Thomas Eichner · Rüdiger Pethig

Self-enforcing environmental agreements and capital mobility

Self-enforcing environmental agreements and capital mobility[*]

Thomas Eichner
Department of Economics, University of Hagen

Rüdiger Pethig
Department of Economics, University of Siegen

Abstract

In a multi-country model with mobile capital and global pollution this paper analyzes the stability of self-enforcing environmental agreements (IEAs) when the coalition formed by the signatory countries plays Nash. In accordance with previous environmental literature we show that there exists a unique self-enforcing IEA consisting of two or three signatory countries if emissions tax rates are strategic substitutes. However, emissions tax rates are strategic complements if the pollution is not too detrimental. In that case we find very small self-enforcing IEAs, as before, but now the socially optimal agreement among all countries may be self-enforcing as well. Special emphasis is placed on the investigation and interpretation of the conditions which render stable the grand coalition.

JEL classification: H23, H77, Q58
Key words: capital mobility, self-enforcing environmental agreements, emissions tax, Nash behavior

[*]Eichner: Department of Economics, University of Hagen, Universitätsstr. 41, 58097 Hagen, Germany, email: thomas.eichner@fernuni-hagen.de; Pethig: Department of Economics, University of Siegen, Hölderlinstr. 3, 57068 Siegen, Germany, pethig@vwl.wiwi.uni-siegen.de.

1 Introduction

The massive reduction of global carbon emissions necessary to stabilize the world climate at safe levels cannot be achieved without an effective international environmental agreement (IEA). However, more than 20 years of international negotiations towards such an agreement have brought little progress. It is therefore important to investigate the conditions for the successful formation of an IEA that is both encompassing and effective. The economic literature on IEAs since the early 1990s is based on the proposition that sovereign countries are reluctant to join an IEA unless it is in their self-interest. Therefore the concept of self-enforcing agreements is crucial which requires that no signatory has an incentive to leave the IEA and no non-signatory has an incentive to join it. The basic model of an IEA employed by Carraro and Siniscalco (1991), Hoel (1992), Barrett (1994) and others since then is a simple static model of identical countries without international trade. Some studies model climate coalitions[1] as Stackelberg leaders (e.g. Barrett 1994, Diamantoudi and Sartzetakis 2006, Rubio and Ulph 2006) and others portray them as Nash players along with all non-signatories (e.g. Carraro and Siniscalco 1991, Hoel 1992, Finus 2001, Rübbelke and Finus 2013). In both variants of the basic model strong free-rider incentives prevent large stable coalitions and large gains from cooperation.

In order to find out whether the prospects of reaching an effective IEA enhance in more structured models, Eichner and Pethig (2013a, 2013b) extend the basic model of coalition formation by explicitly modeling production, consumption and international trade in fossil fuels and a composite consumption good. When the coalition is assumed to be the Stackelberg leader (Eichner and Pethig 2013a), stable coalitions turn out to comprise up to 60% of all countries. But unfortunately, such coalitions hardly reduce climate damage below its level in the non-cooperative scenario - regardless how large they are. When the coalition plays Nash along with all fringe countries (Eichner and Pethig 2013b), stable conditions are both small and ineffective similar as in the basic model without trade.

The present paper also analyzes the formation of a climate coalition in a world economy with international trade, but in contrast to Eichner and Pethig (2013a, 2013b) our focus is now on capital mobility and capital-related global pollution. That means, we take as our point of departure the branch of the fiscal federalism literature dealing with decentralized policymaking in an economy with spillovers among jurisdictions. In their seminal paper Oates and Schwab (1988) argue that the choice of capital taxes and environmen-

---

[1]In the present paper the terms IEA and (climate) coalition are synonymous because our exclusive focus is on a single coalition. Also, we take as equivalent the terms 'self-enforcing IEA' and 'stable (climate) coalition'.

tal standards is efficient in an economy with identical jurisdictions, mobile capital and local pollution. Ogawa and Wildasin (2009) extend the analysis to account for transboundary pollution (spillovers) and asymmetric countries, and still get efficient capital tax rates. Eichner and Runkel (2012) point out that it is the zero capital supply elasticity which drives Ogawa and Wildasin's result. They adopt a two-period framework employed e.g. by Bukovetsky and Wilson (1991) and Keen and Kotsogiannis (2002) and show that in case of strictly positive capital supply elasticities capital tax rates are inefficiently low in the decentralized equilibrium because the jurisdictions' choice of capital taxes is then distorted by their incentive for tax competition and their disregard of spillover effects.[2]

Ogawa and Wildasin (2009) and Eichner and Runkel (2012) investigate the (in)efficiency of capital tax competition in the presence of transboundary pollution when decision-making is decentralized, i.e. when all jurisdictions/countries act non-cooperatively. Here we will take up the analytical framework of Eichner and Runkel (2012) with some minor simplifications[3] to investigate the formation of stable coalitions when the fringe countries as well as the coalition play Nash. Although our approach shares with Eichner and Pethig (2013b) both Nash behavior on the part of the coalition and international trade, the pertaining models differ significantly. In Eichner and Pethig's one-period model there are world markets for a composite consumption good and fossil fuels; fuels are extracted and consumed by the countries' residents along with a consumption good that is produced without using fuel as an input. In contrast, following Eichner and Runkel (2012) we now model world markets for capital and a composite consumption good in the second period, and capital is an intermediate good in the production of the consumption good.

As reported above, Eichner and Pethig (2013b) found no stable coalitions consisting of three or more countries. Similarly, in the present paper we will demonstrate that there exist small stable coalitions with two or three member countries. However, for a smaller but non-empty subset of parameter values the grand coalition turns out to be also stable. In other words, full cooperation of all countries may be self-enforcing. The crucial necessary condition for this unexpected result are economies (= parameter constellations) in which emissions tax rates are strategic complements. For given preferences and technologies strategic complementarity of taxes *and* stable grand coalitions are the more likely, the

---

[2]In the literature on capital tax competition (e.g. Zodrow and Mieszkowski 1986, Wilson 1986) an additional source of inefficiency is the under-provision of public goods resulting from the requirement to finance public goods by capital tax revenues exclusively. If lumpsum taxation is allowed along with capital taxation, as e.g. in Eichner and Runkel (2012), the provision of public goods is efficient.

[3]We do not model public goods whose allocation is efficient in Eichner and Runkel (2012) anyway because they allow for lumpsum taxation. Second, we restrict the analysis to uniformly dispersed emissions ($\beta = 1$) which approximates the case of climate change.

smaller the total number of countries, the less severe the climate damage of emissions and the smaller the flow of emissions. In Eichner and Pethig (2013b) and, to our knowledge, also in the entire theoretical literature on the formation of IEAs emissions taxes are strategic substitutes which is why grand coalitions are always unstable in that literature.

The paper is organized as follows. Section 2 introduces and describes the formal model and briefly characterizes the benchmark scenarios of global non-cooperation and social optimum. Section 3 analyzes the impact of climate coalitions of different but exogenously given size and investigates analytically and numerically economies in which emissions tax rates are either strategic complements or strategic substitutes. Section 4 then turns to the existence and size of self-enforcing IEAs emphasizing the conditions under which the grand coalition is stable. Section 5 concludes.

2   The model

Consider an economy with $n \geq 2$ identical countries. Each country i hosts a representative firm that employs k_i units of capital to produce good X according to the production function

$$X(k_i) \tag{1}$$

that is increasing and concave, i.e. $X' > 0$ and $X'' < 0$. The production process generates emissions, e_i, in strict proportion to the capital employed which we express by writing $e_i = \psi k_i$ with $\psi > 0$ and constant. Emissions are regulated in each country by means of an emissions tax at rate t_i. Each unit of capital is purchased on the economy-wide capital market at the price $1 + r > 1$. Normalizing the price of good X to one, profit maximization of the after-tax profit $\Pi_i = X(k_i) - (1 + r + t_i \psi)k_i$ of the firm located in country i yields

$$X'(k_i) = 1 + r + t_i \psi. \tag{2}$$

Each country is populated by a representative household who lives for two periods. In the first period, country i's resident has an endowment $\bar{k}$ of first-period income that can either be spent for first-period consumption at rate x_i^1 or saved at rate $s_i = \bar{k} - x_i^1$. In the second period, the resident receives capital income $(1 + r)s_i$ and profit income Π_i earned in her country. The second-period budget is $x_i^2 = (1 + r)s_i + \Pi_i + t_i \psi k_i$ where x_i^2 denotes second-period consumption and where $t_i \psi k_i$ is the lumpsum transfer of tax revenues to the household. The utility of country i's resident is increasing in private consumption and is

4

negatively affected by global pollution[4] $\sum_j e_j$. The utility function is given by

$$U(x_i^1) + x_i^2 - D\left(\sum_j e_j\right).$$ (3)

The subutility function U is increasing and concave and the damage function D is increasing and convex. Maximizing (3) with respect to s_i subject to the budget constraints gives $U'(\bar{k} - s_i) = 1 + r$ which determines savings as a function of the interest rate. Capital and good X are traded on perfectly competitive world markets. The condition

$$\sum_j k_j = \sum_j s_j$$ (4)

clears the capital market.

Observe that combining the definition of after-tax profit with the consolidated budget equation yields

$$x_i^2 = X(k_i) + (1+r)(s_i - k_i).$$ (5)

When we sum (5) over all countries and account for (4) we confirm Walras' Law according to which the market for the second-period consumption good is in equilibrium if and only if (4) is satisfied.[5]

The model introduced above is adopted from Eichner and Runkel (2012). Their model, in turn, combines the proportionality between capital and emissions from Ogawa and Wildasin (2009) with the two-period approach applied by Bucovetsky and Wilson (1991) and by Keen and Kotsogiannis (2002).

For the benefit of more specific results, we will specify throughout the paper the functions X, U and D by

$$U(x_i^1) = ax_i^1 - \frac{b}{2}\left(x_i^1\right)^2, \quad X(k_i) = \alpha k_i - \frac{\beta}{2}k_i^2, \quad D\left(\sum_j e_j\right) = \frac{\delta}{2}\left(\sum_j e_j\right)^2,$$ (6)

where a, b, α, β and δ are positive parameters. Utility maximization and profit maximization yield the savings and capital demand, respectively,

$$s_i = \frac{1 - a + r}{b} + \bar{k}, \quad k_i = \frac{\alpha - 1 - r - t_i \psi}{\beta}.$$ (7)

Inserting (7) in (4) we get after rearrangement of terms

$$r = \frac{(\alpha - 1)b + (a - 1)\beta - b\beta\bar{k}}{b + \beta} - \frac{b\psi}{n(b + \beta)}\sum_j t_j.$$ (8)

---

[4]For convenience of notation, we write $\sum_j$ short for $\sum_{j=1}^{j=n}$.
[5]This observation is the rationale for the price of capital being equal to $1 + r$.

Consideration of r from (8) in (7) gives

$$k_i = \frac{1}{b+\beta}\left[\zeta - \frac{t_i\psi(b+\beta)}{\beta} + \frac{b\psi}{n\beta}\sum_j t_j\right], \quad s_i = \frac{1}{b+\beta}\left[\zeta - \frac{\psi}{n}\sum_j t_j\right], \tag{9}$$

where $\zeta := \alpha - a + b\bar{k}$. Observe that

$$\sum_j s_j = \sum_j k_j = \frac{\sum_j e_j}{\psi} = \frac{n}{b+\beta}\left[\zeta - \frac{\psi}{n}\sum_j t_j\right]. \tag{10}$$

We account for $x_i^1 = \bar{k} - s_i$ and (5) in (3) to get

$$W^i(t_1,\ldots,t_n) := U(\bar{k} - s_i) + X(k_i) + (1+r)(s_i - k_i) - D\left(\sum_j e_j\right) \tag{11}$$

Since the variables s_i, k_i, r and $\sum_j e_j$ are determined by $(t_1,\ldots,t_n)$ via (7) - (10), $W^i(t_1,\ldots,t_n)$ from (11) is the equilibrium welfare of country i for any given tax profile $(t_1,\ldots,t_n)$.

Business as usual. For later use as a benchmark, we briefly characterize the non-cooperative Nash-equilibrium which we refer to as business as usual (BAU). The government of country i chooses the tax rate t_i that maximizes $W^i(t_i,\ldots,t_n)$ for given tax rates $(t_1,\ldots,t_{i-1},t_{i+1},\ldots,t_n)$ of the other countries. The associated first-order condition $W^i_{t_i} = 0$ implicitly determines country i's best reply function[6]

$$t_i = \tilde{R}\left(\sum_{j\neq i} t_j\right) = \frac{n^3\beta\delta\psi\zeta}{N^o} + \frac{[(b+\beta)b - n^2\beta\delta\psi^2]}{N^o}\sum_{j\neq i} t_j, \tag{12}$$

where $N^o := (n^2 - 1)(b+\beta)b + b\beta n^2 + n^2\beta(\beta + \delta\psi^2) > 0$. The reaction function is upward [downward] sloping and characterizes the countries' tax rates as strategic complements [substitutes] if and only if $(b+\beta)b > [<]n^2\beta\delta\psi^2$. Sufficient for the tax rates to be strategic complements are the polar cases in which either emissions are absent ($\psi = 0$) or there is no damage from emissions ($\delta = 0$). Ceteris paribus, the more emissions are generated ($\psi \uparrow$), the more severe the damage from emissions is ($\delta \uparrow$), the larger the group of countries is ($n \uparrow$), the more likely the tax rates are strategic substitutes. Since $\frac{dt_i}{d\left(\sum_{j\neq i} dt_j\right)} < 1$ and constant, there is a unique Nash equilibrium, say $(t_{1o},\ldots,t_{no})$. With all countries being alike, the Nash equilibrium is characterized by $t_{1o} = \ldots = t_{no} \equiv t_o$ and hence also by $k_{1o} = \ldots = k_{no} = s_{1o} = \ldots = s_{no} \equiv s_o \equiv k_o$. Making use of the symmetry assumption in (12) yields the BAU tax rate $t_o = \frac{n^2\beta\delta\psi\zeta}{b^2(n-1)+b\beta(2n-1)+n\beta(\beta+n\delta\psi^2)}$.

---

[6]The best-reply function is derived in the Appendix A.

Social optimum   To evaluate the non-cooperative mitigation efforts of Nash governments we briefly determine the socially optimal tax rates. Maximizing $\sum_j W^j(t_1, \ldots, t_n)$ with respect to t_i yields the first order condition $\sum_j W^j_{t_i} = 0$ which can be solved for the socially optimal tax rate[7] $\hat{t} = \frac{\delta\psi\zeta n^2}{b+\beta+n^2\delta\psi^2}$. Insertion of $\hat{t}$ in (7) - (11) gives the socially optimal solution $(\hat{w}, \hat{e}, \hat{k}, \hat{s}, \hat{r}, \hat{t})$. Comparing that allocation with the BAU allocation $(w_o, e_o, k_o, s_o, r_o, t_o)$ yields $\hat{w} > w_o$, $\hat{e} < e_o$, $\hat{k} = \hat{s} < k_o = s_o$, $\hat{t} > t_o$ and $\hat{r} < r_o$.[8] Hence as expected, all countries suffer a welfare loss in BAU because non-cooperative governments ignore the positive impact of their tax on all other countries.

Note that the optimal tax rate is zero ($\hat{t} = 0$) either if no emissions are generated ($\psi = 0$) or emissions are not harmful ($\delta = 0$). In these cases we also have $t_o = 0$ such that the BAU allocation is efficient. Hence the only cause for inefficiency are the detrimental emissions.[9] The inefficiency of BAU (in case of $\psi > 0$ and $\delta > 0$) gives room for enhancing world welfare through cooperation. Obviously, full international cooperation is the straightforward solution. But the insight from juxtaposing the polar cases of full cooperation and no cooperation (BAU) is limited. Our next step is therefore to study intermediate cases of a cooperative sub-group of countries interacting with the non-cooperative rest of the world.

3   Climate coalitions

Suppose now that some countries are members in a coalition. For the sake of the formal analysis, we lump together the first m countries, $2 \leq m < n$, in one group, denoted group $C := \{1, 2, \ldots, m\}$ with C for coalition, and collect all remaining countries in another group, denoted group $F := \{m+1, \ldots, n\}$ with F for fringe. All fringe countries play Nash against the coalition and against all fellow fringe countries. The coalition countries are assumed to commit to cooperative climate policy such that the coalition now acts as a single player whose payoff is the coalition countries' aggregate welfare $\sum_{j \in C} W^j(t_1, \ldots, t_n)$ and who plays Nash against all fringe countries. Taking advantage of the symmetry assumption we treat all countries equally within their group. Specifically, we set $t_i = t_c$ for all $i \in C$ and $t_i = t_f$ for all $i \in F$ from the outset. With this substantial analytical relief the best reply of the

---

[7]For the derivation of $\hat{t}$ we refer to the Appendix A.

[8]The proof is given in the Appendix A.

[9]In the standard model of capital tax competition (e.g. Zodrow and Mieszkowski 1986) the tax competition is inefficient *without* emissions because it leads to an undersupply of national public goods if these goods are financed via capital taxation exclusively. This channel of inefficiency is disregarded in the present paper.

coalition with m members to the fringe countries' strategies t_f is given by[10]

$$t_c = \frac{mn^3\beta\delta\theta\psi}{N^c} + \frac{m(n-m)[b(b+\beta) - n^2\beta\delta\psi^2]}{N^c}t_f, \tag{13}$$

where $N^c := (n^2 - m^2)b(b+\beta) + n^2\beta(b+\beta+m^2\delta\psi^2)$. Each fringe country's best reply is still given by (12). We account for $t_i = t_c$ for $i \in C$ and $t_i = t_f$ for $i \in F$ and solve (12) for t_f to obtain

$$t_f = \frac{n^3\beta\delta\varphi\psi}{N^f} + \frac{m[b(b+\beta) - n^2\beta\delta\psi^2]}{N^f}t_c, \tag{14}$$

where $N^f := b^2[n(n-1)+m] + b\beta[n(2n-1)+m] + n^2\beta[\beta + (n-m)\delta\psi^2]$.

Clearly, a coalition-fringe equilibrium is the n-tuple $(\underbrace{t_c^*, \ldots, t_c^*}_{m\text{-times}}, \underbrace{t_f^*, \ldots, t_f^*}_{(n\text{-}m)\text{-times}})$ of tax rates solving (13) and (14). The solution turns out to be

$$t_c^* = \frac{mn^3\beta\delta\varphi\psi}{N} \quad \text{and} \quad t_f^* = \frac{n^3\beta\delta\varphi\psi}{N} \tag{15}$$

with $N := b(b+\beta)[n(n-1) - m(m-1)] + n^2\beta[b+\beta+(n-m+m^2)\delta\varphi^2]$. The corresponding coalition-fringe equilibrium allocation is characterized by $(w_c^*, w_f^*, e_c^*, e_f^*, k_c^*, k_f^*, r^*, t_c^*, t_f^*)$. From (15) it is obvious that $t_c^* = mt_f^* > t_f^*$ which implies, in turn, $k_c^* < k_f^*$ and $e_c^* < e_f^*$. Since $s_c^* = s_f^*$ due to (7), it follows from (10) that coalition countries export capital and import the consumption good. The tighter tax rate $(t_c^* > t_f^*)$ drives capital out of coalition countries. Since emissions are tied to capital the reduction of emissions in coalition countries $(e_c \downarrow)$ increases the emissions in fringe countries $(e_f \uparrow)$ which is known as carbon leakage in the climate change literature. In order to further assess the coalition's performance we compare the countries' welfare and their tax rates in the coalition-fringe equilibrium with those in BAU. With the notation t_c and t_f we rewrite the welfare of individual countries as $W^c(t_c, t_f, m) := W^i(\underbrace{t_c, \ldots, t_c}_{m\text{-times}}, \underbrace{t_f, \ldots, t_f}_{(n\text{-}m)\text{-times}})$ for all countries in group C, as $W^f(t_c, t_f, m) := W^i(\underbrace{t_c, \ldots, t_c}_{m\text{-times}}, \underbrace{t_f, \ldots, t_f}_{(n\text{-}m)\text{-times}})$ for all countries in group F, and prove in the Appendix D

Proposition 1. *The coalition-fringe equilibrium with a coalition of size $m \in \{2, \ldots, n-1\}$ compares with the non-cooperative Nash equilibrium (BAU) as follows:*

(i) $W^f(t_c^*, t_f^*, m) > W^c(t_c^*, t_f^*, m) > w_o$,

(ii) $t_c^* > t_o$,

(iii) $t_f^* > t_o$ if and only if tax rates are strategic complements $[b(b+\beta) > n^2\beta\delta\psi^2]$.

---

[10]For the derivation of (13) we refer to the Appendix B.

The principal message of Proposition 1 is that the coalition enhances the welfare of all countries compared to BAU, but grants a free-rider welfare advantage to all fringe countries. This result is independent of whether tax rates are strategic substitutes or complements. But according to Proposition 1(iii) the fringe countries' tax rate (and mitigation effort) is higher [lower] than in BAU, if tax rates are strategic complements [substitutes]. In other words, if tax rates are strategic complements [substitutes], the fringe countries support [counteract] the coalition's mitigation effort.

The asymmetry in outcome for fringe and coalition countries is driven by three effects. The first is the negative tax base effect $\left(\frac{\partial k_c}{\partial t_c}, \frac{\partial k_f}{\partial t_f}\right)$. When the coalition or a fringe country increases its tax rate, domestic investments and domestic emissions decline. The tax base effect of coalition countries is smaller in absolute terms than that of fringe countries $\left(\left|\frac{\partial k_f}{\partial t_f}\right| > \left|\frac{\partial k_c}{\partial t_c}\right|\right)$ which implies that the coalition is more effective in preventing capital outflows than a fringe country.[11] The second effect is the terms of trade effect. Since raising the tax reduces domestic investments, total demand for capital decreases, ceteris paribus, and thus reduces the price of capital and stimulates capital input and emissions in the other countries. It is clearly that the coalition's terms of trade effect is more pronounced than that of fringe countries[12] $\left(\left|\frac{\partial r_f}{\partial t_f}\right| = m \left|\frac{\partial r_c}{\partial t_c}\right|\right)$. Finally, the tax rates impact on the global climate. Tightening domestic mitigation policies ($t_c \uparrow, t_f \uparrow$) does not only reduce domestic emissions, but also causes emissions leakage thus rendering unilateral taxation rather ineffective in the fight against global climate change. The global climate effect $\left(\frac{\partial\left(\sum_j e_j\right)}{\partial t_c}, \frac{\partial\left(\sum_j e_j\right)}{\partial t_f}\right)$ answers the question as to how strong the leakage effect is. Both climate effects are negative such that the leakage effect is always less than 100%. Moreover, due to $\left|\frac{\partial\left(\sum_j e_j\right)}{\partial t_c}\right| > \left|\frac{\partial\left(\sum_j e_j\right)}{\partial t_f}\right|$ the coalition is more successful than individual fringe countries in mitigating climate change.[13]

According to our preceding analysis, the coalition-fringe equilibrium allocation is uniquely determined by - and varies with - the coalition size. In the following proposition proved in the Appendix E we present results on how the equilibrium values of tax rates and welfare depend on the coalition size.

Proposition 2. *Denote by $t_v^* = \mathcal{T}^v(m)$ and $w_v^* = \mathcal{W}^v(m) := W^v\left[\mathcal{T}^c(m), \mathcal{T}^f(m), m\right]$ for $v = c, f$ the tax rates and welfare levels of coalition and fringe countries in the coalition-fringe equilibrium with a coalition of size[14] $m \in [1, n] \in \mathbb{R}_+$.*

---

[11]Follows from $\frac{\partial k_c}{\partial t_c} = -\frac{\psi[(n-m)b+n\beta]}{n\beta(b+\beta)}$ and $\frac{\partial k_f}{\partial t_f} = -\frac{\psi[(n-1)b+n\beta]}{n\beta(b+\beta)}$.

[12]Follows from $\frac{\partial r_c}{\partial t_c} = -\frac{mb\psi}{\beta(b+\beta)}$ and $\frac{\partial r_f}{\partial t_f} = -\frac{b\psi}{\beta(b+\beta)}$.

[13]Follows from $\frac{\partial\left(\sum_j e_j\right)}{\partial t_c} = -\frac{m\psi^2}{n(b+\beta)}$ and $\frac{\partial\left(\sum_j e_j\right)}{\partial t_f} = -\frac{\psi^2}{n(b+\beta)}$.

[14]For analytical convenience we take the interval $[1, n]$ to be the domain of these functions, keeping in mind in our later conclusions that the domain of real-world coalitions is the set of integers $\{1, \ldots, n\}$.

(i) *The welfare levels of coalition countries, $\mathcal{W}^c(m)$,* **and** *of fringe countries, $\mathcal{W}^f(m)$, are strictly increasing in m.*

(ii) *The average global tax rate, $\frac{m\mathcal{T}^c(m)+(n-m)\mathcal{T}^f(m)}{n}$, is strictly increasing in m.*

(iii) *Suppose tax rates are strategic complements $[b(b+\beta) > n^2\beta\delta\psi^2]$. Then the tax rates of coalition countries, $\mathcal{T}^c(m)$,* **and** *of fringe countries, $\mathcal{T}^f(m)$, are strictly increasing in m.*

(iv) *Suppose tax rates are strategic substitutes $[b(b+\beta) < n^2\beta\delta\psi^2]$. Then the tax rates of coalition countries, $\mathcal{T}^c(m)$, are strictly increasing in m for $m < (n+1)/2$ and the tax rates of fringe countries, $\mathcal{T}^f(m)$, are strictly decreasing in m.*

The parts (i) and (ii) of Proposition 2 confirm the intuition that the larger the coalition, the more effective is its fight against global change and the more beneficial it is for all countries: Aggregate tax rates and all countries' welfare are strictly increasing in the coalition size. While these features are independent of whether tax rates are strategic substitutes or complements, the relation between tax rates and the coalition size crucially depends on that distinction. In case of strategic complements all countries step up their mitigation efforts with increasing coalition size - with the coalition countries' tax rate being always higher by the factor m than that of fringe countries. In contrast, if tax rates are strategic substitutes, the fringe countries reduce their tax rate with increasing coalition size and thus obstruct the coalition's mitigation effort. One might expect that in case of strategic substitutes the coalition unambiguously increases its tax when it gets larger. But Proposition 2(iv) establishes that result under a restrictive sufficient condition only.

Recall that we briefly compared the allocations in BAU and in the social optimum at the end of Section 2. If we account for $\frac{\mathrm{d}[m\mathcal{T}^c(m)+(n-m)\mathcal{T}^f(m)]}{n\mathrm{d}m} > 0$ (Proposition 2(iii)) in (8) and (9) it is possible to supplement that comparison by observing that r^*, s^* and $me_c^* + (n-m)e_f^*$ are strictly decreasing in m. Hence

$$r_o > r^* > \hat{r}, \quad s_o > s^* > \hat{s}, \quad \text{and} \quad e_o > \frac{me_c^* + (n-m)e_f^*}{n} > \hat{e}. \tag{16}$$

Proposition 2 provides interesting information about the dependence of the equilibrium values w_c^*, w_f^*, $mw_c^* + (n-m)w_f^*$, t_c^* and t_f^* on the coalition size. To obtain even more specific information on the curvature of the functions $\mathcal{W}^v$, $\mathcal{T}^v$, $v = c, f$, we now proceed with numerical illustrations of the analytical results of Proposition 2.[15] Although the dependence on coalition size of variables other than those investigated in Proposition 2 is also very

---

[15]We have varied the parameters in a large number of examples. Since the curvature turned out to be robust with respect to parameter variations, we find it sufficient to discuss below one example each for the cases of strategic complements and substitutes.

important, we restrict our subsequent discussion to the curves in Proposition 2 to keep focused on the stability issue to be addressed in the next section.

Our numerical Example 1 consists of the parameters $a = 1$, $b = 0.1$, $\alpha = 100$, $\beta = 10$, $\delta = 0.001$ and $\psi = 1$. The parameters of Example 2 are the same as in Example 1 with the exception that $b = 0.1$ is replaces by $b = 10$. In Example 1 tax rates are strategic substitutes because best reply curves are downward sloping. In contrast, in Example 2 tax rates are strategic complements because the corresponding best reply curves are upward sloping.
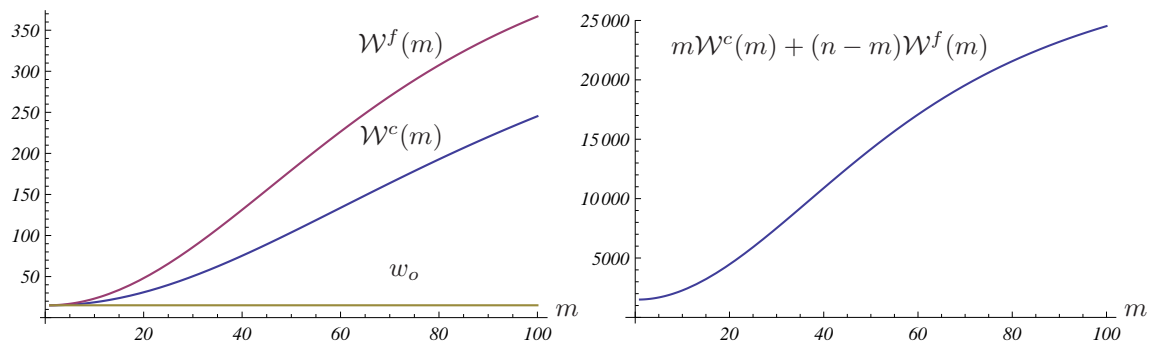

Figure 1: National and total welfare in case of strategic substitutes (Example 1)
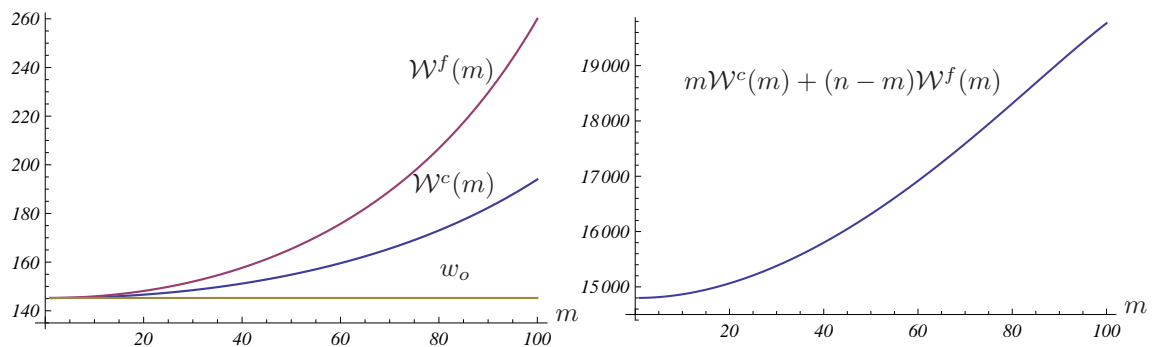

Figure 2: National and total welfare in case of strategic complements (Example 2)

The Figures 1 and 2 illustrate the Propositions 1(i) and 2(i). The welfare curves in both figures deviate in their second derivative, but otherwise there is no significant difference between the cases of strategic substitutes and complements. In Figure 3 the average global tax rate increases with the coalition size as required by Proposition 2(ii), regardless whether

11

tax rates are strategic complements or substitutes. As stated in the Propositions 2(iii) and (iv) and illustrated in the Figures 4 and 5, a remarkable difference between the cases of complements and substitutes emerges, however, with regard to the tax rates of coalition and fringe countries. In both cases, the tax rate of coalition countries is increasing in the coalition size,[16] but in case of substitutes [complements] the fringe countries reduce [increase] their tax rate and mitigation effort. When tax rates are substitutes, the coalition's mitigation effort is thwarted by the fringe countries; otherwise the fringe countries boost the coalition's effort. However, with growing coalition size the increase or decrease, respectively, of the fringe countries' tax rates is so small that it can hardly be identified in the left panels of the Figures 4 and 5. These small tax rate variations become visible in the right panels following a massive amplification the scale of the vertical axes. It can be shown that in the Examples 1 and 2 national emissions respond to taxes as expected. With increasing coalition size, the coalition countries' emissions decline in case of complementarity and substitution, while the fringe countries' emissions vary hardly with the coalition size. As a consequence, total emissions - and climate damage - are strictly decreasing in the coalition size which is confirmed analytically by combining $\sum_j e_j$ from (10) with $\frac{\mathrm{d}[m\mathcal{T}^c(m)+(n-m)\mathcal{T}^f(m)]}{\mathrm{d}m} > 0$ from Proposition 2(ii).
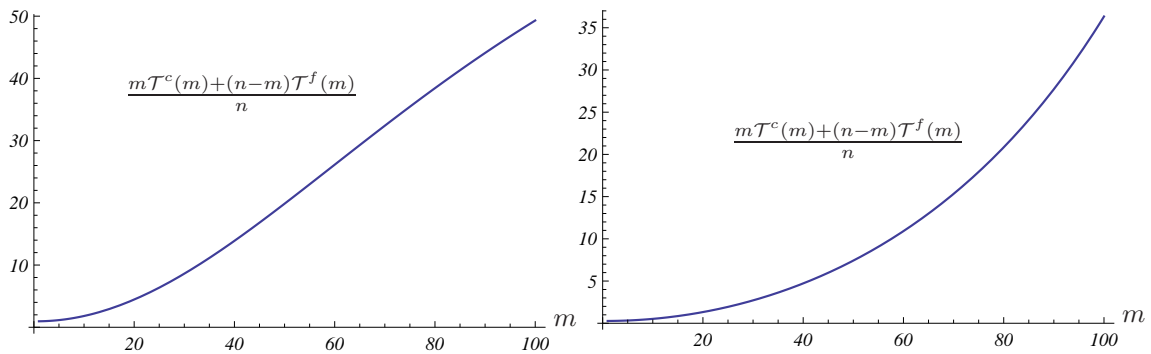

Figure 3: Average tax rates in Example 1 (left panel) and in Example 2 (right panel)

---

[16]Proposition 2(iv) does not exclude the possibility that the tax rate of coalition countries is decreasing in the coalition size for large coalitions. That appears to be an exceptional case, however.
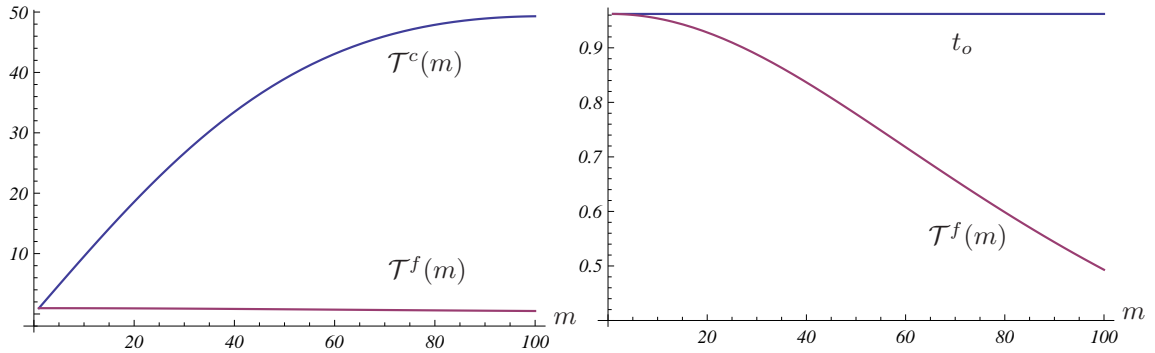
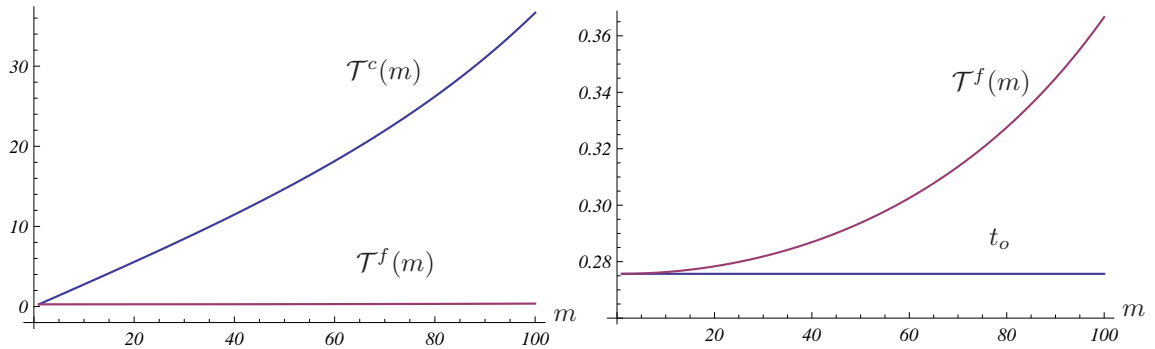Figure 4: Tax rates in case of strategic substitutes (Example 1)

Figure 5: Tax rates in case of strategic complements (Example 2)

4 Self-enforcing IEAs

In the preceding Section 3 we have presupposed the presence of a climate coalition of given size, and our focus has been on characterizing the coalition-fringe equilibrium and its change in case of exogenous variations of the coalition size. Now we turn to the issue of coalition stability. Since supranational authorities for the effective enforcement of IEAs are not available, such agreements will not be concluded unless they are self-enforcing in the sense that no coalition country has an incentive to defect (internal stability) and no fringe country has an incentive to join the tax agreement (external stability).[17] In formal language, an IEA with $m \in \{2, \dots, n\}$ member countries is said to be self-enforcing or stable if it satisfies the internal stability condition

$$\mathcal{W}^c(m) - \mathcal{W}^f(m-1) \geq 0 \tag{17}$$

---

[17]This notion of self-enforcement or stability was originally introduced by D'Asprement et al. (1983) in the context of cartel formation and was first applied to the formation of IEAs by Barrett (1994).

13

and the external stability condition

$$\mathcal{W}^f(m) - \mathcal{W}^c(m+1) \geq 0. \tag{18}$$

In the first step we will apply the stability concept to the Examples 1 and 2. After that we will seek to establish some more general stability results. The Figure 6 plots the graphs of the functions $W^c(m) - W^f(m-1)$ and $W^f(m) - W^c(m+1)$ for the Examples 1 and 2, respectively. Coalition stability requires both graphs to have positive values. In both Figures there is an interval - roughly between 2 and 3 - where both graphs are positive. The integers in that interval are sizes of coalitions that are internally and externally stable. Closer inspection shows that in Example 1 the two-country coalition and in Example 2 the three-country coalition is stable.
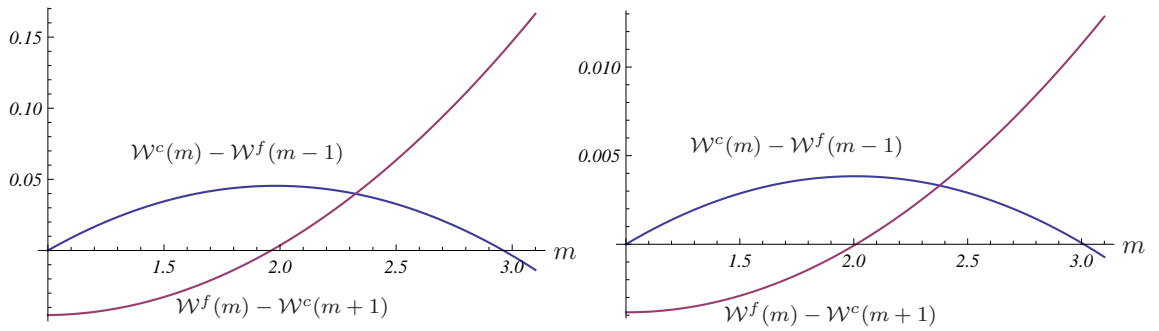

Figure 6: Coalition stability in Example 1 (left panel) and in Example 2 (right panel)

In order to investigate the conditions for the stability of coalitions in a more general way we define in the Appendix F a rather large subspace M of parameters and prove in the Appendix F

Proposition 3. *In all economies belonging to the parameter space M there exists a stable coalition of size $m = 2$ or $m = 3$.*

The proof of Proposition 3 in the Appendix F is constructive. We cannot conclude from Proposition 3 that stable coalitions of size two or three exist for all feasible parameter constellations. But as the set all of all parameters, M, is very large it is unlikely that Proposition 3 does not hold for feasible parameter constellations not contained in M. The total number of countries, n, considered in the parameter space M range from $n = 10$ to

14

$n = 200$. Hence the stable coalitions identified in Proposition 3 are small in both absolute and relative terms and are thus hardly effective in reducing climate damage.

Let us return to the Examples 1 and 2, and to Figure 6, temporarily. It can be shown that in both panels of Figure 6 the graph of the function $W^c(m) - W^f(m-1)$ does not attain positive values anymore for integers m larger than three. Hence the small stable coalitions identified in Figure 6 are the only stable coalitions in the economies of the Examples 1 and 2. This is not a general result, however. To see that we consider another economy, denoted Example 3, that differs from Example 2 only in that the parameter b is increased from $b = 10$ to $b = 50000$. Hence Example 3 represents the case of strategic complements in a more pronounced way than Example 2.
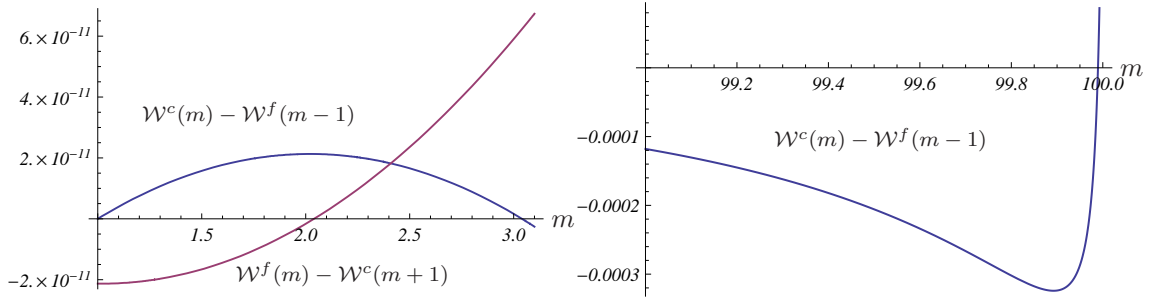

Figure 7: Coalition stability in Example 3

The graphs in the left panel of Figure 7 resemble those of Figure 6 and thus demonstrate that in Example 3 there exists a stable coalition of $m = 3$. Moreover, we can show that (18) is satisfied for all $m \in \{4, \ldots, 99\}$ and that (17) does not hold for all $m \in \{4, \ldots, 99\}$.[18] Hence all coalitions of size $m \in \{4, \ldots, 99\}$ are unstable. But the right panel of Figure 7 reveals the surprising result that the grand coalition is internally stable $[\mathcal{W}^c(100) - \mathcal{W}^f(99)]$.[19] The finding in Example 3 that the grand coalition may be stable along with those small coalitions identified in Proposition 3 is totally unexpected and in stark contrast to the literature on self-enforcing IEA we are aware of. Even in the model of Eichner and Pethig (2013b) which accounts for international trade (in a different way, though), all stable coalitions are

---

[18] We refrain from plotting the graphs of $\mathcal{W}^f(m) - \mathcal{W}^c(m+1)$ and $\mathcal{W}^c(m) - \mathcal{W}^f(m-1)$ on the domain $[4, 99]$ because it is difficult to recognize their sign on that domain unless diagrams are presented for a number of subdomains of $[4, 99]$ employing suitable different scales for $\mathcal{W}^f(m) - \mathcal{W}^c(m+1)$ and $\mathcal{W}^c(m) - \mathcal{W}^f(m-1)$, respectively.

[19] The external stability is irrelevant for $m = 100$ because there is no fringe country that might consider joining or staying outside the coalition.

small. The possibility of full cooperation being self-enforcing calls for a more systematic investigation of the conditions under which that result holds. By definition, the grand coalition is stable if and only if the inequality (17) is satisfied for $m = n$. In the Appendix G, we show that $W^c(n) - W^f(n-1) \geq 0$ is equivalent to

$$F(b, \beta, \delta, \psi, n) := [4(n-1) + n^2(n-3)][(b+\beta)b - n^2\beta\delta\psi^2] - 2n^2(n-3)b(b+\beta) \geq 0. \quad (19)$$

and we prove in the Appendix G

Proposition 4.

(i) *The grand coalition is stable, if and only if $(b, \beta, \psi, \delta, n) \in G$, where $G := \{(b, \beta, \psi, \delta, n)|(19) \text{ is satisfied}\}$. The set G is non-empty and a proper subset of the set of feasible economies.*

(ii) *The strategic complementarity of tax rates is necessary for the grand coalition to be stable.*

(iii) *For all $(b, \beta, \psi, \delta, n) \notin G$ there exists $\tilde{b} > b$ such that*
$$(\check{b}, \beta, \psi, \delta, n) \left\{ \begin{matrix} \in G \\ \notin G \end{matrix} \right\} \iff \check{b} \left\{ \begin{matrix} \geq \\ < \end{matrix} \right\} \tilde{b}.$$

(iv) *For all $(b, \beta, \psi, \delta, n) \in G$ there exists $\tilde{\beta} > \beta$ such that*
$$(b, \check{\beta}, \psi, \delta, n) \left\{ \begin{matrix} \in G \\ \notin G \end{matrix} \right\} \iff \check{\beta} \left\{ \begin{matrix} \leq \\ > \end{matrix} \right\} \tilde{\beta}.$$

(v) *For all $(b, \beta, \psi, \delta, n) \in G$ there exists $\tilde{n} > n$ such that*
$$(b, \beta, \psi, \delta, \check{n}) \left\{ \begin{matrix} \in G \\ \notin G \end{matrix} \right\} \iff \check{n} \left\{ \begin{matrix} \leq \\ > \end{matrix} \right\} \tilde{n}.$$

(vi) *For all $(b, \beta, \psi, \delta, n) \in G$ there exists $\tilde{\delta}\tilde{\psi}^2 > \delta\psi^2$ such that*
$$(b, \beta, \check{\psi}, \check{\delta}, n) \left\{ \begin{matrix} \in G \\ \notin G \end{matrix} \right\} \iff \check{\delta}\check{\psi}^2 \left\{ \begin{matrix} \leq \\ > \end{matrix} \right\} \tilde{\delta}\tilde{\psi}^2.$$

According to Proposition 4 there is a subset of feasible economies in which the grand coalition is stable. In other words, in these economies an all-encompassing IEA is self-enforcing. Proposition 4 does not fully characterize the subset of these economies, but identifies its main properties.

Recall from the discussion of (13) and from (17) and (18) that tax rates are strategic complements, if $b(b+\beta) > n^2\beta\delta\psi^2$. The observation in Proposition 4(ii) that strategic complementarity of tax rates is necessary but not sufficient for the stability of the grand coalition immediately follows from (19). Ceteris paribus, the inequality $b(b+\beta) > n^2\beta\delta\psi^2$

is satisfied the more likely,[20]

- the flatter the household's capital supply curve (the larger b);

- the steeper the firm's capital demand curve (the smaller β);

- the smaller the total number of countries (the smaller n);

- the less severe the climate damage of emissions (the smaller δ);

- the smaller the flow of emissions (the smaller ψ).

It is obvious from the condition $b(b+\beta) > n^2\beta\delta\psi^2$ that tax rates are strategic complements in the absence of the detrimental global externality as is known from the capital tax literature. With increasing climate damage, i.e. with growing parameters n, δ and ψ, the tax rates switch from complements to substitutes. Strategic substitutability is sufficient for unstable grand coalitions. In the extant theoretical literature on the formation of IEAs, emissions taxes are always strategic substitutes, to our knowledge, which is why grand coalitions are always unstable in that literature. Although strategic complementarity is not sufficient for the stability of the grand coalition, the parts (iii) through (vi) of Proposition 4 prove that the conditions for stable grand coalitions are the same in qualitative terms as the conditions for strategic complementarity: For the grand coalition to be stable, the parameter b must be sufficiently large and the parameters β, n, δ and ψ must not be too large. The parameters n, δ and ψ are positively correlated with total climate change and hence their role for the stability of the grand coalition is clear: The less severe the total damage is, the more likely is the grand coalition stable. The role of the parameters b and β for the stability of the grand coalition is less straightforward.

One way to look at the impact of b is to observe that $x_i^1 = (a - 1 - r)/b$ is the amount of good X consumed in period 1 in country i. Presupposing $a > 1 + r$ it follows that $x_i^1 \to 0$ for $b \to \infty$. That is, increasing b shifts increasing shares of production and consumption into period 2. Consequently, our two-period model approaches the one-period model of Ogawa and Wildasin (2009) in which the BAU allocation is efficient.[21] In other words, the potential gains that can be reaped by moving from BAU to the social optimum diminish with increasing parameter b (for large b). We illustrate that finding by plotting in Figure 8 the potential gain measured by the welfare gap $\hat{w} - w_o$ for the Examples 1-3 as a function of the parameter b. The welfare gap function has the following properties: $\lim_{b\to 0}(\hat{w} - w_o) = \lim_{b\to\infty}(\hat{w} - w_o) = 0$ and $\frac{\partial(\hat{w}-w_o)}{\partial b} < 0$ for $b > \tilde{b}$. Hence, the smaller is the

---

[20]Recall that the household's capital supply is given by $U'(\bar{k} - s_i) = 1 + r$ with $\frac{\partial s_i}{\partial r} = -\frac{1}{U''} = \frac{1}{b}$ and firm's capital demand is given by (2) with $\frac{\partial k_i}{\partial r} = \frac{1}{X''} = -\frac{1}{\beta}$.

[21]If b converges to infinity capital is in fixed supply.

welfare loss in BAU compared with the social optimum, the more likely it is that the grand coalition is stable.
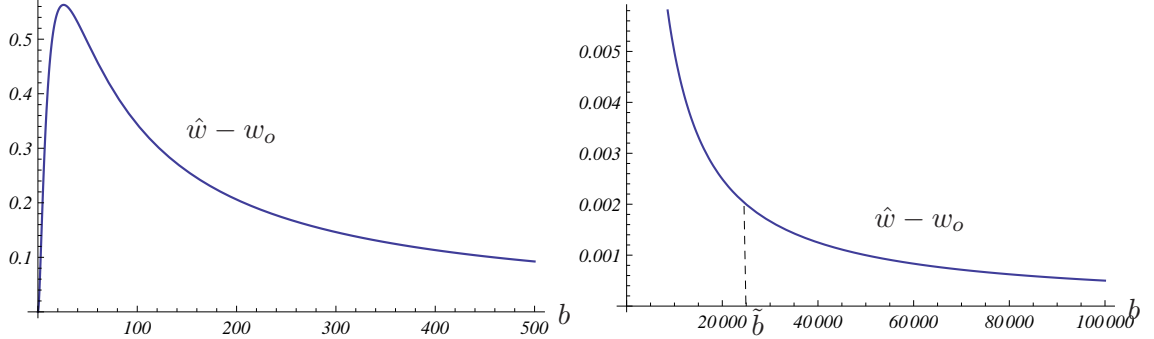

Figure 8: The welfare gap in the Examples 1-3 as a function of the parameter b ($\tilde{b} = 24504.9$)

Ceteris paribus, increasing b reduces total consumption welfare by reducing the utility from consuming good X in the first period. If b gets large the consumer shifts consumption from period 1 into period 2 and uno actu increases the supply of capital which aggravates climate damage, in turn. Thus increasing b (beyond some threshold) reduces all countries' socially optimal welfare, $\hat{w}$. The total capital supply becomes more price-inelastic which implies that abatement becomes more and more expensive. If b is sufficiently large, it is almost impossible to increase welfare by reducing total emissions through taxation below the level of emissions in BAU. We conclude that the relation between the size of parameter b and the stability of the grand coalition corresponds to the conventional view that large and (effective) coalitions are not stable, unless the potential gains from cooperation are small.[22]

The role of the parameter β for the stability of the grand coalition is different. Ceteris paribus, decreasing β increases total consumption welfare by increasing the production - and hence the consumption - of good X in the second period. The consumer shifts consumption from period 2 into period 1 and thus reduces the supply of capital which reduces emissions and climate damage, in turn. We invoke the emissions supply of country i,

$$e_i = \psi k_i = \frac{\psi}{\beta}\left[\alpha - (1 + r + \psi t_i)\right]$$

and verify that the emissions reduction achieved by a small tax increase, $\frac{\partial e_i}{\partial t_i} = -\frac{\psi^2}{\beta}$, is increasing in β, ceteris paribus. Hence an increase in β enhances the abatement costs, and the smaller the abatement costs are the larger the external benefit country i generates by a marginal increase of its tax rate t_i which, in turn, widens the welfare gap $\hat{w} - w_o$. In fact,

[22]See e.g. Barrett (2003) and Finus (2001).

the welfare gap function satisfies $\lim_{\beta \to 0}(\hat{w} - w_o) = \frac{n^2 \delta^2 \bar{k}^4 (\alpha - a + b\bar{k}^2)}{2b^2 (b + n^2 \delta \varepsilon^2)} > 0$, $\lim_{\beta \to \infty}(\hat{w} - w_o) = 0$ and $\frac{\partial(\hat{w} - w_o)}{\partial \beta} < 0$ and is illustrated for Example 3 in Figure 9. The inefficiency of the BAU equilibrium aggravates with increasing parameter β and the stability of the grand coalition becomes more likely. We conclude that the relation between the size of the parameter β and the stability of the grand coalition challenges the generality of the conventional view that large (and effective) coalitions are not stable, unless the potential gains from cooperation are small.
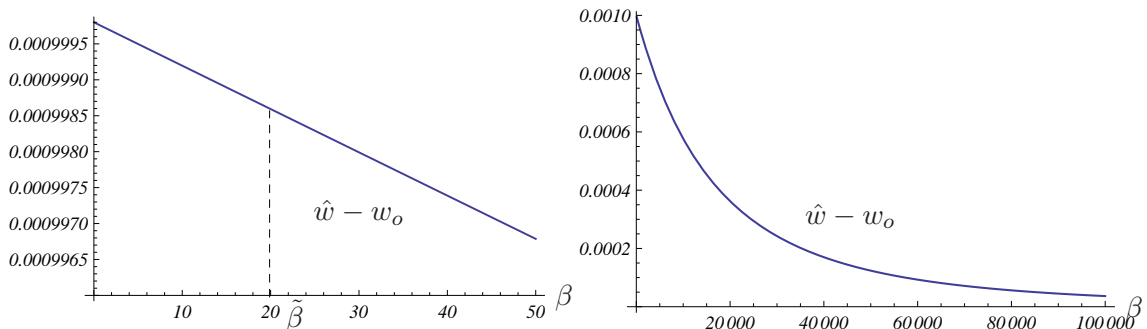

Figure 9: The welfare gap in Example 3 as function of the parameter β ($\tilde{\beta} = 20.5$)

5 Concluding remarks

Our analysis builds on Eichner and Runkel's (2012) two-period model of the world economy in which capital and a consumer good are produced in symmetric countries and traded on (second-period) world markets. Carbon emissions are proportional to capital usage and create global climate damage. While Eichner and Runkel's main message is that the non-cooperative tax competition is inefficient when capital supply elasticities are positive, our focus is on how that inefficiency can be reduced or removed via coalition formation when the coalition and all fringe countries play Nash. Our first result is that, in general, stable coalitions exist with only two or three member countries. That reinforces the disappointing conclusion form the extant theoretical literature on coalition formation: self-enforcing IEAs are not large and (therefore) unable to reduce world emissions significantly below their level in case of global non-cooperation. The second interesting and unexpected finding is that under certain conditions the grand coalition is stable. This result is remarkable because in the extant theoretical literature, we are aware of, all stable coalitions are small, i.e. full cooperation is never self-enforcing. In particular, that holds for the model in Eichner

19

and Pethig (2013b) although it also assumes Nash behavior and accounts for international trade as we do in the present paper. The crucial difference is that in Eichner and Pethig (2013b) the polluting good 'fossil fuel' is a final consumption good whereas here and in Eichner and Runkel (2012) the polluting good 'capital' is an intermediate good used as an input in the production of a consumption good. In the former model tax rates are always strategic substitutes while in the latter they may turn into strategic complements under certain conditions. The case of strategic complements is a necessary condition for the grand coalition to be stable.

Although our model has more 'economic' structure than the basic model of the theoretical literature on self-enforcing IEAs (e.g. Finus 2001), it is still very stylized for the benefit of deriving informative results. Severely restrictive assumptions are a high price to be paid for tractability that we share with the extant theoretical literature on climate coalition formation. We have been able to fully determine analytically the conditions for stable grand coalitions. The second type of stable coalitions consisting of two or three member countries could not be completely characterized. But we proved their existence for a large subset of the space of feasible parameters gibing very strong support to the conjecture that the result is general. Leaving limitations of generality and restrictive assumptions aside, the principle positive message of the present paper is that the prospects for a successful encompassing IEA may not so bleak as suggested by other studies on self-enforcing IEAs that do not model international trade with polluting intermediate goods. It would be desirable, of course, to know whether the conditions fo stable grand coalitions (Proposition 4) are empirically relevant. We refrain from such an assessment based on our analysis above, however, because we think our model is far too simple for that kind of application.

On the agenda for future research are checks of the internal stability of the grand coalition when economies are modeled in a more complex and realistic way using computable large-scale models. Our analysis suggest that such models need to place a strong focus and much detail on international trade in intermediate polluting goods. Another obvious necessity is to drop the assumption of symmetric countries because it abstracts unduly from the real-world complexity and likely underestimates the difficulties of forming stable coalitions.

References

Barrett, S. (2003): *Environment and Statecraft.* Oxford University Press, Oxford UK.

Barrett, S. (1994): Self-enforcing international environmental agreements. *Oxford Eco-*

nomic Papers 46, 878-894.

Bucovetsky, S. and J.D. Wilson (1991): Tax competition with two instruments. *Regional Science and Urban Economics* 21(3), 333-350.

Carraro, C. and D. Siniscalco (1991): Strategies for the international protection of the environment. *CEPR discussion papers* 568.

D'Aspremont, C., Jacquemin, A., Gabszewicz, J.J. and J.A. Weymark (1983): On the stability of collusive price leadership, *Canadian Journal of Economics* 16, 17-25.

Diamantoudi, E. and E. Sartzetakis (2006): Stable international environmental agreements: An analytical approach. *Journal of Public Economic Theory* 8, 247-263.

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

Eichner, T. and R. Pethig (2013b): Is trade liberalization conducive to the formation of climate coalitions. manuscript.

Eichner, T. and M. Runkel (2012): Interjurisdictional spillovers, decentralized policy making and the elasticity of capital supply. *American Economic Review* 102(5), 2349-2357.

Finus, M. (2001): Game Theory and International Environmental Cooperation. Cheltenham, Edward Elgar.

Hoel, M. (1992): International environmental conventions: The case of uniform reductions of emissions. *Environmental and Resource Economics* 2, 141-159.

Keen, M.J. and C. Kotsogiannis (2002): Does federalism lead to excessively high taxes? *American Economic Review* 92(1), 363-370.

Oates, W.E. and R.M. Schwab (1988): Economic competition among jurisdictions: Efficiency enhancing or distortion inducing. *Journal of Public Economics* 35(3): 333-354.

Ogawa, H. and D.E. Wildasin (2009): Think locally, act locally: Spillovers, spillbacks and efficient decentralized policymaking. *American Economic Review* 99(4), 1206-1217.

Rübbelke, D.T.G. and M. Finus 2013: Public good provision and ancillary benefits: The case of climate agreements. *Environmental and Resource Economics*, in press.

Rubio, S. and A. Ulph (2006): Self-enforcing international environmental agreements revisted. *Oxford Economic Papers* 58, 233-263.

Wilson, J.D. (1986): A theory of interregional tax competition. *Journal of Urban Economics* 19(3): 296-315.

Zodrow, G.R. and P. Mieszkowski (1986): Pigou, property taxation, and the underprovision of local public goods. *Journal of Urban Economics* 19(3): 356-370.

Appendix

A: Derivation and comparison of BAU and social optimum

Business as usual. Maximizing $W^i(t_1, \ldots, t_n)$ from (11) with respect to t_i yields the first-order condition

$$W_{t_i}^i = t_i \psi \frac{\partial k_i}{\partial t_i} + (s_i - k_i)\frac{\partial r}{\partial t_i} - D'\left(\sum_j e_j\right) \cdot \frac{\partial \left(\sum_j e_j\right)}{\partial t_i} = 0, \tag{A1}$$

where

$$\frac{\partial r}{\partial t_i} = \frac{\partial r}{\partial t_j} = -\frac{\psi b}{n(b + \beta)} < 0, \tag{A2}$$

$$\frac{\partial k_i}{\partial t_i} = -\frac{\psi[(n-1)b + n\beta]}{(b + \beta)n\beta} < 0, \quad \frac{\partial k_i}{\partial t_j} = \frac{\psi b}{n\beta(b + \beta)} > 0, \tag{A3}$$

$$\frac{\partial s_i}{\partial t_i} = \frac{\partial s_i}{\partial t_j} = -\frac{\psi}{n(b + \beta)} < 0, \quad \frac{\partial \left(\sum_k s_k\right)}{\partial t_i} = \frac{\partial \left(\sum_k s_k\right)}{\partial t_j} = -\frac{\psi}{n(b + \beta)} < 0. \tag{A4}$$

Making use of (A2) - (A4) the first-order condition (A1) can be rearranged to read

$$W_{t_i}^i = -\frac{t_i[(n-1)b + n\beta]}{n\beta} - \frac{b}{n\beta}\left(t_i - \frac{\sum_j t_j}{n}\right) + \frac{\delta \psi n}{b + \beta}\left(\zeta - \frac{\psi}{n}\sum_j t_j\right) = 0. \tag{A5}$$

Solving (A5) with respect to t_i establishes (12).

Social optimum. Maximizing $\sum_j W^j(t_1, \ldots, t_n)$ with respect to t_i we obtain

$$\sum_j W_{t_i}^j = \sum_j \left[t_j \psi \frac{\partial k_j}{\partial t_i} + (s_j - k_j)\frac{\partial r}{\partial t_i} - D'\left(\sum_j e_j\right) \cdot \frac{\partial \left(\sum_j e_j\right)}{\partial t_i} \right] = 0. \tag{A6}$$

Making use of (A2) - (A4) in (A6) and the symmetry condition, we obtain $\hat{t} = \frac{\delta \psi \zeta n^2}{b + \beta + n^2 \delta \psi^2}$.

Comparison of BAU and social optimum. Comparing BAU and social optimum levels reveals

$$\hat{t} - t_o = \frac{n^2 \delta \psi \zeta (n-1)(b + \beta)^2}{(b + \beta + n^2 \delta \psi^2)[b^2(n-1) + b\beta(2n-1) + n\beta(\beta + n\delta\psi^2)]} > 0,$$

$$\hat{w} - w_o = \frac{(n-1)^2 n^4 (b + \beta)^2 \delta^2 \psi^4 \zeta^2}{2(b + \beta + n^2 \delta \psi^2)[b^2(n-1) + b\beta(2n-1) + n\beta(\beta + n\delta\psi^2)]} > 0.$$

$\hat{k} = \hat{s} < k_o = s_o$ and $\hat{r} < r_o$ follows from accounting for $\hat{t} > t_o$ in (8) - (10).

B: Derivation of the coalition's best reply

We set $t_i = t_c$ for all $i \in C$ and $t_i = t_f$ from the outset. Then the coalition maximizes joint welfare

$$m \left[U(\bar{k} - s_c) + s_c + X(k_c) + r_c(s_c - k_c) - D \left(\sum_j e_j \right) \right], \qquad \text{(B1)}$$

where

$$r = \frac{(\alpha - 1)b + (a - 1)\beta - b\beta\bar{k}}{b + \beta} - \frac{b\psi}{b + \beta} \cdot \frac{mt_c + (n - m)t_f}{n}, \qquad \text{(B2)}$$

$$k_c = \frac{1}{b + \beta} \left[\zeta - \frac{(b + \beta)\psi}{\beta} \cdot t_c + \frac{b\psi}{\beta} \cdot \frac{mt_c + (n - m)t_f}{n} \right], \qquad \text{(B3)}$$

$$s_c = \frac{1}{b + \beta} \left(\zeta - \psi \cdot \frac{mt_c + (n - m)t_f}{n} \right), \qquad \text{(B4)}$$

$$\sum e_j = \frac{n\psi}{b + \beta} \left(\zeta - \psi \cdot \frac{mt_c + (n - m)t_f}{n} \right) \qquad \text{(B5)}$$

with respect to t_c. The pertaining first-order condition is

$$t_c \psi \frac{\partial k_c}{\partial t_c} + (s_c - k_c) \frac{\partial r_c}{\partial t_c} - D' \left(\sum_j e_j \right) \cdot \frac{\partial \sum_j e_j}{\partial t_c} = 0 \qquad \text{(B6)}$$

or equivalently

$$\frac{t_c[(n - m)b + n\beta]}{n\beta} = -\frac{mb}{n\beta} \left(t_c - \frac{mt_c + (n - m)t_f}{n} \right) + \frac{m\delta\psi n}{(b + \beta)} \left(\zeta - \psi \cdot \frac{mt_c + (n - m)t_f}{n} \right). \text{ (B7)}$$

Solving (B7) with respect to t_c yields (13).

C: Feasible constraints on the parameter space

Throughout the paper we restrict our attention to economies with positive interest rates and emissions. Due to $\min \{r_o, r^*, \hat{r}\} = \hat{r}$ and $\min \{e_o, e_c^*, e_f^*, \hat{e}\} = \min \{e_c^*, \hat{e}\}$ we require $\hat{r} > 0$, $e_c^* > 0$ and $\hat{e} > 0$ or equivalently

$$b(\alpha - 1) - (1 - \alpha + b\bar{k})(\beta + n^2\delta\psi^2) > 0, \qquad \text{(C1)}$$

$$b[n(n - 1) - m(m - 1)] + n^2\beta - (m - 1)(n - m)n^2\delta\psi^2 > 0, \qquad \text{(C2)}$$

$$\alpha - a + b\psi > 0. \qquad \text{(C3)}$$

D: Proof of Proposition 1:

Solving (13) and (14) with respect to t_c and t_f yields

$$t_c^* = \frac{mn^3\beta\delta\psi\zeta}{N}, \tag{D1}$$

$$t_f^* = \frac{n^3\beta\delta\psi\zeta}{N}, \tag{D2}$$

$$mt_c^* + (n-m)t_f^* = \frac{(n-m+n^2)n^3\beta\delta\psi\zeta}{N}, \tag{D3}$$

where $N := b(b+\beta)[n(n-1)-m(m-1)] + n^2\beta[b+\beta+(n-m+m^2)\delta\varphi^2] > 0$.

Comparing t_c^*, t_f^* and t_o yields

$$t_c^* - t_f^* = \frac{(m-1)n^3\beta\delta\psi\zeta}{N}, \tag{D4}$$

$$t_c^* - t_o = \frac{(m-1)n^2\beta\delta\psi\zeta\,[b(b+\beta)(n^2-n+m)+\beta(b+\beta)n^2+(n-m)n^2\beta\delta\psi^2]}{[b(b+\beta)(n-1)+\beta(b+\beta)n+n^2\beta\delta\psi^2]\,N}, \tag{D5}$$

$$t_f^* - t_o = \frac{(m-1)mn^2\beta\delta\psi\zeta\,[b(b+\beta)-n^2\beta\delta\psi^2]}{[b(b+\beta)(n-1)+\beta(b+\beta)n+n^2\beta\delta\psi^2]\,N}. \tag{D6}$$

(D4) and (D5) imply $t_c^* > t_f^*$ and $t_c^* > t_o$.

Subtracting $W^f(t_c, t_f; m)$ from $W^c(t_c, t_f; m)$ yields

$$W^c(t_c, r_f; m) - W^f(t_c, t_f; m) = \frac{(t_f^2 - t_c^2)\psi^2}{2\beta}. \tag{D7}$$

Accounting for $t_c^* > t_f^*$ in (D6) establishes $W^f(t_c^*, t_f^*; m) > W^c(t_c^*, t_f^*; m)$.

E: Proof of Proposition 2:

Differentiation of $t_c^* = \mathcal{T}^c(m)$ and $t_f^* = \mathcal{T}^f(m)$ from (D1) and (D2) with respect to m yields

$$\mathcal{T}_m^c = \frac{n^3\beta\delta\psi\zeta\,[b(b+\beta)[n(n-1)+m^2]+\beta(b+\beta)n^2+(n-m^2)n^2\beta\delta\psi^2]}{N^2}, \tag{E1}$$

$$\mathcal{T}_m^f = \frac{(2m-1)n^3\beta\delta\psi\zeta\,[b(b+\beta)-n^2\beta\delta\psi^2]}{N^2}, \tag{E2}$$

$$\frac{\partial\left[m\mathcal{T}^c+(n-m)\mathcal{T}^f\right]}{\partial m} = \frac{(2m-1)n^5\beta^2\delta\psi\zeta(b+\beta)^2}{N^2} > 0. \tag{E3}$$

If tax rates are strategic complements $[b(b+\beta)-n^2\beta\delta\psi^2 > 0]$, then $\mathcal{T}_m^c > 0$ and $\mathcal{T}_m^f > 0$. If tax rates are strategic substitutes $[b(b+\beta)-n^2\beta\delta\psi^2 < 0]$, then $\mathcal{T}_m^f < 0$. In addition, from the parameter constraint (C2) we infer

$$(b+\beta)n^2 > b\,[n+m(m-1)] + (m-1)(n-m)n^2\delta\psi^2$$

$$\iff \quad b+\beta+(n-m^2)\delta\psi^2 > \frac{b\,[n+m(m-1)]}{n^2} + \left[n-m^2+(m-1)(n-m)\right]\delta\psi^2$$

$$\iff \quad b+\beta+(n-m^2)\delta\psi^2 > \frac{b\,[n+m(m-1)]}{n^2} + [n+1-2m]\,m\delta\psi^2. \tag{E4}$$

Making use of $b + \beta + (n - m^2)\delta\psi^2 > 0$ if $m < \frac{n+1}{2}$ we obtain $\mathcal{T}_m^c > 0$ if $m < \frac{n+1}{2}$.

Next, we insert (D1) and (D2) in $W^c(t_c, t_f; m)$ and $W^f(t_c, t_f; m)$ and differentiate with respect to m to get after rearrangement of terms

$$
\begin{aligned}
\mathcal{W}_m^c &= \frac{n^6\beta\delta^2\psi^4\zeta^2 \left[b(b+\beta)[nm(n+1) - n^2 - m^3]\right]}{N^3} \\
&+ \frac{n^6\beta\delta^2\psi^4\zeta^2 \left[n^2\beta[(b+\beta)(m-1) + m(m^2-n)\delta\psi^2]\right]}{N^3},
\end{aligned} \tag{E5}
$$

$$
\mathcal{W}_m^f = \frac{(2m-1)n^6\beta\delta^2\psi^4\zeta^2 \left[b(b+\beta)(n^2-1) + \beta n^2(b+\beta) + n^2\beta\delta\psi^2\right]}{N^3} > 0. \tag{E6}
$$

Again, we use (C2) to obtain

$$
\begin{aligned}
(b+\beta) &> \frac{b[n + m(m-1)]}{n^2} + (m-1)(n-m)\delta\psi^2 \\
\Longleftrightarrow \quad (b+\beta)(m-1) + m(m^2-n)\delta\psi^2 &> \frac{b(m-1)[n + m(m-1)]}{n^2} \\
&\quad + \left[(m-1)^2(n-m) + m(m^2-n)\right]\delta\psi^2 \\
\Longleftrightarrow \quad (b+\beta)(m-1) + m(m^2-n)\delta\psi^2 &> \frac{b(m-1)[n + m(m-1)]}{n^2} \\
&\quad + \left[mn(m-3) + n^2(n-m)\right]\delta\psi^2. \tag{E7}
\end{aligned}
$$

Observe that $mn(m-3) + n^2(n-m) = mn(m-1) + n^2\left(n - m - \frac{2m}{n}\right) > 0$ for all m and $n \geq 3$. Then accounting for $(b+\beta)(m-1) + m(m^2-n)\delta\psi^2 > 0$ and

$$
nm(n+1) - n^2 - m^3 = m(n^2 - m^2) - n(n-m) = (n-m)[m(n+m) - n] > 0
$$

in (E5) establishes $\mathcal{W}_m^c > 0$ for all $n \geq 3$.

F: Proof of Proposition 3:

Define the parameter subspaces:[23]

$$
\begin{aligned}
M^b &= \{(b,\beta,\psi,\delta,n)|\, 0.1 \le b \le 100000, \beta = 10, \psi = 1, \delta = 0.001, n = 100\} \\
M^\beta &= \{(b,\beta,\psi,\delta,n)|\, b = 0.1, 0.1 \le \beta \le 100000, \psi = 1, \delta = 0.001, n = 100\} \\
\breve{M}^\beta &= \{(b,\beta,\psi,\delta,n)|\, b = 100000, 0.1 \le \beta \le 100000, \psi = 1, \delta = 0.001, n = 100\} \\
M^\psi &= \{(b,\beta,\psi,\delta,n)|\, b = 0.1, \beta = 10, 0.1 \le \psi \le 2, \delta = 0.001, n = 100\} \\
\breve{M}^\psi &= \{(b,\beta,\psi,\delta,n)|\, b = 100000, \beta = 10, 0.5 \le \psi \le 100, \delta = 0.001, n = 100\} \\
M^\delta &= \{(b,\beta,\psi,\delta,n)|\, b = 0.1, \beta = 10, \psi = 1, 0.0001 \le \delta \le 0.1, n = 100\} \\
\breve{M}^\delta &= \{(b,\beta,\psi,\delta,n)|\, b = 100000, \beta = 10, \psi = 1, 0.01 \le \delta \le 100, n = 100\} \\
M^n &= \{(b,\beta,\psi,\delta,n)|\, b = 0.1, \beta = 10, \psi = 1, \delta = 0.001, 10 \le n \le 200\} \\
\breve{M}^n &= \{(b,\beta,\psi,\delta,n)|\, b = 100000, \beta = 10, \psi = 1, \delta = 0.001, 10 \le n \le 200\} \\
M &:= M^b \cup M^\beta \cup \breve{M}^\beta \cup M^\psi \cup \breve{M}^\psi \cup M^\delta \cup \breve{M}^\delta \cup M^n \cup \breve{M}^n
\end{aligned}
$$

It can be shown that the stable coalition is invariant with respect to changes of the parameter a, α and $\bar{k}$. Exemplarily we show stability for the set M^b. In the Table 1 we have determined the stable coalition for some values of b. Recall that the coalition of size m is stable if $\Phi(m) > 0$ (internal stability) and $\Phi(m+1) \le 0$ (external stability) where $\Phi(m) := \mathcal{W}^c(m) - \mathcal{W}^f(m-1)$ is denoted as stability function. The three-dimensional graph in Figure 9 illustrates how $\Phi(m)$ changes if m and b are varied. Thus, Figure 9 proves that the stable coalition is two if the parameters belong to the set M^b and $0.1 \le b \le 1$. Similar graphs can be plotted for the set M^b and $1 \le b \le 100000$ or for the other sets (if the parameter β, ψ, δ or n is varied) introduced above. All these graphs show that the stable coalition consists of either two or three countries. Due to space constraints these three-dimensional graphs are omitted.

parameters					stable coalition
n	b	β	ψ	δ	
100	**0.1**	10	1	0.001	2
100	**1**	10	1	0.001	2
100	**10**	10	1	0.001	2
100	**1000**	10	1	0.001	3
100	**100000**	10	1	0.001	3

---

[23]The reason for the low values of ψ in M^ψ and δ in M^δ is that for higher values the coalition's emissions become negative.
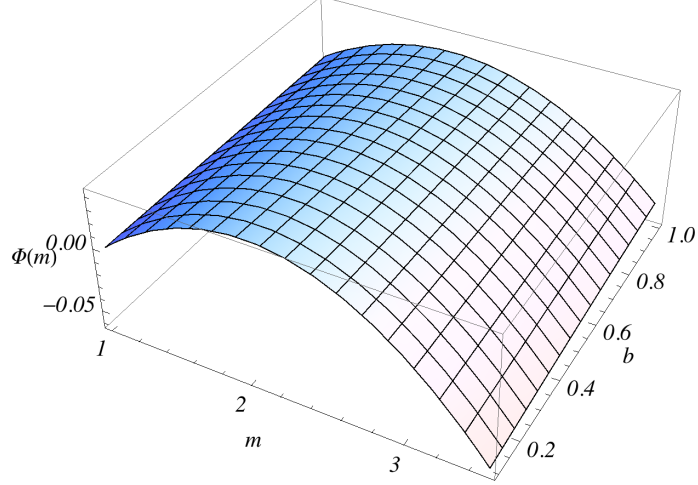
Figure 10: Coalition stability for the set M^b with $0.1 \leq b \leq 1$

G: Proof of Proposition 4:

Making use of $\hat{w} = \frac{(a-\alpha)^2 + 2\alpha b \bar{k} + (2a - b\bar{k})\bar{k}(\beta + n^2 \delta \psi^2)}{2(b + \beta + n^2 \delta \psi^2)}$ and inserting t_c^* and t_f^* from (D1) and (D2) in $W^f(t_c, t_m; m)$ yields after rearrangement of terms

$$
\begin{aligned}
\hat{w} \quad &- \quad \mathcal{W}^f(n-1) \\
&= \quad \frac{(n-1)n^4 \delta^2 \psi^4 \zeta^2 F(b, \beta, \psi, \delta, n)}{2(b + \beta + n^2 \delta \psi)\left[2b^2(n-1) + b\beta(n^2 + 2n - 2) + n^2\beta[\beta + (n^2 + 2 - 2n)\delta\psi^2]\right]^2}, \quad \text{(G1)}
\end{aligned}
$$

where

$$
\begin{aligned}
F(b, \beta, \psi, \delta, n) \quad =: \quad & 4b^2(n-1) - b\beta[n^2(n-3) - 4(n-1)] - n^2\beta^2(n-3) \\
& - \quad n^2\beta\delta\psi^2[n^2(n-3) + 4(n-1)]. \quad \text{(G2)}
\end{aligned}
$$

Since the denominator in (G1) is positive, setting $\hat{w} - \mathcal{W}^f(n-1) \gtreqless 0$ is equivalent to $F(b, \beta, \psi, \delta, n) \gtreqless 0$.

Further rearranging yields

$$
\begin{aligned}
F \quad &= \quad 4b(b+\beta)(n-1) - 4(n-1)n^2\beta\delta\psi^2 - n^2(n-3)(b\beta + \beta^2 + n^2\beta\delta\psi^2) \\
&= \quad 4(n-1)[(b+\beta)b - n^2\beta\delta\psi^2] - n^2(n-3)[(b+\beta)\beta + n^2\beta\delta\psi^2] \\
&= \quad [4(n-1) + n^2(n-3)][(b+\beta)b - n^2\beta\delta\psi^2] - 2n^2(n-3)b(b+\beta) \quad \text{(G3)}
\end{aligned}
$$

and hence we obtain

$$F \gtrless 0 \quad \Longleftrightarrow \quad \frac{(b+\beta)}{\beta}\left[4b - \frac{n^2(n-3)\beta}{(n-1)}\right] \gtrless n^2\delta\psi^2\left[\frac{n^2(n-3)}{n-1} + 4\right]. \tag{G4}$$

Observe that

$$F(b, \beta, \psi, \delta, 2) = 4(b+\beta)^2.$$

The Propositions 4(iii) through 4(vi) follow from (G4). Part (iv) of Proposition is less straightforward than the parts (iii), (v) and (vi). But simple calculation yields

$$\frac{\partial}{\partial\beta}\left\{\frac{b+\beta}{\beta}\left[4b - \frac{n^2(n-3)\beta}{n-1}\right]\right\} = -\left[\frac{4b^2}{\beta^2} + \frac{n^2(n-3)}{n-1}\right]$$

which is negative for $n \geq 3$.