstad, Bård. 2012b. Climate contracts: A game of emissions, investments, negotiations, and renegotiations. Review of Economic Studies 79 (4): 1527–1557.

Hassler, John, and Per Krusell. 2012. Economics and climate change: integrated assessment in a multi-region world. Journal of the European Economic Association 10 (5): 974–1000.

Hassler, John, Per Krusell, and Conny Olovsson. 2021. Directed Technical Change as a Response to Natural Resource Scarcity. Journal of Political Economy 129 (11): 000–000.

Hoel, Michael, and Kerstin Schneider. 1997. Incentives to participate in an international environmental agreement. Environmental and Resource economics 9 (2): 153–170.

Hoel, Michael, and Aart de Zeeuw. 2010. Can a focus on breakthrough technologies improve the performance of international environmental agreements? Environmental and Resource Economics 47 (3): 395–406.

Hong, Fuhai, and Larry Karp. 2012. International environmental agreements with mixed strategies and investment. Journal of Public Economics 96 (9-10): 685–697.

Hope, Chris. 2011. The PAGE09 integrated assessment model: A technical description. Cambridge Judge Business School Working Paper 4, no. 11.

Hyndman, Kyle, and Debraj Ray. 2007. Coalition formation with binding agreements. The Review of Economic Studies 74 (4): 1125–1147.

Jones, Charles I. 2005. The shape of production functions and the direction of technical change. The Quarterly Journal of Economics 120 (2): 517–549.

Karp, Larry, and Leo Simon. 2013. Participation games and international environmental agreements: A non-parametric model. Journal of Environmental Economics and Management 65 (2): 326–344.

Kortum, Samuel S. 1997. Research, patenting, and technological change. Econometrica: Journal of the Econometric Society, pp. 1389–1419.

Lessmann, Kai, Ulrike Kornek, Valentina Bosetti, Rob Dellink, Johannes Emmerling, Johan Eyckmans, Miyuki Nagashima, Hans-Peter Weikard, and Zili Yang. 2015. The stability and effectiveness of climate coalitions. Environmental and Resource Economics 62 (4): 811–836.

Lessmann, Kai, Robert Marschinski, and Ottmar Edenhofer. 2009. The effects of tariffs on coalition formation in a dynamic global warming game. Economic Modelling 26 (3): 641–649.

Maskin, Eric, and Jean Tirole. 2001. Markov perfect equilibrium: I. Observable actions. Journal of Economic Theory 100 (2): 191–219.

Matthews, H Damon, Nathan P Gillett, Peter A Stott, and Kirsten Zickfeld. 2009. The proportionality of global warming to cumulative carbon emissions. Nature 459 (7248): 829–832.

Miller, Eric. 2008. An assessment of CES and Cobb-Douglas production functions. Congressional Budget Office Washington, DC.

Minarova, Nikoletta. 2014. The fibonacci sequence: Nature's little secret. CRIS-Bulletin of the Centre for Research and Interdisciplinary Study 2014 (1): 7–17.

Nordhaus, William. 2014. Estimates of the social cost of carbon: concepts and results from the DICE-2013R model and alternative approaches. Journal of the Association of Environmental and Resource Economists 1 (1/2): 273–312.

Nordhaus, William. 2015. Climate clubs: Overcoming free-riding in international climate policy. American Economic Review 105 (4): 1339–70.

Nordhaus, William D. 1993. Optimal greenhouse-gas reductions and tax policy in the" DICE" model. The American Economic Review 83 (2): 313–317.

Nordhaus, William D, and Zili Yang. 1996. A regional dynamic general-equilibrium model of alternative climate-change strategies. The American Economic Review, pp. 741–765.

Peters, Michael, Alp Simsek, and Daron Acemoglu. 2009. Solutions manual for introduction to modern economic growth. Princeton University Press Princeton, NJ.

Ploeg, Frederick van der, and Armon Rezai. Forthcoming. Optimal carbon pricing in general equilibrium. Journal of Environmental Economics and Management.

Plouffe, Simon. 1993. Exact formulas for integer sequences. Notes.

Ramsey, Frank Plumpton. 1928. A mathematical theory of saving. The economic journal 38 (152): 543–559.

Ray, Debraj. 2007. A game-theoretic perspective on coalition formation. Oxford University Press.

Ray, Debraj, and Rajiv Vohra. 1997. Equilibrium binding agreements. Journal of Economic theory 73 (1): 30–78.

Ray, Debraj, and Rajiv Vohra. 1999. A theory of endogenous coalition structures. Games and economic behavior 26 (2): 286–336.

Ray, Debraj, and Rajiv Vohra. 2001. Coalitional power and public goods. Journal of Political Economy 109 (6): 1355–1384.

Ray, Debraj, and Rajiv Vohra. 2019. Maximality in the farsighted stable set. Econometrica 87 (5): 1763–1779.

Romer, Paul M. 1986. Increasing returns and long-run growth. Journal of political economy 94 (5): 1002–1037.

Rosen, Kenneth H. Discrete Mathematics and Its Applications.

Sinha, Sudipta. 2017. The Fibonacci Numbers and Its Amazing Applications. International Journal of Engineering Science Invention 6 (9): 7–14.

Stern, Nicholas, and Nicholas Herbert Stern. 2007. The economics of climate change: the Stern review. cambridge University press.

Tol, Richard SJ. 2001. Climate coalitions in an integrated assessment model. Computational Economics 18 (2): 159–172.

Vosooghi, Sareh. 2021. Information Design In Coalition Formation Games.

Yang, Zili, et al. 2008. Strategic bargaining and cooperation in greenhouse gas mitigations: an integrated assessment modeling approach. MIT Press.

Yi, Sang-Seung, and Hyukseung Shin. 2000. Endogenous formation of research coalitions with spillovers. International Journal of Industrial Organization 18 (2): 229–256.

Zeeuw, Aart de. 2008. Dynamic effects on the stability of international environmental agreements. Journal of environmental economics and management 55 (2): 163–174.

A Appendix

A.1 The decision making of a signatory in the action stage

Since every country $i \in M$ internalises its emission that affects payoffs of other members in coalition M in any period $\tau \geq t$, using the Lagrange method, the problem of planner of country $i \in M$ can be written as:

$$\max_{\{E_{it+\tau}\}_{\tau=0}^{\infty}} \sum_{i \in M} \sum_{\tau=0}^{\infty} \beta^{\tau} \ln(C_{it+\tau}) \tag{A.1}$$

subject to (3.2), (3.4), (3.5), (3.6), (3.8) and non-negativity constraints.

By Walras law, the feasibility constraint of the final good and the resource constraint are the only market-clearing conditions to be checked. Let $\beta^{\tau}\lambda_{it+\tau}$ be present value Lagrange multiplier for final output feasibility constraint (3.5), $\beta^{\tau}\mu_{it+\tau}$ be present value Lagrange multiplier for resource constraint in (3.2) and $\beta^{\tau}\zeta_{ijt+\tau}$ be present value Lagrange multiplier associated with non-negativity constraints. The first-order condition of E_{it} gives:

$$\lambda_{it}[\frac{\nu Y_{it}}{E_{it}}] - \sum_{i \in M} \sum_{\tau=0}^{\infty} \lambda_{it+\tau}\beta^{\tau}\gamma\xi Y_{it+\tau} = \mu_{it} \tag{A.2}$$

Then the planner of each country independently decides about the consumption, investment in the capital stock and the resource extraction in country i, i.e.

$$\max_{\{C_{it+\tau}, K_{it+\tau+1}, R_{it+\tau+1}\}_{\tau=0}^{\infty}} \sum_{\tau=0}^{\infty} \beta^{\tau} \ln(C_{it+\tau}) \tag{A.3}$$

again, subject to (3.2), (3.4), (3.5), (3.6), (3.8) and the non-negativity constraints. The first-order condition of C_{it} gives:

$$\lambda_{it} = \frac{1}{C_{it}(M, \mathbb{M})} \tag{A.4}$$

The first-order condition of K_{it+1} and (A.4) give the Euler equation of consumption:

$$\frac{1}{C_{it}(M, \mathbb{M})} = \beta \frac{1}{C_{it+1}(M, \mathbb{M})} \frac{Y_{it+1}(M, \mathbb{M})}{K_{it+1}(M, \mathbb{M})}(1 - \nu) \tag{A.5}$$

Using $C_{it}(M, \mathbb{M}) = (1 - s_{it})Y_{it}(M, \mathbb{M})$, and therefore $K_{it+1}(M, \mathbb{M}) = s_{it}Y_{it}(M, \mathbb{M})$, the Euler equation reduces to:

$$\frac{s_{it}}{1 - s_{it}} = \beta \frac{1}{1 - s_{it+1}}(1 - \nu) \tag{A.6}$$

The unique solution to this problem is: $s_{it} = s = \beta(1 - \nu)$, for all t and all i.

Given the first-order condition of emissions in equation (A.2), and using equation (A.4), we have $(1-s_{it})\sum_{i\in M}\sum_{\tau=0}^{\infty}\beta^{\tau}\gamma\xi\frac{1}{1-s_{it+\tau}}$ as the per-unit SCC of each member of coalition of size m. Given the constant saving rate, equation (A.2) simplifies to:

$$\frac{\nu}{E_{it}(m)}-\hat{\Lambda}_i(m)=[1-\beta(1-\nu)]\mu_{it} \tag{A.7}$$

where $\hat{\Lambda}_{it}(m)=\hat{\Lambda}(m)\equiv\frac{\xi\gamma m}{1-\beta}$ is the SCC of each member of coalition of size m; and μ_{it} is the shadow value of the resource. This equation can be re-arranged to

$$E_{it}(m)=\frac{\nu}{\mu_{it}[1-\beta(1-\nu)]+\hat{\Lambda}(m)} \tag{A.8}$$

Equation (A.14) shows the solution of energy consumption in each country in coalition M. Under the assumption of symmetry and strict concavity of the final output and utility function, the equilibrium emission strategy, $E_{it}(m)$, is unique and is the same for all members of the coalition.

Finally the first-order condition for R_{t+1} is $\mu_{it}=\beta\mu_{it+1}$. Using this and equation (A.7), give the Euler equation of energy consumption:

$$\frac{\nu}{E_{it}(m)}-\hat{\Lambda}(m)=\beta(\frac{\nu}{E_{it+1}(m)}-\hat{\Lambda}(m)) \tag{A.9}$$

which simplifies to:

$$\frac{\beta\nu E_{it}(m)}{\nu-(1-\beta)\hat{\Lambda}(m)E_{it}(m)}=E_{it+1}(m) \tag{A.10}$$

It can be shown that

$$E_{it+\tau}(m)=\frac{\beta^{\tau}\nu E_{it}(m)}{\nu-(1-\beta^{\tau})\hat{\Lambda}(m)E_{it}(m)} \tag{A.11}$$

for any $\tau\geq 1$.

A.2 Two benchmarks

The first benchmark corresponds to the singleton coalition structure, i.e. if all $m=1$, where the strategies of the countries coincide with the non-cooperative emission level and the planner of each country chooses its energy consumption non-cooperatively. The unique MPE level of emissions is given in the following corollary.

Corollary 2 . *The non-cooperative SCC is* $\hat{\Lambda}_{it}(1)=\hat{\Lambda}(1)\equiv\frac{\gamma\xi}{1-\beta}$ *for all* t *and* i. *The corresponding fossil fuel use or equivalently emissions are*

$$E_{it}(1)=\nu/[\mu_{it}(1-\beta(1-\nu))+\hat{\Lambda}(1)] \tag{A.12}$$

The second benchmark corresponds to the grand coalition, where $m = N$, and policies are set to the internationally cooperative level corresponding to the global social optimum, which a hypothetical utilitarian supra-national planner would choose in our multi-country setting. She would thus maximise life-time utility of the sum of representative households of all countries,

$$\sum_{i \in I} \sum_{t=0}^{\infty} \beta^t \{\ln(C_{it})\} \tag{A.13}$$

subject to constraints (3.5) and (3.2) for each i.

Corollary 3 . *If a grand coalition forms, the optimal international cooperative SCC is* $\hat{\Lambda}_{it}(N) = \hat{\Lambda}(N) \equiv \frac{\gamma \xi N}{1-\beta}$ *for all* t*, and* i*. The corresponding level of fossil fuel consumption or equivalently flow of emissions is*

$$E_{it}(N) = \nu / [\mu_{it}(1 - \beta(1 - \nu)) + \hat{\Lambda}(N)] \tag{A.14}$$

Note that $\hat{\Lambda}(N) \geq \hat{\Lambda}(m) \geq \hat{\Lambda}(1)$, since the larger the coalition, the larger the per-unit SCC and the smaller energy use and emissions.

A.3 Optimum value function of a signatory

Let $V_i(S_t, K_{it}, \mu_{it}, M, \mathbb{M})$ be the optimum value function of a signatory in coalition M of size m in coalition structure $\mathbb{M}$. By substituting the solutions in the summation of flow and continuation utility of the representative consumer of country $i \in M$, we obtain:

$$\begin{aligned} V_i(S_t, K_{it}, \mu_{it}, M, \mathbb{M}) &= ln(C_{it}(M, \mathbb{M})) + \beta ln(C_{it+1}(M, \mathbb{M})) + ... \\ &= \frac{ln(1-s)}{1-\beta} + \{ln(Y_{it}(M, \mathbb{M})) + \beta ln(Y_{it+1}(M, \mathbb{M})) + ...\} \\ &= \frac{ln(1-s)}{1-\beta} + \{ln[e^{-\gamma \xi S_t - \gamma T_0} A_i K_{it}^{1-\nu} E_{it}(m)^{\nu}] \\ &+ \beta ln[e^{-\gamma \xi S_{t+1} - \gamma T_0} A_i K_{it+1}(M, \mathbb{M})^{1-\nu} E_{it+1}(m)^{\nu}] + ...\} \\ &= \frac{(1-\nu) ln(K_{it}) + H_1 + H_2 + H_3}{1-s} \end{aligned} \tag{A.15}$$

where H_j are defined as

$$H_1 \equiv \frac{s ln(s) - s ln(1-s) + ln(A_i) - \gamma T_0}{1-\beta} \tag{A.16}$$

$$H_2 \equiv -\gamma \xi [S_t + \beta S_{t+1} + \beta^2 S_{t+2} + ...] \tag{A.17}$$

and

$$H_3 \equiv \nu[ln(E_{it}(m)) + \beta ln(E_{it+1}(m)) + \beta^2 ln(E_{it+2}(m)) + ...] \tag{A.18}$$

The second expression can be expanded to a function of the summation of the past, current, future emissions of all countries:

$$\begin{aligned} H_2 = -\frac{\gamma\xi}{1-\beta}\{&\sum_i \sum_{s=1}^{t} E_{it-s} + \sum_{j\notin M} E_{jt} + \sum_{i\in M} E_{it}(m) \\ &+ \sum_{j\notin M}[\beta E_{jt+1} + \beta^2 E_{jt+2} + ...] + \sum_{i\in M}[\beta E_{it+1}(m) + \beta^2 E_{it+2}(m) + ...]\} \end{aligned} \tag{A.19}$$

A.4 The small-coalition paradox under cartel stability

Assume that there is a single coalition M of m countries, and the $N-m$ non-signatories form the fringe. Furthermore, in this section, we assume that the countries are ex-ante symmetric, but after the membership stage they may end up in asymmetric situations.

Definition 5 . *Cartel stability is a state at which no coalition member wishes to leave the coalition (internal stability), and no fringe country wishes to join the coalition (external stability).*

In our model the external stability condition is automatically satisfied whenever $m^* > 1$, because the non-participating countries always gain from free riding and have no incentives for joining the climate coalition.

For the internal stability condition, it is sufficient to check a one-shot deviation. Hence, a coalition of size m is internally stable if the continuation payoff of a signatory is greater or equal to the payoff of a one-shot deviation plus the continuation payoff following the deviation. This condition is a concave function of m. The coalition sizes at which the internal stability condition binds with equality, determine the lower bound and the upper bound of equilibrium coalition sizes, m^*.

As a Nash equilibrium, the internal and external stability conditions imply that the deviating country takes the actions of the other players as given. Here, we follow Battaglini and Harstad (2016), who suggest a more generalised version of this stability condition. They assume that upon a deviation of one period, the remaining participants update their joint climate policies as if $m = m^* - 1$, and then again return to the equilibrium path.[37] This implies that if a country that is supposed to be a signatory considers a deviation, in that period it chooses its best response to the strategy of others. Then, the country will be expected to join the coalition next period. This deviation will therefore affect aggregate emissions and thus the continuation payoff of all countries for ever. As explained above, countries have dominant strategies, thus and the reaction

[37]This is more general than the conventional internal stability which does not require any update of strategies by the remaining signatories upon a deviation by a country.

function of the deviating country is not affected by the number of signatories and it leads to the non-cooperative emission level.

Proposition 9 . *Under the assumption of cartel stability, the largest coalition size is* $m^* = 3$ *for any total number of countries* N*.*

To see this note that a signatory does not have any incentive to leave coalition M of size m if

$$V_i(S_t, K_{it}, \mu_{it}, M) \geq ln(C_{it}^d) + \beta\{V_i(E^t, K_{it+1}, \mu_{it}, M)\} \tag{A.20}$$

where $V_i(S_t, K_{it}, \mu_{it}, M)$ is the optimum value function of a signatory in coalition M as defined in section A.3, and $E^t \equiv (E_t, E_{t-1}, ..., E_0)$. Furthermore, C_{it}^d is the consumption level associated with the deviation period. Note that E^t in $V_i(E^t, K_{it+1}, \mu_{it}, M)$ is impacted by the deviation in period t. More specifically, the right-hand-side of equation (A.20) consists of

$$ln(C_{it}^d) = ln(1-s) + ln(Y_{it}^d)$$

and

$$V_i(E^t, K_{it+1}, \mu_{it}, M) = \frac{(1-\nu)ln(K_{it+1}) + H_1 + H_2' + H_3'}{1-s} \tag{A.21}$$

Accordingly, $ln(K_{it+1}) = ln(1-s) + ln(Y_{it}^d)$, and

$$\begin{aligned} H_2' &\equiv -\gamma\xi[(S_{t+1}) + \beta(S_{t+2}) + ...] \\ &= -\frac{\gamma\xi}{1-\beta}\{\sum_i \sum_{s=1}^{t} E_{it-s} + \sum_{j\notin M\setminus i} E_{jt} + \sum_{j\in M\setminus i} E_{jt}(m-1) \\ &+ \sum_{i\notin M}[E_{it+1} + \beta E_{it+2} + ...] + \sum_{i\in M}[E_{it+1}(m) + \beta E_{it+2}(m) + ...]\} \end{aligned} \tag{A.22}$$

and

$$H_3' \equiv \nu[ln(E_{it+1}(m)) + \beta ln(E_{it+2}(m)) + ...] \tag{A.23}$$

By multiplying both sides of the internal stability condition (A.20) by $1-s$, and cancelling all future emissions of H_2 and H_2' from both sides, and using the symmetry of the emission strategies of signatories and likewise for non-signatories, and the fact that

$$\begin{aligned} ln(Y_{it}^d) = -\gamma\xi\{\sum_i \sum_{s=1}^{t} E_{it-s} + \sum_{j\notin M\setminus i} E_{jt} + \sum_{j\in M\setminus i} E_{jt}(m-1)\} \\ + ln(A_i) + (1-\nu)ln(K_{it}) + \nu ln(E_{it}) \end{aligned} \tag{A.24}$$

we can simplify equation (A.20) to

$$\nu[ln(E_{it}(m)) - ln(E_{it})] + \frac{\gamma\xi}{1-\beta}E_{it} - \frac{\gamma\xi m}{1-\beta}E_{it}(m) + \frac{\gamma\xi(m-1)}{1-\beta}E_{it}(m-1) \geq 0 \quad \text{(A.25)}$$

Using the corresponding emission levels, this condition can be written as a function of variable m and parameters β, ν, γ and the shadow value of resource μ_{it}. Thus the roots of equation (A.25) give the equilibrium coalition size m^*. For any parameter values, the roots are one and maximum three.

Lastly, the internal stability condition is independent of the capital stock or the stock of cumulative emissions or of other state variables but it does depend on current emission levels. Thus it depends on the scarcity rents which in turn indirectly depend on the stocks of fossil fuel. In particular, for low initial stocks of fossil fuel reserves and large values of the per-unit scarcity rent, the stable number of signatories may reduce to one i.e. no coalition can be stable.

A.5 Proof of Proposition 3

The equilibrium coalition structure needs to be defined recursively to ensure the self-enforceability of any deviation and any resulting coalition. Suppose j is the initial proposer. For any N, country j compares the total payoff of the best profitable deviation by forming coalition $M \in \{M_1, M_2, ..., M_k\}$ (which is to be identified) versus the total payoff of the corresponding m members from staying in the grand coalition $\{I\}$. Thus, the planner of country j needs to determine the sign of

$$\sum_{i=1}^{m} V_i^j(S_t, K_{it}, \mu_{it}, M, \mathbb{M}) - \sum_{i=1}^{m} V_i^j(S_t, K_{it}, \mu_{it}, I) \quad \text{(A.26)}$$

Assume the countries are heterogeneous with respect to K_{i0} and/or A_i. But this equation is independent of stocks, in particular independent of K_{i0}. It is also independent of total factor productivities A_i. To see this, note that

$$\begin{aligned} V_i(M,&\{M_1, M_2, ..., M_k\}) - V_i(\{I\}) = \\ &\frac{1}{1-\beta(1-\nu)}\Big\{\nu[ln(\frac{E_{it}(M)}{E_{it}(I)}) + \beta ln(\frac{E_{it+1}(M)}{E_{it+1}(I)}) + ...] \\ &- \frac{\gamma\xi}{1-\beta}\{[\sum_{i\in M_1} E_{it}(M_1) + \sum_{i\in M_2} E_{it}(M_2) + ... + \sum_{i\in M_k} E_{it}(M_k) - \sum_{i\in I} E_{it}(I)] + \\ &\beta[\sum_{i\in M_1} E_{it+1}(M_1) + \sum_{i\in M_2} E_{it+1}(M_2) + ... + \sum_{i\in M_k} E_{it+1}(M_k) - \sum_{i\in I} E_{it+1}(I)] + ...\}\Big\} \end{aligned} \quad \text{(A.27)}$$

Sets $M_p \in \{M_1, M_2, ..., M_k\}$ are included if they are non-empty. Thus, if the source of heterogeneity is either K_{i0} or A_i, then the membership decision of countries in (A.26) is not affected by the heterogeneity across countries.

Furthermore, the above equation is only function of emissions, which only depend on m and $\mathcal{M}$ and not M and $\mathbb{M}$. Hence, $\mathcal{M}^*$ can be characterised independent of heterogeneity with respect to K_{i0} or A_i.

If the countries are heterogeneous with respect to R_{i0} and μ_{it}, then in the limit that $\beta \to 1$, the equation in (A.27) converges to

$$\begin{aligned}\lim_{\beta\to 1} V_i(M,\{M_1, M_2, ..., M_k\}) - V_i(\{I\}) = \\ [ln(\frac{N}{m}) + ln(\frac{N}{m}) + ...] - \{[k-1] + [k-1] + ...\}\end{aligned} \tag{A.28}$$

This equation is independent of μ_{it} of any country and any stocks. Moreover, it only depends on m and $\mathcal{M}$. □

A.6 Proof of Lemma 1

Consider N countries, where the decomposition of N is $D(N) = \{m_1, m_2, ..., m_k\}$, such that $m_1 < m_2 < ... < m_k$. In a public good game, the most profitable and self-enforceable deviation from the grand coalition would lead to $V_i(m_1, \{m_1, m_2, ..., m_k\})$. According to Ray and Vohra (2001) a sufficient condition for the formation of the grand coalition is

$$V_i(m_1, \{m_1, m_2, ..., m_k\}) - V_i(\{N\}) < 0 \tag{A.29}$$

Since the heterogeneity does not affect the equilibrium numerical coalition structure, the above condition is a sufficient condition for every country to prefer the grand coalition to any other coalition structure. In our model, it can be shown that

$$\begin{aligned}V_i(m_1,\{m_1, m_2, ..., m_k\}) - V_i(\{N\}) = \\ \frac{1}{1-\beta(1-\nu)}\Big\{\nu[ln(\frac{E_{it}(m_1)}{E_{it}(N)}) + \beta ln(\frac{E_{it+1}(m_1)}{E_{it+1}(N)}) + ...] \\ - \frac{\gamma\xi}{1-\beta}\{[\sum_{i\in M_1} E_{it}(m_1) + \sum_{i\in M_2} E_{it}(m_2) + ... + \sum_{i\in M_k} E_{it}(m_k) - \sum_{i\in I} E_{it}(N)] + \\ \beta[\sum_{i\in M_1} E_{it+1}(m_1) + \sum_{i\in M_2} E_{it+1}(m_2) + ... + \sum_{i\in M_k} E_{it+1}(m_k) - \sum_{i\in I} E_{it+1}(N)] + ...\}\Big\}\end{aligned} \tag{A.30}$$

Note that emission of each country in equation (A.14) can be written as

$$E_{it}(m) = \frac{\nu(1-\beta)}{(1-\beta)\mu_{it}[1-\beta(1-\nu)] + m\gamma\xi}$$

For $\beta \to 1$, the inequality in equation (A.29) converges to

$$\begin{aligned}\lim_{\beta\to 1} V_i(m_1,\{m_1, m_2, ..., m_k\}) - V_i(\{N\}) = \\ \nu[ln(\frac{N}{m_1}) + ln(\frac{N}{m_1}) + ...] - \nu\{[k-1] + [k-1] + ...\} < 0\end{aligned} \tag{A.31}$$

This is satisfied if

$$ln(\frac{N}{m_1}) < (k-1) \tag{A.32}$$

as required. □

A.7 Proof of Proposition 4

Let $P(n)$ be the proposition that $T_n = T_{n-3} + T_{n-2} + T_{n-1}$ and $T_n \in \mathcal{T}^*$. The proof is by strong induction, which is a standard method of proving propositions of Fibonacci sequences.[38]

We first verify that $P(1)$ holds. As shown before the first three elements, i.e. 1, 2 and 4 satisfy Lemma 1, and the first Tribonacci number, T_n, is indeed 7. Thus for $n = 1$, $P(1)$ is satisfied.

For the strong inductive hypothesis we assume that $P(j)$ is true for all positive integers not exceeding κ, i.e. $P(j)$ is true for all $j = 1, 2, ...\kappa$. That is we assume $T_j = T_{j-3} + T_{j-2} + T_{j-1}$ and $T_j \in \mathcal{T}^*$, for all $j = 1, 2, ...\kappa$. Under this assumption we show that $P(\kappa + 1)$ is true, i.e. $T_{\kappa+1} = T_{\kappa-2} + T_{\kappa-1} + T_\kappa$ and $T_{\kappa+1} \in \mathcal{T}^*$.

We need to show $\lim_{\beta \to 1} V_i(T_{\kappa-2}, \{T_{\kappa-2}, T_{\kappa-1}, T_\kappa\}) - V_i(\{T_{\kappa+1}\}) < 0$. Using inequality (A.32) in Lemma 1, this is equivalent to

$$\frac{T_{\kappa-2} + T_{\kappa-1} + T_\kappa}{T_{\kappa-2}} < e^2 \tag{A.33}$$

equivalently,

$$1 + \frac{T_{\kappa-1}}{T_{\kappa-2}} + (\frac{T_\kappa}{T_{\kappa-1}})/(\frac{T_{\kappa-2}}{T_{\kappa-1}}) < e^2 \tag{A.34}$$

Using the hypothesis, $\frac{T_\kappa}{T_{\kappa-1}}$ and $\frac{T_{\kappa-1}}{T_{\kappa-2}}$ are the Tribonacci constant, which is the ratio towards which consecutive Tribonacci numbers tend as κ gets large. The Tribonacci constant is approximately ≈ 1.83929. Thus the left hand side of (A.34) tends to ≈ 6.22227. Note that for the initial elements of the Tribonacci sequence the ratio of consecutive elements is larger, and if $T_{\kappa-2} = 7$, then the upper bound of the L-H-S is ≈ 6.28571. While the right hand side is $e^2 \approx 7.38905$. This completes the inductive steps. So $P(n)$ is ture for all positive integers n. □

A.8 Proof of Proposition 5

Consider a case where the grand coalition is not stable and country i is the initial proposer of a coalition which leads to the formation of equilibrium numerical coalition structure

[38] As shown by Rosen (2021, p.333) the validity of proof by induction or strong induction can be proved from each other. In other words, a proof using strong induction can be rewritten as a proof by induction and vice versa.

$\{m_1^*, m_2^*, ..., m_k^*\}$, and assume $m_1^* < m_2^* < ... < m_k^*$. Assume that i is the proposer of a non-ultimate coalition with m_{k-1}^* members, and considers between two coalitions of the same size, say M_{k-1} and M'_{k-1}, where the latter includes countries from the set of active players which have the highest scarcity rent, such that at least one member in M'_{k-1} has a scarcity rent which is strictly greater. Thus, at least one country in M'_k (which is a larger coalition) has a scarcity rent which is strictly smaller relative to the scarcity rent of the countries in M_k. Country i itself is in both coalitions. In that case, we have

$$\begin{aligned}
&V_i^i(M_{k-1},\{M_1, M_2, ..., M_k\}) - V_i^i(M'_{k-1}, \{M_1, M_2, ..., M_{k-2}, M'_{k-1}, M'_k\}) = \\
&\quad \frac{1}{1-\beta(1-\nu)}\Big\{\nu[ln(\frac{E_{it}(M_{k-1})}{E_{it}(M'_{k-1})}) + \beta ln(\frac{E_{it+1}(M_{k-1})}{E_{it+1}(M'_{k-1})}) + ...] \\
&\quad - \frac{\gamma\xi}{1-\beta}\{[\sum_{i\in M_{k-1}} E_{it}(M_{k-1}) + \sum_{i\in M_k} E_{it}(M_k)] - [\sum_{i\in M'_{k-1}} E_{it}(M'_{k-1}) + \sum_{i\in M'_k} E_{it}(M'_k)] + \\
&\quad \beta[\sum_{i\in M_{k-1}} E_{it+1}(M_{k-1}) + \sum_{i\in M_k} E_{it+1}(M_k)] - \beta[\sum_{i\in M'_{k-1}} E_{it+1}(M'_{k-1}) + \sum_{i\in M'_k} E_{it+1}(M'_k)] + ...\}\Big\}
\end{aligned} \tag{A.35}$$

Emission of those which remain in coalitions with the same sizes does not affect i's decision. The second line in (A.35) is the direct gain of country i from emitting in M_{k-1} versus in M'_{k-1}. Given that both coalitions have the same size, and country i has the same scarcity rent in both scenarios, the ratio of two emissions in any period is one, and thus the second line is zero. The third and fourth lines are the externality damages. Because both coalition structures correspond to the same numerical coalition structure, the SCC in M_{k-1} is the same as in M'_{k-1}, also both M_k and M'_k have the same SCC. Because of the assumption regarding the scarcity rent of the countries, it is easy to see that for any general β and in any period, the total emissions of M_{k-1} and M_k is larger than the total emissions of M'_{k-1} and M'_k. In other words, in period t for example,

$$[\sum_{i\in M_{k-1}} E_{it}(M_{k-1}) + \sum_{i\in M_k} E_{it}(M_k)] - [\sum_{i\in M'_{k-1}} E_{it}(M'_{k-1}) + \sum_{i\in M'_k} E_{it}(M'_k)] \tag{A.36}$$

is positive. Hence, for any β the difference of payoffs in (A.35) is negative. Furthermore, in the limit that $\beta \to 1$, we have

$$\begin{aligned}
&\lim_{\beta\to 1} V_i^i(M_{k-1},\{M_1, M_2, ..., M_k\}) - V_i^i(M'_{k-1}, \{M_1, M_2, ..., M_{k-2}, M'_{k-1}, M'_k\}) \\
&\qquad = -\{([\frac{m_{k-1}}{m_{k-1}} + \frac{m_k}{m_k}] - [\frac{m_{k-1}}{m_{k-1}} + \frac{m_k}{m_k}]) + ...\} \\
&\qquad = -2(1 - 1 + 1 - 1 + ...)
\end{aligned} \tag{A.37}$$

Note that $\sum_{n=0}^{\infty}(-1)^n$ is the Grandi's series, which is a divergent series. Cesàro summation and Abel summation among other methods conclude that the sum of this

series is $\frac{1}{2}$.[39] For instance, Cesàro summation is based on a limit of an infinite weighted average of the element of the series, such that the weights of odd elements are discounted in the limit. Here the odd elements correspond to the payoff of the less efficient coalition M_{k-1}, and thus the difference in equation (A.37) is negative. This is an intuitive result and it is because in the limit, the heterogeneity *almost* vanishes and not precisely.

Thus, the proposer prefers to form the more efficient coalition of M'_{k-1}. The same analysis applies to any respondent, which considers rejecting i's proposal of M_{k-1} and proposing M'_{k-1} next sub-period. □

A.9 Proof of Proposition 6

Assume that at the beginning of period t, the coalition structure $\mathbb{M}$ is the initial membership state. Note that as stocks of K_{it} and S_t have changed from period $t-1$, by the beginning of period t we have moved to a new state. Without loss of generality assume $\mathbb{M}$ is the coalition structure of singletons. In period t an initial proposer is selected based on the protocol. Because of the binding agreement assumption, the proposed coalition M should guarantee to maximise payoff of its signatories in infinite horizon. This leads to the dominant emission strategies in period t, and in a no-delay MPE if $\beta \to 1$, coalition structure $\mathbb{M}^*$ as prescribed by Proposition 4 and the protocol forms. In period $t+1$ the same initial proposer $i \in M^* \subseteq \mathbb{M}^*$ can make a proposal. Given the binding assumption, the approval committee which consists of the m signatories of M must approve the move. Because the protocol is fixed and deterministic, no party has a profitable deviation. To see this note that from Proposition 4 we know that in a MPE, the new proposal must include m^* number of countries, and that the decision of the proposer is independent of any stocks. In addition, Proposition 4 implies that the equilibrium strategy of the approval committee of a non-ultimate coalition is always rejecting any change of number of signatories, no matter if it is enlarging the coalition (thus increasing their $\hat{\Lambda}$), or if they are offered side-payments to leave the coalition (which would leave them in a larger coalition). So, again the proposer makes offers to m^* signatories and their identity is exactly the same as those in M^*. Therefore, the MPE has an absorbing membership state.

The transfer distributions of equilibrium coalition structure $\mathbb{M}^*$ are renegotiation-proof. To see this, note that with heterogeneous countries, being a proposer of the same coalition is strictly better than being a respondent (though for the characterisation of $\mathcal{M}^*$ that was irrelevant). There can be histories in which the signatories of M^* which were respondents in period t, now in period $t+1$, by rejecting the proposal of i, can change the distribution of transfers of M^*. Because the transfers are budget-balanced, i.e. sum of transfers within the coalition is zero, the renegotiation offer would be rejected by the approval committee. Thus, the equilibrium payoffs of the history by which country i's

[39] For example, see Davis (1989, p.152).

offer is accepted in all periods $\tau > t$ form the only equilibrium path. □

A.10 Proof of Proposition 7

Using the Lagrange method similar to section A.1, and taking into account the market clearing condition of capital, (6.4), the first-order condition of E_{it} for every $i \in M$ gives,

$$\frac{\nu}{E_{iyt}} \leq (1 - s_{it}) \sum_{i \in M} \sum_{\tau=0}^{\infty} \frac{\beta^{\tau} \gamma \xi}{1 - s_{it+\tau+1}} + (1 - s_{it}) \mu_{it} \; ; \; E_{it} \geq 0 \; ; \text{ c.s.} \tag{A.38}$$

This implies that if $E_{it} \geq 0$, then

$$E_{it}(m) = \frac{\nu}{\mu_{it}(1 - s_{it}) + \hat{\Lambda}(m)} - g_{it}(m) \tag{A.39}$$

where

$$\hat{\Lambda}_{it} \equiv (1 - s_{it}) \sum_{i \in M} \sum_{\tau=0}^{\infty} \frac{\beta^{\tau} \gamma \xi}{1 - s_{it+\tau+1}}$$

Hence, the planner of each country maximises the infinite sum of the utility of his/her country given similar constraints as above. The first-order condition of C_{it} is the same as before. Using $C_{it}(M, \mathbb{M}) = (1 - s_{it}) Y_{it}(M, \mathbb{M})$, and $K_{it+1}(M, \mathbb{M}) + \frac{d_i}{2} g_{it}(m)^2 = s_{it} Y_{it}(M, \mathbb{M})$, the Euler equation of saving rate is:

$$\frac{s_{it} Y_{it}(M, \mathbb{M}) - \frac{d_i}{2} g_{it}(m)^2}{(1 - s_{it}) Y_{it}(M, \mathbb{M})} = \frac{\beta(1 - \nu)}{(1 - s_{it+1})} \tag{A.40}$$

First-order condition of g_{it} gives,

$$\frac{\nu Y_{it}(M, \mathbb{M})}{E_{it}(m) + g_{it}(m)} \leq d_i g_{it}(m) \; ; \; g_{it} \geq 0 \; ; \text{ c.s.} \tag{A.41}$$

If $E_{it} \geq 0$ and $g_{it} \geq 0$, solving this with the optimal emission decision of coalition in equation (A.39) results in

$$g_{it}(m) = \frac{\mu_{it}(1 - s_{it}) + \hat{\Lambda}(m)}{d_i} \tag{A.42}$$

$$E_{it}(M, \mathbb{M}) = \frac{\nu - Y_{it}(M, \mathbb{M}) / d_i [\mu_{it}(1 - s_{it}) + \hat{\Lambda}(m)]^2}{\mu_{it}(1 - s_{it}) + \hat{\Lambda}(m)} \tag{A.43}$$

as stated in Proposition 7. □

A.11 Proof of Proposition 8

If the saving rate is a constant at $\bar{s}$, then in the model with green technology, $\hat{\Lambda}_{it}(m) = \frac{\gamma \xi m}{1 - \beta}$. The optimum value function of country $i \in M$ is

$$
\begin{aligned}
V_i^g(S_t, K_{it}, \mu_{it}, m, \mathcal{M}) &= ln(C_{it}(m, \mathcal{M})) + \beta ln(C_{it+1}(m, \mathcal{M})) + ... \\
&= \frac{ln(1-\bar{s})}{1-\beta} + \{ln(Y_{it}(m, \mathcal{M})) + \beta ln(Y_{it+1}(m, \mathcal{M})) + ...\} \\
&= \frac{(1-\nu)ln(K_{it}) + H_1^g + H_2^g + H_3^g}{1-\beta(1-\nu)}
\end{aligned} \tag{A.44}
$$

where are defined as below

$$
H_1^g \equiv \frac{\beta(1-\nu)ln(\bar{s}) - \beta(1-\nu)ln(1-\bar{s}) + ln(A_i) - \gamma T_0}{1-\beta} \tag{A.45}
$$

and other H_j^g, where $j \in \{2, 3\}$, are the same as corresponding H_j derived in section A.3, just replacing E_{it} with E_{iyt}. Furthermore, using Proposition 7, it is easy to show that

$$
E_{it}(m) + g_{it}(m) = \frac{\nu}{\hat{\Lambda}(m) + (1-\bar{s})\mu_{it}} \tag{A.46}
$$

This is indeed the optimal emission level in the model without the green technology. Therefore, the previous analysis of farsighted countries to determine $\mathcal{M}^*$ goes through, i.e. the decision of a country contemplating joining a coalition M of size m versus the grand coalition is independent of heterogeneity with respect to d_i (or any of the other sources of heterogeneity), and it depends on determining the sign of exactly equation (A.30), just replacing E_{it} with E_{iyt}. Thus Lemma 1 holds here too. Clearly the value of saving rate and its limit do not affect the analysis, as long as $0 < \bar{s} < 1$. Likewise, for the internal-external stability conditions, with the same replacement of variables, the analysis in section A.4 stands here for symmetric countries too. □