Self-Enforcing Environmental Federations: Unanimous Climate Coalitions in General Equilibrium

Achim Hagen[†] Christina Roolfs[‡]

July 30, 2020

Abstract

We propose a novel mechanism for coalitional policies in a climate coalition. In our proposed setup, national (unilateral) and federal (multinational) emission policies coexist. A joint federal institution, appointed by all members, sets a joint emission policy for its members and must attain Pareto improvements relative to the non-cooperative outcome. We compare this proposal to the well-established model of an international environmental agreement in which member countries cooperatively decide about a common policy that maximizes joint welfare. We relax the questionably strong assumption of cooperative joint maximization in established models of international environmental agreements and provide a framework that consistently assumes rational governments that maximize national welfare. We provide a micro-founded model with asymmetric countries that leads to moderate coalitional climate policies. These weaker climate policies are in line with the rational self-interest of countries and can help to stabilize larger coalitions. We show that they can lead to lower global emissions and higher global welfare than previously proposed models of international environmental agreements with utilitarian joint welfare maximization.

Keywords: Environmental Regulation; Federalism; Intl. Environmental Agreements
JEL classification: C72; H77; Q58; H23; D62; H87

[†]Resource Economics Group, Humboldt-Universität zu Berlin and Berlin School of Economics, Germany.
[‡]Potsdam Institute for Climate Impact Research and Technical University of Berlin, Germany.

1 Introduction

> "In some instances these [federal] institutions may contain the seeds of their own destruction, while, in other cases, they provide for their own perpetuation—for a "self-enforcing" federalism." (Oates, 2005).

Anthropogenic climate change has long been identified as one of the most challenging environmental problems worldwide. Although international negotiations have aimed at tackling this problem for decades, cooperative efforts to reduce emissions through climate policies are not even close to sufficient in keeping global warming below 1.5°C or 2°C (Climate Analytics, Ecofys, and NewClimate Institute, Climate Analytics et al.) yet. Economic theory suggests that a globally desirable carbon price that accounts for the global social cost of carbon should be implemented to correct the external effects from Greenhouse Gas emissions (GHG). However, in absence of a global institution to enforce such an optimal carbon price, countries need to cooperate and sign self-enforcing environmental agreements to jointly implement climate policies. Unfortunately, protection of the atmosphere through GHG mitigation is a global public good, which gives rise to strong free-rider incentives. Previous economic studies that have analyzed the formation of climate coalitions have assumed that all members jointly implement a climate policy that accounts for all members' climate damages and thus maximizes joint utilitarian welfare. At first sight, this might be a plausible assumption, but it implies a large difference between the (consequently strong) climate policies of member countries and those of the outsiders. This large difference deters countries that are better off as free-riders from joining the coalition, so that the size of such international environmental agreements (IEAs) tends to be small. We propose a novel policy-approach for a climate coalition that restricts the coalitional climate policies to be Pareto improving and allows member countries to complement it with a national climate policy. We exemplify the approach with a simple transfer rule that redistributes federal revenues equally among all member states. We find that this policy arrangement leads them to less ambitious climate policies than traditional models that maximize the sum of members' utilitarian welfare. This reduces the free-rider incentives, so that larger coalitions can be formed and, global emissions can be reduced, and global welfare can be increased, all in comparison with utilitarian coalitional policies.

The economic literature has analyzed the strategic aspects of the formation of IEAs for the provision of a global public good for more than 25 years, starting with seminal papers by Hoel (1992), Carraro and Siniscalco (1993), and Barrett (1994). However, the findings of this strand of literature (see Benchekroun and Van Long, 2012; Marrouch and Ray

Chaudhuri, 2016, for comprehensive overviews) are rather pessimistic and conclude that, in general, only small coalitions can be stabilized and lead to minor emission reductions and welfare improvements, in comparison to the non-cooperative outcome. Most of these studies use a two-stage game setting: countries first choose to be a member in a coalition, after which the coalition acts as one player and implements an emission reduction policy to maximize joint welfare in a non cooperative game against all outsiders [1] (see Hagen et al., 2020, for a formal overview of this approach). These studies are based on the strong assumption that countries change their behavior from being non-cooperatively maximizing their own welfare to cooperatively internalizing all externalities between members by maximizing joint welfare of the coalition. Apart from being a questionable strong assumption about the governments' behavior, this leads to a large difference in climate policies between outsiders and members of the coalition, which consequently creates strong freerider-incentives.

Extensions of this model explore different ways to improve the performance of such an IEA and include support from outsiders (Ansink et al., 2018), the use of trade sanctions against outsiders (Nordhaus, 2015; Hagen and Schneider, 2017) and the use of transfer-payments (Weikard, 2009). Finus and Maus (2008) study the case of what they call modesty in the coalitions' rationale: although all members of the coalition still maximize joint welfare, they only consider a fraction of the damages. This leads to weaker coalitional policies that are found to enable larger coalitions and higher global emission reductions. We provide an alternative approach that also leads to weaker coalitional policies but additionally allows for complementary national climate policies, has a general-equilibrium micro foundation and is rooted in the literature on fiscal federalism.

This strand of literature applies to policies in environmental federations. Federal systems differ from IEAs, as the composition of the federation is exogenously given, and federal states usually do not seek to exit the federation. The literature on environmental federalism can be distinguished into two perspectives, to which we count models with two-layered governmental regulations, but also central regulation and social planner approaches, which impose policies or allocations on states without the option to exit, like inChichilnisky and Heal (1994), Sandmo (2007), Williams (2012), d'Autumne et al. (2016). They all find the importance of a theoretically optimal transfer design. Absent of state actions, Chichilnisky and Heal (1994) show that efficient uniform multinational carbon pricing requires transfers from rich to poor countries. Williams (2012) considers coexisting

[1] Note that this second stage is sometimes modeled as two separate stages in which the coalition acts as a Stackelberg leader and the outsiders as followers.

state and federal authorities which are allowed to regulate emissions simultaneously. He compares different policy instruments and finds, among other things, that a carbon tax implemented by the federal regulator is superior to a federal quantity instrument. Roolfs et al. (2020) keep a two layered governmental structure like Williams (2012) but depart from his set-up by constraining the federal policy-making to ensure unanimity-voting requirements (Pareto improvement). They bridge between Chichilnisky and Heal (1994), Williams (2012), and voluntary public good provision as in Bergstrom and Blume, Lawrence, Varian, Hal (1986), and find that federal minimum prices and commonly used transfers exist which make all states better off relative to the decentralized case. In their model, unanimity ensuring minimum prices are endogenously determined by the richest state's utility.

The present paper is related to both Finus and Maus (2008) and Roolfs et al. (2020). We analyze the formation of a climate coalition, which we call a self-enforcing environmental federation (SEF) with an endogenously determined size. Building on a simplified version of the model developed in Roolfs et al. (2020), we set up a general equilibrium model in which a federal government sets a Pareto improving joint emission policy for its members, which is complemented by state policies. Governments regulate emissions by means of an emission price. The federal government redistributes revenues from emission pricing equally to its members, who apply both the joint federal and a local emission price. Outsiders only apply a local emission price. We compare this to the well-established IEA model in which member countries cooperatively decide about a common policy that maximizes utilitarian joint welfare and thus only use a common emission price. In a four stage game, countries first decide about their membership before the federal government sets the common federal emission price. In stage three, national governments choose their emission pricing policy, and in the final stage, the markets for capital and final goods clear. We find that an SEF can contain more members than an IEA, and performs better both in terms of global emission reductions and in terms of welfare improvements for member and outsider countries.

The paper provides several contributions to the literature. Firstly, by introducing a federal institution we propose a novel mechanism for coalitional policies in a climate coalition that improves the prospects of cooperation: a common policy for members only prescribes a lower benchmark for their effective policy and allows for an additional local policy in the member states. Thereby, we relax the questionably strong assumption of cooperative joint maximization in the coalition and provide a framework that consistently assumes rational governments that maximize national welfare. Further, we provide a general-equilibrium microfounded model with asymmetric countries that leads to modest climate policies. These weaker climate policies that are in line with the rational self-interest of countries can help to stabilize larger coalitions and lead to lower global emissions and higher global welfare than previously proposed models of IEAs with utilitarian joint welfare maximization. Last but not least, we are the first, to the best of our knowledge, to endogenize the size of an environmental federation and show that such a federation does not only lead to Pareto improvements for its members compared to the non-cooperative outcome but it can also be internally stable when members have the option to leave the federation and become free-riders.

The remainder of the paper proceeds as follows: Section 2 introduces the model. Section 3 describes the local economic agents and solves the last stage of the game in which the markets for capital and final goods clear. In the following Section 4, the national policies are determined before Section 5 analyzes international policy scenarios. Section 6 explores the membership decisions of countries and presents the results of the comparison between an SEF and an IEA. Section 7 concludes.

2 Model

We consider a world economy that consists of a set of $N = N_1 \cup N_2$ countries of two types $i \in \{1, 2\}$, with the sets of $N_1 = \{1, 2, ..., n_1\}$ type 1 countries and of $N_2 = \{1, 2, ..., n_2\}$ type 2 countries (note that countries can only be of one type $N_1 \cap N_2 = \emptyset$). The total number of countries is $n \equiv n_1 + n_2$.

The national economy of each country consists of a representative domestic firm and household. The firm produces an identical final good using capital and emissions. Households rent out their capital, which is immobile across countries, to the domestic firms and derive utility from consuming the final good but suffer damages from global emissions. Each country's government regulates local emissions by a national emission price. The revenues from national emission pricing are redistributed to the local households by the state governments.

Additionally, countries can be members of an environmental coalition, which can either be an IEA or an SEF that consists of a subset $M \subseteq N$ of all countries. Both types of countries can be members of the coalition with the number of type 1 members denoted by $m_1 \in \{0, 1, 2, ...n_1\}$ and the number of type 2 countries by $m_2 \in \{0, 1, 2, ...n_2\}$. The total number of members of the coalition is thus given by $m = m_1 + m_2$. In case of an

Figure 1: Game Sequence for an SEF.

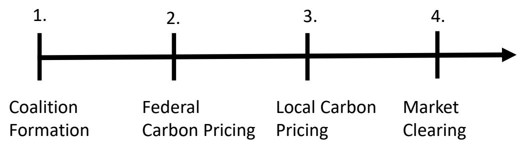
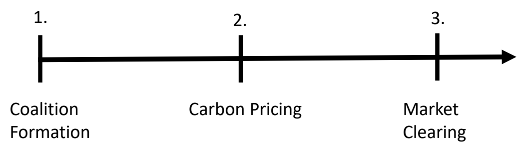


Figure 2: Game Sequence for an IEA.



SEF a federal government sets a uniform price on the emissions of all firms in member countries, ensuring that the federal emission policy leads to Pareto improvements for the members relative to their non-cooperative outcome. The federal government redistributes the revenues from federal emission pricing equally to the households of the member states, which take these as given. In the case of an IEA, all member countries jointly maximize utilitarian welfare by setting a uniform emission price for all members and, as in the case of the SEF, redistribute the revenues equally to its member states' households.

The model is set up as a four-stage game for the case of an SEF, in which (i) countries first decide to join the coalition or remain an outsider. In the second stage (ii), the federal government sets the the federal emission price, acting as a Stackelberg leader before (iii) state governments set the local emission prices simultaneously in the third stage. In the last stage of the game (iv) the production and consumption levels in all states are determined, the markets for capital and final goods clear, the levels of local and global emissions result and the federal government redistributes the revenues from emission pricing to its members. Figure 1 depicts the game sequence in case of an SEF. In the case of an IEA, stages (i) and (iv) are the same as with an SEF, whereas in stage (ii) coalition members choose their joint emission price simultaneously with the outsider countries and stage (iii) does not exist, so the game considered is a three stage game. The sequence of the game for in case of an IEA is depicted in Figure 2. The model is solved by backward induction, solving the last stage of the game first. The following sections refer to the four-stage game for the formation of an SEF, the case of an IEA is covered by Section 5.3.

3 Economic Agents

3.1 Firms

The representative firm located in country i uses capital K_i and emissions E_i with a constant returns to scale Cobb-Douglas-technology to produce the final good Y_i, which is taken as numéraire good. As a price-taker, the firm chooses K_i and E_i to maximize its profits,

$$\max_{K_i,E_i} \left\{ (Y_i - r_i K_i - \rho_i E_i) \,\middle|\, Y_i = A K_i^{\alpha_K} E_i^{\alpha_E} \right\}, \tag{1}$$

with the share parameters of capital and emissions $\alpha_K + \alpha_E = 1$. Parameter $A > 0$ measures efficiency. For country i, r_i represents the rental rate of capital and ρ_i the effective national per unit costs of emissions. The effective national emission price is the sum of the national price p_i and the federal price P, i.e. $\rho_i \equiv p_i + P$. Profit maximization implies setting the marginal product of each factor equal to its price. The marginal cost (mc_i) of producing the final good Y_i is given by $mc_i = r_i^{\alpha_K} \rho_i^{\alpha_E} / \left(\alpha_K^{\alpha_K} \alpha_E^{\alpha_E} A \right)$. Zero profits imply $mc_i = 1$. Conditional demand for capital and emissions follows from the firm's first order condition and reads $K_i = \alpha_K Y_i / r_i$ and $E_i = \alpha_E Y_i / \rho_i$.

3.2 Households

One representative household lives in each country $i = 1, ..., n$ and has capital endowment $\overline{K}_i$. Each household derives utility from final good consumption c_i and suffers damages from global emissions $E = \sum_i^n E_i$.

Utility is an additively separable function. Let $u_x^i \equiv \partial u^i / \partial x$ and $u_{xx}^i \equiv \partial^2 u^i / \partial x^2$. We suppose that $u_{c_i}^i > 0$, $u_{c_i c_i}^i \leq 0$, $u_E^i < 0$, and $u_{EE}^i \leq 0$. To be able to obtain closed form solutions, we assume that the utility function of household i is given by

$$u^i(c_i, E) = c_i - gE. \tag{2}$$

The parameter g denotes constant marginal damages from emissions[2]. Households take as given all prices, aggregate emissions and national and federal governments' transfers and policies. Thus, utility maximization of household i implies choosing its optimal level

[2]As shown in Roolfs et al. (2018) the assumption of linear consumption utility maintains convexity to solve for optimal emission policies with the indirect utility function.

of consumption subject to its income which reads

$$c_i = \begin{cases} r_i\overline{K}_i + p_iE_i & \text{for outsiders} \\ r_i\overline{K}_i + p_iE_i + \pi_iPE^F & \text{for members of EF.} \end{cases} \tag{3}$$

Thereby, the income of household i always consists of the return to capital endowment $r_i\overline{K}_i$ and a transfer of the national emission price receipts p_iE_i. If country i is a member of the SEF, household i receives an additional federal transfer, π_iPE^F. The federal transfer depends on the transfer rule π_i, the federal emission price P and federal emissions $E^F \equiv \sum_m E_j$. In order to illustrate the model and show how it can be solved, we use a particularly simple transfer rule in the following, in which each member receives an equal share of the federal revenues, $\pi_i = 1/m$. We assume that a country of type 2 owns a lager capital endowment than a type 1 country, expressed by

$$\overline{K}_1 = s_1\overline{K} \tag{4}$$

$$\overline{K}_2 = s_2\overline{K} \tag{5}$$

with s_1 normalized to $s_1 = 1$ and $s_2 > s_1$ such that a type 2 country is richer than a type 1 country.

3.3 Market Clearing

Market clearing in final goods implies that aggregate consumption equals aggregate production,

$$\sum_{i=1}^{n} c_i = \sum_{i=1}^{n} Y_i. \tag{6}$$

Capital markets clear when capital demand K_i equals the capital endowment within country i, i.e. $K_i = \overline{K}_i$. All functions that take into account the solutions to consumers and firms' problems and market clearing are represented with **bold** letters. Let $\rho_i \equiv p_1 + P$ if the country is a member of the SEF, and $\rho_i \equiv p_1$ if the country is an outsider of the SEF. Further, let ρ denote the vector of all effective national prices, i.e. $\rho \equiv (p_1+P, ..., p_m+P, p_{m+1}, ..., p_n)$

Lemma 1. *a) All decision variables of national economic agents are fully determined as functions of the effective national policies ρ_i.*
b) Emission levels are given by $\mathbf{E}_i(\rho_i) = \left(\frac{\alpha_E A}{\rho_i}\right)^{\frac{1}{\alpha_K}} \overline{K}_i$, the federal emission level is $\mathbf{E}^F(\rho) \equiv$

$\sum_m \boldsymbol{E}_j\left(\rho_j\right)$ *and global emissions are* $\boldsymbol{E}(\rho) \equiv \sum_n \boldsymbol{E}_j\left(\rho_j\right)$.

Proof. See Appendix A. □

Using Lemma 1 and the zero profit condition, we can rewrite equation (3) in the context of national production as

$$\boldsymbol{c}_i\left(\rho_i\right) = \begin{cases} \boldsymbol{Y}_i\left(\rho_i\right) & \text{for outsiders} \\ \boldsymbol{Y}_i\left(\rho_i\right) + \left(\frac{1}{m}\sum_m \boldsymbol{E}_j\left(\rho_i\right) - \boldsymbol{E}_i\left(\rho_i\right)\right)P & \text{for members of SEF.} \end{cases} \tag{7}$$

While income for outsiders depends only on local production Y_i, members of the SEF receive an increase or decrease in income depending on whether their emission level is below or above the SEF's average emission level $\frac{1}{m}\sum_m E_j$.

Now that we have solved all variables exclusively dependent on national and federal policy, we will use these expressions in the next section to analyze the policies of the states and the federation.

4 National Policies

In stage three of the game, governments take as given the size and composition of the federation as well as the federal emission price. With the results from Lemma 1 we can denote utility as a function of the policies in the country as $\boldsymbol{u}^i(p_i, P) \equiv u^i(\boldsymbol{c}_i(\rho_i), \boldsymbol{E}(\rho))$. All national governments maximize their households' utilities by simultaneously choosing an optimal emission price p_i, depending on its membership status. The objective of the government of country i reads

$$\max_{p_i} \boldsymbol{u}^i(p_i, P) \text{ given } p_j \ \forall j \neq i \text{ and } P. \tag{8}$$

Consider Lemma 1 to see that $\frac{\partial \boldsymbol{E}}{\partial p_i} = \frac{\partial \boldsymbol{E}_i}{\partial p_i}$. Therefore, the first order condition of this problem is given by

$$\frac{\partial \boldsymbol{u}^i}{\partial p_i} = u^i_{c_i}\frac{\partial \boldsymbol{c}_i}{\partial p_i} + u^i_E \frac{\partial \boldsymbol{E_i}}{\partial p_i} = 0. \tag{9}$$

From here we can formulate the following lemma:

Lemma 2. *The emission prices chosen by the national governments are*

$$p_i = \begin{cases} g \equiv p_i^O & \textit{for outsiders} \\ g - \frac{P}{m} \equiv \boldsymbol{p}_i^M(P) & \textit{for members of SEF.} \end{cases} \tag{10}$$

Proof. From equation (2) we get $u^i_{c_i} = 1$ and $u^i_E = -g$. Substitution into (9) yields $\frac{\partial \boldsymbol{c}_i}{\partial p_i} = g \frac{\partial \boldsymbol{E}_i}{\partial p_i}$. Consider equation (7) to get $\frac{\partial \boldsymbol{c}_i}{\partial p_i} = \frac{\partial \boldsymbol{E}_i}{\partial p_i} p_i + \frac{P}{m} \frac{\partial E_i}{\partial p_i}$, substitute this into the previous equation and solve for p_i. □

Since parameter g denotes constant marginal damages from emissions it can be interpreted as the national social cost of carbon. Lemma 2 shows that outsiders have the dominant strategy to set their emission price equal to their national social cost of carbon. Suppose that the federal price P is positive[3], then Lemma 2 states that the local emission price is lower in a member country than in an outsider country, $\boldsymbol{p}_i^M(P) < p_i^O$. Furthermore, member countries of an SEF of a given size have dominant strategies, i.e. the choice of the local emission price in member countries does neither depend on the local emission prices in other member countries nor on the outsiders' prices.

5 International Policy Scenarios

On the international level, we study different policy scenarios. Firstly, we consider a world without any cooperative efforts between countries to mitigate carbon emissions (non-cooperative solution). The second relevant reference scenario is the social optimum, in which a benevolent planer maximizes utilitarian global welfare. Against the backdrop of these two reference scenarios we then continue to focus on the cases with an IEA, in which member countries maximize joint welfare by means of a uniform carbon price, and the case with an SEF, in which the federal government sets a Pareto improving federal carbon price that is complemented by national prices as analyzed in the previous section.

5.1 Non-Cooperative Solution

Our first reference scenario represents the lower benchmark in which we consider a world without cooperation between countries on climate policies. Only national governments are present and thus carry out their maximization problem. Consequently, each country

[3]If P exists, it can actually only be positive. See Roolfs et al. (2020) for the proof.

sets its individually optimal national carbon price exactly as it would do as an outsider of either an IEA or an SEF, so all national prices are given by

$$\tilde{p} \equiv p_i = g = p^O \tag{11}$$

We use a tilde to identify the resulting equilibrium levels. The decentralized utility levels are $\tilde{u}^i$ for all i which we use as the Pareto improving benchmark for the federal scenario.

5.2 Centralized Policy - Social Optimum

Our second reference scenario represents the upper benchmark and we follow Chichilnisky and Heal (1994); Sheeran (2006). We first derive the optimality conditions from a social planner perspective which we then apply to a centralized government using a uniform emission price.

The social planner maximizes global welfare, given by the objective function

$$\max_{c_i, y_i} \sum_n \lambda_i u^i(c_i, E)$$

subject to the market clearing condition $\sum_n y_i = \sum_n c_i$. Since country i's private good production Y_i uses emissions E_i and thus impacts the emission externality E, we define this relationship by $E = E(Y_1, ..., Y_n) \equiv \sum_n E_i(Y_i)$. The Lagrangian reads

$$L(c_1, ..., c_n, Y_1, ..., Y_n, \theta) = \sum_n \lambda_i u^i(c_i, E(Y_1, ..., Y_n)) + \theta \left(\sum_n (Y_i - c_i) \right).$$

The $2n + 1$ first order conditions for a maximum are obtained by differentiating the Lagrangian with respect to each c_i, each Y_i, and the Lagrangian multiplier θ. We get

$$\lambda_i \frac{\partial u^i}{\partial c_i} = \lambda_i = \theta \quad \forall\ i = 1, ..., n, \tag{12}$$

$$\left(-\sum_n \lambda_i \frac{\partial u^i}{\partial E} \right) \frac{\partial E_i}{\partial Y_i} = \theta \quad \forall\ i = 1, ..., n, \tag{13}$$

and

$$\sum_n (y_i - c_i) = 0.$$

With equation (2) follows for (12) that $\lambda_i = \theta$ such that weights must be equal for all countries, $\lambda_i = 1/n$. For (13) follows $g/\partial Y_i/\partial E_i = \theta$. Equating (12) and (13) with these simplifications allows to write

$$ng = \frac{\partial Y_i}{\partial E_i} \; \forall \; i = 1, ..., n. \tag{14}$$

Suppose there is a central government that wants to obtain the social optimum by using a uniform price ρ_{GLO}. Since firms set $\partial Y_i/\partial E_i = \rho_{GLO}$ it follows from equation (14) that

$$\rho_{GLO} = ng. \tag{15}$$

5.3 Policies in an IEA

In case of an IEA we follow the widely used assumption in the literature of IEAs that members of the coalition set their policies to jointly maximize their utilitarian welfare whereas outsiders set their policies non-cooperatively to maximize their own benefits. The policies in an IEA are depicted in Figure 3. IEA members apply only the joint coalitional emission price and do not have an additional local price on emissions. From the maximization problem $max_\rho \sum_m u^i(c_i, E)$ we get the coalitional price of an IEA member whereas we know the national policies of outsiders from (10). The following lemma results.

Lemma 3. *The effective policies in an IEA are given by*

$$\rho_i = \begin{cases} g = p_i^O & \text{for outsiders} \\ (m_1 + m_2)g \equiv \rho^{IEA}(m_1, m_2) & \text{for members of IEA.} \end{cases} \tag{16}$$

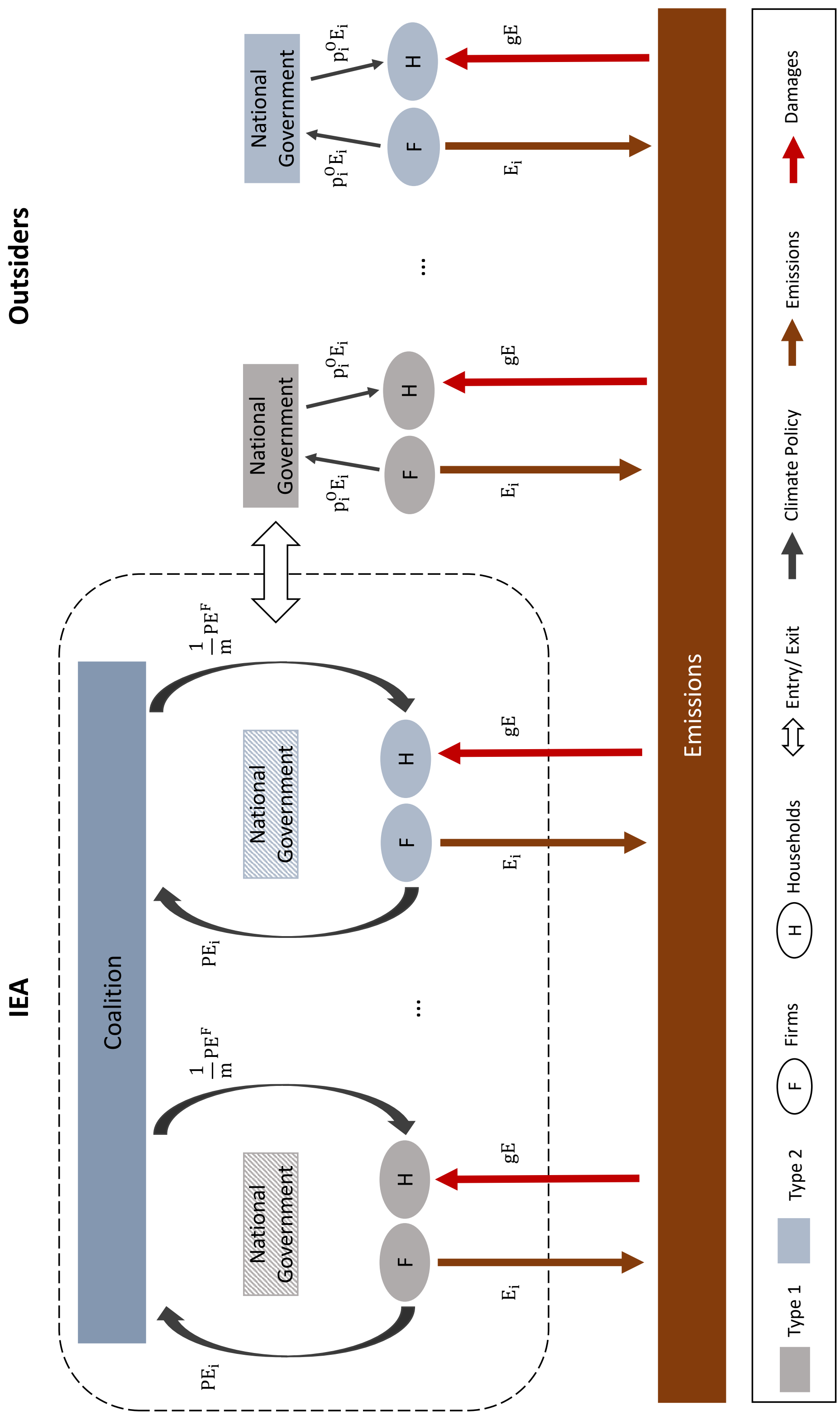
Proof. See Appendix B. □

Equation (16)shows that the state policy remains the same for outsiders of an IEA as in the decentralized case. Consistent with the case of a federation we assume that coalitional revenues are redistributed equally to all member countries[4].

Comparing (11),(15), and (16) results in the following corollary:

[4]This is relevant for the welfare of single members and therefore for the membership decision. However, as members maximize the utilitarian sum of their welfare, this distributional assumption is irrelevant for the choice of the coalitional price.

Figure 3: Institutional Setup and Decisions in an IEA.



Corollary 1. *If $m_1 + m_2 < n$, the level of the international climate policy of an IEA lies below the social optimum but above the decentralized climate policy. If no non-trivial coalition with $m \geq 2$ exists we are in the decentralized case whereas the grand coalition $m = n$ achieves the social optimum.*

5.4 Policies in an SEF

Having established the reference scenarios for comparison, we now turn to the policies in an SEF. In this case, all countries that have chosen to become a member of the coalition in Stage 1 establish a federal institution, which we call federal government. This institution raises an emission price P that is uniformly applied in all member countries. This price is required to be unanimously accepted by all member states so we assume that relative to the non-cooperative solution without any coalition, this uniform federal price must guarantee that at least one member country is better off while no other country is made worse off (Pareto improvement). This setting has been introduced by Roolfs et al. (2020)but absent of endogenized membership decision of states. As in the case of an IEA, we assume that outsiders of the coalition only apply a national emission price to maximize their welfare non-cooperatively. The policies in an SEF are depicted in Figure 4.

The federal government knows the problem of all firms, households, state governments and market clearing conditions. It acts as a Stackelberg Leader for the entire federal economy. Using the notation of $\boldsymbol{u}^i(p_i, P) \equiv u^i(\boldsymbol{c}_i(\rho_i), \boldsymbol{E}(\rho))$ and $\rho_i = p_i + P$ the federal objective is formalized as

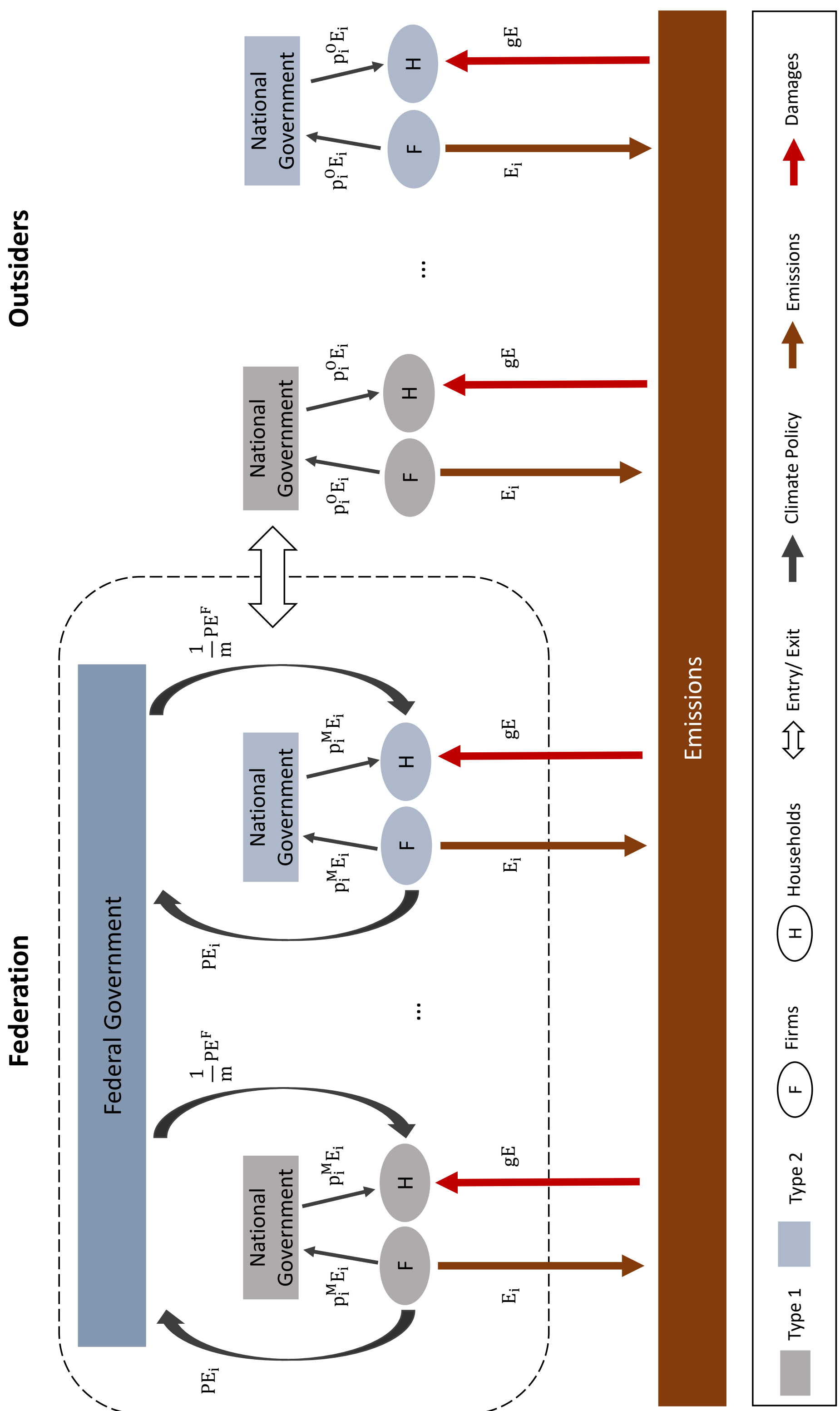
$$\max_P \left\{ \boldsymbol{u}^i(\boldsymbol{p}, P)) | \boldsymbol{u}^j(\boldsymbol{p}, P) \geq \tilde{u}_j \forall j \neq i \right\}, \tag{17}$$

where $\boldsymbol{p} = (\boldsymbol{p^M}_1(P), ..., \boldsymbol{p}^M_m(P))$. The Lagrangian function related to objective (17) reads

$$\mathscr{L}^i(P, \lambda) = \boldsymbol{u}^i(\boldsymbol{p}, P) + \sum\nolimits_{j \neq i} \lambda_j \left(\boldsymbol{u}^j(\boldsymbol{p}, P) - \tilde{u}_j \right) \tag{18}$$

where $\lambda_{j \neq i}$ are the $m-1$ Karush-Kuhn-Tucker multipliers related to the Pareto improvement constraint in objective (17). We follow the proof of Roolfs et al. (2020) and show in Appendix C that if

Figure 4: Institutional Setup and Decisions in an SEF.



$$\sigma_2 = \frac{\overline{K}_2}{\overline{K}_F} = \frac{s_2\overline{K}}{m_1 s_1 \overline{K} + m_2 s_2 \overline{K}} < \frac{1}{m}\frac{m - \alpha_E}{2 - \alpha_E - \frac{1}{m}}$$

then there exist some P that satisfy $\boldsymbol{u}^j > \tilde{u}_j$ for all $j \neq i$. This implies that $\lambda_{j\ \forall j \neq i} = 0$. Thus, by using equation (7), the federal government's first-order conditions are reduced to

$$-\frac{u^i_E}{u^i_{c_i}} = \left(\frac{d\boldsymbol{Y}^i}{dP} + \frac{d(s_i \boldsymbol{E}^F - \boldsymbol{E}^F_i)P}{dP}\right) / \frac{d\boldsymbol{E}^F}{dP} = 0 \tag{19}$$

Let P^1 and P^2 denote the price that maximizes the utility of type 1 and type 2 countries, respectively. The solution to equation (19) for type 1 and type two countries rank such that $P^2 < P^1$ (cf.C). It means that P^2 is the smallest (Pareto dominant) optimal solution for the federal objective: suppose the federal government increases the federal price starting from zero until the first country vetoes against it, this will be at P^2 and the veto will come from the richer type 2 countries. The formulation of equation (17) is functionally equivalent to maximizing the weighted sum of a social welfare function (Krepps, 1990).

In the following we will use this P^2 as the conservative perspective of an agreeable but Pareto dominant federal price which is in line with veto power of all members of the federation. We define $P^{\min} \equiv P^2$. It can now be shown that

Lemma 4. *The uniform federal price is given by*

$$\boldsymbol{P}^{\min}(m_1, m_2) = \boldsymbol{P}^2(m_1, m_2) = \frac{m}{m-1}\left(\theta_2 + \frac{\sigma_2 - 1}{m}\right)\left(\frac{g}{\chi_2 - \theta_2 - \sigma_2 \frac{1}{m}}\right) \tag{20}$$

where $\chi_2 = \left(\frac{1}{m} - \sigma_2\right) + 1$, $\theta_2 = \chi_2 - \alpha_K \sigma_2 - \frac{\alpha_E}{m}$ *and* $\sigma_2 = \frac{s_2}{m_1 s_1 + m_2 s_2} < \frac{1}{m}\frac{m - \alpha_E}{2 - \alpha_E - \frac{1}{m}}$.

Proof. See Appendix C. □

Definition 1. *If a solution to the federal problem exists, the effective national emission prices with the federal minimum price are*

$$\rho_i = \begin{cases} g = p^O_i & \text{for outsiders} \\ \boldsymbol{p}^M_i(P) + \boldsymbol{P}^{\min} \equiv \rho^M_i & \text{for members of EF.} \end{cases} \tag{21}$$

Using (10) allows to express the effective emission price for a member of the SEF solely depending on the number of member states $m = m_i + m_2$. We use bold letters to indicate these functional forms. We get

$$\boldsymbol{\rho}_i^M(m_i, m_2) \quad = \quad g + \boldsymbol{P}^{\min}(m_1, m_2)(1 - \frac{1}{m_1 + m_2}). \tag{22}$$

Using (20) we can, after some algebraic manipulations, write this as

$$\boldsymbol{\rho}_i^M(m_1, m_2) = -\frac{g(m_1 + m_2)(m_1 s_1 + (-1 + m_2)s_2)}{s_2 - m_2 s_2 + m_1(-\alpha_E s_1 - \alpha_K s_2)}. \tag{23}$$

Comparing the effective prices for members and outsiders of a federation with those of an IEA results in the following proposition.

Proposition 1. *The effective prices for a coalition member are higher in an IEA than in an SEF of the same size. Outsiders choose the same price in both scenarios.*

Proof. See Appendix D. □

Further, we find

Proposition 2. *If all members receive the same share of multilateral emission pricing revenues then a) global emissions are lower in an IEA than in an SEF of the same size, and b) global welfare in an IEA is higher than in an SEF of the same size.*

Proof. See Appendix E.

Although we analyze a world economy with asymmetric countries that differ in their capital endowments, we cover also the case of symmetric countries[5]. In this case, we can state the following result. □

Proposition 3. *If countries are symmetric, the federal price in an SEF is equal to the uniform emission price in an IEA and countries do not apply a national emission price. The settings of an IEA and an SEF are equivalent then.*

Proof. Using (16) and (23) we see that the prices are identical with if $s_1 = s_2$ and $m_1 = m_2$. From (10) we see that member countries do not apply a national emission price in that case. □

[5]This assumption of symmetric countries is often made in the literature on IEAs, including Finus and Maus (2008)

Summing up, we see that an SEF of the same size leads to lower global emissions reduction than an IEA because every member has lower effective emission prices in the SEF. A naïve interpretation of this result would be that from a global perspective, IEAs are preferable to SEFs. However, the overall effects on emissions and welfare depend both on the ambition and on the size of the formed coalition. Finus and Maus (2008) have shown that less-ambitious goals of an IEA can increase the size of the coalition. It thus remains to be shown if the lower effective emission prices for members of an SEF can incentivize enough countries to join and thereby outperform an IEA.

6 Membership

In the first stage of the game, all countries simultaneously decide about their membership in the coalition. They know how the size and composition of the coalition will influence their effective emission prices and how these prices in turn result in market equilibrium and emission levels. To solve for the size of a stable coalition we apply the concepts of internal and external stability. These concepts were first developed by D'Aspremont et al. (1983) in the context of cartel formation and are standard in the economic literature on IEAs since the seminal papers of Hoel (1992), Carraro and Siniscalco (1993), and Barrett (1994). Internal stability requires that no member has an incentive to leave the coalition whereas external stability requires that no outsider has an incentive to join. Formally the stability conditions are given by

$$\boldsymbol{u}^{iM}(m_i^*, m_j^*) \geq \boldsymbol{u}^{iO}(m_i^* - 1, m_j^*) \forall i, j \in \{1, 2\}\ i \neq j \text{ (internal stability)} \quad (24)$$

$$\boldsymbol{u}^{iO}(m_i^*, m_j^*) > \boldsymbol{u}^{iO}(m_i^* + 1, m_j^*) \forall i, j \in \{1, 2\}\ i \neq j \text{ (external stability)}, \quad (25)$$

where m_i^* denotes the number of members of type i in a stable coalition, $\boldsymbol{u}^{iM}(m_i^*, m_j^*)$ the welfare of a member country of this coalition, and $\boldsymbol{u}^{iO}(m_i^*, m_j^*)$ the welfare of an outsider country. Following Hoel and Schneider (1997), we write the stability function

$$\boldsymbol{\Phi_i}(m_i, m_j) = \boldsymbol{u}^{iM}(m_i, m_j) - \boldsymbol{u}^{iO}(m_i - 1, m_j) \quad (26)$$

which denotes the change in the payoff for a member country of type i that it would face in the case it leaves the coalition. Using the concept of potential internal stability as defined in Carraro et al. (2006) coalition $S \subseteq N$ is potentially internally stable iff $\sum_m \boldsymbol{\Phi}_i(m_i^*, m_j^*) \geq 0$, which is if the sum of its members' payoffs are larger then the sum of their outside option payoffs[6].

We know from (10) that outsiders of an SEF have dominant strategies, and we can thus write the stability function for country i as

$$\begin{aligned}&\boldsymbol{\Phi}_i(m_i, m_j) = \boldsymbol{Y}_i\left(\boldsymbol{\rho}_i^M(m_i, m_j)\right) + \frac{1}{m}\mathbf{P}^{min}(m_i, m_j)\boldsymbol{E}^F(m_i, m_j)\\ &-\boldsymbol{P}(m_i, m_j)\boldsymbol{E}_i\left(\boldsymbol{\rho}_i^M(m_i, m_j)\right) - g\boldsymbol{E}^F(m_i, m_j) - \boldsymbol{Y}_i(p_i^O)\\ &+g\boldsymbol{E}^F(m_i - 1, m_j) + g\boldsymbol{E}_i(p_i^O) \end{aligned} \tag{27}$$

with $\boldsymbol{E}^F(m_i, m_j) = \sum_m \left(\frac{\alpha_E A}{\boldsymbol{p}_l^F(m_i, m_j)}\right)^{\frac{1}{\alpha_K}} \overline{K}_i$. Similarly, the stability function for country i in the case of an IEA is

$$\begin{aligned}&\boldsymbol{\Phi}_{\boldsymbol{i}}(m_i, m_j) = \boldsymbol{Y}_i\left(\boldsymbol{\rho}^{IEA}(m_1, m_2)\right) + \frac{1}{m}\boldsymbol{\rho}^{IEA}(m_1, m_2)\boldsymbol{E}^{IEA}(m_i, m_j)\\ &-\boldsymbol{\rho}^{IEA}(m_1, m_2)\boldsymbol{E}_i\left(\boldsymbol{\rho}^{IEA}(m_1, m_2)\right) - g\boldsymbol{E}^{IEA}(m_i, m_j) - \boldsymbol{Y}_i(p_i^O)\\ &+g\boldsymbol{E}^{IEA}(m_i - 1, m_j) + g\boldsymbol{E}_{\boldsymbol{i}}(p_i^O) \end{aligned} \tag{28}$$

To proceed with insights on coalition stability, we will resort to numerical examples in the following[7]. We analyze coalition stability of an SEF and compare the size of the SEF as well as the welfare and emission values with those of an IEA. Table 1 shows the parameter values of our numerical examples.

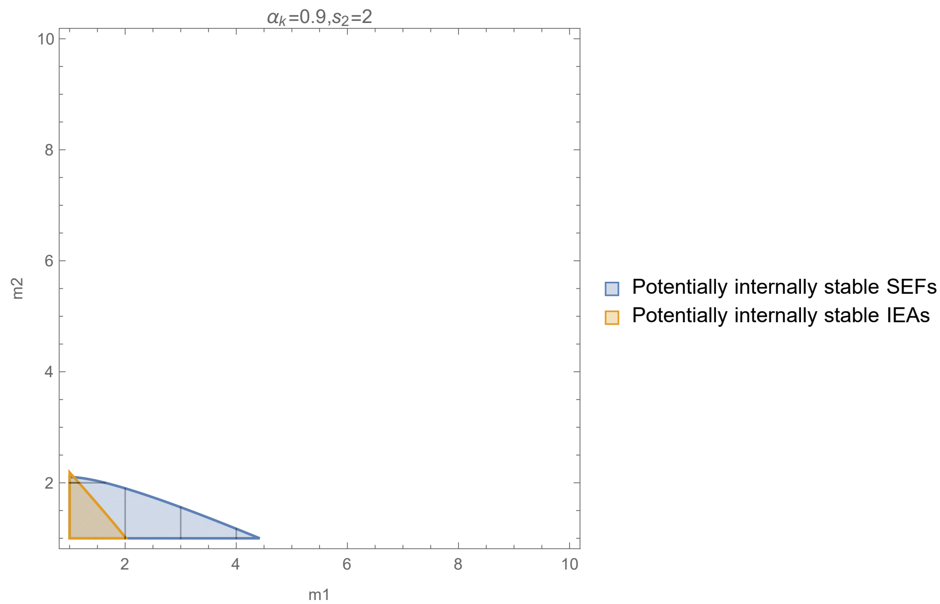
Table 1: Assumptions for numerical examples.

Example	N_1	N_2	α_K	A	g	$\bar{K}$	s_2
1	10	10	0.9	1	1	10	2
2	10	10	0.9	1	1	10	3
3	10	10	0.9	1	1	10	5

[6]For further applications of the concept of potential internal stability see e.g. Pavlova and de Zeeuw (2013)and Lessmann et al. (2015). This stability concept requires that a transfer scheme exists which can redistribute payoffs amongst members of the coalition such that no member has an incentive to leave the coalition. This is a necessary (but not sufficient) condition for coalition stability (see Carraro et al. (2006)).

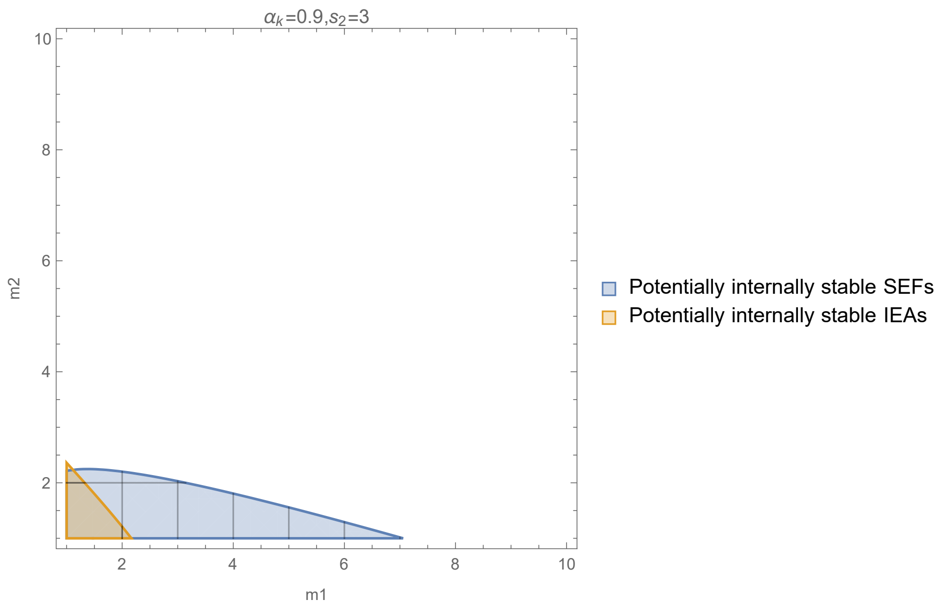
[7]It is recognized in the literature on IEAs that analytical complexity of the models make the use of numerical examples necessary to determine coalition stability, see e.g. the seminal paper of Barrett (1994) and Eichner and Pethig (2013).

Figure 5: Potential Internal Stability in Example 1.



Coalition stability is examined in Figures 5-7, which depict the potential internally stable coalitions for both cases of an IEA and an SEF. We see that several combinations of member countries are potentially internally stable. We are only interested in those coalitions on the frontier of the depicted areas, which are the largest possible coalitions. Coalitions below that frontier are also potentially internally stable, but external stability is not given, as outsiders can still gain from joining. Further, although we depict the results continously, we restrict our interest to the solutions with $m_i \in \mathbb{N}^+ \forall i$, because it follows from Proposition 3 that a comparison of an IEA and an SEF is only of interest for coalitions with assymetric members, i.e. at least one member of each type. Figure 5 shows the results for Example 1, which indicate that in the case of an IEA, two members of one type can form a potentially internally stable coalition with one country of the other type. In the case of an SEF, both the combinations of two type 2 countries with one type 1 country and of four type one with one type 2 countries can be stable. The results for Examples 2 and 3, as depicted in Figures 6 and 7, show that the possible coalitions in the case of an IEA are the same in these Examples, whereas the size of potentially internally stable SEFs increases with the inequality in capital endowments. With a richer country having three times as much capital as a poor country, in Example 2, an SEF can either contain three rich type 2 countries and two poorer type 1 countries or seven members of type 1 and one type 2 member. In Example 3 (Figure 7) this is increased to either 6 members of type 1

Figure 6: Potential Internal Stability in Example 2



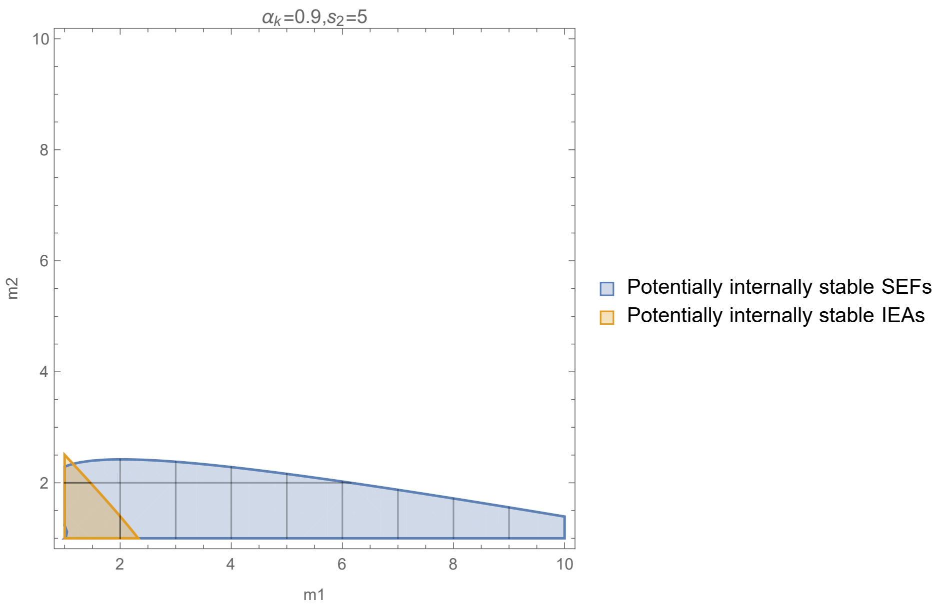
with 2 members of type two or 10 members of type 1 together with 1 type 2 member.

Table 2: Results for stable coalitions in SEF and IEA.

SEF					IEA				
Example	m_1	m_2	E	$\sum_i^n u_i.$	*Example*	m_1	m_2	E	$\sum_i^n u_i.$
1	4	1	20.1677	-175.81	1	2	1	21.0446	-192.171
1	1	2	20.8853	-188.862	1	1	2	20.4987	-182.144
2	7	1	25.6395	-211.609	2	2	1	28.2414	-259.571
2	3	2	26.3802	-225.002	2	1	2	27.1497	-239.517
3	10	1	39.2502	-331.169	3	2	1	42.6350	-394.370
3	6	2	38.1774	-311.932	3	1	2	40.4517	-354.262

Let us now turn to the resulting global levels of emissions and welfare, summarized in Table 2. From Example 1 we see that in the case where both an IEA and an SEF contain the same set of members ($m_1 = 1, m_2 = 2$), global emission levels are higher (lower) and global welfare lower (higher) in the case of an SEF (IEA), which is formally shown in Proposition 2. In all other cases, we see that the larger numbers of members in the cases with an SEF (that has less stringent climate policies than an IEA) helps to reduce global emissions below the level that stable IEAs can achieve and therefore outperfom IEAs in terms of global welfare.

Figure 7: Potential Internal Stability in Example 3.



7 Conclusion

International cooperation to combat climate change and prevent dangerous human interference with the climate system is an extremely challenging task. Due to the public good character of greenhouse gas mitigation, there are strong free-rider incentives for individual states. In the absence of a global authority to enforce first-best climate policies, these incentives both prevent countries from effective unilateral action, and make it difficult to form self-enforcing climate coalitions.

We propose a novel policy-approach for a climate coalition by supposing a federation-like institution. This institution imposes a joint emission price on members, which has to be unanimously accepted by all members. Member states, in addition, complement this federal common policy with their own national emission price. Federal revenues from emissions pricing are distributed equally to all members.

We show that this institutional setup has the potential to increase global efforts to reduce emissions and improves global welfare in comparison to traditional approaches of IEAs, which only set a uniform joint emission price that fully accounts for all members damages in a Pigouvian manner. The mechanism behind this result is the following: The federal coalition structure reduces the difference between the policy-stringency of members and outsiders and thus reduces the incentives to leave the coalition. This enables larger coalitions that overcompensate the weaker policies and improve global welfare.

A central assumption in our model is that federal revenues are distributed equally to all members, which implies that rich countries who produce more (final goods and emissions) than the average member are net donors, while poor member countries are net recipients.

There are a number of ways in which we could extend the current analysis. First, we could analyze other well-established distribution rules and examine how they would affect the size of the coalition and welfare an (see Kverndokk, 2018, for a recent overview of such rules). A second approach would involve calibrated simulations with a computable general equilibrium model. However, the large number of possible coalition structures makes the analysis of coalition formation in such a model a challenging task (c.f. Nordhaus, 2015). Third, one could analyze country differences that go beyond wealth, such as differences in population size or abatement technology. We plan to address some of these issues.

References

Ansink, E., H.-P. Weikard, and C. Withagen (2018). International environmental agreements with support. *Journal of Environmental Economics and Management*.

Barrett, S. (1994). Self-Enforcing International Environmental Agreements. *Oxford Economic Papers 46*, 878–894.

Benchekroun, H. and N. Van Long (2012). Collaborative environmental management: A review of the literature. *International Game Theory Review 14*(04), 1240002.

Bergstrom, T. and Blume, Lawrence, Varian, Hal (1986). On the private provision of public goods. *Journal of Public Economics 29*, 25–49.

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.

Chichilnisky, G. and G. Heal (1994). Who should abate carbon emissions? *Economics Letters 44*, 443–449.

Climate Analytics, Ecofys, and NewClimate Institute. Temperatures - addressing global warming. https://climateactiontracker.org/global/temperatures/.

D'Aspremont, C., A. Jacquemin, J. J. Gabszewicz, and J. A. Weymark (1983). On the Stability of Collusive Price Leadership. *The Canadian Journal of Economics / Revue canadienne d'Economique 16*(1), 17–25.

d'Autumne, A., K. Schubert, and C. A. Withagen (2016). Should the carbon price be the same in all countries? *Jounal of Public Economic Theory*.

Eichner, T. and R. Pethig (2013). Self-enforcing environmental agreements and international trade. *Journal of Public Economics 102*, 37–50.

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

Hagen, A., P. v. Mouche, and H.-P. Weikard (2020). The two-stage game approach to coalition formation: Where we stand and ways to go. *Games 11*(1), 3.

Hagen, A. and J. Schneider (2017). Boon or bane? trade sanctions and the stability of international environmental agreements. *Oldenburg Discussion Papers in Economics* (V–403–17).

Hoel, M. (1992). International environment conventions: The case of uniform reductions of emissions. *Environmental & Resource Economics 2*(2), 141–159.

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

Krepps, D. (1990). *A Course in Microeconomic Theory*. Princeton NY: Princeton University Press.

Kverndokk, S. (2018). Climate Policies, Distributional Effects and Transfers Between Rich and Poor Countries. *International Review of Environmental and Resource Economics 12*(2-3), 129–176.

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

Marrouch, W. and A. Ray Chaudhuri (2016). International Environmental Agreements: Doomed to Fail or Destined to Succeed? A Review of the Literature. *International Review of Environmental and Resource Economics 9*(3-4), 245–319.

Nordhaus, W. (2015). Climate Clubs: Overcoming Free-riding in International Climate Policy. *The American Economic Review 105*(4).

Oates, W. (2005). Toward A Second-Generation Theory of Fiscal Federalism. *International Tax and Public Finance 12*(4), 349–373.

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

Roolfs, C., B. Gaitan, and O. Edenhofer (2020). Make or brake — Rich states in voluntary federal emission pricing. *under review. Working paper available at SSRN*.

Sandmo, A. (2007). Global Public Economics: Public Goods and Externalities: mis en ligne le 17 octobre 2007, consulté le 30 septembre 2016. *Économie publique/Public economics [En ligne] 18-19*.

Sheeran, K. A. (2006). Who Should Abate Carbon Emissions? A Note. *Environmental and Resource Economics 35*(2), 89–98.

Weikard, H.-P. (2009). Cartel stability under an optimal sharing rule. *The Manchester School 77*(5), 575–593.

Williams, R. C. (2012). Growing state–federal conflicts in environmental policy: The role of market-based regulation. *Journal of Public Economics 96*(11-12), 1092–1099.

Appendix

A Proof of Lemma 1

For the derivation of this result see also Roolfs et al. (2018). Using marginal cost and the zero profit condition and solving for r_i, we obtain

$$\boldsymbol{r}_i(\rho_i) = \left(\frac{\alpha_K^{\alpha_K}\alpha_E^{\alpha_E}A}{\rho_i^{\alpha_E}}\right)^{\frac{1}{\alpha_K}}. \tag{29}$$

From conditional capital demand and using (29) and capital market clearing it follows that

$$\boldsymbol{Y}_i(\rho_i) = \left(\frac{\alpha_E^{\alpha_E}A}{\rho_i^{\alpha_E}}\right)^{\frac{1}{\alpha_K}} \overline{K}_i. \tag{30}$$

From conditional emission demand and using equation(29) we obtain

$$\boldsymbol{E_i}(\rho_i) = \left(\frac{\alpha_E A}{\rho_i}\right)^{\frac{1}{\alpha_K}} \overline{K}_i. \tag{31}$$

□

B Proof of Lemma 3

The coalition jointly solves the maximization problem $max_{\rho^{IEA}} \sum_m u^i(c_i, E)$.

Differentiating (31) we get

$$\frac{\partial \boldsymbol{E_i}}{\partial \rho_i} = -\frac{1}{\alpha_K \rho_i}\left(\frac{\alpha_E A}{\rho_i}\right)^{\frac{1}{\alpha_K}} \bar{K}_i = -\frac{1}{\alpha_K \rho_i} E_i. \tag{32}$$

Further, we know from the F.O.C.s of the firm that

$$Y_i = E_i \frac{\rho_i}{\alpha_E} = \bar{K}_i \frac{\rho_i}{\alpha_E}\left(\frac{\alpha_E A}{\rho_i}\right)^{\frac{1}{\alpha_K}}. \tag{33}$$

Let $E^{out} \equiv \sum_O E_i$ and $E^{IEA} \equiv \sum_m E_i|_{\rho_i=\rho^{IEA}}$. The joint welfare in the coalition is given by

$$\pi^{IEA}(\rho) = \sum_m u_i(c_i(\rho^{IEA}), E(\rho)) = \sum_m \left(\boldsymbol{Y}_i\left(\rho^{IEA}\right) + \left(\frac{1}{m} \sum_m \boldsymbol{E}_j\left(\rho^{IEA}\right) - \boldsymbol{E}_i\left(\rho^{IEA}\right) \right) P \right) - gmE(\rho) \tag{34}$$

$$= \sum_M Y_i\left(\rho^{IEA}\right) + m\left(\frac{1}{m} \sum_M E_j\left(\rho^{IEA}\right) \right) - \sum_M E_i\left(\rho^{IEA}\right) - gmE(\rho)$$

$$= \sum_M Y_i\left(\rho^{IEA}\right) - gmE(\rho)$$

$$= \sum_M E_i(\rho^{IEA}) \frac{\rho^{IEA}}{\alpha_E} - gmE^{IEA}(\rho^{IEA}) - gmE^{out}$$

$$= \sum_M \bar{K}_i \frac{\rho^{IEA}}{\alpha_E} \left(\frac{\alpha_E A}{\rho^{IEA}} \right)^{\frac{1}{\alpha_K}} - gm \sum_M \bar{K}_i \left(\frac{\alpha_E A}{\rho^{IEA}} \right)^{\frac{1}{\alpha_K}} - gmE^{out}.$$

Differentiation with respect to ρ^{IEA} gives

$$\frac{\partial \pi^{IEA}(\rho)}{\partial \rho^{IEA}} = \sum_M \bar{K}_i \left(-\frac{1}{\alpha_K \rho^{IEA}} \frac{\rho^{IEA}}{\alpha_E} \left(\frac{\alpha_E A}{\rho^{IEA}} \right)^{\frac{1}{\alpha_K}} + \frac{1}{\alpha_E} \left(\frac{\alpha_E A}{\rho^{IEA}} \right)^{\frac{1}{\alpha_K}} \right) - gm \sum_M \bar{K}_i \left(-\frac{1}{\alpha_K \rho^{IEA}} \left(\frac{\alpha_E A}{\rho^{IEA}} \right)^{\frac{1}{\alpha_K}} \right)$$

$$= \sum_M \bar{K}_i \left(\frac{gm}{\alpha_K \rho^{IEA}} \left(\frac{\alpha_E A}{\rho^{IEA}} \right)^{\frac{1}{\alpha_K}} - \frac{1}{\alpha_K \alpha_E} \left(\frac{\alpha_E A}{\rho^{IEA}} \right)^{\frac{1}{\alpha_K}} + \frac{1}{\alpha_E} \left(\frac{\alpha_E A}{\rho^{IEA}} \right)^{\frac{1}{\alpha_K}} \right)$$

$$= \sum_M \bar{K}_i \left(\frac{\alpha_E A}{\rho^{IEA}} \right)^{\frac{1}{\alpha_K}} \left(\frac{gm}{\alpha_K \rho^{IEA}} - \frac{1}{\alpha_K \alpha_E} + \frac{1}{\alpha_E} \right) = \frac{1}{\alpha_K \alpha_E} \sum_M \bar{K}_i \left(\frac{\alpha_E A}{\rho^{IEA}} \right)^{\frac{1}{\alpha_K}} \left(\frac{gm}{\rho^{IEA}} \alpha_E + \alpha_K - 1 \right). \tag{35}$$

As the last term in brackets in (35) has a single zero at $gm = \rho^{IEA}$whereas all other terms are positive for all positive values of ρ^{IEA}.The second derivative of $\pi^{IEA}(\rho)$ can be rearranged as follows

$$\frac{\partial^2 \pi^{IEA}(\rho)}{\partial \rho^{IEA^2}} = \frac{1}{\alpha_K \alpha_E} \sum_M \bar{K}_i \left(-\frac{1}{\alpha_K \rho^{IEA}} \left(\frac{\alpha_E A}{\rho^{IEA}} \right)^{\frac{1}{\alpha_K}} \left(\frac{gm}{\rho^{IEA}} \alpha_E + \alpha_K - 1 \right) + \left(\frac{\alpha_E A}{\rho^{IEA}} \right)^{\frac{1}{\alpha_K}} gm\alpha_E \left(-\frac{1}{\rho^{IEA^2}} \right) \right) \tag{36}$$

$$= \frac{1}{\alpha_K \alpha_E} \sum_M \bar{K}_i \left(\frac{\alpha_E A}{\rho^{IEA}} \right)^{\frac{1}{\alpha_K}} \left(\frac{1}{\alpha_K \rho^{IEA}} - \frac{mg\alpha_E}{\rho^{IEA^2} \alpha_K} - \frac{1}{\rho^{IEA}} - \frac{mg\alpha_E}{\rho^{IEA^2}} \right)$$

$$= \frac{1}{\alpha_K \alpha_E} \sum_M \bar{K}_i \left(\frac{\alpha_E A}{\rho^{IEA}}\right)^{\frac{1}{\alpha_K}} \left(\frac{\rho^{IEA} - mg\alpha_E - \rho^{IEA}\alpha_K - mg\alpha_E\alpha_K}{\rho^{IEA^2}\alpha_K}\right)$$

. As the numerator of the last term in brackets is negative at $gm = \rho^{IEA}$ and all other terms are positive for all positive values of ρ^{IEA} this is the only maximum of $\pi^{IEA}(\rho)$ for positive values of ρ^{IEA}. □

C Proof of Lemma 4

We follow the proof of Roolfs et al. (2020) with the simplifying assumptions of our model. For the sake of readability, we drop functional dependencies and use bold letters to denote variables.

We replace p_i^M from equation (10) into $\rho_i = p_i + P$ in$\mathbf{E^F}$ from 1,to get

$$\mathbf{E^F} = \left(\frac{\alpha_E A}{g + \left(\frac{m-1}{m}\right) P}\right)^{\frac{1}{\alpha_K}} \overline{K}_F \tag{37}$$

where $\overline{K}_F \equiv \sum_m \overline{K}_i$. Rearranging (37) solves implicitly for P

$$P = \frac{m}{m-1}\left(\alpha_E A \left(\frac{\overline{K}_F}{\boldsymbol{E}^F}\right)^{\alpha_K} - g\right). \tag{38}$$

Substituting equations p_i^M from equation (10) and (38) into (30) , (31), and (7), defines Y_i, E_i, c_i implicitly in terms of P;

$$\boldsymbol{Y}_i = A\left(\frac{\boldsymbol{E}^F}{\overline{K}_F}\right)^{\alpha_E} \overline{K}_i. \tag{39}$$

$$\boldsymbol{E}_i = \left(\frac{\alpha_E A}{g + \left(\frac{m-1}{m}\right) P}\right)^{\frac{1}{\alpha_K}} \overline{K}_i. \tag{40}$$

using (39) we get for consumption

$$\boldsymbol{c_i} = \boldsymbol{Y}_i + \left(\frac{\boldsymbol{E}^F}{m} - \boldsymbol{E}_i\right) P = \boldsymbol{Y}_i + \left(\frac{1}{m} - \frac{\overline{K}_i}{\overline{K}_F}\right) P \boldsymbol{E}^F. \tag{41}$$

Substitution of (37) and (41) into the indirect utility function $\boldsymbol{u}^i(\boldsymbol{c}_i(p_i, P), \boldsymbol{E}^F(p, P) + E^o) = \boldsymbol{u}^i(p_i, P)$ corresponding to (2), we get

$$\mathbf{u^i} = A\left(\frac{\boldsymbol{E}^F}{\overline{K}_F}\right)^{\alpha_E}\overline{K}_i + \left(\frac{1}{m} - \frac{K_i}{K_F}\right)P\boldsymbol{E}^F - g(\boldsymbol{E}^F + \sum_o E_i) \tag{42}$$

$$\tag{43}$$

Substitution of (38) we get

$$\mathbf{u^i} = A\overline{K}_F^{\alpha_K}\left(\mathbf{E^F}\right)^{\alpha_E}\left(\frac{(\alpha_K m - 1)\frac{K_i}{K_F} + \alpha_E}{m-1}\right)\left(\left(\frac{m\frac{K_i}{K_F} - 1}{m-1} - 1\right)\boldsymbol{E}^F - \sum_o E_i\right)g. \tag{44}$$

The first order condition of the federal government's objective from equation above is the one that solves

$$\frac{d\boldsymbol{u^i}}{dP} = \frac{\partial \boldsymbol{u^i}}{\partial \boldsymbol{E}^F}\frac{\partial \boldsymbol{E}}{\partial P} = \boldsymbol{Z}_i\frac{\partial \boldsymbol{E^F}}{\partial P} \stackrel{!}{=} 0 \tag{45}$$

where

$$\boldsymbol{Z}_i = \frac{m}{m-1}\left(\alpha_E A\left(\frac{\overline{K}_F}{\mathbf{E^F}}\right)^{\alpha_K}\left(\chi_i - \theta_i - \frac{1}{m}\frac{K_i}{K_F}\right) + g\left(\frac{1}{m} - \chi_i\right)\right) \tag{46}$$

and

$$\frac{\partial \boldsymbol{E}^F}{\partial P} = -\frac{m-1}{m}\frac{\boldsymbol{E}^F}{\alpha_E\alpha_K A\left(\frac{\overline{K}_F}{\mathbf{E^F}}\right)^{\alpha_K}} < 0. \tag{47}$$

Solving $Z_i = 0$ for $\boldsymbol{E^F}$ yields the optimal level from the point of view of country i

$$\boldsymbol{E}^{Fi} = \left(\frac{m\alpha_E A}{g}\frac{\chi_i - \theta_i - \frac{1}{m}\frac{K_i}{K_F}}{m\chi_i - 1}\overline{K}_F^{\alpha_K}\right)^{\frac{1}{\alpha_K}}. \tag{48}$$

Substituting equation (48) into equation (38) leads to

$$P^i = \frac{mg}{m-1}\left(\frac{\theta_i + (\frac{K_i}{K_F} - 1)\frac{1}{m}}{\chi_i - \theta_i - \frac{K_i}{K_F}\frac{1}{m}}\right) \tag{49}$$

which represents the federal price that maximizes the utility f country i.

We proceed to show when P^i must be positive.

Evaluating Z_i at $P = 0$ for type 1 countries. Note that $1 < \chi_1$

$$Z_i|_{P=0} = -\frac{g}{m-1}\left(m\theta_i + \frac{K_i}{K_F} - 1\right)\Bigg|_{P=0}. \tag{50}$$

Consider type 1 countries, then $\frac{1}{m} > \frac{K_1}{K_F}$. It follows that $Z_1|_{P=0}\Big| < 0$. Let for type 2 countries be that

$$\frac{\overline{K}_2}{\overline{K}_F} < \frac{1}{m}\frac{m-\alpha_E}{2-\alpha_E-\frac{1}{m}} \text{ for } i = 1, ..., m. \tag{51}$$

then $Z_2|_{P=0} < 0$ and consequently $Z_i^*\Big|_{P=0} < 0$ for all i.

We now show that P^i is unique. Let $P^b > P^i$. Since $d\boldsymbol{E}^F/dP < 0$ and using equation (48) it follows that

$$\left(\boldsymbol{E}^F\Big|_{P^b}\right)^{\alpha_K} < \left(\boldsymbol{E}^{Fi}\right)^{\alpha_K} = \frac{m\alpha_E A}{g\gamma}\frac{\chi_i - \theta_i - \frac{\kappa_i}{m}}{m\chi_i - 1}\overline{K}_F^{\alpha_K}. \tag{52}$$

Subtracting $\left(E^F\Big|_{P^b}\right)$ on both sides and after some algebraic manipulation, we manage to arrive at the following inequality

$$0 < \frac{m}{m-1}\left(\alpha_E A\left(\frac{\overline{K}_F}{\boldsymbol{E^F}|_{P^b}}\right)^{\alpha_K}\left(\chi_i - \theta_i - \frac{1}{m}\frac{K_i}{K_F}\right) + g\left(\frac{1}{m} - \chi_i\right)\right). \tag{53}$$

The right-hand side of equation (53) is nothing other than $Z_i|_{P^b}$ from equation (46) and hence $Z_i|_{P^b} > 0$. Therefore, it follows that $\boldsymbol{u}^i$ is a concave function with a unique maximum at $P^i > 0$. Let $\sigma_i = \frac{\overline{K}_i}{\overline{K}_F}$. The P^is can be ranked depending on σ_i by considering equation (48)

$$\frac{\partial \boldsymbol{E}^{Fi}}{\partial \sigma_i} = \frac{m-1}{m}\overline{K}_F\left(\frac{\alpha_E A}{g\left(\chi_i - \sigma_i\right)^{\alpha_K+1}}\right)^{\frac{1}{\alpha_K}}\left(\chi_i - \theta_i - \frac{1}{m}\sigma_i\right)^{\frac{\alpha_E}{\alpha_K}} \tag{54}$$

where $\chi_i - \sigma_i > 1$. Suppose inequality (51) holds, then $\chi_i - \theta_i - \frac{1}{m}\sigma_i > 0$. Hence $\partial\boldsymbol{E}^{Fi}/\partial\sigma_i > 0$ for all i. Thus from $\sigma_2 > \sigma_1$ follows $P^2 < P^1$. □

D Proof of Proposition 1

The difference $\Delta_\rho(m_1, m_2)$ of the effective price in an EF and in an IEA is given by subtracting $\rho^{IEA}(m_1, m_2)$ from $\rho_i^M(m_1, m_2)$. Using (23) and (16) we know that

$$\rho^{IEA}(m_1, m_2) = g(m_1 + m_2) \tag{55}$$

and

$$\rho_i^M(m_1, m_2) = -\frac{g(m_1 + m_2)(m_1 s_1 + (-1 + m_2)s_2)}{s_2 - m_2 s_2 + m1(-\alpha_E s_1 - \alpha_K s_2)}. \tag{56}$$

We can thus write

$$\begin{aligned} \Delta_\rho(m_1, m_2) &= -\frac{g(m_1 + m_2)(m_1 s_1 + (-1 + m_2)s_2)}{s_2 - m_2 s_2 + m1(-\alpha_E s_1 - \alpha_K s_2)} - g(m_1 + m_2) \\ &= -\frac{\alpha_K g m_1 (m1 + m_2)(s_1 - s_2)}{-\alpha_E m_1 s_1 + s_2 - \alpha_K m_1 s2 - m_2 s_2} < 0. \end{aligned} \tag{57}$$

Outsider prices are given by $g = p_i^O$ in both cases (see (16) and (21)).

□

E Proof of Proposition 2

a) Follows from Proposition 1 and (32). For b) remember that $\rho^{IEA}(m_1, m_2)$ maximizes the utilitarian welfare of the coalition members (see proof of Lemma 3), whose welfare is thus lower with any other effective price including ρ_i^M. From (16) and (21) we know that outsiders choose the same policies in both cases so that their consumption levels are equal too. From a) follows that they suffer less damages in the IEA case and are therefore better off.