CESifo WORKING PAPERS

9768
2022

Original Version: May 2022
This Version: October 2023

Self-Enforcing Climate Coalitions for Farsighted Countries: Integrated Analysis of Heterogeneous Countries

Sareh Vosooghi, Maria Arvaniti, Federick van der Ploeg

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9768*

Self-Enforcing Climate Coalitions for Farsighted Countries: Integrated Analysis of Heterogeneous Countries

Abstract

We study formation of international climate coalitions. Countries are farsighted and rationally predict the consequences of their membership decisions in climate negotiations. Within the context of an integrated assessment model of the economy and the climate, we characterise the equilibrium number of coalitions and their number of signatories independent of certain types of heterogeneity, and show that the resulting treaties are robust to renegotiation. With a richer structure of energies we investigate possible coalition outcomes for a calibrated model. We confirm our heterogeneity results and in contrast to earlier approaches based on internal and external stability, much larger coalitions can be sustained in equilibrium.

JEL-Codes: Q540, D700, D500.

Keywords: climate economics, international environmental agreements, coalition formation, heterogeneous countries, integrated assessment models.

Sareh Vosooghi
Department of Economics
KU Leuven / Belgium
sareh.vosooghi@kuleuven.be

Maria Arvaniti
Department of Economics
University of Bologna / Italy
maria.arvaniti@unibo.it

Frederick van der Ploeg
Department of Economics
University of Oxford / United Kingdom
rick.vanderploeg@economics.ox.ac.uk

October 2023

We thank Bard Harstad, Tim Worrall, Margaret Meyer, Marek Pycia, Piotr Dworczak, Michal Kobielarz, Paula Onuchic, Ludvig Sinander, Aart de Zeeuw, Simon Cowan and conference, seminar and workshop participants at the European Winter Meeting of the Econometric Society, the Royal Economic Society Conference, EEA-ESEM conference at Bocconi, University of Oxford, the Transatlantic Theory Workshop, Nuffield Economic Theory Workshop, EAERE conference, CRETE, FAERE conference, ASSET conference, Environmental Protection and Sustainability Forum and the CEC Paris Dauphine Annual Conference for their helpful comments and conversations. We also thank the editor and the anonymous referees for all their valuable comments and suggestions, and thank Amin Mirabdollah for his excellent research assistance. Any errors are our own.

1 Introduction

The biggest planetary tragedy is the failure of countries to work together to curb anthropogenic greenhouse gases (GHG) emissions and global warming. After three decades of climate negotiations, there is still no effective and self-enforcing internationally cooperative climate policy. One reason might be the focus on formation of the unlikely grand climate coalition of all countries concerned and the lack of a more sophisticated approach regarding countries' motives to join a coalition.

We explore the idea that allowing for multiple climate coalitions among smaller groups of more sophisticated countries might result in self-enforcing and overall more ambitious agreements than what we observe today. We thus consider farsighted countries and their motives to join a climate treaty and study these in an integrated assessment model of the economy and the climate (IAM).

Our main contribution is to characterise equilibrium coalition structures in a well-defined IAM which incorporates the general-equilibrium effects in climate coalition formation with farsighted agents, taking into account certain types of heterogeneity among countries and the possibility of renegotiation of agreements. In addition to the analytical characterisation of the equilibrium coalition structure for such an IAM, to the best of our knowledge, this is the first paper which delivers numerical results on equilibrium coalition formation with farsighted players in a calibrated IAM. We show that with farsighted countries larger climate coalitions can be sustained, and multiple climate coalitions of different sizes can be sustained alongside each other.

Most of the literature on international climate coalition formation abstracts from details of macroeconomic outcomes and their underlying determinants (see the review in Battaglini and Harstad (2016)). Instead, climate economists have developed multi-country IAMs which are growth models that allow for examining the effects of the economy on global warming and vice versa, and are used to analyse different scenarios such as business-as-usual (BAU), and the global social optimum, (e.g., Hassler and Krusell (2012)), but they are typically not used to analyse the strategic interactions of countries seeking climate agreements. We build a bridge between the literature on climate coalition formation and the one on optimal climate policies from IAMs. We thus capture the broader incentives of policymakers in climate negotiations and analyse international climate coalition within the context of macroeconomic, multi-country IAMs.

We allow for heterogeneity across countries with respect to the initial stocks of capital, total factor productivities, and the initial levels of exhaustible fossil fuel reserves. With such asymmetries one cannot usually use the conventional coalition formation methodologies developed for symmetric countries. We offer an approach to *decouple* the problem of characterising equilibrium number of coalitions and their number of signatories from their heterogeneity which relies on the fact that the source of heterogeneity affects the payoffs of countries in an additive-separable form. Although this treatment depends on the payoff structure, it provides a stepping stone in studying heterogeneity in the theory of coalition formation with farsighted agents.

To keep matters tractable, we use a multi-country IAM of economic growth and climate dynamics based on Golosov et al. (2014) and van der Ploeg and Rezai (2021) and integrate it with an analysis of the decision of individual countries to participate in international climate coalition formation. The optimal SCC is then proportional to current output and independent of future values of output or consumption, and the model replicates the properties of the Dynamic Integrated Climate-Economy (DICE) model of Nordhaus (1993) reasonably well.

We let policy makers of each country be farsighted: negotiating countries rationally consider all self-enforceable unilateral and multilateral deviations from their membership decisions, and predict the entire structure of conceivable coalitions. This contrasts with the most commonly used solution concept of internal and external cartel stability (a Nash equilibrium), which assumes that countries are myopic and are only concerned with the immediate gain or losses of their unilateral deviations and do not take account of reactions of other countries. Ignoring retaliation by other countries after breaking off climate negotiations increases incentives for free riding. Deviation from a coalition is then less costly since it is assumed that the rest of the coalition remains intact and the agreed policy will not deteriorate a lot. Hence, use of cartel stability conditions result in very small coalitions (see seminal papers by Carraro and Siniscalco (1993) and Barrett (1994)) Since this results from the unrealistic assumption that countries are myopic, we investigate how robust the finding of small coalitions is if countries are farsighted.

To find the equilibrium number of signatories, we check whether the membership strategy under consideration is immune to deviations. Here farsightedness implies that deviations from equilibrium strategies which are not themselves self-enforceable must be excluded. Thus, characterisation of the equilibrium structure of coalitions relies on an algorithm which recursively identifies the set of the total number of countries for which a grand coalition forms in equilibrium. The recursion starts from the smallest total number of countries that can form a self-enforceable coalition and continues to some finite integer. This set determines the possible farsighted deviations. Furthermore, the number of members of equilibrium coalition(s) is a subset of this set. If countries have a one-off payoff, the comparison of payoffs and the characterisation in each step of the algorithm is not too demanding. In an infinite-horizon IAM the recursion process can be onerous and generally requires one to resort to numerical simulations. However, with our IAM we can obtain intuitive and analytical results in some cases.

In particular, for our model with only exhaustible energy, we can obtain analytical results if policymakers are very patient. We find a simple condition that characterises the equilibrium number of coalitions and their number of signatories. We show that the number of signatories of climate coalitions is a Tribonacci number if the total number of countries is fewer than 78.[1] Moreover, we show that our results are robust to countries renegotiating their membership decisions.

For more general versions of our IAM, we calibrate to match the current world situation and simulate the recursive algorithm to find the equilibrium coalition structure of farsighted countries. We confirm that large self-enforceable coalitions are feasible and multiple coalitions can co-exist together in equilibrium. Furthermore, the degree of internalisation of global warming externalities is much higher than under the conventional solution concept of cartel stability. Hence, climate policy is more ambitious with a higher SCC, and thus emissions and temperature are curbed rather more with smaller drops in consumption and capital due to global warming. Furthermore, we obtain robust coalitional equilibria under heterogeneity, across different versions of our model, that confirm our decoupling result. Given the derived equilibrium coalitional structure, we can back out the macroeconomic policies, global temperature, growth rate, energy consumption, and the optimal climate policy for the various countries associated with self-enforceable climate treaties. This enriches the usual economic approaches that have been used in the literature on international climate coalition formation.

Section 2 reviews related literature. Section 3 presents our multi-country IAM. Section 4 derives optimal decisions for a given coalition structure. Section 5 analyses climate coalition formation, for symmetric and asymmetric countries. Section 5.4 discusses reversible agreements, where countries can renegotiate any existing agreement. The calibration of our IAM and the quantitative analysis of coalition formation is presented in section 6. Section 7 concludes. All proofs are relegated to the Appendix.

2 Related Literature

Research on International Environmental Agreements (IEAs) or climate governance and international cooperation by forming climate coalitions has led to an extensive literature (e.g., Carraro and Siniscalco, 1993; Barrett, 1994; and the reviews in Battaglini and Harstad, 2016, and Benchekroun and Long, 2012). This literature has provided inputs into the design of international climate treaties, including the Paris Climate Accord.

Most studies employ the solution concept of cartel stability which implies that only unilateral deviations are checked while taking the membership decision of the complementary set of players as given. This leads to the *small-coalition paradox*, which states that the maximise size of any coalition is a small number (e.g., 3 countries). This is a robust result.[2] To overcome this paradox, many remedies have been explored: international transfers (Carraro and Siniscalco, 1993; Hoel and Schneider, 1997; Carraro et al., 2006); a breakthrough green technology (Barrett, 2006); 'modest' agreements (Finus and

[1] Cf. the Fibonacci sequence in Ray and Vohra (2001).

[2] See Battaglini and Harstad (2016) for the literature on the robustness of the small coalition paradox.

Maus, 2008); a refunding club (Gersbach et al., 2021); trade-off between R&D costs and costs of adopting breakthrough technology (Hoel and De Zeeuw, 2010); markets for fuel and tradable rights (Harstad, 2012); non-quadratic functional forms (Karp and Simon, 2013); asymmetric countries (Pavlova and De Zeeuw, 2013)[3]; trade clubs (Nordhaus, 2015); and incomplete contracts of green technologies (Battaglini and Harstad, 2016).[4]

However, allowing for formation of only one climate coalition is unnecessarily restrictive. We therefore allow the formation of multiple coalitions. The literature on IEAs with Nash equilibrium and open membership allows for multiple coalitions under cartel stability (e.g., Yi and Shin, 2000;Asheim et al., 2006; Finus and Rundshagen, 2003, 2009; Finus et al. (2009)). However, we use a different solution concept for coalition formation and assume all countries are farsighted (e.g., Aumann and Myerson, 1988; Dutta et al., 1989; Chwe, 1994; Bloch, 1996; Ray and Vohra, 1997; Chatterjee et al., 1993). Ray and Vohra (1999) generalise farsighted coalition formation to allow for externalities, and Ray and Vohra (2001) allow for public goods, in an IEA context.[5] Vosooghi (2017) uses the notion of farsighted stability in a stochastic IEA setting, and Diamantoudi and Sartzetakis (2018) and De Zeeuw (2008) analyse it in deterministic IEA settings.[6] We examine a *dynamic* game extension of Ray and Vohra (2001) within the context of an IAM. Our analysis allows for heterogeneous countries and reversible agreements, which the above studies abstract from.

Our paper also relates to the literature on IAMs which while abstracting from international climate agreements try to capture the global economy and geophysics using a numerical analysis (e.g., Nordhaus, 2014; Anthoff and Tol, 2013). These IAMs have been used to obtain analytical expressions for the optimal SCC and climate policies (e.g., Golosov et al., 2014; Hassler and Krusell, 2012; van der Ploeg and Rezai, 2021; Van den Bremer and Van der Ploeg, 2021). Only a small subset of the literature combines IEAs and IAMs (e.g., Tol, 2001; Eyckmans and Tulkens, 2006; Yang et al., 2008: Buchner and Carraro, 2009) but in contrast to our approach, these papers use a cooperative game-theoretic approach.[7] Lessmann et al. (2009, 2015) and Bosetti et al. (2013) do use a non-cooperative game-theoretic approach and cartel stability within the context of IAMs. They offer numerical simulations to examine the stability of an IEA which in the absence of any remedies always results in *small* climate coalitions. Our paper uses both analytical and numerical approaches, and generalises the stability concept to allow for farsightedness and thereby departs from the small-coalition paradox.

[3]But the larger coalitions can only be sustained with small gains from cooperation.

[4]See also de Zeeuw (2015) for the different approaches and assumptions used in the literature.

[5]Ray and Vohra (2019) and Dutta and Vohra (2017) study farsighted sets. Since they use the cooperative approach and rely on the characteristic function, they do not allow for externalities.

[6]De Zeeuw (2008) studies the effect of gradual adjustment of emission reductions 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.

[7]Due to the externalities inherent in climate games, cooperative game theory in such settings has been criticised (Ray and Vohra, 2001).

3 Integrated Assessment Model

Our IAM framework has N countries; each country is indicated by the subscript $i \in I$, where $I \equiv \{1, 2, ..., N\}$. Time is discrete and indexed by $t = 0, 1, 2, ...$ Each country is represented by a social planner, who can implement any desired policy in a competitive market economy. A coalition structure is a partition of the set of countries, I, into coalitions, $\mathbb{M} \equiv \{M_1, M_2, ..., M_k\}$. Let $m \leq N$ be a positive integer showing the cardinality (i.e., the number of members) of coalition M. A *numerical coalition structure*, $\mathcal{M} \equiv \{m_1, m_2, ..., m_k\}$, is a partition of N into the sizes of coalitions.

3.1 The Economy

Each country has a representative infinitely-lived household with lifetime utility

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

where $\beta \in (0, 1)$ is a constant discount factor, C_{it} is consumption of the final good, and $U(C_{it}) = \ln(C_{it})$ is the instantaneous utility function. The intertemporal elasticity of substitution is thus equal to one.[8]

Each country i has an energy sector, E_{it}, and a final good sector, Y_{it}. Production of the final good uses labour, capital and energy. Following the DICE model of Nordhaus (1993) and the RICE model of Nordhaus and Yang (1996), global temperature negatively affects aggregate production of final output and global warming damages are proportional to aggregate output. Golosov et al. (2014) show that an exponential functional form for damages related to the stock of atmospheric carbon approximates the ratio of global warming damages to aggregate output of the DICE and RICE models reasonably well. With a Cobb-Douglas technology and constant returns to scale, aggregate output is

$$Y_{it} = \exp(-\gamma T_t) A_{0i} K_{it}^{\alpha} E_{it}^{\nu} L_{0it}^{1-\alpha-\nu} \tag{3.2}$$

with L_{0it} labour use, K_{it} the aggregate capital stock at the start of period t, and α and ν the shares of capital and energy in output, respectively.[9] Initial capital, K_{i0}, may differ across countries. The energy composite E_{it} follows from the CES production function

[8] Golosov et al. (2014) argue that this is a reasonable assumption in long-run economic growth models. Chetty (2006) estimates the coefficient of relative risk aversion and shows that the mean estimate is about one. Gandelman and Hernández-Murillo (2015) use a mega database of 75 countries and also obtain an estimate of around one.

[9] Jones (2005) provides a micro-foundation for the Cobb-Douglas aggregate production function at the macroeconomic level if the parameters of the production technology are drawn from a Pareto distribution. Moreover, Hassler et al. (2021) using historical data to estimate an aggregate production function show that the long-run input shares are stationary, which also suggests a Cobb-Douglas production technology. 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.

$$E_{it} = (\kappa_1 E_{1it}^{\rho} + \kappa_2 E_{2it}^{\rho} + \kappa_3 E_{3it}^{\rho})^{1/\rho}, \tag{3.3}$$

where E_{1it} is fossil fuel with exhaustible reserves (oil and gas), E_{2it} is fossil fuel in abundant supply (coal), and E_{3it} is green energy. Here $1/(1-\rho)$ is the constant elasticity of substitution between the three types of energy, and κ_j, for $j = 1, 2, 3$, are their respective shares in the energy production function with $\sum_{j=1}^{3} \kappa_i = 1$.[10]

Oil and gas, E_{1it}, has zero marginal extraction cost but has a shadow price, μ_{it}, which follows from the resource constraint $E_{1it} = R_{it} - R_{it+1}$ or

$$R_{it+1} = R_{i0} - \sum_{s=0}^{t} E_{1it-s}, \tag{3.4}$$

where R_{it} is the exogenous stock of reserves of oil and gas of country i at the start of period t. Initial oil and gas reserves, R_{i0}, may differ across countries. Coal production, E_{2it}, and green energy production, E_{3it}, use only labour and no capital so

$$E_{jit} = A_{jit} L_{jit} \quad \text{for} \quad j = 2, 3. \tag{3.5}$$

These energies are abundant and have no scarcity rent, and A_{2it} and A_{3it} are the labour productivities for coal and green energy production. The production function for final output is then

$$Y_{it} = \exp(-\gamma T_t) A_{0i} K_{it}^{\alpha} (\kappa_1 E_{1i}^{\rho} + \kappa_2 (A_{2it} L_{2it})^{\rho} + \kappa_3 (A_{3it} L_{3it}^{\rho}))^{\nu/\rho} L_{0it}^{1-\alpha-\nu}. \tag{3.6}$$

Total Factor Productivity (TFP) of final output has two multiplicative terms, a constant, A_{0i}, which can vary across countries, and a negative exponential function of global temperature, T_t, where γ is the damage coefficient.[11] The energy-specific factor productivities are exogenous. We assume they are constant, but relax this in the quantitative analysis of Section 6. The labour force in each country L_{it} is exogenous.

In competitive equilibrium, the labour and the final good markets have to clear and the fossil fuel resource constraints have to be satisfied for each country. We assume zero adjustment cost of capital and full capital depreciation in each period.[12] We assume no international trade in fossil fuel and abstract from any other international interactions, hence in a decentralised economy all markets clear at the national level. The only factor which links countries is the externality resulting from global warming damages. Hence,

[10] Note that κ_j measures the relative energy efficiency of the different energy sources. Since coal produces more carbon emissions than oil or gas per unit of energy and E_{1it} and E_{2it} are measured in the same units (carbon amount emitted), $\kappa_1 > \kappa_2$.

[11] The damage coefficient can be assumed to be an uncertain parameter. For example, Golosov et al. (2014) replace the damage coefficient with the expectation of a fixed and common distribution of γ. We ignore that the damage coefficient can differ across countries.

[12] Barrage (2014) using numerical methods shows robustness of the social cost of carbon characterisation to the assumption of 100% in Golosov et al. (2014). van der Ploeg and Rezai (2021) show that allowing for logarithmic depreciation does not affect the SCC.

$$C_{it} + K_{it+1} = Y_{it}, \quad L_{0it} + L_{2it} + L_{3it} = L_{it}, \quad E_{1it} = R_{it} - R_{it+1} \tag{3.7}$$

for every $i = 1, .., N$. Henceforth, we label the above model as 'General Model' (GM), while we consider two simplified versions of the economic part of the model.

Model with no exhaustible energy (NEEM): oil and gas are bundled together with coal and the resulting fossil fuel has inexhaustible reserves, so the exhaustible resource constraint (3.4) is irrelevant. There are thus two broad types of energies: fossil fuel, $E^f_{it} \equiv E_{1it} + E_{2it}$, and green energy, E_{3it}. Hence, the energy equation (3.3) is

$$E_{it} = (\kappa(E^f_{it})^\rho + (1-\kappa)(E_{3it})^\rho)^{1/\rho}. \tag{3.8}$$

Model with only exhaustible energy (EEM): exhaustible fossil fuel is the only type of energy, so $\kappa_1 = 1$ and $\kappa_2 = \kappa_3 = 0$. No labour is required for extraction of fossil fuel. Labour is only used for production of final output. To obtain an AK model of economic growth, we use units of effective labour $\bar{K}_{it} L_{it}$ with $\bar{K}_{it}$ the economy-wide capital stock (efficiency of labour). The production function for final output is thus

$$Y_{it} = \exp(-\gamma T_t) A_{0i} K^\alpha_{it} E^\nu_{it} (\bar{K}_{it} L_{it})^{1-\alpha-\nu}. \tag{3.9}$$

Labour is supplied inelastically and fixed at unity. In equilibrium, economy-wide capital equals firm-level capital, i.e. $\bar{K}_{it} = K_{it}$, and hence

$$Y_{it} = \exp(-\gamma T_t) A_{0i} K^{1-\nu}_{it} E^\nu_{it}. \tag{3.10}$$

3.2 Dynamics of temperature and cumulative emissions

Global temperature increases linearly in cumulative emissions of CO_2, S_t (e.g., Allen et al., 2009; Matthews et al., 2009; van der Ploeg and Rezai, 2021; Dietz et al., 2021)

$$T_t = T_0 + \xi S_t, \tag{3.11}$$

where T_0 is initial temperature and ξ the transient climate response to cumulative emissions. The stock of cumulative emissions is the sum of past and current emissions:

$$S_t = S_{t-1} + E^f_t, \tag{3.12}$$

where $E^f_t \equiv \sum_{i=1}^N (E_{1it} + E_{2it})$ is the flow of emissions produced by all countries at time t (oil/gas and coal). With cumulative emissions at time zero equal to zero, we obtain

$$S_t = \sum_{i=1}^N \sum_{s=0}^t (E_{1it-s} + E_{2it-s}). \tag{3.13}$$

4 Optimal Carbon Pricing for a Given Coalition

When choosing their optimal climate policy, the m members of coalition M internalise the emissions externality they impose on other coalition members, while acting non-cooperatively against other coalitions. The members of each coalition maximise their joint discounted infinite-horizon payoff or the total worth of coalition M given by

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

with respect to E_{1it} and E_{2it} subject to (3.6) and (3.7), for each $i \in M$, and the temperature dynamics (3.11) and (3.13).

For the members of coalition M, the optimality condition for exhaustible energy is[13]

$$\frac{\nu \kappa_1 Y_{it}}{E_{1it}^{1-\rho} E_{it}^{\rho}} = \mu_{it} C_{it} + \hat{\Lambda}_{it}(m) Y_{it}, \quad i \in M, \tag{4.2}$$

where m denotes the number of countries in coalition M (Appendix A.1). Equation (4.2) states that the marginal productivity of exhaustible fossil fuel is set to its marginal cost which equals the scarcity rent, $\mu_{it} C_{it}$ (the shadow price in utility units, μ_{it}, divided by the marginal utility of consumption, $1/C_{it}$, to convert to final goods units) *plus* the SCC, $\hat{\Lambda}(m) Y_{it}$, with the SCC per unit of aggregate output given by

$$\hat{\Lambda}_{it}(m) = \hat{\Lambda}(m) \equiv \frac{\gamma \xi m}{1-\beta}, \quad i \in M. \tag{4.3}$$

The SCC per unit of output of each signatory for any period t is the present value of the sum of discounted climate damages for all m members of coalition M from emitting one unit of carbon today.[14] It follows that the SCC is proportional to aggregate economic activity in the coalition, i.e., mY_{it}. Hence, the larger the coalition, the larger aggregate economic activity and internalised global warming damages, and thus the larger the SCC. In equilibrium all members of coalition M have the same per unit SCC. The SCC also increases in the damage coefficient, γ, and the transient climate response to cumulative emissions, ξ. Furthermore, the SCC is higher and climate policy more vigorous with more patient policy makers (higher β).

The case of *full international cooperation* corresponds to a grand coalition with $m = N$ and $k = 1$. This gives the first-best outcome where the SCC is proportional to world economic activity and all global warming externalities are fully internalised. The fully *non-cooperative outcome* corresponds to $k = N$ singletons. The SCC is then proportional

[13] C_{it}, Y_{it} and other variables with subscript i refer to country i and not to the whole coalition to which i belongs.

[14] The term *social* here is from the point of view of coalition M. The SCC can be implemented in a decentralised economy using for example a Pigouvian carbon tax or permit market where the revenues are rebated in lump-sum fashion.

to the output of one country, since only global warming externalities affecting the own country are internalised. Carbon prices are low and climate policy is lacklustre, and as a result temperatures are higher. For any $1 < m < N$, we have $\hat{\Lambda}(1) < \hat{\Lambda}(m) < \hat{\Lambda}(N)$.

The optimality condition for the exhaustible fossil fuel (4.2) can be rearranged as

$$E_{1it} = E_{it}^{-\frac{\rho}{1-\rho}} \epsilon_{1it}(m), \tag{4.4}$$

where

$$\epsilon_{1it}(m_i) \equiv \left(\frac{\nu \kappa_1}{\mu_{it}(1-\alpha\beta) + \hat{\Lambda}(m)} \right)^{\frac{1}{1-\rho}} \tag{4.5}$$

and the scarcity rent per unit of consumption is

$$\mu_{it} = \beta^{-t} \mu_{i0} \tag{4.6}$$

for $i \in M$. Here μ_{i0} is the initial shadow price of exhaustible resources that exactly satisfies equations (3.4), (4.4), (4.5) and (4.6) for a given stock of initial reserves, R_{i0}.

Similarly, energy use for non-exhaustible fossil fuel, E_{2it}, and green energy, E_{3it}, are

$$E_{2it} = E_{it}^{-\frac{\rho}{1-\rho}} \epsilon_{2it}(m) \quad \text{and} \quad E_{3it} = E_{it}^{-\frac{\rho}{1-\rho}} \epsilon_{3it}, \quad i \in M, \tag{4.7}$$

where

$$\epsilon_{2it}(m_i) \equiv \left(\frac{\nu \kappa_2 A_{2i} L_{0it}}{1 - \alpha - \nu + \hat{\Lambda}(m) A_{2i} L_{0it}} \right)^{\frac{1}{1-\rho}} \quad \text{and} \quad \epsilon_{3it} \equiv \left(\frac{\nu \kappa_3 A_{3i} L_{0it}}{1-\alpha-\nu} \right)^{\frac{1}{1-\rho}}, \quad i \in M. \tag{4.8}$$

Here ϵ_{3it} does not depend on m, because green energy does not generate emissions. For a given value of E_{1it}, equations (4.7) together with the labour market-clearing condition and the energy production functions (3.5) can be solved for E_{2it}, E_{3it}, and thus

$$E_{it} = \left(\kappa_1 \epsilon_{1it}^{\rho}(m) + \kappa_2 \epsilon_{2it}^{\rho}(m) + \kappa_3 \epsilon_{3it}^{\rho} \right)^{\frac{1-\rho}{\rho}}, \quad i \in M. \tag{4.9}$$

The optimality condition for optimal coal use sets the marginal product of non-exhaustible fossil fuel (coal) to its SCC (as there is no scarcity rent for this type of fossil fuel). An important consequence of our functional assumptions is that the per-unit SCC of E_{1it} and E_{2it} is independent of all stocks and of future values of output, consumption and cumulative emissions. The emission strategies are dominant, so that emissions of complementary coalitions do not affect emission strategies of any coalition.[15] Although

[15]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 ($\hat{\Lambda}(m) Y_{it}$) are proportional to output, and logarithmic utility implies that the marginal utility of consumption is inversely proportional to output.

emission strategies are dominant, their payoffs depend on global cumulative emissions and the entire coalition structure, and the associated energy use of all countries.

To pin down the path of μ_{it} and E_{1it}, we first guess E_{1i0} for any given level of R_{i0} and solve for all other values recursively. To see this, notice first that equation (4.2) together with the optimality condition $\mu_{it} = \beta\mu_{it+1}$ gives

$$\frac{\nu\kappa_1}{E_{1it}^{1-\rho}E_{it}^{\rho}} - \hat{\Lambda}(m) = \beta\left(\frac{\nu\kappa_1}{E_{1it+1}^{1-\rho}E_{it+1}^{\rho}} - \hat{\Lambda}(m)\right). \tag{4.10}$$

Since energy levels in period zero can be computed as function of an initial guess for E_{1i0}, equation (4.10) is in terms of E_{1i1} and E_{i1}. It can be used to solve for all the energy levels in period 1, again using equations (4.7), for period 1, and so on. This delivers the entire sequence of energy inputs and, hence, cumulative emissions, temperature, output, consumption, and investment. One then needs to verify whether $R_{i0} = \sum_t R_{it}$. If fossil fuel use exceeds (falls short of) initial reserves, the initial guess for E_{1i0} is adjusted downwards (upwards). Hence, exhaustible fossil fuel is fully depleted asymptotically.

Indeed, as the per-unit shadow price of exhaustible fossil fuel increases over time at the rate $1/\beta$, demand for exhaustible fossil fuel decreases over time. The scarcity rent (in units of final goods) grows at a rate equal to the marginal product of capital, r_{it}, which is $\alpha Y_{it}/K_{it}$ in the GM and the NEEM, and $(1-\nu)Y_{it}/K_{it}$ in the EEM. It equals the rate of interest plus the depreciation rate (i.e., 1) in the market economy. This rule for the scarcity rent is the *Hotelling rule*. The initial shadow price in utility units, μ_{i0}, is such that cumulative fossil fuel use exhausts initial reserves for each country either in finite time or asymptotically, $\lim_{t\to\infty}\sum_{s=0}^{t} E_{1it-s} = R_{i0}$. Hence, at time t, after joining a (non-singleton) coalition, and by committing to a new and higher per-unit SCC, the shadow price in utility units in each country in coalition M is adjusted.

Corollary 1 . *The larger the size of the coalition, m, the smaller the shadow price of fossil fuel in utility units, μ_{i0}, for $i \in M$, after joining coalition M.*

The shadow price of fossil fuel in utility units μ_{it} in countries which are signatories to larger coalitions is lower after the membership stage, since internalising the global warming externality implies that such countries deplete their given reserves at a later time. By joining a coalition, the emission level of signatories is affected by two counteracting factors: the decrease in the shadow price and the increase in coalition size and hence the per-unit SCC. However, the former is a result of the latter and it is plausible to assume that the former does not dominate the effect of the higher SCC on emissions. We allow μ_{it} to be heterogeneous across countries and assume that the participating countries in climate negotiations have a finite scarcity rent.

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

Proposition 1 . *In the GM, aggregate output, consumption, and the capital stock, and the economic growth rate of each member i of the coalition M (of size m) in coalition structure $\mathbb{M}$ at time t are*

$$
\begin{aligned}
&Y_{it}(\mathbb{M}) = \exp(-\gamma T_t(\mathbb{M}))A_{0i}K_{it}(\mathbb{M})^{\alpha}\left(\kappa_1\epsilon_{1it}^{\rho}(m)+\kappa_2\epsilon_{2it}^{\rho}(m)+\kappa_3\epsilon_{3it}^{\rho}\right)^{\frac{(1-\rho)\nu}{\rho}} L_{0it}^{1-\alpha-\nu},\\
&C_{it}(\mathbb{M}) = (1-s)Y_{it}(\mathbb{M}), \quad K_{it+1}(\mathbb{M}) = sY_{it}(\mathbb{M}), \quad \text{and} \qquad (4.11)\\
&\frac{Y_{it}(\mathbb{M})}{Y_{it-1}(\mathbb{M})} = \exp(-\gamma\xi E_t^f(\mathbb{M}))\left(\frac{s\,r_{it-1}(\mathbb{M})}{\alpha}\right)^{\alpha}\left(\frac{\kappa_1\epsilon_{1it}^{\rho}(m)+\kappa_2\epsilon_{2it}^{\rho}(m)+\kappa_3\epsilon_{3it}^{\rho}}{\kappa_1\epsilon_{1it-1}^{\rho}(m)+\kappa_2\epsilon_{2it-1}^{\rho}(m)+\kappa_3\epsilon_{3it-1}^{\rho}}\right)^{\frac{(1-\rho)\nu}{\rho}},
\end{aligned}
$$

respectively, where $s = \alpha\beta$ is the countries' common and constant saving rate.

Aggregate output, consumption and investment increase in TFP, but decrease in warming and the fossil fuel price.[16] So via the per-unit SCC, aggregate output falls in the number of signatories. Consumption and investment are constant shares of output. The savings rate is high if society is more patient (higher β) and the share of capital (α) is high. Consumption and capital choices are non-stationary, because they depend through aggregate output on the time-varying paths of μ_{it} and S_t.

5 Formation of Self-Fulling Climate Coalitions

Before any production or consumption takes place, countries have the choice to participate in climate negotiations and form climate coalitions to collectively decide about their emissions and cut damages from global warming. The game begins when all countries are singletons. Each period has two stages: the *membership* stage and the *action* stage. After the membership stage all countries enter the action stage, where the signatories set their climate policy as agreed at the membership stage and their individually-determined policies. Finally, at the end of each period, countries observe emissions of all countries and payoffs for each country are realised.

We assume that formation of coalitions is costless and open, so that no country can be excluded from joining and no country can be forced to join a coalition. Joining a climate coalition requires signing a binding agreement with the other signatories of the coalition. Upon signing an agreement, the signatories tie their hands and act cooperatively as a block in deciding on their common climate policy summarised by the per unit SCC implemented by this coalition, for all $t \in \{0, 1, ..., \infty\}$ and all $i \in M$. Implementation of a climate treaty is costless because upon formation of an equilibrium coalition compliance is not a problem. After the membership stage, each country $i \in M$, in addition to the collectively-chosen fossil fuel uses, determines its individual green energy, consumption, next period capital stock and resource extrac-

[16] Since temperature depends on global emissions, it depends on the entire coalition structure, $\mathbb{M}$ and thus so do aggregate output, consumption, investment and the capital stock for all $i \in M$.

tion for the infinite horizon: $\{E_{1it+\tau}(M,\mathbb{M}), E_{2it+\tau}(M,\mathbb{M}), E_{3it+\tau}(M,\mathbb{M}), C_{it+\tau}(M,\mathbb{M}), K_{it+\tau+1}(M,\mathbb{M}), R_{it+\tau+1}(M,\mathbb{M})\}_{\tau=0}^{\infty}$.

Assumption 1 . *Membership decisions are irreversible.*

This ensures that countries cannot renegotiate. If at the start of a period a full coalition structure, $\mathbb{M}$, has been negotiated and agreed from previous periods, the membership stage is skipped in that period and all future periods. We relax this assumption in section 5.4 where we allow for renegotiation at the start of each period.

5.1 A bargaining model of climate coalition formation

The climate negotiation stage is modeled as a non-cooperative bargaining process with proposals and responses in sub-periods of the membership stage.[17] In each sub-period, one country is chosen as the initial *proposer*. We assume that the duration of a sub-period relative to the length of any period is fixed, and as in Rubinstein (1982) there is a cost of delay for rejections in a sub-period which is captured by the discount factor σ. Our farsighted methodology holds for $\sigma \to 1$ from below, so that in the limit bargaining in period t is frictionless. We assume $\sigma < 1$ to avoid multiplicity of equilibria.

Definition 1 . *The rules of bargaining are set by a protocol.*

The protocol is exogenous and is set at the very beginning of the game. According to the protocol, a proposer makes a *proposal* to form a coalition to a group of *respondent* countries which are in the negotiation room (those who have not joined any other binding coalition yet). Such respondents are the *active* players in the negotiation room.

The proposal consists of the identity of the members (thus of the size m too) and the optimal per-unit SCC of the coalition signatories and the corresponding emission plans for the members of the treaty. We do not allow for transfers across coalition members so the only purpose of joining a coalition is to internalise the externality of emissions. The proposal is conditioned on the complementary coalition structure and thus the proposed emission plan is conditioned on the emission plans of other coalitions, at every contingency. However, in our setup coalitions have dominant strategies as both the marginal cost and benefit of countries are proportional to cumulative emissions, when determining their optimal emissions, hence there is no need to condition on the emission plans of other coalitions. If some countries have left the negotiation room on binding agreements, the proposal must be conditioned on those coalition structures which are consistent with this. We define the information set on which the proposal is conditioned in equilibrium in section 5.2. After a proposal is made, it is the turn of the respondents to either accept or reject the proposal. If $m = 1$, the proposer exits the negotiations as a singleton coalition, and if $m > 1$, the proposer must at least include herself.[18]

[17] This captures that usually, climate negotiations take a couple of months.

[18] Committing to staying alone is a reasonable assumption in a public good game. We show that a

The order in which countries take action in the negotiation stage is deterministic and determined by the protocol. This includes the order of the initial proposers and all chosen respondents. We focus on protocols which require the unanimity of members for a coalition to form: if the proposal is rejected by even one country, no coalition forms in that sub-period. The next proposer may or may not include the initial proposer in its proposal. We focus on *rejector-friendly* protocols, where rejectors are not excluded from counterproposing. We use a special class of such protocols where the first rejector is the next proposer of a coalition M. Excluding countries in a public-good game is never beneficial and the assumption of rejector-friendly protocols is in line with what is observed in climate negotiations.

If a proposal is unanimously accepted, a binding coalition of size m forms and irreversibly leaves the negotiation room. Negotiation continues among the remaining active countries in the negotiation room. Once all treaties are concluded, the coalition structure $\mathbb{M}$ which corresponds to a numerical coalition structure, $\mathcal{M}$, is established.[19]

5.2 Equilibrium climate membership decisions

The incentives of countries in the climate negotiation stage are determined by the optimum value function of a country in a given coalition M. To ensure sequential rationality, we solve the model backwards in time. Thus, to characterise the equilibria of the entire game, we build on the results from section 4 which characterise the optimal decisions of the action stage and we move to the membership stage.

We use the equilibrium notion of Ray and Vohra (1999) which extends Rubinstein (1982) and Chatterjee et al. (1993), and is used when countries write binding agreements to act cooperatively within the coalition while there is non-cooperative play across coalitions. Farsightedness is central to this equilibrium notion. The proposer and the respondents of an ongoing proposal rationally predict the complementary coalition structure that may form among the active players in the negotiation room, as well as their contributions, and in equilibrium such predictions are correct. If a coalition forms, it is the best option for all members of such a coalition. Then the complementary coalition(s) (if any) forms (form) in a similar way. Thus, the equilibrium coalition structure forms sequentially and endogenously. More precisely, before signing any binding agreement, a potential *group* of deviating countries has to consider further possible deviations by the deviating group (the deviating group can split further before signing their binding agreement) as well as deviations from the active players in the negotiation room, which may disband too. Hence, each group of players which contemplates forming a coalition makes a rational prediction about the entire coalition structure $\mathbb{M}$.

singleton coalition, if formed in equilibrium, has the highest payoff. If it joins any other coalition, it must set the same per-unit SCC as the other signatories of that coalition, and thus has a larger per-unit SCC.

[19] If bargaining continues forever, we assume that all countries receive a (normalised) payoff of zero. This assumption is also made in the IEA model of Ray and Vohra (2001).

Since potential deviations from a treaty must be constrained to be farsighted, the set of equilibrium coalitions must be defined recursively. In each step of the recursion, we need to identify for which number of countries, a grand coalition forms in equilibrium. Let us denote the equilibrium coalition structure by $\mathbb{M}^*$ and the equilibrium numerical coalition structure by $\mathcal{M}^*$. If countries are identical, the identity of any particular country is indeterminate in equilibrium and characterising the equilibrium membership strategies involves only the sizes of the coalitions and the number of coalitions (the equilibrium numerical coalition structure, $\mathcal{M}$). 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.

Dynamic games often have a large number of subgame-perfect equilibria. To refine these equilibria, we focus on pure strategy Markov Perfect equilibria (MPEs). A strategy profile is a MPE, if it is a subgame-perfect equilibrium in which all countries use Markovian strategies.[20] 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 with no state or stocks, investigating MPEs in dynamic games is common.[21] The use of MPEs can be restrictive as it prevents players from punishing other players for past decisions, but in a framework with binding agreements as ours, there is no compliance issue in the action stage, so there is no incentive to punish a country for its past actions as all coalition members will comply with the agreed policy. Maskin and Tirole (2001) argue that MPEs are simple, robust and consistent with rationality and state that 'bygones are bygones more completely than does the concept of subgame-perfect equilibrium'. As Harstad (2016) argues, since the state includes only payoff-relevant variables and does not depend on the history in arbitrary ways, the MPEs are robust to miscoordination among players.

The current state in our framework includes the formed coalitions (if any), the number and identity of countries that are negotiating (if any), the proposal (ongoing or signed) and the identity of the proposing country, the capital stocks K_{it}, the stock of cumulative emissions S_t, and the shadow price of exhaustible fossil fuel in utility units, μ_{it}.

Before presenting our results under farsightedness, we briefly discuss the outcomes of our model under the more commonly used cartel (or internal and external) stability conditions for symmetric countries. Under cartel stability, the largest coalition size is $m^* = 2$ for any total number of countries N in our GM (Appendix A.3), and the remainder are singletons. This is the small-coalition paradox. The cartel stability concept considers only unilateral deviations and upon a deviation, either totally disregards the possibility of updating the *membership* strategies by the active countries, or disregards

[20] Focusing on pure strategies is a common 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.

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

an optimal update of strategies (as in Appendix A.3). Under such myopic assumptions, as put by Chwe (1994), the countries act naively and, having a higher incentive to free ride, can only form a small coalition with a maximum size of 3 countries. This is true for any number of countries N.

Relative to the internal and external stability concept, where only *unilateral* deviations are considered, the *coalition-proofness* stability concept generalises the Nash equilibrium in that respect and examines *multilateral* deviations too. However, upon a deviation by a potential coalition, the membership decisions of the complementary set are assumed to be fixed. Instead, as explained above, we will use an equilibrium notion which is based on farsightedness to relax this restrictive assumption.

5.2.1 Numerical coalition structure for symmetric countries in the EEM

As the case of asymmetric countries follows directly from the case of symmetric countries, we start with the symmetric setup and discuss asymmetries in Section 5.3. Moreover, we first work with the EEM presented in Section 3 as in this case we can derive analytical results and gain intuition for the dynamics of our more complicated models (GM and NEEM) which are solved numerically in Section 6.

Let $V_i(S_t, K_{it}, \mu_{it}, m, \mathcal{M})$ denote the optimum value function of a signatory of a coalition M with size m in a numerical coalition structure $\mathcal{M}$. Henceforth, we suppress all arguments not directly relevant for the analysis of characterising the equilibrium numerical coalition structure, so we use the shorthands $V_i(m, \{\mathcal{M}\})$ and $V_i(\{N\})$ for the optimum value functions of a country in a coalition of size m in $\mathcal{M}$ and of a country in a grand coalition $\{N\}$ respectively.

As mentioned earlier, we identify the equilibrium numerical coalition structure recursively. For completeness, if $N = 1$, a singleton coalition forms. Next, we need to find $\mathcal{M}^*$ if $N = 2$. Given that, we then find $\mathcal{M}^*$ if $N = 3$, and continue the recursion until we have reached the total number of countries N that are in the global economy. If $N = 2$, the problem reduces to whether $\{1, 1\}$ or $\{2\}$ forms. This depends on the sign of

$$\begin{aligned} &V_i(1, \{1,1\}) - V_i(\{2\}) = \\ &\frac{1}{1-\beta(1-\nu)} \left\{ \nu\{ln\left(\frac{E_{it}(1)}{E_{it}(2)}\right) + \beta ln\left(\frac{E_{it+1}(1)}{E_{it+1}(2)}\right) + ...\} \right. \\ &\left. -\frac{2\gamma\xi}{1-\beta}\{[E_{it}(1) - E_{it}(2)] + \beta[E_{it+1}(1) - E_{it+1}(2)] + ...\} \right\}. \end{aligned} \tag{5.1}$$

Here $E_{it} = E_{1it}$. So, $V_i(1, \{1,1\}) - V_i(\{2\})$ is independent of the capital stocks and cumulative emissions, and only depends on the emission paths under the two scenarios. The second line in equation (5.1) is the discounted infinite sum of the ratio of the benefit of emitting in a singleton relative to the benefit of emitting in a grand coalition, and is positive. The third line captures the discounted infinite sum of the losses resulting from the damages of emitting in a singleton relative to the damages of emitting in a grand coalition, and is negative.

In general, determining the sign of equation (5.1) requires a numerical analysis with specific parameter values. However, for the case of $\sigma \to 1$ and $\beta \to 1$, we can characterise the equilibrium analytically.[22] For the EEM, it is easy to show that with two countries $\lim_{\beta \to 1}(V_i(1, \{1, 1\}) - V_i(\{2\})) < 0$ so that the grand coalition forms in equilibrium, and the equilibrium numerical coalition structure is $\mathcal{M}^* = \{2\}$.

Continuing to the case $N = 3$, there are three possible numerical coalition structures with symmetric countries: $\{3\}$, $\{2, 1\}$, or $\{1, 1, 1\}$. From the case $N = 2$, we 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 $\{1, 1, 1\}$.

In a public-good game, considering deviations implies splitting N (or any active number of players in the negotiation room) into coalitions where their sizes result from breaking up N into the largest possible integers at which a grand coalition was stable in previous stages of the recursion. Ray and Vohra (2001) show that in a public-good game with symmetric countries, it is sufficient to check the deviation by the smallest element of $\mathcal{M}$ when upon this deviation N countries have to split into the largest possible coalitions, i.e., those that result from the *decomposition* of N. Since we are in a symmetric setup checking such a deviation from the grand coalition is a sufficient condition for every country to prefer the grand coalition to any other coalition structure.

We proceed with defining two fundamental concepts for our analysis.

Definition 2 . *$\mathcal{T}^*$ is defined as the set of the total number of countries, N, for which a grand coalition forms in equilibrium.*

Definition 3 . *For any integer N, the decomposition $D(N)$ is defined as $\{m_1, m_2, ..., m_k\}$, such that m_k is the largest integer in $\mathcal{T}^*$ that is strictly smaller than N. Then any other m_i in $D(N)$, is the largest integer in $\mathcal{T}^*$ that is no greater than $N - \sum_{j=i+1}^{k} m_j$.*

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

As mentioned earlier, at each stage of the recursion, the optimum value of only the smallest coalition in a coalition structure must be compared with that of the grand coalition which significantly cuts 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. Hence, as negotiations start, if the grand coalition is not stable, a proposer makes an acceptable offer first to the smallest equilibrium coalition, that in a public-good game has the highest payoff, and without any delay the offer is accepted and the coalition forms. And a similar process continues among the remaining countries.

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

For example, if $N = 3$, it is sufficient to check the sign of $V_i(1, \{2, 1\}) - V_i(\{3\})$. 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 singleton and a coalition of size two, so that the equilibrium numerical coalition structure is $\mathcal{M}^* = \{2, 1\}$. Similarly, for $N = 4$, the only conceivable decomposition consistent with farsightedness is $\{2, 2\}$ which has to be considered against $\{4\}$, and it can be shown that here, a grand coalition forms again, i.e., $\mathcal{M}^* = \{4\}$. Hence, $\mathcal{T}^*$ expands to $\{1, 2, 4\}$.[23]

Comparing the optimum value function of the smallest coalition in the decomposition with the value function of the grand coalition at each stage of the recursive procedure (i.e., for each N) can be demanding. We show that in our EEM, the recursion process can be simplified as there is a general rule for the inequality (5.1).

Proposition 2 . *Let $D(N) = \{m_1, m_2, ..., m_k\}$ be the decomposition of N, such that $m_1 < m_2 < ... < m_k$. For $\beta \to 1$ and symmetric countries, a grand coalition forms in equilibrium in the EEM if*

$$\ln(\frac{N}{m_1}) < k - 1. \tag{5.2}$$

This proposition is proved in Appendix A.4 and offers a simple sufficient condition for the full characterisation of the set $\mathcal{T}^*$ for our EEM and $\beta \to 1$.[24] The left-hand side of (5.2) is the gain from emitting in the smallest coalition versus emitting in the grand coalition. The right-hand side of (5.2) is the externality damage resulting from forming $D(N)$ versus the grand coalition. Since in the limit as $\beta \to 1$, emissions are almost stationary, it is sufficient to compare the gains and losses of one period only: if damages are higher than the gains from emitting, a grand coalition forms in equilibrium.[25]

Corollary 2 . *For the EEM and $\beta \to 1$, a grand coalition occurs in equilibrium if the total number of countries N is an element of*

$$\mathcal{T}^* = \{1, 2, 4, 7, 13, 24, 44, 79, 146, 268, 482, 873, 1580, ...\}. \tag{5.3}$$

If $N \in \mathcal{T}^*$, then $\mathcal{M}^* = \{N\}$; if $N \notin \mathcal{T}^*$, then $\mathcal{M}^* = D(N)$, given $\mathcal{T}^*$. Thus to characterise $\mathcal{M}^*$ in our EEM with farsighted and patient countries there are two simple steps. First, generate the set $\mathcal{T}^*$ using Proposition 2, and if the total number of countries belongs to the set $\mathcal{T}^*$ then a grand coalition forms in equilibrium. If not, then the equilibrium coalition structure is the decomposition of N using $\mathcal{T}^*$. For example, going back to the case of $N = 3$, since $3 \notin \mathcal{T}^*$, the grand coalition does not form and

[23] If there are two numerical coalition structures $\{m\}$ and $\{m', m'\}$ with the same payoffs, we break the tie in favour of the larger coalition in line with the convention in the literature.

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

[25] From equation (4.6) the shadow price of fossil fuel rises at an infinitesimally small rate and thus from equations (4.4) and (4.5) emissions decline at an infinitesimally small rate in the limit.

its decomposition, using the elements of $\mathcal{T}^*$ that are smaller than 3, determines the equilibrium coalition structure, i.e., $\mathcal{M}^* = \{2, 1\}$. The decomposition structure implies that the total efficiency loss should be small, since even in cases where the grand coalition does not emerge, the biggest coalition is still large, and the number of coalitions is small.

In our EEM, the equilibrium number of signatories to a climate treaty (corresponding to the grand coalition or a smaller coalition), is a Tribonacci number (i.e., the sum of the two preceeding numbers in the sequence) if $N \leq 78$.[26] We note that for values of up to $N = 500$, the Tribonacci sequence is a close approximation of $\mathcal{T}^*$.[27] The methodology of Ray and Vohra (2001) can be applied to any public-good setting and can lead to any set of total number of countries for which a grand forms in equilibrium. They show that with a quadratic cost function, $\mathcal{T}^*$ follows the Fibonacci sequence (of order 2) up to the 7th element. In our case too the series at which the formation of a grand coalition is an equilibrium diverges from Tribonacci sequence (or Fibonacci sequence of order 3) after the 7th element. The reason that we obtain a Tribonacci set instead of a Fibonacci set relates to functional forms. Ray and Vohra (2001) have a quadratic-linear payoff function, while we have a log-linear optimum value function (for further details and discussion, see Appendix B).

Ray and Vohra (2001) assume that countries are symmetric and have a one-off payoff, so after bargaining countries receive their agreed payoffs and the game ends. The recursion is not demanding and leads to an analytical solution. We use and extend their methodology to answer questions in an infinite-horizon IAM with time-varying payoffs.[28] Proposition 2 gives a sufficient condition to analytically characterise $\mathcal{M}^*$. Furthermore, given Corollary 2, for the EEM there is no need to check payoffs at each stage of the recursion to find the integers at which a grand coalition forms in equilibrium if $N < 78$.

Proposition 2 results from the solution concept and the special features of our IAM. In particular the structure of our IAM yields a per-unit SCC that is independent of aggregate economic outcomes, a constant saving rate and dominant emission strategies. The assumption $\beta \to 1$ leads to unambiguous tractable outcomes and can be justified by a normative approach.[29] However, Corollary 2 is not robust to lower values of β.

Proposition 2 shows that the equilibrium number of signatories in climate coalitions

[26] They occur in the natural world if an efficient way of packing elements together is called for. For example, the number of petals of flowers, bracts of pine cones, and trees branches tend to be from a Fibonacci sequence (Campbell, 2020; Minarova, 2014; Sinha, 2017). Tribonacci numbers were first found by Charles Darwin in the *Origin of Species*.

[27] The Tribonacci sequence starting with $\{0, 0, 1\}$ is $\{1, 2, 4, 7, 13, 24, 44, 81, 149, 274, 504, 927, 1705, ...\}$. For our IAM, $T_j = T_{j-3} + T_{j-2} + T_{j-1}$ and $T_j \in \mathcal{T}^*$, for all $j = 1, 2, ..., 7$.

[28] Generally, numerical simulation is called for at each stage of the recursion. De Zeeuw (2008) is the only infinite-horizon public-good game with farsighted countries, but derives results only numerically.

[29] Many have argued that the social is smaller than the private discount rate. Arrow et al. (2003) argue that because of market imperfections, especially in long-run, using market observables such as the interest rate to identify the social discount rate can be misleading. Following Ramsey (1928) who argues that it is unethical to discount the welfare of future generations, climate economists have often used a near-zero rate of time preference (e.g. Stern, 2007; Dietz and Stern, 2015).

can be large. This contrasts with the small-coalition paradox resulting from the Nash equilibrium solution concept and its assumption of a single coalition found in the literature on IAM and IEA (Lessmann et al., 2009, 2015; Bosetti et al., 2013). Adopting the more realistic farsighted assumption without any of the known 'remedies' to increase cooperation, we have shown that the number of signatories depends on the number of countries N and the countries' payoffs and can be significantly larger. If a grand coalition does not form, the largest stable climate coalition in $\mathcal{M}^*$ can still be large (the largest integer in the set $\mathcal{T}^*$ that is smaller than N). Moreover, multiple (non-singleton) climate coalitions can form with more ambitious climate policies than the singleton coalitions that occur under cartel stability.

Our equilibrium notion does not assume that if a country breaks off the negotiations, other countries do not react as under cartel stability. This implies that in equilibrium, countries have to rationally predict the reaction of all active countries prior to their membership decisions. This, in turn, reduces their free-riding incentives and leads to formation of larger coalitions.

5.2.2 Example: the EEM with 195 symmetric countries

Consider $N = 195$. Since $195 \notin \mathcal{T}^*$ (from Corollary 2) , we verify $\mathcal{M}^* = \{146, 44, 4, 1\}$ with $146 + 44 + 4 + 1 = 195$. Hence, 4 coalitions form. The coalition of 1 signatory forms first, then the coalition with 4, then the one with 44 signatories leaves the negotiation room, and lastly the largest coalition with 146 members forms.[30] There is no small-coalition paradox and multiple coalitions emerge. The equilibrium numerical coalition structure is not too complicated, since only 4 coalitions form. The large coalition of $m^* = 146$ has more ambitious climate policies than the coalition of 44 countries. The coalition of 44 countries has more ambitious policies than the coalition of 4 countries, which is more ambitious than the singleton. Furthermore, the set $\mathcal{M}^*$ does not have many coalitions with a small number of countries. Our results imply that a group of 146 countries forms a stable coalition and sets the per-unit SCC at $\hat{\Lambda}(m^*) = \frac{146\gamma\xi}{1-\beta}$ for all future periods. The other three smaller coalitions set their climate policy in their binding agreement accordingly leading to a smaller per-unit SCC corresponding to fractions 44/146, 4/146 and 1/146 of the per-unit SCC of the coalition of 146 countries.

[30]By using MPE, we abstract from the order of coalition formation, but as mentioned earlier it is well known that in a public-good game the smallest coalitions have the highest payoff and form first.

5.3 Equilibrium coalition formation of heterogeneous countries

In our framework, countries may differ in their initial capital levels, K_{i0}, total factor productivities, A_{0i}, and shadow price, μ_{it} depending on fossil fuel stocks, R_{i0}.[31] [32] These asymmetries are important topics for climate negotiations. We first investigate the impact of heterogeneity with respect to initial capital and TFP in the GM with three sources of energy and then discuss heterogeneity with respect to initial stocks of fossil fuel, and thus the scarcity rent in the EEM.

Starting from the smallest set of countries ($N = 2$), we should find $\mathbb{M}^*$ for each group of two countries. Then, knowing which group of two countries can strike a deal, all possible $\mathbb{M}^*$ have to be found for $N = 3$ and the process continues for $N = 4$, etc. With heterogeneous countries, there will be path-dependency as the equilibrium outcome depends on which countries are chosen in earlier stages of the recursive process. Multiplicity of equilibria will occur and the analysis can be tedious.

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

Suppose j is the initial proposer and has approached country i (which can be j itself). For any N, a farsighted country i must identify the most profitable deviation from the grand coalition. It is sufficient to compare the payoff of the best profitable deviation by forming coalition $M \in \{M_1, M_2, ..., M_k\}$ versus that of staying in the grand coalition $\{I\}$. So, country i needs to determine the sign of

$$V_i^j(S_t, K_{it}, \mu_{it}, M, \mathbb{M}) - V_i^j(S_t, K_{it}, \mu_{it}, I). \tag{5.4}$$

This expression is independent of A_{0i} and the initial capital stock for any discount factor β (see Appendix A.5), so membership decisions are unaffected by heterogeneity with respect to K_{i0} and A_{0i}.

Expression (5.4) is a log-linear function of energies only. As discussed in the previous section, because of their dominant strategies, emission plans of signatories of coalition M depend only on its *own* size, m, and need not be conditioned on the entire coalition structure. Also, all members of a coalition of size m have the same per-unit SCC, $\hat{\Lambda}(m)$. Although V_i^j depends on the equilibrium coalition structure, emission plans in the proposal and thus (5.4) depends only on m and total energy used in each country

[31] In a Supplementary Appendix (available here) we allow green energy and exhaustible fossil fuel to be perfect substitutes and we study heterogeneity with respect to cost of the green energy.

[32] To the best of our knowledge, only Ray (2007) studies non-cooperative games and coalition formation with heterogeneous agents and externalities. He derives sufficient conditions for existence of equilibria without delay: (i) coalitions which form subsequently have a lower average worth; (ii) the more active countries in the negotiation room, the larger equilibrium payoff; and (iii) the equilibrium payoff of being a proposer exceeds the equilibrium payoff of being proposed to (without any lapse of time or discounting). In our climate game with free-riding incentives of the countries, all these conditions are satisfied.

(through equations (4.4) and (4.7)). With different K_{i0}'s and A_{0i}'s (and symmetry in all other parameters), all different energy sources and total energy of each country are symmetric in a given coalition. With dominant strategies and in the absence of transfers, heterogeneity thus does not affect the sign of (5.4). The identity of the initial proposer is irrelevant too. We can thus characterise the numerical equilibrium coalition structure using the method developed for symmetric countries. Hence, to determine the sign of (5.4), we can determine the sign of

$$V_i^j(S_t, K_{it}, \mu_{it}, m, \mathcal{M}) - V_i^j(S_t, K_{it}, \mu_{it}, N) \quad \text{for} \quad i \in M, \tag{5.5}$$

where $\mathcal{M}$ is a numerical coalition structure. This holds for all our three models.[33]

Heterogeneity in initial fossil fuel stocks, R_{i0}, impacts the value function via its shadow price in utility units, μ_{it}, which are (via (4.4) and (4.5)) negatively related. Heterogeneity in fossil fuel stocks μ_{it}'s gives heterogeneity in emissions, even within a given coalition.[34] This heterogeneity does not affect the countries' payoffs in an additive-separable fashion, so it is not straightforward to conclude whether it affects the equilibrium numerical coalition structure or not. However in case of the EEM ($\kappa_1 = 1$) and $\beta \to 1$, we show that the difference of payoffs in (5.4) is independent of μ_{it}. The decision-making of farsighted countries in joining climate coalitions is then also independent of heterogeneity in scarcity rents and the identity of the proposer in this model specification, hence we can characterise the equilibrium numerical coalition structure also here.[35] Proposition 3 simplifies the characterisation of the equilibrium coalition structure for any number of heterogeneous countries N. No matter how heterogeneous countries are, the equilibrium numerical coalition structure is unique (see Appendix A.5 for proof).

Proposition 3 . *$\mathcal{M}^*$, can be characterised independently of heterogeneities in the K_{i0}'s and A_{0i}'s that enter the value functions in an additive-separable form, and is thus independent of the identity of initial proposers. For the EEM, $\mathcal{M}^*$ can also be characterised independently of heterogeneities in the R_{i0}'s (and μ_{it}'s) if $\beta \to 1$.*

We thus decouple the problem of cardinality of coalitions in equilibrium from the composition of countries in each $M^* \subseteq \mathbb{M}^*$, where $\mathbb{M}^*$ is the equilibrium coalition structure (including identities of members). So, the numerical coalition structure can be characterised without considering heterogeneity. Then after finding $\mathcal{M}^*$, we focus on

[33]In fact, it holds for any other public-good payoff structure with additive-separable heterogeneity, dominant strategies, and no transfers.

[34]Emission levels can differ for two reasons. First, countries differ in initial fossil fuel stocks. Second, by joining coalitions with different sizes, their emission path affects the trajectory path of their shadow price of fossil fuel. In the membership stage of reversible coalition formation, we focus on the effect of ex-ante asymmetry on membership decisions.

[35]As showed in the previous section, for the EEM the analytical characterisation of the equilibrium numerical coalition structure relies on the assumption of $\beta \to 1$ even for symmetric countries.

which m^* countries an initial proposershould propose to, considering global emissions and efficiency improvements.

Although this decoupling result depends on the particular payoff specification of our IAM, it allows us to characterise equilibrium numerical coalition structures among heterogeneous countries. This result is stronger with heterogeneities in K_{i0}'s and A_{0i}'s, since it does not require β to be close to 1.[36] The decoupling result for heterogeneous R_{0i} and μ_{it} does depend on functional forms and requires β to be close to 1. Section 6 explores robustness of the decoupling result with regards to these heterogeneities. Since equilibrium payoffs and emissions, and thus temperature depend on the identity of the initial proposer and the composition of the countries in each coalition, equilibrium payoffs may differ across countries. No matter what the protocol ordering of initial proposers is, every proposer selects the number of members which $\mathcal{M}^*$ prescribes.

5.3.1 Global Emissions with Heterogeneous Countries

We now discuss how successful these coalitions are in curbing global emissions and how this depends on the identity of the members in each coalition.

The fully globally efficient outcome arises if the grand coalition forms and all global warming damages are fully internalised. Any other equilibrium coalition structure will lead to some inefficiency as not all damages are internalised. If the grand coalition is not stable, how much of damages are internalised depends on the numerical coalition structure (sizes of the coalitions) and the identity of the coalition members.

Consider the EEM and assume that countries differ only in initial fossil fuel reserves and associated scarcity rents. For example, consider countries $I = \{1, 2, 3, 4, 5, 6\}$, where for all $i \in I$ we have $\mu_{it} > \mu_{i+1t}$. This ensures a strict order, so country 1 has lowest initial fossil fuel reserves, highest scarcity rent, and lowest emissions while country 6 has highest reserves, lowest scarcity rent, and highest emissions. No two countries have equal scarcity rents. Proposition 2 gives the equilibrium numerical coalition structure $\mathcal{M}^* = \{2, 4\}$. Consider two alternative compositions for the coalition structure, i.e., $\{\{1, 2\}, \{3, 4, 5, 6\}\}$ and $\{\{5, 6\}, \{1, 2, 3, 4\}\}$. Global emissions in period t are then

$$E_{1t}(2) + E_{2t}(2) + E_{3t}(4) + E_{4t}(4) + E_{5t}(4) + E_{6t}(4) \tag{5.6}$$

for the first composition and

$$E_{5t}(2) + E_{6t}(2) + E_{1t}(4) + E_{2t}(4) + E_{3t}(4) + E_{4t}(4) \tag{5.7}$$

for the second composition, where emissions are

[36] This result can be used with any reduced-form payoffs where heterogeneity affects countries' payoffs in an additive-separable fashion. For any $V_i(E_i, b_i) = f(E_i) + g(b_i)$, where $f(.)$ and $g(.)$ are two independent functions, b_i is heterogeneous across players, and E_i is the choice variable among all coalition members, the decoupling result holds. In our case, $g(.)$ is logarithmic, and b_i is either K_{i0} or A_{0i}.

$$E_{it}(m) = \frac{\nu(1-\beta)}{(1-\beta)\mu_{it}[1-\beta(1-\nu)] + m\gamma\xi}. \tag{5.8}$$

For $0 < \beta < 1$, the first composition results in lower emissions than the second one if

$$E_{1t}(2) + E_{2t}(2) + E_{5t}(4) + E_{6t}(4) < E_{5t}(2) + E_{6t}(2) + E_{1t}(4) + E_{2t}(4). \tag{5.9}$$

Since global emissions are lower if countries with large reserves (and lower scarcity rents) belong to the larger coalition with 4 members and countries with small reserves belong to the smaller coalition with 2 members, this inequality holds and the first composition gives lower emissions than the second composition. Also, the value of the coalition with the first composition is higher than that of the second composition. Appendix A.6 generalises this result for any set of countries I: a country prefers the coalition structure with lowest global emissions.

Note from inequality (5.9) that if $\beta \to 1$ (for which we have established an equilibrium), $E_{it}(m)$ becomes infinitesimally small and positive asymptotically while the effect of the shadow price, μ_{it}, becomes infinitesimally small too. Hence, the composition of the coalitions has only a vanishingly small effect on global emissions as $\beta \to 1$.

Proposition 4 *. Assume that the grand coalition is not stable. For $\kappa_1 = 1$ and $\beta \to 1$, global emissions in the equilibrium numerical coalition structure are asymptotically independent of the identity of the coalitions' members.*

Hence, if $\beta \to 1$, the only thing that matters for global emissions is the size of the equilibrium coalitions via its effect on the SCC while the effect of the scarcity rent is infinitesimally small. But if $0 < \beta < 1$, initial fossil fuel reserves and scarcity rents do play a role for global emissions.

5.4 Reversible climate agreements

Now we relax Assumption 1 and assume that countries can renegotiate agreements at no cost. To get analytical solutions, we focus on the EEM ($\kappa_1 = 1$), but the analysis can be generalised for the GM and NEEM too. We allow countries to have different K_{0i}'s and/or A_{0i}'s. We use the *weak renegotiation-proof* concept of Farrell and Maskin (1989), so a subgame-perfect equilibrium is renegotiation-proof if there do not exist two histories such that all players strictly prefer the continuation equilibrium in the one to the continuation equilibrium in the other. We continue to focus on Markovian strategies. The change of the coalition structure over time under reversible coalition formation is like moving from one Markov state to another. There is a fixed protocol for all periods.

Renegotiation starts while coalition structure $\mathbb{M}^*_{t-1}$ from $t-1$ is in place at the start of period $t \geq 1$, and the countries in coalition M^*_{t-1} face a proposal to form coalition M_t of size m_t. If the proposal is accepted by all, and M_t forms, then in the action stage of period t, countries consider all future renegotiations and so all possible future coalition structures when deciding about emissions. Thus, the signatories jointly maximise

$$\sum_{i \in M_t} \sum_{\tau=0}^{\infty} \beta^{\tau} \, \mathbb{E}\{\ln(C_{it+\tau}) \mid \mathbb{M}_t\} \tag{5.10}$$

subject to the usual constraints. We abstract from presenting other variables in the Markov state which are not directly relevant here. Appendix A.7 shows that this gives

$$E_{it}(m_t) = \frac{\nu(1-\beta)}{\mathbb{E}(\mu_{it} \mid M_t)(1-\beta(1-\nu))(1-\beta) + \gamma\xi[m_t(1-\beta) + \beta\,\mathbb{E}(m_{t+\tau} \mid M_t)]}, \tag{5.11}$$

where $\tau \geq 1$. With reversible agreements, emissions (and μ_{it}) depend on the expected value of the size of future coalitions. Now we go backward to the membership stage of period t. In a reversible setting, binding agreements are justified if there is an *approval committee* (including all parties of an existing binding coalition) that can approve the move to another Markov state (Hyndman and Ray, 2007). Some members of this agreement are affected by the new state: either their membership is affected, and/or their payoffs are directly affected.[37] The approval committee ensures that the rights of those in binding agreements are protected. Hence, in each sub-period of period t, 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.

Because the protocol is fixed, at the start of the first sub-period of the membership stage in t, the same initial proposer as in the first sub-period of $t-1$ is selected. Recall that in a public-good game, the smallest coalition forms first. The initial proposer is thus a member of either the smallest coalition in $t-1$ or the grand coalition (if formed in $t-1$). If $D(N) = \{m_1, m_2, ..., m_k\}$, such that m_1 is the size of the smallest coalition in the decomposition of N, then in the first sub-period, the initial proposer j and the approval committee of M^*_{1t-1} consider a proposal about the formation of M^j_t, and each must determine the sign of

$$\mathbb{E}\, V_i(M^j_t, \mathbb{M}_t, \mathbb{M}^*_{t-1}) - \mathbb{E}\, V_i(I_t, \mathbb{M}_t, \mathbb{M}^*_{t-1}). \tag{5.12}$$

If $m^j_t > m_{1t-1}$, the above is upon approval by the approval committees of other relevant coalitions in $\mathbb{M}^*_{t-1}$. Note that the left-hand-side of the difference in (5.12) is the expected payoff of the best possible deviation for country i, and the right-hand-side is the expected payoff of the grand coalition in period t. The inequality (5.12) is independent of the various types of heterogeneity that we consider, so characterisation of the equilibrium

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

is reduced to that of $\mathcal{M}^*$. Appendix A.7 shows that this inequality only depends on emissions, which depend on future sizes of coalitions.

A key observation is thus that any renegotiation can affect only future emissions. Hence, if the equilibrium coalition structure in $t-1$ is not a grand coalition, the approval committee of any non-ultimate coalition, will reject inclusion of any new members in the coalition in any period $t \geq 1$, because inclusion implies 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. If a grand coalition was formed in $t-1$, any approval committee would reject any exit by free riders.

Therefore, if countries sign a binding agreement to form M_{t-1}, there is no expectation that the size of coalition will change ex-post. Thus $\mathbb{E}(m_{t+\tau} \mid M_t, \mathbb{M}_t, \mathbb{M}^*_{t-1},) = m_t$ for $\tau \geq 1$. Countries expect this to be the case for all other binding coalitions. Therefore, we can show that the inequality (5.12) reduces to the inequality in Proposition 2.

Proposition 5 *. For the EEM ($\kappa_1 = 1$) and reversible coalition formation, $\mathcal{M}^*_t$ is independent of heterogeneity for any t. If $\beta \to 1$, $\mathcal{M}^*_t$ in Proposition 2 for irreversible agreements, is renegotiation-proof with same equilibrium coalition structure $\mathbb{M}^*_t$ for all t.*

Proposition 5 (proof in Appendix A.7) extends Proposition 2 and states that even if countries have the option to renegotiate every period, the equilibrium numerical coalition structure is the same as with irreversible agreements. In addition, the approval committee and the fixed protocol ensure that not only the same numerical coalition structure $\mathcal{M}^*$ forms, but also the same coalition structure $\mathbb{M}^*$ forms upon every renegotiation. So, if within a coalition there is no ex-post Pareto improvement,[38] there will be no renegotiation. Whenever the result of renegotiation is fully predictable, they can write renegotiation-proof agreements.[39]

6 Quantitative analysis

So far, most of our focus has been on the EEM and very patient countries ($\beta \to 1$) to obtain analytical results for the equilibrium coalition structure. Here we consider lower values of β and use the more realistic NEEM and GM (see Section 3). We use a numerical approach to offer robustness checks for the discount rate and the type of model, and to allow heterogeneity, also from a policy perspective. We first describe our calibration and then discuss our numerical results.

[38]Here, a Pareto improvement is *only* from the perspective of members of the binding coalition who are considering the renegotiation proposal.

[39]In the EEM, N includes those countries which contribute to the externality and want to internalise it. So, the membership strategies are renegotiation-proof as long as countries have a finite scarcity rent.

Name	Value	Unit	Description
β	0.999ˆ10 or 0.985ˆ10	-	Decadal Discount Factor
α	0.3	-	Share of Capital in Production
γ	0.007	T$/°C	Temperature Damage Coefficient
ν	0.043	-	Share of Energy in Production
ξ	2	°C/TtC	TCRE
s	0.297 or 0.2579	-	Saving Rate ($\alpha\beta$)
ρ	-0.058	-	Elasticity of Substitution of Energy sources
T_0	1.3	°C	Initial Temperature
Y_0	800	T$/decade	Initial Global Decadal GDP
K_0	128	T$	Initial Global Capital Stock
R_0	0.3	TtC	Initial Global Oil & Gas Reserves
	2	%/year	Annual Growth Rate of TFP
General Model (GM)			
κ_1	0.508	-	Share of Oil & Gas in Total Energy
κ_2	0.08916	-	Share of Coal in Total Energy
κ_3	0.41004	-	Share of Green Energy in Total Energy
A_2	7683	-	Labour Efficiency in Coal Production
A_3	1311	-	Labour Efficiency in Green Energy Production
E_0	0.119	TtC/decade	Initial Global Decadal Energy Consumption
A_0	2063624	-	TFP
No Exhaustible Fossil Fuel Model (NEEM)			
κ_2	0.42	-	Share of Oil, Gas and Coal in Total Energy
κ_3	0.58	-	Share of Green Energy in Total Energy
A_F	3.635	-	Labour Efficiency in Fossil Fuel Energy
A_G	0.1045	-	Labour Efficiency in Green Energy Production
E_0	0.0143	TtCe/decade	Initial Global Decadal Energy Consumption
A_0	2260194	-	TFP

Table 6.1: Calibration

6.1 Calibration and parameter values

We first calibrate the three specifications of our EEM, GM, and NEEM for the business as usual case (BAU), where BAU refers to when none of the countries take climate action and is our benchmark for calibration to the world economy. We take each period to be 10 years and assume full capital depreciation.[40] We consider two discount factors. We first check our analytical results for case $\beta \to 1$ with a discount factor of 0.999^{10} per decade (Stern's β). We also provide results for $\beta = 0.985^{10}$ per decade (Nordhaus's β).

We have zero population growth and set $L_{it} = 1$. Following Golosov et al. (2014), we use a capital share of 30% and an energy share of 4.3%. The elasticity of substitution

[40]To approximate our infinite-horizon IAM, we run our simulation for 1500 decades. This is long enough, since our results do not change for any horizon between 1000 to 3000 decades.

between energy types is set to 0.945 ($\rho = -0.058$), so all energy factors are essential and cannot be phased out completely. For the GM we follow Golosov et al. (2014) and set the energy share parameters to $\kappa_1 = 0.5008$, $\kappa_2 = 0.08916$ and $\kappa_3 = 0.041004$, and the labour-efficiency parameters to $A_2 = 7683$ and $A_3 = 1311$.[41] Initial global total energy use is set to $E_0 = 0.119$TtC, the global capital stock to $K_0 = 128\$$, and initial global GDP to $Y_0 = 800$T\$, all per decade, which leads to $A_0 = 2063624$. To obtain, values for individual countries, we divide these numbers by the number of countries N. We set the annual growth rate for A_t to 2%.

The NEEM merges coal and oil to one energy source and we use the BP Statistical Review to get decadal oil, gas, coal, and green energy consumption, and convert total energy in Exajoules to TtCe which gives $E_0 = 0.0143$. We then set $\kappa_2 = 0.42$, $\kappa_3 = 0.58$, $A_F = 3.635$ and $A_G = 0.1045$ to fit initial global total energy consumption and match a 4°C increase in temperature by 2100 under BAU.[42]

Finally, we set the temperature response to cumulative emissions to $\xi = 2$°C/TtC. We use the estimate of Nordhaus (2017) of a 2.1% loss of GDP at 3°C and fit our exponential damage function to obtain $\gamma = 0.007$.

6.2 Numerical Results and Equilibrium Coalition Structure

We define the Internalisation Index (II) as the *weighted* average SCC for the underlying coalition structure. This indicates how much of the climate change externality is internalised. To illustrate, II equals 1 for the grand coalition and $1/N$ for the fully non-cooperative outcome (all singletons) for the case of N countries. Hence, for any coalition structure, II $\in [1/N, 1]$ while II is equal to zero for BAU.[43]

6.2.1 The case of 195 countries ($N = 195$)

With 195 countries in the world, the first row of Table 6.2 gives $\mathcal{T}^*$ for two discount factors for the NEEM and the GM: this is the set of numbers for which a grand coalition forms in equilibrium. We see that for both models, more patience (higher β) implies that there are more cases in which a grand coalition forms. Intuitively, the more countries take account of future payoffs, the more costly it is to leave a grand coalition. Note that the $\mathcal{T}^*$ for $\beta = 0.999^{10}$ in both models follow partly the Tribonacci sequence, which offers some robustness for our analytical result for the EEM.

The second row of Table 6.2 shows the equilibrium coalition structure, also for $N =$ 195. For the GM with $\beta = 0.999^{10}$ per decade, we find that $\mathcal{M}^* = \{177, 15, 2, 1\}$, so there are 4 coalitions with the largest of them having 177 members and the smallest being a

[41] Golosov et al. (2014) use relative prices and demands for coal and green energy, and extraction cost for coal to calibrate these parameters.

[42] Details of these conversions and intermediate steps are available upon request.

[43] Note II $= m_1^2/N^2 + m_2^2/N^2 + m_3^2/N^2 + .. + m_k^2/N^2$ for k coalitions in the coalition structure, since the weight of coalition i is m_i/N and its SCC equals m_i/N times the SCC under the grand coalition.

singleton. For the NEEM with the same discount factor, $\mathcal{M}^* = \{168, 27\}$ so there is a large coalition of 168 countries and a smaller coalition with 27 countries.[44] Despite a different equilibrium coalition structure for the two models, we get a large coalition and a few smaller ones.

Row 3 of Table 6.2 shows the Internalisation Index (II) for the different specifications. For both models, more patience (high β) gives fewer and larger coalitions which results in a higher II: larger coalitions internalise more of the climate change externality as they have to take account of the payoffs in the far future. Note that for the EEM, $\mathcal{M}^* = \{146, 44, 4, 1\}$, and II=0.6119 which is not too different from the II in the GM.

It is instructive to compare our results if cartel stability is used as the equilibrium concept for coalition formation. In line with this literature, our numerical analysis shows that for the NEEM and the GM and both discount factors, a stable coalition with a maximum of 2 members emerges so that $m^* \leq 2$ and all other countries are singletons. This confirms the small-coalition paradox. With cartel stability II ≤ 0.0002 which is way lower than the II we get with farsighted countries.

	GM		**NEEM**	
	$\beta = 0.999^{10}$	$\beta = 0.985^{10}$	$\beta = 0.999^{10}$	$\beta = 0.985^{10}$
$\mathcal{T}^*$	{1,2,4,8,15,29,52,96,177}	{1,4,16,64}	{1,2,4,8,15,27,50,91,168}	{1,5,25,125}
$\mathcal{M}^*$	{177,15,2,1}	{64,64,64,1,1,1}	{168,27}	{125,25,25,5,5,5,5}
II	0.83	0.3232	0.7614	0.4464

Table 6.2: Equilibrium Coalition Structure, $\mathcal{M}^*$, and Internalisation Index (II), for 195 countries ($N = 195$)

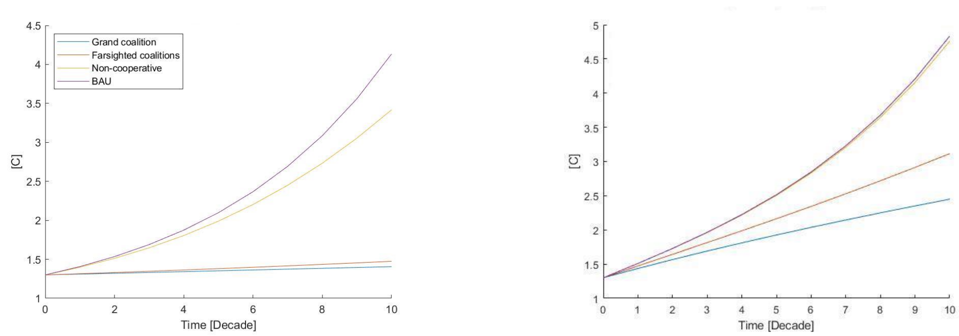
Finally, we numerically check outcomes for heterogeneous countries to confirm the decoupling result that we have obtained analytically.[45] We find that heterogeneity in the initial capital stocks and TFP's does not affect the equilibrium coalition structure for the GM and NEEM and for both discount factors. As the K_{i0}'s and A_{0i}'s do not affect total energy E_i, emissions and temperature paths also remain unchanged. With heterogeneity in the R_{0i}'s, we find numerically that for $\beta \to 1$, the equilibrium coalition structure for the GM model remains unchanged.

Time paths for temperature and coal use: The two panels of Figure 1 show the temperature path for the GM under BAU, no cooperation, the grand coalition, and the equilibrium coalition structure for Stern's and Nordhaus's β. The temperature path for the fully non-cooperative outcome (all singletons) is the highest of all paths (but lower than temperature under BAU) while the fully cooperative (grand coalition) gives

[44] With the much more restrictive model EEM, we found that for $\beta \to 1$, $\mathcal{M}^* = \{146, 44, 4, 1\}$.

[45] We let one country have twice as high K_{i0}, A_{0i} and R_{0i} as the other countries. We show that whether this country is in the grand or smallest coalition m_1 does not affect the equilibrium coalition structure. The exercise can be generalised for any number of heterogeneous countries.

the lowest temperature. In between lies the temperature trajectory for the equilibrium coalition structure. We see that for very patient policy makers (high β), the fully cooperative and the equilibrium coalition paths are very close. This is in line with our previous finding that for high β we get larger coalitions and a higher II.

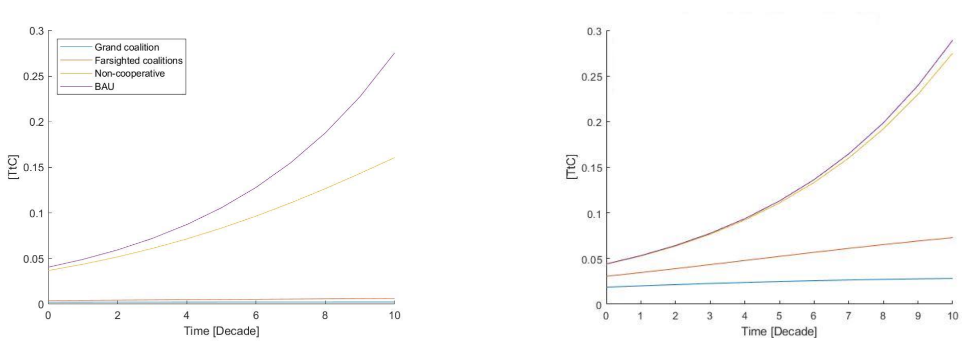


(a) $\beta = 0.999^{10}$/decade (Stern)

(b) $\beta = 0.985^{10}$/decade (Nordhaus)

Figure 1: Temperature trajectory for the GM model

The two panels of Figure 2 show the trajectory of aggregate coal use under the different scenarios. As expected, the grand coalition has the lowest coal trajectory but for a high β, the fully cooperative and the equilibrium coalitions paths are very close. Further details of time paths for the different outcomes are given in Appendix C.



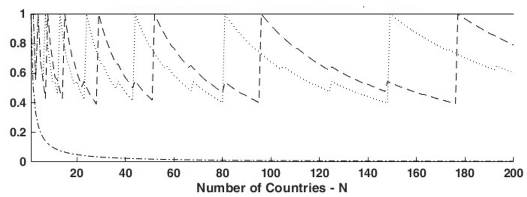
(a) $\beta = 0.999^{10}$/decade (Stern)

(b) $\beta = 0.985^{10}$/decade (Nordhaus)

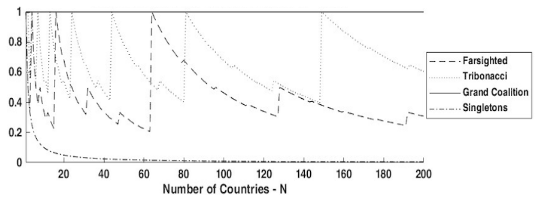
Figure 2: Aggregate Coal use for the GM model

6.2.2 A general number of countries in the world (N)

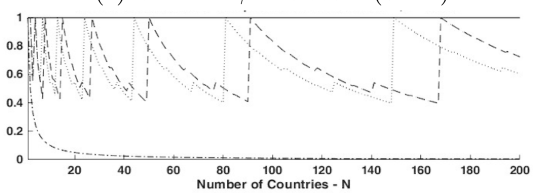
Let us now consider the exercise where the world consists of a fixed number of N countries, where N can be any number from 1 to 200 including 195. Figure 3 illustrates the benchmarks for the NEEM and GM: the grand coalition and the fully non-cooperative outcome (all singletons). The II for the grand coalition is always 1 and $1/N$ for the fully non-cooperative outcome, so the II converges to zero as the number of countries increases (and note that the II under BAU equals zero). The figure indicates that the II for the equilibrium coalition structure is non-monotonic in the number of countries in the world economy. For $N \in \mathcal{T}^*$, a grand coalition forms in equilibrium and the path of II hits the upper limit of 1. For $N \notin \mathcal{T}^*$, we get an equilibrium coalition structure, $\mathcal{M}^*$, which equals $D(N)$ and the II decreases until we reach a grand coalition again. Following a similar pattern, we see that with more patient policy makers (higher β), larger coalitions in equilibrium and a larger II emerge. We also reach a grand coalition more often, recognised by II = 1.



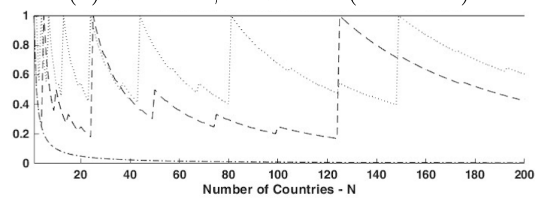
(a) GM with $\beta = 0.999^{10}$ (Stern)



(b) GM with $\beta = 0.985^{10}$ (Nordhaus)



(c) NEEM with $\beta = 0.999^{10}$ (Stern)



(d) NEEM with $\beta = 0.985^{10}$ (Nordhaus)

Figure 3: Internalisation Index for the equilibrium coalition structure for different N.

Interestingly, the equilibrium II qualitatively is not too dissimilar from the saw-tooth pattern of the II produced by the Tribonacci sequence, but it does not give the exact coalition structures and internalisation indices of the equilibrium coalition structure.

7 Conclusion

We have examined the formation of climate coalitions with farsighted countries and certain types of heterogeneity and have put forward an approach to characterise the equilibrium number of coalitions and their number of signatories (numerical coalition structure) which are independent of their heterogeneity. By studying sources of heterogeneity which affect the payoffs of countries in additive-separable ways (the case of asymmetric initial capital stocks or TFP) or sources of heterogeneity which become asymptotically irrelevant if the policymakers are patient enough (the case of asymmetric stocks of initial fossil fuel reserves), we have obtained unique predictions for the equilibrium coalition structure in both analytical and numerical analyses. We have also shown that the resulted treaties are robust if agreements are renegotiable, i.e., if countries can renegotiate, they do not have any incentive to walk away from the agreed treaty.

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 consumption/saving decisions and management of fossil fuel resources.

In a simplified version of our IAM where only exhaustible fossil fuel is used and policy makers are very patient, we have fully characterised the unique equilibrium numerical coalition structure. We have shown that farsighted countries that foresee the consequences of their climate membership decisions, form treaties where the number of participating countries is a Tribonacci number in equilibrium provided there are less than 78 countries. We have also analysed numerically equilibrium numerical coalition structure in a fully calibrated IAM with a more realistic energy and production sector. We confirm robustness of our analytical heterogeneity results in our general IAM. Moreover, with sufficiently patient policymakers, we have shown that the equilibrium Social Cost of Carbon (SCC) and our Internalisation Index (II) move closely with its Tribonacci counterpart.

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. Unfortunately, there is a big gap between current pledges of countries and the targets of the Paris Agreement. There has been too much focus on the formation of the unlikely grand climate coalition of all countries concerned. It may be more worthwhile to search under the umbrella of these IEA's for multiple climate coalitions among smaller group of countries that are overall more ambitious than what we observe today.

We have therefore departed from the single-coalition architecture of climate coalitions and the cartel equilibrium solution concept. By assuming that policy makers are farsighted and predict all consequences of their decisions, we have shown that if the grand coalition does not form in equilibrium, multiple climate coalitions can form with different levels of ambition regarding their emission mitigation strategies. Furthermore, the number of signatories can be large and in particular much larger than three (predicted by the cartel stability solution concept). Thus, we need to move away from myopic policy makers which increase free-riding incentives in climate negotiations.

The only link among countries stems from global emissions and temperature negatively affecting production in all countries in our IAM. Future research could investigate other factors linking countries such as international trade for fossil fuels and final goods, an international capital market, or international labour migration. It might also examine how international climate treaties are hampered by political economy constraints on the international transfers that are needed to sustain coalitions.

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 resembles the increasing emphasis 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, and now the Task

Force, a working group of the Intergovernmental Panel on Climate Change, is responsible for developing and implementing a unified methodology in measuring and reporting emissions and abatement of each country.

References

Allen, M. R., D. J. Frame, C. Huntingford, C. D. Jones, J. A. Lowe, M. Meinshausen, and N. Meinshausen (2009). Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature 458*(7242), 1163–1166.

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

Arrow, K. J., P. Dasgupta, and K.-G. Mäler (2003). Evaluating projects and assessing sustainable development in imperfect economies. *Environmental and Resource Economics 26*(4), 647–685.

Asheim, G. B., C. B. Froyn, J. Hovi, and F. C. Menz (2006). Regional versus global cooperation for climate control. *Journal of Environmental Economics and Management 51*(1), 93–109.

Aumann, R. and R. Myerson (1988). Endogenous formation of links between coalitions and players: An application of the shapley valueb. *The Shapley Value, Cambridge University Press, Cambridge*.

Barrage, L. (2014). Sensitivity analysis for golosov, hassler, krusell, and tsyvinski (2013): Optimal taxes on fossil fuel in general equilibrium. *Econometrica Supplementary Material*.

Barrett, S. (1994). Self-enforcing international environmental agreements. *Oxford economic papers*, 878–894.

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

Battaglini, M. and B. Harstad (2016). Participation and duration of environmental agreements. *Journal of Political Economy 124*(1), 160–204.

Benchekroun, H. and N. V. Long (2012). Collaborative environmental management: A review of the literature. *International Game Theory Review 14*(04).

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

Bosetti, V., C. Carraro, E. De Cian, E. Massetti, and M. Tavoni (2013). Incentives and stability of international climate coalitions: An integrated assessment. *Energy Policy 55*, 44–56.

Buchner, B. and C. Carraro (2009). Parallel climate blocs, incentives to cooperation in international climate negotiations. *The Design of Climate Policy*, 137–163.

Campbell, S. C. (2020). *Growing patterns: Fibonacci numbers in nature*. Astra Publishing House.

Carraro, C., J. Eyckmans, and M. Finus (2006). Optimal transfers and participation decisions in international environmental agreements. *The Review of International Organizations 1*(4), 379–396.

Carraro, C. and D. Siniscalco (1993). Strategies for the international protection of the environment. *Journal of Public Economics 52*(3), 309–328.

Chatterjee, K., B. Dutta, D. Ray, and K. Sengupta (1993). A noncooperative theory of coalitional bargaining. *The Review of Economic Studies 60*(2), 463–477.

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

Chwe, M. S.-Y. (1994). Farsighted coalitional stability. *Journal of Economic theory 63*(2), 299–325.

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

de Zeeuw, A. (2015). International environmental agreements. *Annu. Rev. Resour. Econ. 7*(1), 151–168.

Diamantoudi, E. and E. S. Sartzetakis (2018). International environmental agreements the role of foresight. *Environmental and Resource Economics 71*(1), 241–257.

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

Dietz, S., F. van der Ploeg, A. Rezai, and F. Venmans (2021). Are economists getting climate dynamics right and does it matter? *Journal of the Association of Environmental and Resource Economists 8*(5), 895–921.

Dixit, A. and M. Olson (2000). Does voluntary participation undermine the coase theorem? *Journal of public economics 76*(3), 309–335.

Dutta, B., D. Ray, K. Sengupta, and R. Vohra (1989). A consistent bargaining set. *Journal of economic theory 49*(1), 93–112.

Dutta, B. and R. Vohra (2017). Rational expectations and farsighted stability. *Theoretical Economics 12*(3), 1191–1227.

Eyckmans, J. and H. Tulkens (2006). Simulating coalitionally stable burden sharing agreements for the climate change problem. In *Public goods, environmental externalities and fiscal competition*, pp. 218–249. Springer.

Farrell, J. and E. Maskin (1989). Renegotiation in repeated games. *Games and economic behavior 1*(4), 327–360.

Finus, M. and S. Maus (2008). Modesty may pay! *Journal of Public Economic Theory 10*(5), 801–826.

Finus, M. and B. Rundshagen (2003). Endogenous coalition formation in global pollution control: a partition function approach. *Endogenous Formation of Economic Coalitions, Edward Elgar, Cheltenham, UK*, 199–243.

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

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

Gandelman, N. and R. Hernández-Murillo (2015). Risk aversion at the country level.

Gersbach, H., N. Hummel, and R. Winkler (2021). Long-term climate treaties with a refunding club. *Environmental and Resource Economics*, 1–42.

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

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

Harstad, B. (2016). The market for conservation and other hostages. *Journal of Economic Theory 166*, 124–151.

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

Hassler, J., P. Krusell, and C. Olovsson (2021). Directed technical change as a response to natural resource scarcity. *Journal of Political Economy 129*(11), 000–000.

Hoel, M. and A. De Zeeuw (2010). Can a focus on breakthrough technologies improve the performance of international environmental agreements? *Environmental and Resource Economics 47*(3), 395–406.

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

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

Hyndman, K. and D. Ray (2007). Coalition formation with binding agreements. *The Review of Economic Studies 74*(4), 1125–1147.

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

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

Lessmann, K., U. Kornek, V. Bosetti, R. Dellink, J. Emmerling, J. Eyckmans, M. Nagashima, H.-P. Weikard, and Z. Yang (2015). The stability and effectiveness of climate coalitions. *Environmental and Resource Economics 62*(4), 811–836.

Lessmann, K., R. Marschinski, and O. Edenhofer (2009). The effects of tariffs on coalition formation in a dynamic global warming game. *Economic Modelling 26*(3), 641–649.

Maskin, E. and J. Tirole (2001). Markov perfect equilibrium: I. observable actions. *Journal of Economic Theory 100*(2), 191–219.

Matthews, H. D., N. P. Gillett, P. A. Stott, and K. Zickfeld (2009). The proportionality of global warming to cumulative carbon emissions. *Nature 459*(7248), 829–832.

Miller, E. (2008). *An assessment of CES and Cobb-Douglas production functions*. Congressional Budget Office Washington, DC.

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

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

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

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

Nordhaus, W. D. (2017). Revisiting the social cost of carbon. *Proceedings of the National Academy of Sciences 114*(7), 1518–1523.

Nordhaus, W. D. and Z. Yang (1996). A regional dynamic general-equilibrium model of alternative climate-change strategies. *The American Economic Review*, 741–765.

Pavlova, Y. and A. De Zeeuw (2013). Asymmetries in international environmental agreements. *Environment and Development Economics 18*(1), 51–68.

Pierre, D. A. (1986). *Optimization theory with applications.* Courier Corporation.

Ramsey, F. P. (1928). A mathematical theory of saving. *The economic journal 38*(152), 543–559.

Ray, D. (2007). *A game-theoretic perspective on coalition formation.* Oxford University Press.

Ray, D. and R. Vohra (1997). Equilibrium binding agreements. *Journal of Economic theory 73*(1), 30–78.

Ray, D. and R. Vohra (1999). A theory of endogenous coalition structures. *Games and economic behavior 26*(2), 286–336.

Ray, D. and R. Vohra (2001). Coalitional power and public goods. *Journal of Political Economy 109*(6), 1355–1384.

Ray, D. and R. Vohra (2019). Maximality in the farsighted stable set. *Econometrica 87*(5), 1763–1779.

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica: Journal of the Econometric Society*, 97–109.

Sinha, S. (2017). The fibonacci numbers and its amazing applications. *International Journal of Engineering Science Invention 6*(9), 7–14.

Stern, N. H. (2007). *The economics of climate change: the Stern review.* cambridge University press.

Tol, R. S. (2001). Climate coalitions in an integrated assessment model. *Computational Economics 18*(2), 159–172.

Van den Bremer, T. and F. Van der Ploeg (2021). The risk-adjusted carbon price. *American Economic Review 111*, 2782–2810.

van der Ploeg, F. and A. Rezai (2021). Optimal carbon pricing in general equilibrium: Temperature caps and stranded assets in an extended annual dsge model. *Journal of Environmental Economics and Management 110*, 102522.

Vosooghi, S. (2017). Information design in coalition formation games. *FEEM Working Paper No. 28*.

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

Yi, S.-S. and H. Shin (2000). Endogenous formation of research coalitions with spillovers. *International Journal of Industrial Organization 18*(2), 229–256.

Appendices

A Proofs and Model Details

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

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

$$\max_{\{E_{1it+\tau}, E_{2it+\tau}\}_{\tau=0}^{\infty}} \sum_{i \in M} \sum_{\tau=0}^{\infty} \beta^{\tau} \ln(C_{it+\tau}) \tag{A.1}$$

subject to (3.6), (3.7), (3.11), (3.13) and non-negativity constraints.

Let $\beta^{\tau}\lambda_{it+\tau}$ be present value Lagrange multiplier for final output feasibility constraint (3.1) and $\beta^{\tau}\mu_{it+\tau}$ be present value Lagrange multiplier for resource constraint in (3.4).[46] Thus the Lagrange function of each member of coalition can then be written as,

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

$$\lambda_{it+\tau}(\exp(-\gamma(T_0 + \xi(S_0 + \sum_{i=1}^{N} \sum_{s=0}^{t} (E_{1it+\tau-s} + E_{2it+\tau-s})))) \tag{A.2}$$

$$A_{0i} K_{it+\tau}^{\alpha} (\kappa_1 E_{1it+\tau}^{\rho} + \kappa_2 (A_{2i} L_{2it+\tau})^{\rho} + \kappa_3 (A_{3i} L_{3it+\tau})^{\rho})^{\nu/\rho}$$

$$(L_{it+\tau} - L_{2it+\tau} - L_{3it+\tau})^{1-\alpha-\nu}) - C_{it+\tau} - K_{it+\tau+1} + \mu_{it+\tau}(R_{it+\tau} - E_{1it+\tau} - R_{it+\tau+1})] \tag{A.3}$$

The first-order optimality condition for E_{1it} is

$$\lambda_{it}[\frac{\nu\kappa_1 Y_{it}}{E_{1it}^{1-\rho} E_{it}^{\rho}}] - \gamma\xi \sum_{i \in M} \sum_{\tau=0}^{\infty} \lambda_{it+\tau} \beta^{\tau} Y_{it+\tau} = \mu_{it}, \tag{A.4}$$

where μ_{it} is the shadow cost of exhaustible fossil fuel defined in units of utility. The first-order optimality condition for E_{2it} is,

$$\lambda_{it}[\frac{\nu\kappa_2 Y_{it}}{E_{2it}^{1-\rho} E_{it}^{\rho}}] - \gamma\xi \sum_{i \in M} \sum_{\tau=0}^{\infty} \lambda_{it+\tau} \beta^{\tau} Y_{it+\tau} = 0. \tag{A.5}$$

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

$$\max_{\{C_{it+\tau}, K_{it+\tau+1}, R_{it+\tau+1}, E_{3it+\tau}\}_{\tau=0}^{\infty}} \sum_{\tau=0}^{\infty} \beta^{\tau} \ln(C_{it+\tau}) \tag{A.6}$$

[46] For simplicity of exposition, we omit the non-negativity constraints from the presentation here.

again, subject to (3.6), (3.7), (3.11), (3.13) and non-negativity constraints.

The first-order optimality condition for C_{it} is[47]

$$\lambda_{it} = \frac{1}{C_{it}(M, \mathbb{M})}. \tag{A.7}$$

The first-order optimality condition for K_{it+1} and (A.7) give the Euler equation

$$\frac{1}{C_{it}(M, \mathbb{M})} = \alpha\beta \frac{1}{C_{it+1}(M, \mathbb{M})} \frac{Y_{it+1}(M, \mathbb{M})}{K_{it+1}(M, \mathbb{M})}. \tag{A.8}$$

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

$$\frac{s_{it}}{1 - s_{it}} = \alpha\beta \frac{1}{1 - s_{it+1}}. \tag{A.9}$$

with unique solution $s_{it} = s = \alpha\beta$, for all t and all i.

Next, we compute the SCC associated with any coalition with m members as the present discounted value of the utility cost of all m members caused by an additional unit of emissions. As we have already noted, emissions are associated with energy produced from fossil fuel and their effect on global temperature. Using the expression for first-order condition of E_{1it} in (A.4) (or E_{2it}), the shadow cost of emissions in utility units can be computed from

$$\Lambda_{it}(m) \equiv \gamma\xi \sum_{i \in M} \sum_{\tau=0}^{\infty} \frac{\lambda_{it+\tau}}{\lambda_{it}} \beta^{\tau} Y_{it+\tau}, \tag{A.10}$$

which using equation equation (A.7) and $s = \alpha\beta$ can be simplified to

$$\Lambda_{it}(m) = Y_{it} \frac{\xi\gamma m}{1 - \beta}, \tag{A.11}$$

$$\Lambda_{it}(m)/Y_{it} \equiv \hat{\Lambda}_{it}(m) = \hat{\Lambda}(m) \equiv \frac{\xi\gamma m}{1 - \beta}. \tag{A.12}$$

Then the first-order optimality condition for E_{1it} can be written as

$$\frac{\nu\kappa_1 Y_{it}}{E_{1it}^{1-\rho} E_{it}^{\rho}} = \frac{\mu_{it}}{\lambda_{it}} + \Lambda_{it}(m), \tag{A.13}$$

which using equation (A.7) and $s_{it} = s = \alpha\beta$ can be simplified to

$$\frac{\nu\kappa_1}{E_{1it}^{1-\rho} E_{it}^{\rho}} = \mu_{it}(1 - \alpha\beta) + \hat{\Lambda}(m). \tag{A.14}$$

Therefore,

[47] Notice that the choices of R_{it+1}, C_{it}, K_{it+1}, do not affect the payoffs of the other members of the coalition to which country i belongs.

$$E_{1it} = E_{it}^{-\rho/1-\rho} \epsilon_{1it}(m), \tag{A.15}$$

where

$$\epsilon_{1it}(m) \equiv \left(\frac{\nu\kappa_1}{\mu_{it}(1-\alpha\beta) + \hat{\Lambda}(m)} \right)^{1/1-\rho}. \tag{A.16}$$

Using the first-order optimality condition for R_{it+1} ($\mu_{it} = \beta\mu_{it+1}$), (A.14) implies

$$\frac{\nu\kappa_1}{E_{1it}^{1-\rho} E_{it}^{\rho}} - \hat{\Lambda}(m) = \beta \left(\frac{\nu\kappa_1}{E_{1it+1}^{1-\rho} E_{it+1}^{\rho}} - \hat{\Lambda}(m) \right). \tag{A.17}$$

This is a version of Hotelling's rule corrected for the exogenous externality term $\hat{\Lambda}(m)$.

Similarly the first-order optimality conditions for L_{2it} and L_{3it} (using the first-order conditions of E_{2it} and E_{3it}) are

$$A_{2it} \left(\frac{\nu\kappa_2}{E_{2it}^{1-\rho} E_{it}^{\rho}} - \hat{\Lambda}(m) \right) = \frac{1-\alpha-\nu}{L_{0it}}, \tag{A.18}$$

$$A_{3it} \frac{\nu\kappa_2}{E_{2it}^{1-\rho} E_{it}^{\rho}} = \frac{1-\alpha-\nu}{L_{0it}}, \tag{A.19}$$

which give

$$E_{2it} = E_{it}^{-\rho/1-\rho} \epsilon_{2it}(m), \tag{A.20}$$

$$E_{3it} = E_{it}^{-\rho/1-\rho} \epsilon_{3it}, \tag{A.21}$$

where

$$\epsilon_{2it}(m) \equiv \left(\frac{\nu\kappa_2 A_{2it} L_{0it}}{1-\alpha-\nu+\hat{\Lambda}(m) A_{2it} L_{0it}} \right)^{1/1-\rho}, \tag{A.22}$$

$$\epsilon_{3it} \equiv \left(\frac{\nu\kappa_3 A_{3it} L_{0it}}{1-\alpha-\nu} \right)^{1/1-\rho}. \tag{A.23}$$

A.2 Optimum value function of a signatory

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

$$
\begin{aligned}
V_i(S_t, K_{it}, \mu_{it}, M, \mathbb{M}) &= \ln(C_{it}(M,\mathbb{M})) + \beta \ln(C_{it+1}(M,\mathbb{M})) + ... \\
&= \frac{\ln(1-s)}{1-\beta} + \{\ln(Y_{it}(M,\mathbb{M})) + \beta \ln(Y_{it+1}(M,\mathbb{M})) + ...\} \\
&= \frac{\ln(1-s)}{1-\beta} + \{\ln[e^{-\gamma\xi S_t - \gamma T_0} A_{0i} K_{it}^{\alpha} E_{it}(m)^{\nu} L_{0it}^{1-\alpha-\nu}] \\
&+ \beta \ln[e^{-\gamma\xi S_{t+1} - \gamma T_0} A_{0i} K_{it+1}^{\alpha} E_{it+1}(m)^{\nu} L_{0it+1}^{1-\alpha-\nu}] + ...\} \\
&= \frac{\alpha \ln(K_{it}) + H_1 + H_2 + H_3}{1-s},
\end{aligned} \tag{A.24}
$$

where the H_j are defined as

$$
H_1 \equiv \frac{s\ln(s) - s\ln(1-s) + \ln(A_{0i}) - \gamma T_0 + (1-\alpha-\nu)\ln(L_{0i})}{1-\beta} \tag{A.25}
$$

$$
H_2 \equiv -\gamma\xi[S_t + \beta S_{t+1} + \beta^2 S_{t+2} + ...] \tag{A.26}
$$

and

$$
H_3 \equiv \nu[\ln(E_{it}(m)) + \beta \ln(E_{it+1}(m)) + \beta^2 \ln(E_{it+2}(m)) + ...]. \tag{A.27}
$$

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

$$
\begin{aligned}
H_2 = -\frac{\gamma\xi}{1-\beta}\{&\sum_i \sum_{s=0}^{t} (E_{1it-s} + E_{2it-s}) + \sum_{i \notin M} (E_{1it} + E_{2it}) + \sum_{i \in M} (E_{1it}(m) + E_{2it}(m)) \\
&+ \sum_{i \notin M} [\beta(E_{1it+1} + E_{2it+1}) + \beta^2(E_{1it+2} + E_{2it+2}) + ...] \\
&+ \sum_{i \in M} [\beta(E_{1it+1}(m) + E_{2it+1}(m)) + \beta^2(E_{1it+2}(m) + E_{2it+2}(m)) + ...]\}.
\end{aligned} \tag{A.28}
$$

A.3 The small-coalition paradox under cartel stability

Consider the general model presented in Section 3, with three types of energy. Assume that there is a single coalition M of m countries, and the $N-m$ non-signatories form the fringe. Furthermore, in this section, we assume that the countries are ex-ante symmetric, but after the membership stage they may end up in asymmetric situations.

Definition 4 . *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.

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.[48] 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 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 6 . *Under the assumption of cartel stability, the largest coalition size is* $m^* = 2$ *for any total number of countries* N*.*

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

$$V_i(S_t, K_{it}, \mu_{it}, M) \geq \ln(C_{it}^d) + \beta\{V_i(E^t, K_{it+1}, \mu_{it}, M)\}, \tag{A.29}$$

where $V_i(S_t, K_{it}, \mu_{it}, M)$ is the optimum value function of a signatory in coalition M as defined in section A.2, and $E^t \equiv (E_t, E_{t-1}, ..., E_0)$. Furthermore, C_{it}^d is the consumption level associated with the deviation period. Note that E^t in $V_i(E^t, K_{it+1}, \mu_{it}, M)$ is impacted by the deviation in period t. More specifically, the right-hand-side of equation (A.29) consists of

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

and

$$V_i(E^t, K_{it+1}, \mu_{it}, M) = \frac{\alpha \ln(K_{it+1}) + H_1 + H_2' + H_3'}{1 - s}. \tag{A.30}$$

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

[48]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.

$$
\begin{aligned}
H_2' &\equiv -\gamma\xi[(S_{t+1}) + \beta(S_{t+2}) + ...] \\
&= -\frac{\gamma\xi}{1-\beta}\{\sum_i \sum_{s=1}^{t} (E_{1it-s} + E_{2it-s}) + \sum_{j\notin M\setminus i} (E_{1jt} + E_{2jt}) + \sum_{j\in M\setminus i} (E_{1jt}(m-1) + E_{2jt}(m-1)) \\
&+ \sum_{i\notin M} [(E_{1it+1} + E_{2it+1}) + \beta(E_{1it+2} + E_{2it+2}) + ...] \\
&+ \sum_{i\in M} [(E_{1it+1}(m) + E_{2it+1}(m)) + \beta(E_{1it+2}(m) + E_{2it+2}(m)) + ...]\}
\end{aligned} \tag{A.31}
$$

and

$$
H_3' \equiv \nu[\ln(E_{it+1}(m)) + \beta \ln(E_{it+2}(m)) + ...]. \tag{A.32}
$$

The internal stability condition (A.29) can be further simplified by multiplying both sides by $1-s$, and cancelling all future emissions of H_2 and H_2' from both sides, and using the symmetry of the emission strategies of signatories and likewise for non-signatories, and the fact that

$$
\begin{aligned}
\ln(Y_{it}^d) = -\gamma\xi\{\sum_i \sum_{s=1}^{t} E_{it-s} + \sum_{j\notin M\setminus i} (E_{1jt} + E_{2jt}) + \sum_{j\in M\setminus i} (E_{1jt}(m-1) + E_{2jt}(m-1))\} \\
+ \ln(A_{0i}) + \alpha \ln(K_{it}) + \nu \ln(E_{it}) + (1-\alpha-\nu)L_{0it}.
\end{aligned} \tag{A.33}
$$

However, note that the shadow price, μ_{it}, in the two sides of the internal stability condition (A.29) are not the same after the deviation. Using the corresponding emission levels, the internal stability condition forms a concave function of m and the parameters of the model. 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 usual, the roots of the internal stability condition are found numerically. Here we use the parameter values from the calibrated model in section 6, and with broad robustness checks, the solutions are 1 and 2.

Lastly, the internal stability condition is independent of the capital stock or the stock of cumulative emissions or of other state variables but it does depend on current emission levels. Thus it depends on the scarcity rents which in turn indirectly depend on the stocks of fossil fuel. In particular, for low initial stocks of exhaustible 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.

Proofs

A.4 Proof of Proposition 2

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

$$V_i(m_1, \{m_1, m_2, ..., m_k\}) - V_i(\{N\}) < 0. \tag{A.34}$$

In the EEM where $\kappa_1 = 1$, and the production function of final output is given by equation (3.10), it can be shown that the sufficient condition for every country to prefer the grand coalition to any other coalition structure is

$$\begin{aligned}
&V_i(m_1, \{m_1, m_2, ..., m_k\}) - V_i(\{N\}) = \\
&\quad \frac{1}{1-\beta(1-\nu)}\Big\{\nu[\ln(\frac{E_{it}(m_1)}{E_{it}(N)}) + \beta ln(\frac{E_{it+1}(m_1)}{E_{it+1}(N)}) + \beta^2 ...] \\
&\quad - \frac{\gamma\xi}{1-\beta}\{[\sum_{i\in M_1} E_{it}(m_1) + \sum_{i\in M_2} E_{it}(m_2) + ... + \sum_{i\in M_k} E_{it}(m_k) - \sum_{i\in I} E_{it}(N)] + \\
&\quad \beta[\sum_{i\in M_1} E_{it+1}(m_1) + \sum_{i\in M_2} E_{it+1}(m_2) + ... + \sum_{i\in M_k} E_{it+1}(m_k) - \sum_{i\in I} E_{it+1}(N)] + \beta^2 ...\}\Big\} < 0.
\end{aligned} \tag{A.35}$$

Note that in the EEM, the optimal emission of each country in equation (A.15) and (A.16) can be written as

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

The inequality in (A.35) can thus be rewritten as

$$\begin{aligned}
&V_i(m_1, \{m_1, m_2, ..., m_k\}) - V_i(\{N\}) = \\
&\frac{1}{1-\beta(1-\nu)}\Big\{\nu[\ln(\frac{(1-\beta)\mu_{it}[1-\beta(1-\nu)] + N\gamma\xi}{(1-\beta)\mu_{it}[1-\beta(1-\nu)] + m_1\gamma\xi}) + \beta\ln(\frac{(1-\beta)\mu_{it+1}[1-\beta(1-\nu)] + N\gamma\xi}{(1-\beta)\mu_{it+1}[1-\beta(1-\nu)] + m_1\gamma\xi}) + \beta^2 ...] \\
&- \frac{\gamma\xi}{1-\beta}\{[\frac{m_1\nu(1-\beta)}{(1-\beta)\mu_{it}[1-\beta(1-\nu)] + m_1\gamma\xi} + \frac{m_2\nu(1-\beta)}{(1-\beta)\mu_{it}[1-\beta(1-\nu)] + m_2\gamma\xi} + ... \\
&+ \frac{m_k\nu(1-\beta)}{(1-\beta)\mu_{it}[1-\beta(1-\nu)] + m_k\gamma\xi} - \frac{N\nu(1-\beta)}{(1-\beta)\mu_{it}[1-\beta(1-\nu)] + N\gamma\xi}] + \\
&\beta[\frac{m_1\nu(1-\beta)}{(1-\beta)\mu_{it+1}[1-\beta(1-\nu)] + m_1\gamma\xi} + \frac{m_2\nu(1-\beta)}{(1-\beta)\mu_{it+1}[1-\beta(1-\nu)] + m_2\gamma\xi} + ... \\
&+ \frac{m_k\nu(1-\beta)}{(1-\beta)\mu_{it+1}[1-\beta(1-\nu)] + m_k\gamma\xi} - \frac{N\nu(1-\beta)}{(1-\beta)\mu_{it+1}[1-\beta(1-\nu)] + N\gamma\xi}] + \beta^2 ...\}\Big\} < 0.
\end{aligned} \tag{A.36}$$

In the limit that $\beta \to 1$, the inequality in equation (A.36) converges to

$$\nu[\ln(\frac{N}{m_1}) + \ln(\frac{N}{m_1}) + ...] - \nu\{[k-1] + [k-1] + ...\} < 0. \tag{A.37}$$

This is satisfied if

$$ln(\frac{N}{m_1}) < (k-1), \tag{A.38}$$

as required. □

A.5 Proof of Proposition 3

Consider the general model with three sources of energy, presented in section 3, and assume the countries are heterogeneous with respect to K_{i0} and/or A_{0i}. 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 and has approached country i (which can be j itself). For any N, country i compares the payoff of the best profitable deviation by giving rise to the formation of coalition $M \in \{M_1, M_2, ..., M_k\}$ (which is to be identified) versus the payoff of staying in the grand coalition $\{I\}$. Thus, the planner of country i needs to determine the sign of

$$V_i^j(S_t, K_{it}, \mu_{it}, M, \mathbb{M}) - V_i^j(S_t, K_{it}, \mu_{it}, \{I\}). \tag{A.39}$$

But this equation is independent of stocks, in particular independent of K_{i0}. It is also independent of TFP A_{0i}. To see this, note that

$$\begin{aligned}
&V_i(S_t, K_{it}, \mu_{it}, M, \{M_1, M_2, ..., M_k\}) - V_i(S_t, K_{it}, \mu_{it}, \{I\}) = \\
&\frac{1}{1-\beta(1-\nu)}\Big\{\nu[\ln(\frac{E_{it}(M)}{E_{it}(I)}) + \beta\ln(\frac{E_{it+1}(M)}{E_{it+1}(I)}) + \beta^2...] \\
&-\frac{\gamma\xi}{1-\beta}\{[\sum_{i\in M_1} E_{it}(M_1) + \sum_{i\in M_2} E_{it}(M_2) + ... + \sum_{i\in M_k} E_{it}(M_k) - \sum_{i\in I} E_{it}(I)] + \\
&\beta[\sum_{i\in M_1} E_{it+1}(M_1) + \sum_{i\in M_2} E_{it+1}(M_2) + ... + \sum_{i\in M_k} E_{it+1}(M_k) - \sum_{i\in I} E_{it+1}(I)] + \beta^2...\}\Big\}.
\end{aligned} \tag{A.40}$$

Coalitions in $\{M_1, M_2, ..., M_k\}$ are included if they are non-empty. Clearly if the source of heterogeneity is K_{i0} and/or A_{0i}, then the membership decision of countries in (A.40) is not affected by the heterogeneity across countries. Furthermore, the above equation is only a function of emissions, which only depend on size of coalitions and $\mathcal{M}$ and not the identity of countries in any coalition and $\mathbb{M}$. Hence, $\mathcal{M}^*$ can be characterised independent of heterogeneity with respect to K_{i0} or A_{0i}. This proves the first part of the proposition.

Now consider the EEM, where $\kappa_1 = 1$, and the production function of final output is given by equation (3.10). If the countries are heterogeneous with respect to R_{i0} and μ_{it}, as shown in (A.36), in the limit that $\beta \to 1$, the difference in (A.39) converges to

$$\lim_{\beta\to 1}\big(V_i(S_t, K_{it}, \mu_{it}, M, \{M_1, M_2, ..., M_k\}) - V_i(S_t, K_{it}, \mu_{it}, \{I\})\big) = \\ [\ln(\frac{N}{m}) + \ln(\frac{N}{m}) + ...] - \{[k-1] + [k-1] + ...\}. \tag{A.41}$$

This equation is independent of μ_{it} of any country and any stocks. Moreover, it only depends on k, and m, which can be determined by characterising $\mathcal{M}^*$.

A.6 Proof of Section 5.3.1

Consider the EEM, where $\kappa_1 = 1$, and the production function of final output is given by equation (3.10). Assume 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. Thus, at least one country in M'_k (which is a larger coalition) has a scarcity rent which is strictly smaller relative to the scarcity rent of the countries in M_k. Country i itself is in both coalitions. In that case, we have

$$V_i^i(M_{k-1},\{M_1, M_2, ..., M_k\}) - V_i^i(M'_{k-1}, \{M_1, M_2, ..., M_{k-2}, M'_{k-1}, M'_k\}) = \\ \frac{1}{1-\beta(1-\nu)}\Big\{\nu[\ln(\frac{E_{it}(M_{k-1})}{E_{it}(M'_{k-1})}) + \beta\ln(\frac{E_{it+1}(M_{k-1})}{E_{it+1}(M'_{k-1})}) + ...] \\ - \frac{\gamma\xi}{1-\beta}\{[\sum_{i\in M_{k-1}} E_{it}(M_{k-1}) + \sum_{i\in M_k} E_{it}(M_k)] - [\sum_{i\in M'_{k-1}} E_{it}(M'_{k-1}) + \sum_{i\in M'_k} E_{it}(M'_k)] + \\ \beta[\sum_{i\in M_{k-1}} E_{it+1}(M_{k-1}) + \sum_{i\in M_k} E_{it+1}(M_k)] - \beta[\sum_{i\in M'_{k-1}} E_{it+1}(M'_{k-1}) + \sum_{i\in M'_k} E_{it+1}(M'_k)] + ...\}\Big\}. \tag{A.42}$$

Emission of those which remain in coalitions with the same sizes does not affect i's decision. The expression in (A.42) is independent of capital stocks or TFPs. The second line in (A.42) is the direct gain of country i from emitting in M_{k-1} versus in M'_{k-1}. Given that both coalitions have the same size, and country i has the same scarcity rent in both scenarios, the ratio of two emissions in any period is one, and thus the second line is zero. The third and fourth lines are the externality damages. Because both coalition structures correspond to the same numerical coalition structure, the SCC in M_{k-1} is the same as in M'_{k-1}, also both M_k and M'_k have the same SCC. It is easy to see that for any general β, the total emissions of M_{k-1} and M_k is larger than the total emissions of M'_{k-1} and M'_k. In other words, in period t for example,

$$[\sum_{i\in M_{k-1}} E_{it}(M_{k-1}) + \sum_{i\in M_k} E_{it}(M_k)] - [\sum_{i\in M'_{k-1}} E_{it}(M'_{k-1}) + \sum_{i\in M'_k} E_{it}(M'_k)] \qquad (A.43)$$

is positive. Hence, for any $\beta \in (0,1)$, the difference of payoffs in (A.42) is negative. Therefore, when comparing the value of any two coalition structures in any given numerical coalition structure, a country prefers the coalition structure with the lower global emissions. □

A.7 Proof of Proposition 5

Consider the EEM where $\kappa_1 = 1$, and the production function of final output in equation (3.10). We begin in period $t \geq 1$, when the countries in a coalition M^*_{t-1}, formed in $t-1$, are considering to renegotiate M^*_{t-1} and are facing a proposal to form coalition M_t of size m_t.

In period t, countries solve the problem backwards from the action stage, and now they have to consider all future renegotiations and so all possible future coalition structures when deciding about emissions. In the action stage of period t, after M_t is formed, the signatories jointly maximise

$$\sum_{i\in M_t}\sum_{\tau=0}^{\infty}\beta^{\tau}\,\mathbb{E}\{\ln(C_{it+\tau}) \mid \mathbb{M}_t\}. \qquad (A.44)$$

For ease of readability of the information set, here we abstract from the other variables in the Markov state which are not directly relevant here. The maximisation problem is subject to the usual non-negativity constraints, and the expected value of final output production function (3.10), the market-clearing conditions without the labour market constraint (3.7), the climate dynamics constraints (3.11) and (3.13), for each $i \in M$.

The Lagrange function is now,

$$\mathbb{E}\Bigg\{\sum_{i\in M_t}\sum_{\tau=0}^{\infty}\beta^{\tau}[\ln(C_{it+\tau})+$$

$$\lambda_{it+\tau}(\exp(-\gamma(T_0+\xi(S_0+\sum_{i=1}^{N}\sum_{s=0}^{t}(E_{it+\tau-s})))) \qquad (A.45)$$

$$A_iK_{it+\tau}^{1-\nu}E_{it+\tau}^{\nu} - C_{it+\tau} - K_{it+\tau+1} + \mu_{it+\tau}(R_{it+\tau} - E_{1it+\tau} - R_{it+\tau+1})] \mid \mathbb{M}_t\Bigg\}. \qquad (A.46)$$

The first-order optimality condition for E_{it} is

$$\lambda_{it}[\frac{\nu Y_{it}}{E_{it}}] - \mathbb{E}\Bigg\{\sum_{i\in M}\sum_{\tau=0}^{\infty}\lambda_{it+\tau}\beta^{\tau}\gamma\xi Y_{it+\tau} - \mu_{it} \mid \mathbb{M}_t\Bigg\} = 0. \qquad (A.47)$$

The first-order optimality condition for C_{it} gives

$$\lambda_{it} = \frac{1}{C_{it}(M, \mathbb{M})}. \tag{A.48}$$

The first-order condition of K_{it+1} and (A.48) give the Euler equation

$$\frac{1}{C_{it}(M, \mathbb{M})} = \beta\, \mathbb{E}\left\{ \frac{1}{C_{it+1}(M, \mathbb{M})} \frac{Y_{it+1}(M, \mathbb{M})}{K_{it+1}(M, \mathbb{M})} (1-\nu) \mid \mathbb{M}_t \right\}. \tag{A.49}$$

The Euler equation reduces to

$$\frac{s_{it}}{1-s_{it}} = \beta(1-\nu)\, \mathbb{E}\left\{ \frac{1}{1-s_{it+1}} \mid \mathbb{M}_t \right\} \tag{A.50}$$

Using the solution $s_{it} = s = \beta(1-\nu)$, for all t and all i, and equations (A.48) and (A.47), we get

$$E_{it}(m_t) = \frac{\nu(1-\beta)}{\mathbb{E}(\mu_{it} \mid M_t)(1-\beta(1-\nu))(1-\beta) + \gamma\xi[m_t(1-\beta) + \beta\, \mathbb{E}(m_{t+\tau} \mid M_t)].} \tag{A.51}$$

where $\tau \geq 1$. We go backwards to the membership stage. Since the protocol is fixed and deterministic, the same initial proposer j, is selected. In a public-good game, as it is here, the smallest coalition which has the highest payoff forms first, thus at the beginning of renegotiation stage in period t, the initial proposer is a member either of the smallest coalition of $t-1$ (where $D(N) = \{m_1, m_2, ..., m_k\}$) or of the grand (if it was stable in the last period). Thus the initial proposer and any member i of the approval committee of M^*_{1t-1} consider a proposal about the formation of M_t, each need to determine the sign of the following difference:

$$\mathbb{E}\, V_i(M^j_t, \mathbb{M}_t, \mathbb{M}^*_{t-1}) - \mathbb{E}\, V_i(I_t, \mathbb{M}_t, \mathbb{M}^*_{t-1}). \tag{A.52}$$

If $m^j_t > m_{1t-1}$, the above is upon approval by the approval committees of all other relevant coalitions in $\mathbb{M}^*_{t-1}$. Similar to Proposition 3, this difference is independent of all sources of heterogeneity with respect to the capital stock and TFP constant, and the difference in (A.52) is only a function of expected value of emissions, which are independent of the identity of countries. Thus, characterising the equilibrium reduces to the characterisation of $\mathcal{M}^*$. If the approval committee of M^*_{1t-1} allows country i, then joining the smallest coalition of size m_{1t} in the decomposition of N is the best profitable deviation. And given the dominant strategies in emissions, the grand coalition forms in period t if

$$\mathbb{E}\, V_i(m_{1t}, \{m_{1t}, m_{2t}, ..., m_{kt}\}, m^*_{1t-1}) - \mathbb{E}\, V_i(\{N\}_t, m^*_{1t-1}) \leq 0, \tag{A.53}$$

where $\{N\}_t$ refers to the formation of the grand coalition in period t. The above inequality can be expanded to

$$\frac{1}{\beta(1-\nu)}\Big\{\nu\Big(\mathbb{E}\left[\ln(E_{it}(m_{1t}))+\beta\ln(E_{it+1}(m_{it+1}))+\beta^2\ln(E_{it+2}(m_{it+2}))+...\mid m^*_{1t-1}\right]$$
$$-\mathbb{E}\left[\ln(E_{it}(N))+\beta\ln(E_{it+1}(m_{it+1}))+\beta^2\ln(E_{it+2}(m_{it+2}))+...\mid m^*_{1t-1}\right]\Big)$$
$$-\frac{\gamma\xi}{1-\beta}\mathbb{E}\left[m_{1t}E_{it}(m_{1t})+m_{2t}E_{it}(m_{2t})+...+m_{kt}E_{it}(m_{kt})-NE_{it}(N)\mid m^*_{1t-1}\right]\Big\}\leq 0, \tag{A.54}$$

where $\tau = 1, 2, ...$ All future values of size of coalitions are unknown at time t, thus all future values of emissions in the first line and second line cancel each other out. For the same reason, in the third line, only the current values of emissions are written here. Hence, the inequality is simplified to

$$\frac{1}{\beta(1-\nu)}\Big\{\nu\Big(\mathbb{E}[\ln(E_{it}(m_{1t}))-\ln(E_{it}(N))\mid m^*_{1t-1}]\Big)$$
$$-\frac{\gamma\xi}{1-\beta}\mathbb{E}\left[m_{1t}E_{it}(m_{1t})+m_{2t}E_{it}(m_{2t})+...+m_{kt}E_{it}(m_{kt})-NE_{it}(N)\mid m^*_{1t-1}\right]\Big\}\leq 0. \tag{A.55}$$

Thus the only uncertainty in this decision making concerns the uncertainty regarding the future coalition sizes in the SCC in equation (A.51).[49] Given that the agreements are binding, and the fact that the renegotiation is only affecting the SCC, the approval committee of M_{1t-1} (and in general of any non-ultimate coalition) always rejects any change of number of signatories: clearly no one would be interested to leave any non-ultimate coalition in a public-good game, as the subsequent coalitions are larger, and no approval committee would allow enlarging the coalition as it would increase their $\hat{\Lambda}$. If a grand coalition was formed in $t-1$, then any approval committee would also reject any exit by free riders.

Thus, there will be no ex-post incentive to change size of existing coalitions, and even their identity (initially determined by the protocol) remains the same. Thus $\mathbb{E}(m_{t+\tau} \mid m_t, \mathbb{M}_t, \mathbb{M}_{t-1}) = m_t$ for $\tau \geq 1$. So, the emission in equation (A.51) is

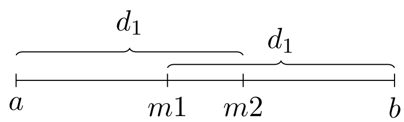
$$E_{it}(m_t) = \frac{\nu(1-\beta)}{\mu_{it}(1-\beta(1-\nu))(1-\beta)+\gamma\xi m_t}. \tag{A.56}$$

Furthermore, they expect this is going to be the case for any other (if any) formed binding coalitions, and the result of renegotiation is fully predictable. So, the uncertainty in the inequality in (A.55) is resolved, and the problem is identical to the problem of irreversible case. Therefore, the MPE has an absorbing membership state, with the same equilibrium coalition structure $\mathbb{M}^*_t$ for any t, i.e. from any initial equilibrium coalition structure, the equilibrium moves to the same membership state. And, in the limit analysis where $\beta \to 1$, the inequality in (A.55) reduces to the inequality in Proposition 2. So $\mathcal{M}^*_t$ is renegotiation-proof for any t.□

[49] The uncertainty about μ_{it} is also due to the uncertainty about the future coalition sizes.

B Discussion of Fibonacci and Tribonacci Results

Obtaining a series of Fibonacci sequences by optimising payoffs relates to the *Fibonacci search method* in mathematics (Pierre (1986)). That is a bracketing method to find the optimum of a single-valued function, say $V(m)$, over an (initial) interval, say $[a, b]$, and it does not require any differentiation. The bracketing methods involve finding the optimum over several iterations, where each iteration leads to elimination of a part of the interval, until the optimum is found, given a set tolerance error. More precisely, in each iteration, two values of the interval, say $m1$ and $m2$ (which do not need to be integers) are selected such that they have the 'same' distance d_p from the two sides of the boundaries a, and b. The distance d_p is a ratio, q_p, of the interval such that $d_p \equiv q_p(b-a) = m2 - a = b - m1$. For example the following figure shows a bracketing method in the first round.



Then the values of the function at $m1$ and $m2$ are compared and if we aim to find the maximum of the function and for example $V(m1) \geq V(m2)$, then in the next round, the interval is reduced to $[a, m_2]$ and the process continues.

In the Fibonacci search method, the distance d_p is not made as a fixed ratio of the initial interval but in each round p of iteration, q_p is a Fibonacci constant (the ratio of two consecutive Fibonacci numbers), i.e. $q_p = F_p/F_{p+1}$, where

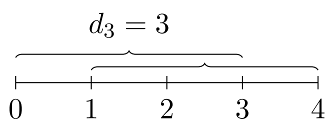
$$F_0 = F_1 = 1, \qquad F_p = F_{p-1} + F_{p-2}, \quad p = 2, 3, ... \tag{B.1}$$

For large number of iterations, q_p converges to the golden ratio, 0.618, so q_p becomes a constant. The problem of finding set $\mathcal{T}^*$ in the farsightedness approach is equivalent to comparing the coalitions' values for a country i, $V_i(m)$, in a round of iteration p. There are differences though: m is a positive integer here; and here building up set $\mathcal{T}^*$ over different rounds of the farsightedness recursions, can be seen as if we add to N, and thus expand the interval. So we are going in the opposite direction and

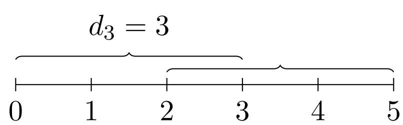
$$d_p = \frac{F_p}{F_{p+1}} d_{p+1}. \tag{B.2}$$

Since m can only be an integer, if d_{p+1} is not an integer, we need to round it to the closest integer. Given (B.2), it is not surprising that if the initial element of $m2$ coincides with Fibonacci numbers itself, all subsequent $m2$'s at which we need to compare values of $V_i(.)$ will be a Fibonacci number. For example, trivially $d_1 = 1$ then, $d_2 = (F_1/F(2))d_2$, so $d_2 = 2$; and $d_3 = 3$, etc.

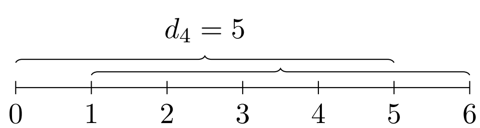
For any N, the two m's that the Fibonacci search method finds to compare the values of V_i, are the two relevant values of decomposition of N, $D(N)$.[50] If one maximises a decreasing function $V_i(m)$, the lower bound of the interval (a_p) is fixed, as we expect $V_i(m1) > V_i(m2)$. But an important feature of coalition formation games is that $V_i(.)$ function shifts across set m, as we increase N, and thus its maximum can move. So, if N is a number at which a grand coalition forms, then we get the boundary optimum and d_p moves up to d_{p+1}. To see this in an example, if N=4, then using (the Fibonacci sequence in) $\mathcal{T}^*$ of Ray and Vohra (2001), we know that $D(4) = \{3, 1\}$, and the Fibonacci search method also suggests $d_3 = 3$ and so comparing values of $m1 = 1$ and $m2 = 3$,



From the quadratic-linear specification of $V_i(.)$ in Ray and Vohra (2001), we know that the maximum is at $m1 = 1$. If $N = 5$, using $d_3 = 3$ implies comparing $V_i(2)$ and $V_i(3)$,



Here from the specification of $V_i(.)$, we know that there is a boundary optimum at $N = 5$, so for the next round, i.e. for $N = 6$, we use $d_4 = 5$, which leads to $m1 = 1$ and $m2 = 5$,



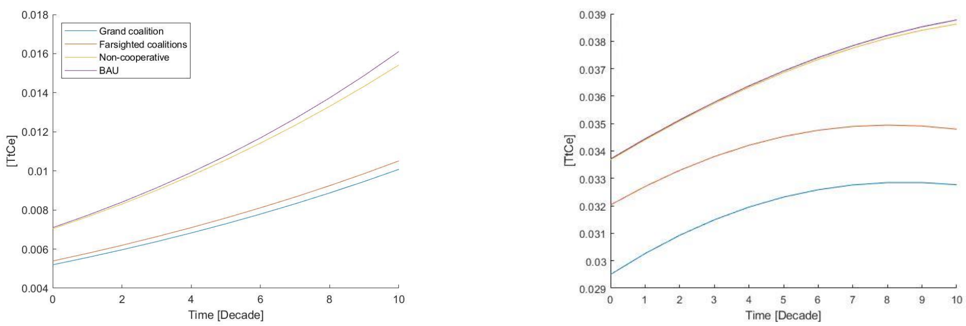
Here $V_i(1)$ dominates $V_i(5)$ and $V_i(6)$. So, no grand forms here and $d_4 = 5$ is used for the next round too, and it continues.

To the best of our knowledge there is no Tribonacci search method, but the same logic applies here. While we point out that the emergence of the Fibonacci sequences in an optimisation problem is well-known in mathematics, and the Fibonacci search method is an efficient and quick way of finding the optimum, and it is particularly useful if the function is not differentiable, we do not claim that the equilibrium number of signatories in a climate coalition can be found using a Tribonacci sequence, and indeed in section 6 we show that the divergence of $\mathcal{T}^*$ from a Tribonacci sequence in NEEM and GM can

[50] Note that if $D(N)$ contains more than two elements, then the search misses the intermediate elements of $D(N)$, however in the coalition theory of public goods, we know that they are irrelevant in finding the maximum of $V_i(m)$, and thus irrelevant for characterisation of $\mathcal{T}^*$.

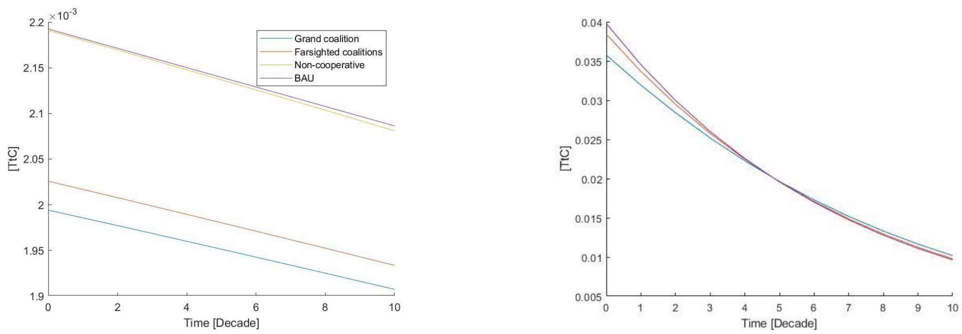
happen very quickly, in the case of $\beta \to 1$, and for lower values of β they do not emerge at all. Divergence of $\mathcal{T}^*$ from the Fibonacci sequences can have different reasons including rounding d_{p+1}'s into integers.

C Time Paths of Energy and Macroeconomic Outcomes



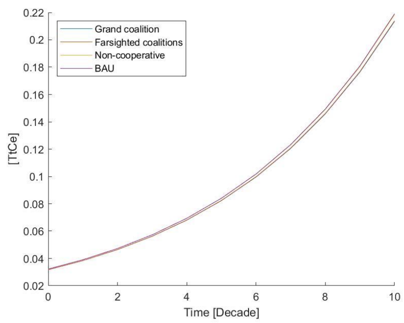
(a) $\beta = 0.999^{10}$ (Stern)   (b) $\beta = 0.985^{10}$ (Nordhaus)

Figure 4: Aggregate energy use for the GM model

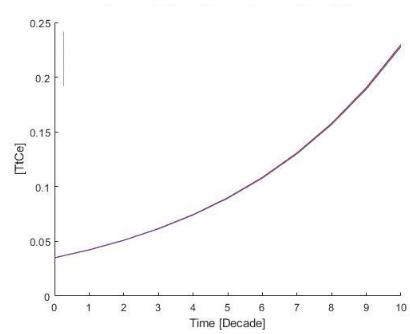


(a) $\beta = 0.999^{10}$ (Stern)   (b) $\beta = 0.985^{10}$ (Nordhaus)

Figure 5: Aggregate fossil fuel use for the GM model

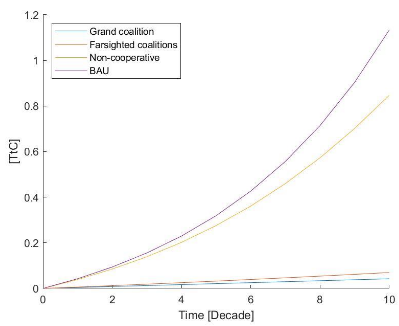


(a) $\beta = 0.999^{10}$ (Stern)

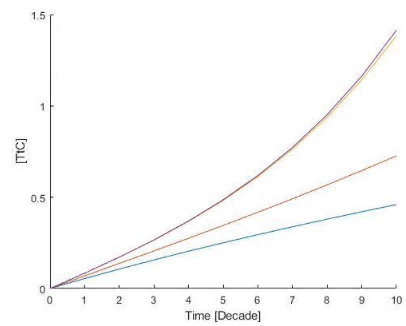


(b) $\beta = 0.985^{10}$ (Nordhaus)

Figure 6: Aggregate Renewable energy use for the GM model

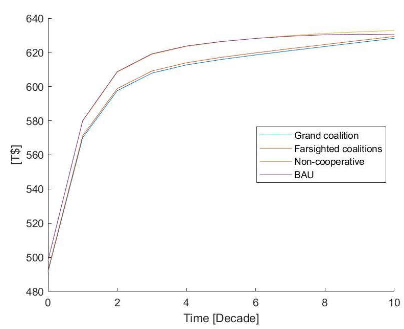


(a) $\beta = 0.999^{10}$ (Stern)

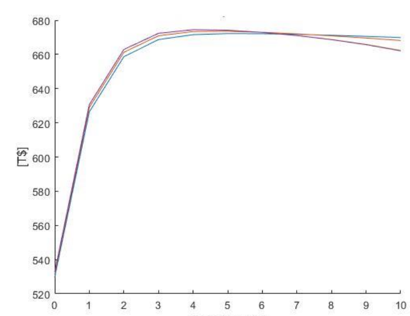


(b) $\beta = 0.985^{10}$ (Nordhaus)

Figure 7: Aggregate cumulative emissions for the GM model



(a) $\beta = 0.999^{10}$ (Stern)



(b) $\beta = 0.985^{10}$ (Nordhaus)

Figure 8: Aggregate consumption for the GM model