Self-enforcing climate coalitions for farsighted countries: integrated analysis of heterogeneous countries∗†

Sareh Vosooghi‡ Maria Arvaniti§ Frederick van der Ploeg¶

November 2021

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 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 Tim Worrall, Meg Meyer, Piotr Dworczak, Paula Onuchic, Ludvig Sinander and seminar and workshop participants at 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, Department of Economics, University of Bologna, 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

Curbing anthropogenic greenhouse gas (GHG) emissions, known as climate change mitigation, is currently the primary policy for combating global warming. The biggest planetary tragedy has been the failure of countries to work together to combat global warming effectively. Unfortunately, after three decades of climate negotiations, there is still no effective and self-enforcing international climate policy with a real chance of being implemented. Still, the only way to meet the Paris agreement targets is for countries to cooperate and set an ambitious climate policy. In the international arena there has been too much focus on the formation of a grand climate coalition. However, it has been shown that such a coalition may not be ambitious enough to tackle global warming. Instead, it might be the case that under the umbrella of Paris agreement, multiple climate coalitions can be stable and they can overall be more ambitious than what we observe today.

In this spirit, this paper models countries’ decision-making in joining climate treaties and attempts to suggest a pragmatic approach. We develop an integrated assessment model 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 the climate agreements. Signatories of climate coalitions commit to choose climate policies jointly such that they maximise the benefits of the block. There is a common consensus in the literature that the larger the number of signatories of a climate coalition, the more ambitious their climate policy is; however not all large coalitions are self-enforceable. In this paper, we address three questions. First, how do we model the problem of coalition formation of heterogeneous countries? Second, do multiple climate coalitions form in equilibrium? Third, how many signatories commit to such 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 integrated assessment model with farsighted countries, we show that in equilibrium, the number of signatories of a climate treaty 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 wherever an efficient way of packing elements together is called for.\footnote{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 found by Charles Darwin in the *Origin of Species*.\footnote{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).}

More technically, if countries were symmetric, coalitional equilibria would need to be characterised only in terms of number of coalitions and number of signatories, since
the identity of symmetric countries is indeterminate in equilibrium. However, with 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 productivity, the initial levels of fossil fuel and the associated scarcity rents of the fossil fuel resources. With asymmetric countries it is not possible to directly use the conventional coalition formation methodologies which 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, since in our setting the heterogeneity affects 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, while the heterogeneity does indeed affect the countries equilibrium payoffs and the composition of coalitions. 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 in an affine form.

The literature on climate coalition formation abstracts from the details of macroeconomic outcomes and their underlying determinants. Hence, it works with very stylised models without micro foundations and restrictive assumptions. A main objective of this paper is to capture broader incentives for the policy makers in climate negotiations. In another strand of the literature, climate economists, have developed global and multi-country Integrated Assessment Model (IAMs). These are macroeconomic growth models which allow for the effects of economy on global warming and vice versa. Hence, they have been able to achieve a wide range of economic analyses of climate policy. However, in this strand of the literature global cooperation is assumed from the outset, and the strategic interactions of countries within the climate agreements is not considered.

The two strands of literature have developed almost independently. We build a bridge between the strand of climate coalition formation literature and the IAM literature. But this comes at a cost: there are few macroeconomic growth models with analytical solutions, and our analysis needs to accommodate both climate dynamics and coalition formation. Golosov et al. (2014) add simple climate dynamics to the Brock and Mirman (1973) model of economic growth and find a closed-form solution for the social cost of carbon (SCC) which 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) and the Regional Integrated Climate-Economy (RICE) model of Nordhaus and Yang (1996) relatively well. Here, in capturing the countries’ incentives in climate negotiations, we use payoff specifications from a well-tested multi-country version of Golosov et al. (2014). We modify its climate dynamics and temperature structure based on recent advances in climate science, and we integrate it with the participation in international climate
agreements.

On the strategic thinking side of the coalition formation, we assume that the countries are farsighted. In other words, we allow for negotiating countries to rationally consider all self-enforceable unilateral and multilateral deviations from their membership decisions, and thus predict the entire conceivable coalition structure. This is in sharp contrast to the cartel stability solution concept, which corresponds to a Nash equilibrium, and relies on 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 reaction by others upon breaking off climate negotiations increases incentives for free riding. Thus, as widely confirmed in the literature, the use of cartel stability solution concept results in the formation of small coalitions.

In finding the equilibrium number of signatories, immunity of the equilibrium to deviations must be checked, and farsightedness implies that deviations from equilibrium strategies which are not themselves farsighted must be excluded. Hence, in conventional farsighted methodologies, the characterisation of the equilibrium coalition structure relies on algorithms which recursively identifies the set of 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 and continues to a finite \( N \). This set determines the possible farsighted deviations; also, 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 undemanding. Nevertheless, in an infinite-horizon IAM, the recursion process can be onerous and potentially would require numerical simulations. However, for the integrated assessment framework that we analyse we are able to obtain analytical results.

We show that in our model, the set of total number of countries for which the equilibrium is the formation of a grand coalition is the set of Tribonacci numbers. This is a known set; hence, there is no need to check the payoffs of the countries recursively. Thus we suggest a tractable and intuitive algorithm to characterise the equilibrium number of climate coalitions and their number of signatories analytically in an IAM. Importantly, this algorithm does not require any recursions.

Furthermore, the equilibrium number of signatories in any coalition is 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 the incentive of countries to free ride.

Fixing the equilibrium number of signatories, our decoupling result allows us to 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. And, provided there are adequate transfers to sustain the equilibrium, coalitions form without any delay if the number of signatories obeys the Tribonacci rule.

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 policy-oriented 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 climate coalition formation.

The remainder of the paper is organised as follows. Related literature is reviewed in the next section, and the model is presented in section 3. We analyse how heterogeneous countries arrive at climate treaty memberships using our general equilibrium framework 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 an 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 sits in the intersection of these two strands of literature.

Research on IEAs and cooperation by forming climate agreements has led to an extensive literature on climate economics, and it has provided some valuable inputs into the design of international climate treaties, including the Paris Climate Accords. Seminal papers on the strand of the literature concerned with coalition formation and IEAs include Carrao and Siniscalco (1993) and Barrett (1994). Benchekroun and Long (2012) and Battaglini and Harstad (2016), among others, provide extensive reviews of the literature on IEAs.

The dominant part of this literature employs the solution concept of cartel stability also known as the internal and external stability, where 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.\footnote{For example, see Battaglini and Harstad (2016) for the literature on the robustness of this result.}

To achieve the formation of larger coalitions within 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); heterogeneity (Botteon and Carraro, 1997); the adoption 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).

Using the d’Aspremont et al. (1983) notion of cartel stability implies focusing 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. 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 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, our paper uses a different solution concept and assumes that all the countries are farsighted. Note that the internal and external stability conditions, although necessary, are 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. In the strand of IEAs, Vosooghi (2017) uses the assumption of farsighted stability in a stochastic setting while Diamantoudi and Sartezetakis (2018) and de Zeeuw (2008) use it in a deterministic set-
tings. 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

Our paper examines a dynamic extension of the game analysed in Ray and Vohra (2001) and does this within the context of integrated assessment models of the economy and the climate. Furthermore, our analysis generalises to the case of heterogeneous countries, and to reversible agreements, both of which 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 some main climate regions too. These IAMs are often very large, detailed and too complicated to be solved analytically, so that typically numerical methods are used to analyse them. These IAMs address a wide range of analyses of climate policy; for example, the seminal IAM on estimating the SCC and optimal carbon price called DICE (Nordhaus, 2014), the IAM referred to as the FUND model with effects of uncertainties and different climate regions (Anthoff and Tol, 2013), and the IAM referred to as the PAGE model with a regional temperatures leading to global average temperature (Hope, 2011).

More recently, analytical expressions for the optimal SCC and climate policies have been obtained from IAMs. The pioneer works are the global economy of Golosov et al. (2014); and the multiple country model of Hassler and Krusell (2012). Van der Ploeg and de Zeeuw (2016) study and compare fully cooperative and non-cooperative climate policies in a North-South model of the global economy; Van der Ploeg and de Zeeuw (2019) suggest the adjustment of the carbon tax to the risk of one or multiple climate catastrophes; Van der Ploeg and Rezai (2017) investigate the sensitivity of climate policy to the choice of model. Van den Bremer and Van der Ploeg (2021) use perturbation theory to obtain a tractable expression for the optimal risk-adjusted SCC.

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 in modelling coalition formation, these papers use a cooperative game-theoretic approach. However, 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.4

---

3There 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.

4For more discussions see Rosenthal, (1971) and Ray and Vohra (2001).
Only a few authors combine these two strands using a non-cooperative game-theoretic approach to coalition formations. They add a coalition-formation stage to known numerical IAM models 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. It uses the assumptions of Brock and Mirman (1973) to obtain a tractable general equilibrium model of economic growth and climate. 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.

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

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

(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 (2013) explores the sensitivity of the optimal SCC in the IAM of Golosov et al. (2014) to more generalised preference
structures with elasticities of intertemporal substitution different from one and shows numerically that their results are robust.\(^5\)

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.

\[
E_{it} = R_{it} - R_{it+1}
\]  

(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), it can be verified that,

\[
R_{it+1} = R_{i0} - \sum_{s=0}^{t} E_{it-s}
\]

(3.3)

where \(R_{i0}\) is the exogenous stock of reserves of fossil fuel of country \(i\) in \(t = 0\), and it can be heterogeneous across countries.

Production of the final output or the aggregate output 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 thus assume that global warming damages are proportional to aggregate output. Golosov et al. (2014) show that an exponential functional form for damages by stock of atmospheric carbon describes the ratio of global warming damages to aggregate output of the DICE and RICE models reasonably well.\(^6\) The production follows a Cobb-Douglass technology so that the aggregate output of country \(i\) at time \(t\) is given by

\[
Y_{it} = \exp(-\gamma T_t) A_{iy} K_{it}^{1-v} E_{it}^\nu
\]

(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; \(\nu\) is the output elasticity of energy; and \(E_{it}\) is the energy use in the production of the final good. Th current level of capital, \(K_{it}\), and the initial

\(^5\)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.

\(^6\)We will assume instead that this damage ratio decreases linearly (rather than quadratically) in temperature.
level of capital, $K_{i0}$, can vary across countries.\footnote{The model can be interpreted as an AK growth model in the spirit of Romer (1986): by assuming $Y_t = \exp(-\gamma T_t)A_t K_{i0}^{1-a-v}E_{it}(K_{it}L_{it})^\alpha$ where $K_{it}L_{it}$ is the effective labour and $K_{it}$ is the economy wide capital stock (e.g. human capital, RD, 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. $K_{i1} = K_{it}$ and hence the expression in (3.4) results.} Total factor productivity (TFP) has two multiplicative terms, a constant, $A_{iy}$, which can vary across countries, and a negative exponential function of global temperature, $T_t$, where $\gamma$ is the damage coefficient.\footnote{The damage coefficient can be assumed to be an uncertain parameter. In the literature it is common to replace it with the expectation of a fixed and common distribution of $\gamma$. We ignore that this damage coefficient may differ across countries; e.g., developing countries in the south typically have higher damage coefficients than developed countries in the north.} Jones (2005) provides a micro-founded theoretical 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.\footnote{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.} Moreover, Hassler et al. (2021) using historical data to calibrate an IAM, estimate an aggregate production function and show that in long run the input shares tend to be stationary. This 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.

The feasibility constraint for the final good requires that aggregate consumption plus investment equals aggregate production in each country, so that

$$C_{it} + K_{it+1} = Y_{it}$$

Note that capital and energy are used only in the final goods sector. It is assumed that the adjustment cost of capital is zero. In a decentralised economy the markets for the final good clears at national level. The fossil fuel depletion constraints (3.2) must also be satisfied for each country. 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 a tractable analytical solutions, we assume full capital depreciation in each period. Barrage (2013) shows that the characterisation of the optimal SCC in Golosov et al. (2014) in the long run is numerically robust with respect to depreciation rates that are less than 100%.

### 3.2 Climate dynamics

The use of fossil fuel in the production of final goods inevitably results in carbon emissions. Here we depart from Golosov et al. (2014) and based on recent insights in climate
science,\textsuperscript{10} we assume a simpler, but more realistic linear relationship between temperature and cumulative emissions of $CO_2$ or GHGs, instead of with the stock of atmospheric carbon.\textsuperscript{11} Global temperature is thus given by

$$T_t = T_0 + \xi S_t$$

(3.6)

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

$$S_t = S_{t-1} + E_t$$

(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

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

(3.8)

where $S_0$ is the pre-industrial level of cumulative emissions, typically set to zero.

3.3 Climate Coalition Formation

We allow the countries to form climate coalitions to collectively reduce their emissions and thus their damages from global warming. At the beginning of period $t$, they have the choice of participating in climate negotiations. Following convention, we assume that if the negotiations do not come to a conclusion, all countries suffer an infinite loss. This is to ensure 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, $\mathcal{M} \equiv \{M_1, M_2, ..., M_k\}$. Let $m \leq N$ be a positive integer showing the cardinality (the number of members) of coalition $M$. A numerical coalition structure, $\mathcal{M} \equiv \{m_1, m_2, ..., m_k\}$, is a partition of $N$ into the sizes of coalitions. If countries were symmetric, the identity of any particular country would be indeterminate in equilibrium and characterising the equilibrium membership strategies would therefore involve only the size 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 no country can be excluded from joining and no country can be forced to join. But joining a climate

\textsuperscript{10}For example see Allen et al. (2009); and Matthews et al. (2009).

\textsuperscript{11}Mathematically (but not in physics) this is analogous to having a model of a permanent component of atmospheric carbon only with no rate of decay. Cumulative emissions and stock of atmospheric carbon are indeed the same up to a constant if there is no decay of atmospheric carbon, which is due to the absorption of $CO_2$ by the oceans.
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 for this coalition, for all \( t \in \{0, 1, ..., \infty\} \) and all \( i \in M \). Thus, implementing the decisions of a climate treaty is costless. Furthermore,

**Assumption 1.** Membership decisions are irreversible.

In other words, countries do not have any chance of renegotiation.\(^{12}\) 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.

\[
\{E_{it+\tau}(m, M), C_{it+\tau}(m, M), K_{it+\tau+1}(m, M), R_{it+\tau+1}(m, M)\}_{\tau=0}^{\infty}.
\]

If at the beginning of period \( t \) an agreement is already in place, then the membership stage is skipped. At the end of each period, they observe emission \( E_{it} \) of each country and payoffs for each country are realised.

### 3.3.1 Climate negotiations

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 length of time of sub-periods to length of period \( t \) is fixed, and there is cost of delay in sub-periods which is captured by discount factor \( \sigma \).\(^{13}\)

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 which have not joined any other binding coalition yet.

The proposal consists of the identity of the members (thus 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. The proposal can potentially be conditioned on the complementary formed coalition structure. In other words, the proposed emission plan should be conditioned on the emission plans of other coalitions in the coalition structure, \( M \). However, as will become clear in the next section, the coalitions have dominant strategies, and the cumulative stock of emissions does not affect the marginal cost and benefit of the countries in determining their optimal emissions. If \( m = 1 \), the proposer exits the negotiations as a singleton coalition, and if

\(^{12}\)We relax this assumption in section 5.

\(^{13}\)Although the farsighted methodology that we use holds for \( \sigma \to 1 \) and in the limit bargaining is frictionless, assuming \( \sigma \neq 1 \) is needed to avoid multiplicity of equilibria.
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 it is set at the very beginning of the negotiation stage. We focus on a special class of rejector-friendly protocols, where the first rejector is the next proposer of a coalition $M$. Furthermore, we assume the order determined by the protocol is a deterministic order. 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 $\mathcal{M}$ 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).\textsuperscript{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. In contrast to repeated games where there is no state and no stocks, investigating MPEs in dynamic games is very common.\textsuperscript{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 of their fossil fuel reserve, $\mu_{it}$. In the next section, we show why these variables are payoff relevant. Finally, we focus on the farsighted stability concept of ‘equilibrium binding agreements’ of Ray and

---

\textsuperscript{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.

\textsuperscript{15}Our dynamic game presented in the last section has $2N + 1$ state variables.
To ensure sequential rationality, we solve the model backward.

### 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 play across coalitions. The members of the coalition maximise their joint discounted infinite-horizon payoff, which we call the total worth of coalition $M$ and is given by

$$
\sum_{i \in M} \beta^{\tau} \{ \ln(C_{i,t+\tau}) \}
$$

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

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

which implies that the marginal productivity of fossil fuel is set equal to its price which is the scarcity rent $\mu_{it} C_{it}$ (as mentioned earlier, $\mu_{it}$ 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

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

The per-unit SCC is the SCC per unit of output of each signatory $i \in M$ for any period $t$ 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.

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. The per-unit SCC also increases in the damage coefficient $\gamma$; the transient climate response of temperature to cumulative emissions $\xi$; and the discount factor, $\beta$. 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 per-unit SCC are both proportional to aggregate economic activity. In the Appendix we show that equation (4.2) in conjunction with the constant saving rate result from our general equilibrium analysis and leads to our first result.

---

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

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

(4.4)

\[
\mu_{it} = \beta^{-t}\mu_{i0}
\]

(4.5)

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 against what other coalitions choose, in the sense that the per-unit SCC of complementary coalitions does not affect the per-unit SCC strategy 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, marginal damages, i.e. \( \hat{\Lambda}(m)Y_{it} \) are proportional to output while 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. Although the emission strategies are dominant, their payoffs depend on global cumulative emissions and thus on the entire coalition structure, and the associated energy use of all countries.

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. \( (1 - \nu)Y_{it}/K_{it} \). This rule for the actual scarcity rent is known as the Hotelling rule. The initial scarcity rent, \( \mu_{i0} \), is such that cumulative fossil fuel use exhausts all of 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.

\(^{18}\)If fossil fuel reserves are abundant, as in the case of coal, the per-unit and actual scarcity rents would be zero.
**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 expected to decrease more 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 \( \mu_{it} \) to be heterogeneous across the countries. We, realistically, assume at the membership stage 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 causing externality.

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. The direction of change in emissions depends on the magnitude of each of these factors. However, our analysis of coalition formation by far-sighted countries makes the assumption of frictionless bargaining and patient countries, where patient refers to a discount factor infinitesimally close to unity. This leads to the following corollary.

**Corollary 2.** If countries are patient, i.e. if \( \beta \to 1 \), emissions of signatories unambiguously decrease upon joining a non-singleton coalition.

To determine the sign of the change in emissions we do not necessarily need the assumption of no discounting in the limit. Depending on the parameter values, one can find a discount factor \( \beta^* \) such that if \( \beta \in (\beta^*, 1) \), then emissions decrease. However, Corollary 2 fixes ideas for our future analysis.

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

**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

\[
Y_{it}(m, \mathcal{M}) = \exp(-\gamma T_t)A_yK_{it}^{1-\nu}(\nu/(\mu_{it}(1-s) + \hat{\Lambda}(m)))^\nu
\]

\[
C_{it}(m, \mathcal{M}) = (1-s)Y_{it}(m, \mathcal{M})
\]

\[
K_{it+1}(m, \mathcal{M}) = sY_{it}(m, \mathcal{M})
\]

\[
Y_{it}(m, \mathcal{M})/Y_{it-1}(m, \mathcal{M}) - 1 = \exp(-\frac{\gamma E_t(\mathcal{M})}{\nu}s^{1-\nu}\left(\frac{r_{it-1}}{1-\nu}\right)^{1-\nu}\left(\frac{\beta \mu_{it}(1-s) + \hat{\Lambda}(m)}{\mu_{it}(1-s) + \hat{\Lambda}(m)}\right)^\nu - 1
\]

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

Note that the marginal product of capital is equivalent to the interest rate in a decentralised economy. Aggregate output increases in the saving rate \( s \) and total factor productivity, but decreases in global warming (cumulative emissions) and the price of

\[19\text{Note that } \lim_{\beta \to 1} dE_{it}(m) = \lim_{\beta \to 1}\left\{ \frac{\partial E_{it}(m)}{\partial \mu_{it}} d\mu_{it} + \frac{\partial E_{it}(m)}{\partial \Lambda_{it}} \frac{\partial \Lambda_{it}}{\partial m} d\Delta m \right\} < 0.\]
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 these strategies depend on the stock of cumulative emissions, $S_t$, through the global temperature. Moreover, consumption and capital choices are non-stationary as they depend on the time-varying paths of $\mu_{it}$ 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, $\mu_{it}$, 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.\(^{21}\)

### 4.1.1 Two benchmarks

The first benchmark corresponds to the singleton coalition structure, i.e. if $m = 1$, 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 3.** 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

$$E_{it}(1) = \frac{\nu}{\left[\mu_{it}(1 - \beta(1 - \nu)) + \hat{\Lambda}(1)\right]}$$

(4.6)

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,

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

(4.7)

\(^{20}\)Although there is full deprecation of capital by the end of each period, heterogeneity with respect to $K_{it}$ in this model with a constant and common saving rate leads to weak heterogeneity with respect to $K_{it}$ at the beginning of each subsequent period.

\(^{21}\)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, the model would exhibit a positive growth rate. The objective of our paper is not to suggest methods to achieve a positive growth rate; for such treatments see for example Golosov et al. (2014). The analysis of coalition formation of countries is not sensitive to their growth rate as long as initially the participating countries have a finite per-unit scarcity rent.
subject to constraints (3.5) and (3.2) for each $i$.

**Corollary 4.** If a grand coalition forms, the optimal international cooperative SCC is 
$\hat{\Lambda}_it(N) = \hat{\Lambda}(N) = \frac{\xi N}{1-\beta}$ for all $t$, and $i$. The corresponding level of fossil fuel consumption or equivalently flow of emissions is

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

(4.8)

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.

### 4.2 Climate membership decisions

Given the above analysis, the proposers and respondents of climate treaties decide about their memberships. The incentives of countries in the climate negotiation stage are determined by the resulting optimum value function of a country in a coalition $M$. Before presenting our results under the farsightedness assumption, let us first examine the outcomes of our model under the more commonly used cartel (or internal and external) stability conditions.

#### 4.2.1 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 the countries are ex-ante symmetric, but after the membership stage they may end up in asymmetric situations.

**Definition 2.** 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.\(^{22}\) 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 function of the deviating country is not affected by the number of signatories and it leads to the non-cooperative emission level.

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

The proof is in the Appendix. Proposition 3 echoes the well-known result of the small-coalition paradox. Even with the more generalised internal stability version, 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 for the case we have analysed here). Hence, under such assumptions the countries naively have a higher incentive to free ride and can only form a small coalition with a maximum size of three countries. This is true for any number of countries \( N \).

Lastly, as shown in the Appendix, 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.

### 4.2.2 Farsighted countries

For the internal and external stability concept, 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 3.** 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).

\(^{22}\)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.
In other words, from the point of view of a farsighted country, a potential group of deviating countries 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 $M$. Let us denote the equilibrium coalition structure by $M^*$ 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., where $N = 2$, $M^*$ for each group of two countries should be found. Then knowing which group of two countries can strike a deal, for $N = 3$ all possible $M^*$ is found. And the process continues. Clearly multiplicity of equilibria is expected and the analysis can be tedious.

There are not many studies on coalition formation of heterogeneous farsighted countries. To the best of our knowledge, only Ray (2007, p.130) studies such games with externalities. He derives the sufficient conditions for existence of equilibria without delay.\(^{23}\)

In our model, the countries are heterogeneous in various ways: different initial capital levels, $K_{i0}$; different total factor productivity constant, $A_{iy}$; different scarcity rents, $\mu_{it}$ or different initial stocks of fossil fuel resource, $R_{i0}$ across countries. All of these sources of asymmetry have been important topics in climate negotiations. The nature of which these sources of heterogeneity affect the countries’ payoffs is different. 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 $\mathcal{M}$ with $V^j_i(S_t, K_{it}, \mu_{it}, M, \mathcal{M})$, and the optimum value function of the country in a grand coalition $\{I\}$ with $V^j_i(S_t, K_{it}, \mu_{it}, I)$. 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 in adequate. 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

\(^{23}\)These sufficient conditions are (i) as coalitions form over time, their equilibrium average worth decreases, (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.

20
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

$$\sum_{i=1}^{m} V^j_i(S_t, K_{it}, \mu_{it}, m, M) - \sum_{i=1}^{m} V^j_i(S_t, K_{it}, \mu_{it}, I)$$  \hspace{1cm} (4.9)$$

In the appendix we show that this difference is independent of stocks and $A_{iy}$ for any discount factor $\beta$. Hence, the membership decisions are independent of heterogeneity with respect to $K_{i0}$ and $A_{iy}$.

Furthermore, the difference of payoffs in (4.9) 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 $V^j_i$ 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 here. Thus, the problem in (4.9) reduces to determining sign of

$$V^j_i(S_t, K_{it}, \mu_{it}, m, M) - V^j_i(S_t, K_{it}, \mu_{it}, N)$$  \hspace{1cm} (4.10)$$

for each $i \in M$. As shown in the literature, although transfers are allowed, they play no role in characterising the equilibrium numerical coalition structure. Likewise the identity of the initial proposer is irrelevant too. Regarding heterogeneity with respect to the initial stock of fossil fuel note that $R_{i0}$ affects the optimum value function indirectly through the associated scarcity rent, $\mu_{it}$.

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. Given corollary 1, when the discount factor converges to unity, the effect of number of signatories dominates the effect of post-membership change in the scarcity rent. Hence, heterogeneity with respect to stock of fossil fuel and $\mu_{it}$ causes heterogeneity with respect to emissions, even within a coalition $M$ with $m$ members.

This source of heterogeneity does not affect the countries’ payoffs linearly. However, in the limit that $\beta \rightarrow 1$, the difference of payoffs in (4.9) becomes independent of $\mu_{it}$. 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

---

24In 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 size, their emission path affects the trajectory path of their scarcity rent (similar to the case of ex ante symmetric countries), which will differ for countries with different initial fossil fuel reserves. In the membership stage of a reversible coalition formation, our focus is on the effect of ex-ante asymmetry on membership decisions.
proposer. Thus independently of this source of heterogeneity, we can again characterise
the numerical equilibrium coalition structure.\textsuperscript{25}

The following proposition summarises the discussion above and the proof is in the
Appendix.

**Proposition 4.** The equilibrium numerical coalition structure $M^*$ can be characterised
independent of the heterogeneity with respect to $K_{i0}$ and $A_{iy}$. It can also be characterised
independent of the heterogeneity with respect to $R_{i0}$ and $\mu_{it}$, if $\beta \to 1$.

In other words, we can decouple the problem of the cardinality of coalitions in equi-
librium from the actual composition of countries in each $M^* \subseteq M^*$. Hence, the numerical
coalition structure can be characterised, while we can keep our focus off their heterogene-
ity. Then after finding $M^*$, 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 $A_{iy}$ is
stronger, as it does not rely on the patience of policymakers. 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 $\mu_{it}$ 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
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 $M^*$ 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.

### 4.2.2.1 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 $\mathcal{M}$ with $V_i(m, \mathcal{M})$, and the optimum value function
of a country in a grand coalition $\{N\}$ with $V_i(N)$.

\textsuperscript{25}We show in the next section, that in our IAM, the characterisation of equilibrium numerical coalition
structure relies on the assumption of patience of countries in the limit even for symmetric countries.
Here, again, the equilibrium numerical coalition structure needs to be identified recursively. For completeness, note that if \( N = 1 \), a singleton coalition forms. Then, we need to find \( M^* \) if \( N = 2 \). Given that, we then find \( M^* \) if \( N = 3 \), and continue the recursion until we have reached the total number of countries \( N \) under consideration.

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

\[
V_i(1, \{1, 1\}) - V_i(\{2\}) = \frac{1}{1 - \beta(1 - \nu)} \left\{ \nu \ln \left( \frac{E_{it}(1)}{E_{it}(2)} \right) + \beta \ln \left( \frac{E_{it+1}(1)}{E_{it+1}(2)} \right) + \ldots \right\} - \frac{2\gamma \xi}{1 - \beta} \left\{ [E_{it}(1) - E_{it}(2)] + \beta [E_{it+1}(1) - E_{it+1}(2)] + \ldots \right\}
\]

(4.11)

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 (4.11) is a ratio of the discounted infinite sum 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 as a free rider 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 as we focus on patient countries in the limit, i.e. \( \sigma \to 1 \), and \( \beta \to 1 \), it is easy to show that with two countries \( \lim_{\beta \to 1} (V_i(1, \{1, 1\}) - V_i(\{2\})) < 0 \) and the grand coalition forms in equilibrium, i.e. \( M^* = \{2\} \).

Continuing to the case of \( N = 3 \), 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. Considering farsighted deviations implies splitting \( N \) (or any active number of players in the negotiation room) into coalitions where their sizes result from the decomposition of \( N \), i.e. from breaking up \( N \) into the largest possible integers at which a grand coalition was stable in previous stages of the recursion.

**Definition 4.** \( T^* \) is the set of the total number of countries, \( N \), for which a grand coalition forms in equilibrium.

**Definition 5.** For any integer \( N \), the decomposition \( D(N) \) is \( \{m_1, m_2, \ldots, m_k\} \), such that \( m_k \) is the largest integer in \( T^* \) that is strictly smaller than \( N \). Then any other \( m_i \) in \( D(N) \), is the largest integer in \( T^* \) that is no greater than \( N - \sum_{j=i+1}^k m_j \).

For example, for the case of \( N = 3 \), we know from previous stages of the recursion that \( T^* = \{1, 2\} \), and thus the decomposition of \( N \) is \( \{2, 1\} \).
Furthermore, Ray and Vohra (2001) show that in a public-good game, it is sufficient to check the deviation by the smallest element of the decomposition of $N$ in $\mathcal{M}$. More specifically, the most profitable deviation is forming the smallest farsighted coalition, when upon this deviation, $N$ countries have to split into the largest possible farsighted coalitions, i.e. those that result from the decomposition of $N$. 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 3.3.1. 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 of $N = 3$, 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, $T^*$ expands to $\{1,2,4\}$.

Comparing the optimum value function of the smallest coalition in the decomposition with the value function of the grand coalition can be demanding. In the Appendix, we show that in our IAM, the recursion process can be simplified and there is a general rule for inequalities like (4.11).

**Lemma 1.** Let $D(N) = \{m_1, m_2, \ldots, m_k\}$ be the decomposition of $N$, such that $m_1 < m_2 < \ldots < m_k$. If countries are patient in the limit, then in our IAM, a grand coalition forms in equilibrium if

$$\ln\left(\frac{N}{m_1}\right) < (k - 1)$$

This Lemma provides a simple sufficient condition for full characterisation of the set $T^*$ in the limit for our IAM. The LHS is the gain from emitting in the small coalition versus emitting in the grand coalition. The RHS is the externality damage resulting from forming $D(N)$ versus the grand coalition. Since in the limit, emissions are 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.

---

26Ray and Vohra (2001) have shown that in general an equilibrium numerical coalition structure never consists of two coalitions with the same size. This holds for continuum action spaces, and not for binary action spaces.
Using this lemma and continuing the recursion to find more elements of $T^*$, we obtain our next result that the set $T^*$ is a Tribonacci sequence. This is a novel result. While the full proof can be found in the Appendix, intuitively, we show by induction that this is a general property and that the sequence does not diverge from the Tribonacci numbers.\footnote{Tribonacci numbers are elements of a Fibonacci sequence of order 3.} In particular we show that an element $T_{n+1} \in T^*$ is also the sum of the last three elements of the set given that this is the case for $T_n$, and we note that given the predetermined elements the proposition holds for $n = 1$ too.

**Proposition 5.** In our IAM with farsighted and patient countries in the limit, for any number of heterogeneous countries, a grand coalition occurs in equilibrium if the total number of countries $N$ is an element of

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

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

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

The result in Proposition 5 simplifies greatly the characterisation of the equilibrium coalition structure for any number of countries $N$. For our IAM, there is no need to use any recursions to find the set $T^*$. In Ray and Vohra (2001) countries are assumed to have a one-off payoff, and the recursion thus leads to an analytical solution. We use and extend their methodology 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 obtains his results only numerically. However, it turns out that in our model 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 5 is laborious, the formalisation of Tribonacci set by mathematician Plouffe (1993) can be particularly helpful:

\[
T_n = \left\lfloor 3b \left( \frac{1}{3} (a_+ + a_- + 1) \right)^n \right\rfloor
\]

where \(a_+ = \frac{3}{2} \sqrt{19 \pm 3\sqrt{33}}\) and \(b = \frac{3}{2} \sqrt{586 + 102\sqrt{33}}\), and \(\lfloor \cdot \rfloor\) is the nearest integer function.

The economic insight behind Proposition 5, in addition to the solution concept, is due to the aspects of the model which lead to Lemma 1. These are in particular the special structure of our IAM, which results to a per-unit SCC independent of aggregate economic outcomes. In addition to the payoff specification, our general equilibrium model leads to a constant saving rate and the resulting dominant emission strategies are the other element of the model which derived Proposition 5. Importantly, our assumption of patient policymakers 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 larger 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 interest rate to identify social discount rate is misleading. Our assumption here that the discount factor of policymakers converges to unity is realistically the case to be interested in the context of global warming.

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 (Lessmann et al. (2009; and 2015), Bosetti et al. (2013)) is 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, \(T^*\), that is no greater 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, the farsighted set of coalitions \(\mathcal{M}^*\) that we have characterised in Proposition 5 passes more self-enforceability tests because 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.

4.2.2.2 Example: equilibrium numerical coalition structure with 195 countries

Let us assume $N = 195$. Based on the Tribonacci sequence in Proposition 5, because $195 \notin T^*$, we can verify that $M^* = \{149, 44, 2\}$, where $149 + 44 + 2 = 195$. In other words, three coalitions can form with 149, 44 and 2 signatories each. Clearly, there is no small-coalition paradox. 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 $M^*$ is not too refined, i.e. does not have a lot of small elements, and it only has three elements. This increases the total SCC of all coalitions internalised by all coalitions.

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 $\Lambda(m^*) = \frac{149\gamma\xi_1 - \beta}{1-\beta}$ 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. $1.03 = \frac{(192 \times 1 + 3 \times 3)}{195} < \frac{(149 \times 149 + 44 \times 44 + 2 \times 2)}{195} = 123.8$. In words, the average SCC under farsightedness is a whopping 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, i.e. 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.

4.2.2.3 Efficiency

After we show that the heterogeneous countries follow the equilibrium numerical coalition structure suggested by Tribonacci numbers, we turn to the question of composition of countries in coalitions, and the equilibrium coalition structure. Because of multiplicity of equilibria we cannot answer many questions here, but we try to address a narrower
question: whether the countries, knowing the equilibrium number of signatories, are willing to take any step to improve efficiency.

The efficient outcome for the global economy is the case when there is a fully internationally cooperative outcome. This corresponds to the formation of a grand coalition. In this section we refer to constrained efficiency where the grand coalition is not stable, but when an individual country does its best to internalise the emission externality. With heterogeneity, it is possible to specify an efficient coalitional matching but the equilibrium might depend on the bargaining protocol, i.e. the order in which a country gets to be the proposer. Although the (ordering by the) protocol is irrelevant in characterising $\mathcal{M}^*$, it is important for $\mathcal{M}^*$ of course. In this section we investigate the problem of an initial proposer (and later its respondents) to a non-ultimate coalition. In other words, the proposer knows that the next coalition is going to be larger and examines whether it is worthwhile to include less emitting countries in the smaller coalition of its own.

Among different sources of heterogeneity, only heterogeneity with respect to initial fossil fuel reserves and thus scarcity rents affect the countries’ emissions. Thus, here we assume countries are heterogeneous with respect to these factors. Therefore, the efficiency of different equilibria depends on the identity (initial fossil fuel reserves and scarcity rent) of the initial proposer and of the other potential signatories.

If the grand coalition is not stable, then 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 large 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 externality problem, 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.

Furthermore, 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. To fix ideas, in this section, we only focus on the question of efficiency while keeping all other factors, especially any element that can affect the transfers fixed.\(^{28}\)

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

\(^{28}\)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.
rent of the countries such that country 1 is the country that emits least and country 6 is the country that emits most, and no two countries have equal scarcity rents. From Proposition 5 we know that the equilibrium numerical coalition structure is $M^* = \{2, 4\}$. 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 efficiency, then by rejecting the offer of country 2, the most 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 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 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 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 if countries are patient, then (keeping everything else fixed) country 4 prefers the formation of more 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 $\{\{1, 2\}, \{3, 4, 5, 6\}\}$. Therefore, if country 4 adequately compensates, country 3 accepts the proposal. We can show that this is a general property. In other words, keeping everything fixed, permuting $\mu_i$ leads to an efficiency gain in the limit.

**Proposition 6.** Assume that the grand coalition is not stable. Then, all patient proposers and respondents in the limit among all equilibrium coalition structures, $M^*$, which only differ in the scarcity rent of the countries, prefer the more efficient coalitions. Furthermore, every initial proposer of a coalition $M^*$, at every history that it proposes, makes an acceptable offer to $m^*$ number of countries, and all offers are accepted without any delay.

The efficiency preferences of countries here need to be interpreted with care. The proposition is not stating that countries prefer the most 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.
The formal proof is in the Appendix. Here, fixing the equilibrium numerical coalition structure, $\mathcal{M}^*$, characterised by Proposition 5, 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. It may sound striking that in the limit, although the heterogeneity with respect to the scarcity rent of all countries in the value function of the proposer vanishes (in long run they are the same in that respect), damages under any such two coalitions are not the same. But note that the heterogeneity almost vanishes and not precisely. Hence, any such proposer would prefer the most efficient coalition. This is an intuitive result and supports our analysis in the limit. 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. But no proposer should lose the opportunity of being a proposer, 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. Thus any initial proposer in equilibrium makes acceptable offers 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 countries, the stock of global cumulative emissions $S_t$ and the individual capital stocks $K_{it}$, and individual scarcity rents $\mu_{it}$ for all $i \in I$. We assume that the negotiations start while coalition structure $\mathcal{M}$ is in place at the beginning of period $t$ (this can be the coalition structure of singleton). Reversible coalition formation can be thought 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 coalition structure other than the grand or the coalition structure of singleton, then the countries are heterogeneous with respect to $K_{it}$, $R_{it}$ and $\mu_{it}$ (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.\footnote{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.} As before, it is a deterministic protocol. The assumption of binding agreements and reversibility can coexist. In fact, we continue to 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 include all parties of an existing binding treaty can approve the move to another state. This treaty will be affected by the new state: those whose membership will be affected, and those whose payoffs will be directly affected.\footnote{By direct affect on payoffs we do not refer to the indirect affect through the channel of externally of global temperature on the signatories of other coalitions, but we refer to the effect of change of their per-unit SCC and emissions.} 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. It 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.

A final point that in the model with fossil fuels as the only source of energy, 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 the countries have a finite scarcity rent.

**Proposition 7.** With reversible coalition formation of farsighted and patient countries at any time $t \in \{0, 1, 2, \ldots\}$ a grand coalition forms in equilibrium if the total number of countries is an element of the Tribonacci set, i.e. $T^* = \{1, 2, 4, 7, 13, 24, \ldots\}$. Furthermore, the MPE has an absorbing membership state with the same equilibrium 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 from any initial coalition structure, the equilibrium converges to the same membership state. The proof is in the Appendix. Proposition 7 states that even if the countries have the option to renegotiate every period, the MPE numerical coalition structure is the same as in case of irreversible agreements. This is an extension of Proposition 5. In addition the approval committee ensures that not only the same numerical coalition structure, \( M^* \), forms, but also the same coalition structure \( M^* \) forms upon every renegotiation. Thus, if coalition structure \( M^* \) is formed in period \( t - 1 \), moving to period \( t \), we expect the same equilibrium coalition structure if they were to renegotiate.

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 \( M^* \) 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 integrated assessment model 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, \( E_{it} \), can be sourced from either fossil fuel or 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) with

\[
Y_{it} = \exp(-\gamma T_t) A_{iy} K_{it}^{1-\nu} E_{iyt}^{\nu} \tag{6.1}
\]

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. Let us assume therefore 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
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 $A_{iy}$. The feasibility constraint for the final good, thus becomes

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

where the RHS side is the sum of consumption and the investment in capital used in the final goods and green energy sectors of each country $i$ at $t$.

We continue to assume that 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, $E_{igt}$, to the price 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, $E_{igt}$. Then, each country $i \in M$ finds its optimal level of investment in green technologies by equating the marginal productivity of $E_{igt}$ 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 to 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 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 rule, 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, then growth rate of output and thus growth rate of consumption can remain positive, while \( \mu_{it} \) remains finite.

See the analytical counterpart of this discussion in Appendix 8.9, and the resulting solutions for \( E_{it} > 0 \) and \( g_{it} > 0 \) in the first phase of decarbonisation, given in the following proposition

**Proposition 8.** In the integrated assessment model with green as well fossil fuel energy, the unique interior emission level and investment in green technology in each country \( i \in M \) are, respectively,

\[
E_{it}(m) = \frac{\nu}{\mu_{it}(1 - s_{it}) + \hat{\Lambda}(m)} - g_{it}
\]

\[
= \frac{\nu - Y_{it}(m)/d_i(\mu_{it}(1 - s_{it}) + \hat{\Lambda}(m))^2}{\mu_{it}(1 - s_{it}) + \hat{\Lambda}(m)} \quad (6.5)
\]

\[
g_{it}(m) = \frac{\mu_{it}(1 - s_{it}) + \hat{\Lambda}(m)}{d_i} \quad (6.6)
\]

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, 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 \((1 - s_{it}) \sum_{i \in M} \sum_{\tau=0}^{\infty} \beta^\tau \gamma \xi \frac{1}{s_{it + \tau + 1}}\) is the per-unit SCC, but it is impossible 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 also 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 integrated assessment model put forward by Golosov et al (2014). This assumption also resonates with Peters et al. (2009), which show that in long-run the saving rate
is relatively constant over time. Given this assumption, the per-unit SCC is \( \hat{\Lambda}_i(m) = \hat{\Lambda}(m) = \frac{\gamma m}{1-\beta} \) for all \( i \in M \) at any time \( t \).

The countries which participate in climate negotiation are assumed to have a finite \( \mu_{it} \). Here, the nature of heterogeneity with respect to \( d_i \) is similar to \( K_{i0} \) and \( A_{iy} \), as the independence of (4.9) 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 it 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 \( E_{igt} \) 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 4, we get the following proposition.

**Proposition 9.** If we assume that the saving rate is a constant in our model 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 \( M^* \) can be characterised independent of the heterogeneity with respect to \( d_i \). Furthermore, the grand coalition occurs in equilibrium if the total number of countries, \( N \), is a member of the Tribonacci set, i.e. \( \mathcal{T}^* = \{1, 2, 4, 7, 13, 24, \ldots \} \).

The proof is given in the Appendix. 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 9 implies that the equilibrium numerical coalition structure can be characterised in the same way described in the model without the green technologies. And, again here the number of signatories of climate coalitions in equilibrium is a Tribonacci number.

### 7 Conclusion

We have examined the formation of climate coalitions with heterogeneous countries. We offer an approach to characterise the equilibrium numerical coalition structure of countries independent of their heterogeneity. We fully characterise the unique equilibrium number of coalitions and their number of signatories, and we show these are independent

---

\(^{31}\)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 \( 0 < \bar{s} < 1 \).
of the different fossil fuel reserves, and capital stocks held, total factor productivity and scarcity rents faced by these countries.

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 (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 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 departed from the cartel model of climate coalition formation and we have allowed 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, that the more commonly used cartel stability solution concept predicts. In fact, 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. Our results are independent of relying on any remedy to fix the small-coalition paradox resulting from the cartel stability assumption in the literature.

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 included in the analysis. 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.
8 Appendix

8.1 The decision-making of a signatory

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:

\[
\max_{\{E_{it+\tau}\}_{\tau=0}^{\infty}} \sum_{i \in M} \sum_{\tau=0}^{\infty} \beta^\tau \ln(C_{it+\tau})
\]

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 \(\beta^\tau \zeta_{it+\tau}\) be present value Lagrange multiplier associated with non-negativity constraints. The first-order condition of \(E_{it}\) gives:

\[
\lambda_{it} \left[ \frac{\nu Y_{it}}{E_{it}} \right] - \sum_{i \in M} \sum_{\tau=0}^{\infty} \lambda_{it+\tau} \beta^\tau \gamma \xi_{it+\tau} = \mu_{it}
\]  

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.

\[
\max_{\{C_{it+\tau}, K_{it+\tau+1}, R_{it+\tau+1}\}_{\tau=0}^{\infty}} \sum_{\tau=0}^{\infty} \beta^\tau \ln(C_{it+\tau})
\]

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

\[
\lambda_{it} = \frac{1}{C_{it}(m, M)}
\]

First-order condition of \(K_{it+1}\) and (8.4) give the Euler equation of consumption:

\[
\frac{1}{C_{it}(m, M)} = \beta \frac{1}{C_{it+1}(m, M)} \frac{Y_{it+1}(m, M)}{K_{it+1}(m, M)} (1 - \nu)
\]

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

\[
\frac{s_{it}}{1 - s_{it}} = \beta \frac{1}{1 - s_{it+1}} (1 - \nu)
\]

The unique solution to this problem is: \(s_{it} = s = \beta (1 - \nu)\), for all \(t\) and all \(i\).
Given first-order condition of emissions in (8.2), and using (8.4), 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, (8.2) simplifies to:

\[
\frac{\nu}{E_{it}(m)} \frac{\nu}{E_{it+1}(m)} = [1 - \beta(1 - \nu)] \mu_{it} \tag{8.7}
\]

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 \(\mu_{it}\) is the shadow value of the resource. This equation can be re-arranged to

\[
E_{it}(m) = \frac{\nu}{\mu_{it}[1 - \beta(1 - \nu)] + \hat{\Lambda}(m)} \tag{8.8}
\]

Equation (8.8) 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 of \(R_{t+1}\) is \(\mu_{it} = \beta \mu_{it+1}\). Using this and equation (8.7), gives the Euler equation of energy consumption:

\[
\frac{\nu}{E_{it}(m)} \frac{\nu}{E_{it+1}(m)} = \beta(\frac{\nu}{E_{it+1}(m)} - \hat{\Lambda}(m)) \tag{8.9}
\]

which simplifies to:

\[
\frac{\beta \nu E_{it}(m)}{\nu - (1 - \beta) \hat{\Lambda}(m) E_{it}(m)} = E_{it+1}(m) \tag{8.10}
\]

It can be shown that

\[
E_{it+\tau}(m) = \frac{\beta^\tau \nu E_{it}(m)}{\nu - (1 - \beta^\tau) \hat{\Lambda}(m) E_{it}(m)} \tag{8.11}
\]

for any \(\tau \geq 1\).

### 8.2 Optimum value function of a signatory

Let \(V_i(S_i, K_{it}, \mu_{it}, m, M)\) be the optimum value function of a signatory in coalition \(M\) of size \(m\) in numerical coalition structure \(M\). By substituting the solutions in the summation of flow and continuation utility of the representative consumer of country \(i \in M\):
\[ V_i(S_t, K_{it}, \mu_{it}, m, M) = \ln(C_{it}(m, M)) + \beta \ln(C_{it+1}(m, M)) + \ldots \]
\[ = \frac{\ln(1 - s)}{1 - \beta} + \{ \ln(Y_{it}(m, M)) + \beta \ln(Y_{it+1}(m, M)) + \ldots \} \]
\[ = \frac{\ln(1 - s)}{1 - \beta} + \{ \ln(e^{-\gamma \xi S_t - \gamma T_0} A_{iy} K_{it}^{1 - \nu} E_{it}(m)') \]
\[ + \beta \ln[e^{-\gamma \xi S_{t+1} - \gamma T_0} A_{iy} K_{it+1}(m, M)^{1 - \nu} E_{it+1}(m)'] + \ldots \} \]
\[ = \frac{(1 - \nu) \ln(K_{it}) + H_1 + H_2 + H_3}{1 - s} \]

where \( H_j \) are defined as below:

\[ H_1 \equiv \frac{s \ln(s) - s \ln(1 - s) + \ln(A_{iy}) - \gamma T_0}{1 - \beta} \] (8.13)

and,

\[ H_2 \equiv -\gamma \xi [S_t + \beta S_{t+1} + \beta^2 S_{t+2} + \ldots] \] (8.14)

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

\[ H_2 = -\frac{\gamma \xi}{1 - \beta} \left( \sum_{i} \sum_{s=1}^{t} E_{it-s} + \sum_{j \notin M} E_{jt} + \sum_{i \in M} E_{it}(m) \right) \]
\[ + \sum_{j \notin M} \{ \beta E_{jt+1} + \beta^2 E_{jt+2} + \ldots \} + \sum_{i \in M} \{ \beta E_{it+1}(m) + \beta^2 E_{it+2}(m) + \ldots \} \] (8.15)

Finally,

\[ H_3 \equiv \nu \{ \ln(E_{it}(m)) + \beta \ln(E_{it+1}(m)) + \beta^2 \ln(E_{it+2}(m)) + \ldots \} \] (8.16)

8.3 Internal stability condition

In this section, assume that there is a single coalition of size \( m \), and the rest are non-signatories (fringe). A signatory does not have any incentive to leave a coalition of size \( m \) if

\[ V_i(S_t, K_{it}, \mu_{it}, m) \geq \ln(C_{it}^d) + \beta \{ V_i(E_t^t, K_{it+1}, \mu_{it}, m) \} \] (8.17)

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 8.2, and \( E_t^t \equiv (E_t, E_{t-1}, \ldots, E_0) \). Furthermore, \( C_{it}^d \) is the consumption level associated with the deviation period. Note that \( E_t^t \) in \( V_i(E_t^t, K_{it+1}, \mu_{it}, m) \) is impacted by the deviation in period \( t \). More specifically, the right-hand-side of (8.17) consists of,
\[ \ln(C_{it}^d) = \ln(1 - s) + \ln(Y_{it}^d) \]

and

\[ V_i(E^t, K_{it+1}, \mu_{it}, m) = \frac{(1 - \nu)\ln(K_{it+1}) + H_1 + H'_2 + H'_3}{1 - s} \quad (8.18) \]

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

\[ H'_2 \equiv -\gamma \xi \left[ (S_{t+1}) + \beta (S_{t+2}) + \ldots \right] \]

\[ = -\frac{\gamma \xi}{1 - \beta} \left\{ \sum_i \sum_{s=1}^t E_{it-s} + \sum_{j \notin M \setminus i} E_{jt} + \sum_{j \in M \setminus i} E_{jt}(m - 1) \right\} + \sum_{i \notin M} [E_{it+1} + \beta E_{it+2} + \ldots] + \sum_{i \in M} [E_{it+1}(m) + \beta E_{it+2}(m) + \ldots] \quad (8.19) \]

and,

\[ H'_3 \equiv \nu \left[ \ln(E_{it+1}(m)) + \beta \ln(E_{it+2}(m)) + \ldots \right] \quad (8.20) \]

By multiplying both sides of the internal stability condition (8.17) 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:

\[ \ln(Y_{it}^d) = -\gamma \xi \left\{ \sum_i \sum_{s=1}^t E_{it-s} + \sum_{j \notin M \setminus i} E_{jt} + \sum_{j \in M \setminus i} E_{jt}(m - 1) \right\} + \ln(A_{iy}) + (1 - \nu) \ln(K_{it}) + \nu \ln(E_{it}) \quad (8.21) \]

then (8.17) can be simplified to:

\[ \nu \left[ \ln(E_{it}(m)) - \ln(E_{it}) \right] + \frac{\gamma \xi}{1 - \beta} E_{it} - \frac{\gamma \xi m}{1 - \beta} E_{it}(m) + \frac{\gamma \xi (m - 1)}{1 - \beta} E_{it}(m - 1) \geq 0 \quad (8.22) \]

Using the corresponding emission levels, this condition can be written as a function of variable \( m \) and parameters \( \beta, \nu, \gamma \) and the shadow value of resource \( \mu_{it} \). Thus the roots of (8.22) give the equilibrium coalition size \( m^* \). For any parameter values, the roots are one and maximum three.

### 8.4 Proof of Proposition 4

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 \( M \in \{ M_1, M_2, \ldots, M_k \} \) (which is to be identified) versus the total
payoff of the corresponding \( m \) members from staying in the grand coalition \( \{I\} \). Thus, \( j \) needs to determine the sign of

\[
\sum_{i=1}^{m} V_i^j(S_t, K_{it}, \mu_{it}, M, \mathbb{M}) - \sum_{i=1}^{m} V_i^j(S_t, K_{it}, \mu_{it}, I) \tag{8.23}
\]

Assume the countries are heterogenous with respect to \( K_{i0} \) and/or \( A_{iy} \). But this equation is independent of stocks, in particular independent of \( K_{i0} \). It is also independent of \( A_{iy} \). To see this, note that

\[
V_i(M_i, \{M_1, M_2, \ldots, M_k\}) - V_i(\{I\}) =
\frac{1}{1 - \beta (1 - \nu)} \left\{ \nu \left( \ln \left( \frac{E_{it}(M)}{E_{it}(I)} \right) + \beta \ln \left( \frac{E_{it+1}(M)}{E_{it+1}(I)} \right) + \ldots \right) \right.
\]

\[
- \frac{\gamma \xi}{1 - \beta} \left\{ (\sum_{i \in M_1} E_{it}(M_1) + \sum_{i \in M_2} E_{it}(M_2) + \ldots + \sum_{i \in M_k} E_{it}(M_k) - \sum_{i \in I} E_{it}(I)) \right. + \beta \left[ \sum_{i \in M_1} E_{it+1}(M_1) + \sum_{i \in M_2} E_{it+1}(M_2) + \ldots + \sum_{i \in M_k} E_{it+1}(M_k) - \sum_{i \in I} E_{it+1}(I) \right] \right. \}
\]

\[
\right. \}
\right. \}
\right. \}
\right. \}
\]

\[
(8.24)
\]

Sets \( M_p \in \{M_1, M_2, \ldots, M_k\} \) are included if they are non-empty. Thus, if the source of heterogeneity is either \( K_{i0} \) or \( A_{iy} \), then the membership decision of countries in (8.23) is not affected by the heterogeneity.

Furthermore, the above equation is only function of emissions, which only depend on \( m \) and \( M \) and not \( M \) and \( \mathbb{M} \). Hence \( M^* \) can be caracterised independent of heterogeneity with respect to \( K_{i0} \) or \( A_{iy} \).

If the countries are heterogeneous with respect to \( R_{i0} \) and \( \mu_{it} \), in the limit that \( \beta \to 1 \), the equation in (8.24) converges to

\[
\lim_{\beta \to 1} V_i(M_i, \{M_1, M_2, \ldots, M_k\}) - V_i(\{I\}) =
\left[ \ln \left( \frac{N}{m} \right) + \ln \left( \frac{N}{m} \right) + \ldots \right] - \left\{ \left[ k - 1 \right] + \left[ k - 1 \right] + \ldots \right\}
\]

\[
(8.25)
\]

This equation is independent of \( \mu_{it} \) of any country and any stocks. Moreover, it only depends on \( m \) and \( M \). □

8.5 Proof of Lemma 1

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

\[
V_i(m_1, \{m_1, m_2, \ldots, m_k\}) - V_i(\{N\}) < 0 \tag{8.26}
\]
In our model, it can be shown that

\[
V_i(m_1, \{m_1, m_2, ..., m_k\}) - V_i(\{N\}) = \\
\frac{1}{1 - \beta (1 - \nu)} \left\{ \nu \ln \left( \frac{E_i(m_1)}{E_i(N)} \right) + \beta \ln \left( \frac{E_i+1(m_1)}{E_i+1(N)} \right) + ... \right\} \\
- \frac{\gamma \xi}{1 - \beta} \left\{ \sum_{i \in M_1} E_i(m_1) + \sum_{i \in M_2} E_i(m_2) + ... + \sum_{i \in M_k} E_i(m_k) - \sum_{i \in I} E_i(N) \right\} \\
+ \beta \left[ \sum_{i \in M_1} E_i+1(m_1) + \sum_{i \in M_2} E_i+1(m_2) + ... + \sum_{i \in M_k} E_i+1(m_k) - \sum_{i \in I} E_i+1(N) \right] + ... \right\} \\
\] (8.27)

For \( \beta \to 1 \), the inequality in (8.26) converges to

\[
\lim_{\beta \to 1} V_i(m_1, \{m_1, m_2, ..., m_k\}) - V_i(\{N\}) = \\
\nu \left[ \ln \left( \frac{N}{m_1} \right) + \ln \left( \frac{N}{m_1} \right) + ... \right] - \nu \left[ [k - 1] + [k - 1] + ... \right] < 0 \\
\] (8.28)

This is satisfied if

\[
\ln \left( \frac{N}{m_1} \right) < (k - 1) \] (8.29)

as required. □

### 8.6 Proof of Proposition 5

If \( T^* = \{1, 2, 4, ..., T_n, T_{n+1}, ...\} \), we need to show that \( T_{n+1} \) is also sum of the last three elements of the set. This proof is by induction. Assume \( T^* = \{1, ..., T_{n-3}, T_{n-2}, T_{n-1}, T_n, ...\} \), and \( T_n = T_{n-3} + T_{n-2} + T_{n-1} \), and \( T_{n-3} > 1 \). We need to show \( T_{n+1} \in T^* \) is such that \( T_{n+1} = T_{n-2} + T_{n-1} + T_n \).

The assumption implies: \( \lim_{\beta \to 1} V_i(T_{n-3}, \{T_{n-3}, T_{n-2}, T_{n-1}\}) - V_i(\{T_n\}) < 0 \). Using inequality (8.29) in Lemma 1, this is equivalent to

\[
\frac{T_{n-3} + T_{n-2} + T_{n-1}}{T_{n-3}} < e^2 \] (8.30)

and we need to show

\[
\frac{T_{n-2} + T_{n-1} + T_n}{T_{n-2}} < e^2 \] (8.31)

Replacing for \( T_n = T_{n-3} + T_{n-2} + T_{n-1} \), inequality (8.31) is equivalent to

\[
2 \frac{T_{n-3} + T_{n-2} + T_{n-1}}{T_{n-3}} - 1 < e^2 \frac{T_{n-2}}{T_{n-3}} \] (8.32)

The assumption (8.30) implies that,
\[
\frac{2T_{n-3} + T_{n-2} + T_{n-1}}{T_{n-3}} - 1 < 2e^2 - 1
\]  
(8.33)

Thus a sufficient condition for (8.32) to be satisfied is to show \(2e^2 - 1 \leq e^2 \frac{T_{n-2}}{T_{n-3}}\); equivalently to show,

\[
\frac{2e^2 - 1}{e^2} \leq \frac{T_{n-2}}{T_{n-3}}
\]  
(8.34)

The R-H-S is the Tribonacci constant, which is the ratio towards which consecutive Tribonacci numbers tend. By rounding to the second decimal place, both sides of the inequality in (8.34) converge to 1.85. Finally, given the predetermined elements of the Tribonacci set, i.e. \(\{0, 0, 1\}\), for \(n = 1\) the rule of \(T_n = T_{n-3} + T_{n-2} + T_{n-1}\) is satisfied too. □

8.7 Proof of Proposition 6

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 \(\{m^*_1, m^*_2, ..., m^*_k\}\), 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. Country \(i\) itself is in both coalitions. In that case,

\[
V^i(M_{k-1}, \{M_1, M_2, ..., M_\}) - V^i(M'_{k-1}, \{M_1, M_2, ..., M_{k-2}, M'_{k-1}, M_\}) = \\
\frac{1}{1 - \beta(1 - \nu)} \left\{ \nu \left[ \ln \frac{E_{it}(M_{k-1})}{E_{it}(M'_{k-1})} \right] + \beta \ln \left( \frac{E_{it+1}(M_{k-1})}{E_{it+1}(M'_{k-1})} \right) + ... \right\} \\
- \frac{\gamma \xi}{1 - \beta} \left[ \sum_{i \in M_{k-1}} E_{it}(M_{k-1}) + \sum_{i \in M_\} \right] - \left[ \sum_{i \in M_{k-1}} E_{it}(M'_{k-1}) + \sum_{i \in M'_{k-1}} E_{it}(M') \right] + \\
\beta \left[ \sum_{i \in M_{k-1}} E_{it+1}(M_{k-1}) + \sum_{i \in M_\} \right] - \left[ \sum_{i \in M_{k-1}} E_{it+1}(M'_{k-1}) + \sum_{i \in M'_{k-1}} E_{it+1}(M') + ... \right] \right\}
\]  
(8.35)

Emission of those which remain in coalitions with the same sizes does not affect \(i\)'s decision. The second line in (8.35) 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, the ratio two emissions is one, and thus the second line is zero. The third and fourth lines are the externality damages, and in the limit,
\[
\lim_{\beta \to 1} V_i^i(M_{k-1}, \{M_1, M_2, \ldots, M_k\}) - V_i^i(M'_{k-1}, \{M_1, M_2, \ldots, M_{k-2}, M'_{k-1}, M'_k\}) \\
= -\{(\frac{m_{k-1}}{m_{k-1}} + \frac{m_k}{m_k}) - (\frac{m_{k-1}}{m_{k-1}} + \frac{m_k}{m_k}) + \ldots\} \\
= -2(1 - 1 + 1 - 1 + \ldots \quad (8.36))
\]

Note that \(\sum_{n=0}^{\infty} (-1)^n\) is the Grandi’s series, which is a divergent series. But (relatively) recent mathematical methods assign the summation of \(\frac{1}{2}\) to this series. Cesàro summation and Abel summation are among methods which conclude that the sum is \(\frac{1}{2}\).\(^{32}\) Hence, the difference equation in (8.36) is negative.

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. Finally, in coalition formation in a public good game, no initial proposer loses chance of being a proposer and makes acceptable offers which are accepted without any delay. □

8.8 Proof of Proposition 7

Assume that at the beginning of period \(t\), the coalition structure \(\mathbb{M}\) is the initial membership state. Note that as stocks of \(K_t\) and \(S_t\) have changed from period \(t - 1\), by the beginning of period \(t\) we have moved to a new state. In period \(t\) the same initial proposer \(i \in M^* \in \mathbb{M}\) can make a proposal. Given the binding assumption, the approval committee which consists of the \(m\) signatories of \(M\) must approve the move. Given the fixed deterministic protocol, no party has a profitable deviation. To see this note that from Proposition 5 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 5 implies that the equilibrium strategy of the approval committee 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 \(M^*\) 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 \(M^*\) that was irrelevant). There can be histories in which the signatories of \(M^*\) which were respondents in period \(t - 1\), now in period \(t\), 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

\(^{32}\)For example, see Davis (1989, p.152).
offer is accepted in all periods $\tau > t - 1$ form the only equilibrium path. □

### 8.9 Proof of Proposition 8

Using the Lagrange method similar to section 8.1, 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,

$$\frac{\nu}{E_{igt}} \leq (1-s_{it}) \sum_{i \in M} \sum_{\tau=0}^{\infty} \frac{\beta^\tau \gamma \xi}{1-s_{it+\tau+1}} + (1-s_{it}) \mu_{it} ; E_{it} \geq 0 ; \text{complementary slackness condition}$$

(8.37)

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

$$E_{it}(m) = \frac{\nu}{\mu_{it}(1-s_{it}) + \Lambda(m)} - g_{it}(m)$$

(8.38)

where

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

Then, 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 $C_{it}(m) = (1-s_{it})Y_{it}(m)$, and $K_{it+1}(m) + \frac{d_i}{2} g_{it}(m)^2 = s_{it} Y_{it}(m)$, the Euler equation of saving rate is:

$$s_{it} Y_{it}(m) - \frac{d_i}{2} g_{it}(m)^2 \leq \frac{\beta(1-\nu)}{(1-s_{it+1})}$$

(8.39)

First-order condition of $g_{it}$ gives,

$$\frac{\nu Y_{it}(m)}{E_{it}(m) + g_{it}(m)} \leq d_i g_{it}(m) ; g_{it} \geq 0 ; \text{complementary slackness condition}$$

(8.40)

If $E_{it} \geq 0$ and $g_{it} \geq 0$, solving this with optimal emission decision of coalition in (8.38) results in

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

(8.41)

$$E_{it}(m) = \frac{\nu - Y_{it}(m)/d_i[\mu_{it}(1-s_{it}) + \hat{\Lambda}(m)]^2}{\mu_{it}(1-s_{it}) + \Lambda(m)}$$

(8.42)

as stated in the Proposition. □

### 8.10 Proof of Proposition 9

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,
\[ V_i^g(S_t, K_{it}, \mu_{it}, m, M) = \ln(C_{it}(m, M)) + \beta \ln(C_{it+1}(m, M)) + ... \]
\[ = \frac{\ln(1 - \bar{s})}{1 - \beta} + \{\ln(Y_{it}(m, M)) + \beta \ln(Y_{it+1}(m, M)) + ...\} \]
\[ = \frac{(1 - \nu)\ln(K_{it}) + H_1^g + H_2^g + H_3^g}{1 - \beta(1 - \nu)} \]  

where are defined as below:
\[ H_1^g = \frac{\beta(1 - \nu)\ln(\bar{s}) - \beta(1 - \nu)\ln(1 - \bar{s}) + \ln(A_{iy}) - \gamma T_0}{1 - \beta} \]  

and other \( H_j^g \), where \( j \in \{2, 3\} \), are the same as corresponding \( H_j \) derived in section 8.2, just replacing \( E_{it} \) with \( E_{iyt} \). Furthermore, using Proposition 8, it is easy to show that
\[ E_{it}(m) + g_{it}(m) = \frac{\nu}{\Lambda(m) + (1 - \bar{s})\mu_{it}} \]  

this is indeed the optimal emission level in the model without the green technology. Therefore, the previous analysis of farsighted countries to determine \( M^* \) 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 (8.29), just replacing \( E_{it} \) with \( E_{iyt} \). 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 8.3 stands here for symmetric countries. □
References


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


