



Volkswirtschaftliche Diskussionsbeiträge  
Discussion Papers in Economics

No. 165-14

July 2014

Thomas Eichner · Rüdiger Pethig

**Self-enforcing international environmental agreements  
and trade: taxes versus caps**

Universität Siegen  
Fakultät III  
Wirtschaftswissenschaften, Wirtschaftsinformatik und Wirtschaftsrecht  
Fachgebiet Volkswirtschaftslehre  
Hölderlinstraße 3  
D-57068 Siegen  
Germany

<http://www.wiwi.uni-siegen.de/vwl/>

ISSN 1869-0211

Available for free from the University of Siegen website at  
<http://www.wiwi.uni-siegen.de/vwl/research/diskussionsbeitraege/>

Discussion Papers in Economics of the University of Siegen are indexed in RePEc  
and can be downloaded free of charge from the following website:  
<http://ideas.repec.org/s/sie/siegen.html>

# Self-enforcing international environmental agreements and trade: taxes versus caps\*

Thomas Eichner

Department of Economics, University of Hagen

Rüdiger Pethig

Department of Economics, University of Siegen

## Abstract

This paper studies within a multi-country model with international trade the stability of international environmental agreements (IEAs) when countries regulate carbon emissions either by taxes or caps. Regardless of whether coalitions play Nash or are Stackelberg leaders the principal message is that the choice of caps or taxes matters. International trade and tax regulation are necessary conditions for the existence of the *encompassing* self-enforcing IEA, and that the latter is attained the more likely, the less severe the climate damage. Hence, cap regulation is inferior to tax regulation insofar as in case of the former there exist no large and *effective* self-enforcing IEAs, in particular not the encompassing self-enforcing IEA. Further results are that for the formation of encompassing self-enforcing IEAs it does not matter whether climate coalitions play Nash or are Stackelberg leaders or whether fossil fuel is modeled as a consumer good or an intermediate good.

JEL classification: C72, F02, Q50, Q58

Key words: cap, tax, international trade, self-enforcing environmental agreements, Nash, Stackelberg

---

\*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 substantial reduction of global carbon emissions necessary to stabilize the world climate at safe levels calls for an effective international environmental agreement (IEA). The first legally binding IEA, the Kyoto Protocol, stipulated rather unambitious commitments for a small number of countries and therefore accomplished only little more than global non-cooperation (Buchner et al. 2002). The Kyoto Protocol expired in 2012, and the prospects are uncertain for reaching a new IEA with many signatories and strong emission reductions. The tedious practical negotiations on the one hand and the threat of serious climate damage in case of insufficient action on the other hand require further efforts to improve our understanding of the conditions for successful and effective IEAs.

The economic literature on the formation of 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. Self-interest is captured by the concept of self-enforcing agreements which requires that no signatory has an incentive to defect from the IEA and no non-signatory has an incentive to join. The main objective of the present paper is to investigate the conditions for the stability of climate coalitions<sup>1</sup> in the world economy with international trade, when all non-cooperative countries and the coalition choose emission taxes as their climate policy instrument.

Early studies on self-enforcing IEAs employ a simple model, called *basic model* hereafter, of identical countries in autarky with national climate policies taking the form of emission cap (rather than emission tax) regulation (Finus 2003). That model specifies the welfare of each country as a function that is increasing and concave in domestic and progressively declining in aggregate greenhouse gas emissions.<sup>2</sup> Some studies using the basic model portray coalitions as Stackelberg leaders, or as *Stackelberg coalitions*, for short, and others model them as Nash players, or as *Nash coalitions*, for short.<sup>3</sup> It turns out that no large stable coalitions exist regardless of whether Nash or Stackelberg coalitions are assumed.<sup>4</sup>

One of the reasons why the basic model offers only limited insights in the determinants

---

<sup>1</sup>In the present paper the terms self-enforcing IEA and stable (climate) coalition are used synonymously.

<sup>2</sup>Throughout the paper we assume climate damage to be progressively increasing in aggregate emissions. Linear damage functions are also considered in the literature. See e.g. Finus (2001).

<sup>3</sup>Stackelberg coalitions are assumed e.g. by Barrett (1994), Diamantoudi and Sartzetakis (2006) and Rubio and Ulph (2006) and Nash coalitions e.g. by Carraro and Siniscalco (1991), Hoel (1992), Finus (2001), Finus and Rübbelke (2013).

<sup>4</sup>In the basic model, stable coalitions have at most two members (Carraro and Siniscalco 1991) with Nash coalitions and at most four members in case of Stackelberg coalitions, if negative emissions are excluded (Diamantoudi and Sartzetakis 2006; Rubio and Ulph 2006) - no matter how large the total number of countries is.

of coalition formation is the neglect of international trade.<sup>5</sup> Eichner and Pethig (2012, 2013) add structure to the national economies of the basic model and allow for trade of fossil fuel and a composite consumer good on world markets. In the case of cap-and-trade regulation they find that stable coalitions are very small and hence ineffective in reducing climate damage, if coalitions play Nash (Eichner and Pethig 2012), and that stable coalitions may be relatively large but also ineffective, if coalitions are modeled as Stackelberg leaders (Eichner and Pethig 2013). These studies appear to suggest that international trade does not improve the conditions for the formation of *effective* stable climate coalitions - neither with Nash nor with Stackelberg coalitions. We will demonstrate, however, that such a conclusion is not robust.

The analytical framework of the present paper is, like that of Eichner and Pethig (2012, 2013), the basic model of coalition formation extended by international trade, but it deviates from these studies by replacing caps with emission taxes. The aim is to investigate how that modification changes the conditions for stable climate coalitions. At first cursory sight this objective may seem to be odd because both instruments are known to be equivalent in perfectly competitive economies in which all agents including governments are price takers. That equivalence holds, indeed, in the benchmark autarky regime of our model. But in the trade regime governments and coalitions choose their taxes or caps strategically by accounting for the terms-of-trade effects of their own regulatory action. In various analytical settings without coalitions Ulph (1996), Pizer (2002), Ishikawa and Kiyono (2006), Strand (2010), Karp and Zhang (2012), Kiyono and Ishikawa (2013) and others find that caps and taxes are not equivalent if governments fix their taxes or caps strategically.<sup>6</sup> In a setting similar to our scenario of global non-cooperation (business as usual, BAU) Kiyono and Ishikawa (2013) identify a carbon-leakage effect that works under tax regulation only and causes the global BAU emissions to be larger with taxes than with caps.

To our knowledge there is no literature on the strategic non-equivalence of caps and taxes that addresses the issue of stable coalitions. The present paper aims to fill that gap. In Section 2 we set up the model and derive the differential impact of taxes and caps in BAU confirming the leakage effect described by Kiyono and Ishikawa (2013). Answering the question as to how the non-equivalence of taxes and caps identified in BAU changes the determinants of coalition stability requires working through the model with taxes assuming

---

<sup>5</sup>In recent years the basic model has been modified and extended in various directions. See e.g. Hoel and Schneider (1997), Na and Shin (1998), Kolstad (2007), de Zeeuw (2008), Carbone et al. (2009) and Hong and Karp (2012).

<sup>6</sup>Other reasons for the non-equivalence between taxes and caps are uncertainty and asymmetric information (Weitzman 1974), imperfect competition (Mansur 2013), incomplete enforcement (Montero 2002) and lobbying and political pressure (Finkelshtain and Kislev 1997).

Type of regulation		Cap and trade		Emission taxes	
		Trade	Autarky	Trade	Autarky
Type of coalition		1	2	3	4
Nash	1	Eichner and Pethig (2012) $m^* \leq 2$	<i>Basic model</i> of literature $m^* \leq 2$	Present paper: $m^* \leq 3$ <b>or</b> $m^* \leq 3$ and $m^* = n$	Same as Box (2.1)
Stackelberg	2	Eichner and Pethig (2013) $m^* \leq 0.6 \cdot n$	<i>Basic model</i> of literature $m^* \leq 4$	Present paper: $m^* \leq 3$ <b>or</b> $m^* \leq 3$ and $m^* = n$	Same as Box (2.2)

Table 1: Cap versus tax regulation: Overview<sup>7</sup>

first Nash coalitions (Section 3) and then Stackelberg coalitions (Section 4).

The results of tax regulation and their comparison with cap regulation are summarized in Table 1 in a self-explanatory way. The striking and unexpected result is that in case of international trade and either Nash or Stackelberg coalitions there are certain conditions to be specified later under which the grand coalition is stable, if climate policy takes the form of emission taxation. Another important result is that international trade is crucial for the stability of grand coalitions because all stable grand coalitions that may exist become unstable when borders are closed (Section 5). Thus, we identify both tax regulation and trade as necessary conditions for the stability of grand coalitions while modeling coalitions as either Nash players or Stackelberg leaders turns out to be inessential for the stability of grand coalitions. Our results challenge the disturbing view of some previous literature that the formation of stable and effective climate coalitions is bound to fail. On the other hand, the robustness of our results is unclear, as already mentioned, due to restrictive assumptions made for reasons of tractability.

One of our simplifying assumptions is the structure of the economy of individual countries adopted from Eichner and Pethig (2012, 2013) which is special because fossil fuel is assumed to be a final consumer good. In real economies, many fossil energy taxes and/or cap and trade schemes, e.g. that of the EU, are imposed on industries. To see whether the structure of the economy matters, we modify in Section 6 the model analyzed so far by treating fossil fuel as an intermediate good. The modified model of Section 6 turns out to be isomorphic to the model in the first part of the paper so that Table 1 applies to both models,

<sup>7</sup> $m^*$  = number of countries in the stable coalition;  $n$  = total number of countries; Box  $(i, j)$  = box in the  $i$ th column and  $j$ th row of Table 1.

qualitatively speaking. Hence our results are not sensitive with regard to both approaches of modeling fossil fuel.

## 2 The theoretical framework

**The model.**<sup>8</sup> The world economy consists of  $n$  identical countries. Each country produces two consumer goods. The first is a standard composite good, called *good X* (quantity  $x_i$ ) and the second is a fossil energy carrier (quantity  $e_i$ ), e.g. oil, gas or coal extracted from domestic fossil reserves. We refer to that good simply as *fuel*. Each country's production technology is represented by the production possibility frontier<sup>9</sup>

$$x_i^s = T(e_i^s) \quad i = 1, \dots, n, \quad (1)$$

where the function  $T$  is decreasing and strictly concave in  $e_i^s$ . The transformation function (1) implies that both commodities are produced by means of domestic productive factors (e.g. labor and capital) whose endowments are given. The utility<sup>10</sup>

$$V(e_i^d) + x_i^d - D\left(\sum_j e_j^d\right) \quad (2)$$

of the representative consumer of country  $i$  is additive separable in all arguments and linear in the consumption  $x_i^d$  of good  $X$ .  $V$  is increasing and concave, and  $D$  is increasing and convex in its argument. The consumption of fuel generates the greenhouse gas carbon dioxide whose emission is proportional to fuel consumption. Emission units are chosen such that  $e_i^d$  denotes both fuel demanded by consumer  $i$  and carbon emissions from burning fuel. There is no abatement technology for emission reduction. The function  $D$  captures the climate damage caused by worldwide carbon emissions from burning fuel.

For the sake of more specific results, we will specify the functions  $T$ ,  $V$  and  $D$  from (1) and (2) throughout the paper by the following quadratic functional forms:

$$T(e_i^s) = \bar{x} - \frac{\alpha}{2}(e_i^s)^2, \quad V(e_i^d) = ae_i^d - \frac{b}{2}(e_i^d)^2, \quad D\left(\sum_j e_j^d\right) = \frac{\delta}{2}\left(\sum_j e_j^d\right)^2, \quad (3)$$

where  $\bar{x}$ ,  $a$ ,  $b$ ,  $\alpha$  and  $\delta$  are positive parameters.

---

<sup>8</sup>The equations (1)-(5) below describe the basic model of Eichner and Pethig (2013) who then proceed with taking the domestic emissions (= fuel demand  $e_i^d$ ) as politically determined emission caps. Here we deviate from their study, beginning with equation (6), by turning to tax rather than cap regulation.

<sup>9</sup>The superscript  $s$  indicates quantities supplied. Upper-case letters denote functions. Subscripts attached to them indicate partial derivatives.

<sup>10</sup>The superscript  $d$  indicates quantities demanded.

There are perfectly competitive world markets for good  $X$  (price  $p_x \equiv 1$ ) and for fuel (producer price  $p$ ), and the markets are in equilibrium if

$$\sum_j x_j^s = \sum_j x_j^d \quad \text{and} \quad \sum_j e_j^s = \sum_j e_j^d. \quad (4)$$

Taking prices as given, the (aggregate) producer  $i$  maximizes profits  $x_i^s + pe_i^s$  subject to (1) which yields the first-order condition

$$p = -T'(e_i^s) \quad \text{or} \quad e_i^s = \frac{p}{\alpha} \quad \text{for} \quad i = 1, \dots, n. \quad (5)$$

Equation (5) represents the (inverse) fuel supply function.

Emissions are regulated in each country by means of an emission tax at rate  $t_i$  such that the consumer price of fuel is  $p + t_i$  in country  $i$ . The representative consumer  $i$  ignores the impact of her emissions on climate damage and maximizes her (consumption) utility  $V(e_i^d) + x_i^d$  subject to her budget constraint

$$x_i^d + (p + t_i)e_i^d = y_i, \quad \text{where} \quad y_i := x_i^s + pe_i^s + t_ie_i^d \quad (6)$$

is consumer  $i$ 's income (= profit income plus recycled tax revenues). Utility maximization yields the fuel demand function

$$V'(e_i^d) = p + t_i \quad \text{or} \quad e_i^d = \frac{a - p - t_i}{b} \quad \text{for} \quad i = 1, \dots, n. \quad (7)$$

Inserting (5) and (7) in the fuel equilibrium condition from (4) determines the unique fuel price

$$p = \frac{\alpha}{(\alpha + b)n} \left( na - \sum_j t_j \right). \quad (8)$$

Making use of (8) in (5) and (7) establishes the fuel supply and demand, respectively,

$$e_i^s = \frac{1}{(\alpha + b)n} \left( na - \sum_j t_j \right) \quad \text{and} \quad e_i^d = \frac{a}{b} - \frac{t_i}{b} - \frac{\alpha}{(\alpha + b)bn} \left( na - \sum_j t_j \right) \quad (9)$$

and

$$\sum_j e_j^s = \sum_j e_j^d = \frac{1}{(\alpha + b)} \left( na - \sum_j t_j \right). \quad (10)$$

When combined with (9), the equations

$$x_i^s = T(e_i^s) \quad \text{and} \quad x_i^d = T(e_i^s) + p(e_i^s - e_i^d) \quad (11)$$

determine the equilibrium supplies and demands on the market for good  $X$ .



Finally, we account for (11) in (2) to get

$$W^i(t_1, \dots, t_n) := V(e_i^d) + T(e_i^s) + p(e_i^s - e_i^d) - D\left(\sum_j e_j^d\right). \quad (12)$$

Since the variables  $p$ ,  $e_i^s$  and  $e_i^d$  on the right side of (12) are determined by (8) and (9), equation (12) represents the equilibrium welfare of country  $i$  for any given tax profile  $(t_1, \dots, t_n)$ .

**Business as usual (BAU).** For later use as a benchmark, we briefly characterize the non-cooperative Nash equilibrium. The government of country  $i$  chooses the tax rate  $t_i$  that maximizes  $W^i(t_i, \dots, t_n)$  for given tax rates  $(t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_n)$  of the other countries. The associated first-order condition  $W_{t_i}^i = 0$  implicitly determines country  $i$ 's best reply function<sup>11</sup>

$$t_i = \tilde{R}\left(\sum_{j \neq i} t_j\right) = \frac{n^3 ab \delta}{N^o} + \frac{\alpha(\alpha + b) - n^2 b \delta}{N^o} \sum_{j \neq i} t_j, \quad (13)$$

where  $N^o := b^2 n^2 + (n^2 - 1)\alpha^2 + b[(2n^2 - 1)\alpha + n^2 \delta] > 0$ . The reaction function is upward [downward] sloping and hence characterizes the countries' tax rates as strategic complements [substitutes], if and only if  $\alpha(\alpha + b) > n^2 b \delta$  [ $\alpha(\alpha + b) < n^2 b \delta$ ].

Since  $\frac{dt_i}{d(\sum_{j \neq i} dt_j)} < 1$  and constant, there is a unique Nash equilibrium, say  $(t_{1o}, \dots, t_{no})$ . With all countries being alike, the Nash equilibrium is characterized by  $t_{1o} = \dots = t_{no} \equiv t_o$  and hence also by  $e_{1o}^s = \dots = e_{no}^s = e_{1o}^d = \dots = e_{no}^d \equiv e_o$ . Making use of the symmetry assumption in (13) yields the BAU tax rate  $t_o := \frac{n^2 ab \delta}{n^2 b + (n-1)\alpha^2 + b[(2n-1)\alpha + n^2 \delta]}$  which in turn inserted in (8) and (12) yields BAU emissions  $e_o := \frac{abn + (n-1)\alpha}{n^2 b + (n-1)\alpha^2 + b[(2n-1)\alpha + n^2 \delta]}$  and BAU welfare  $w_o = \frac{a^2 [bn + (n-1)\alpha] [b^2 n + (n-1)\alpha(\alpha - n^2 \delta) + b((2n-1)\alpha - (n-2)n^2 \delta)]}{[n^2 b + (n-1)\alpha^2 + b((2n-1)\alpha + n^2 \delta)]^2}$ .

**Social optimum.** To evaluate the non-cooperative mitigation efforts of Nash governments we briefly determine the socially optimal solution. Maximizing  $\sum_j W^j(t_1, \dots, t_n)$  with respect to  $t_i$  yields the first order condition  $\sum_j W_{t_i}^j = 0$  which can be solved for the socially optimal tax rate<sup>12</sup>  $\hat{t} := \frac{n^2 a \delta}{b + \alpha + n^2 \delta}$ . Insertion of  $\hat{t}$  in (9) and (12) gives the socially optimal emissions  $\hat{e} := \frac{a}{b + \alpha + n^2 \delta}$  and socially optimal welfare  $\hat{w} := \frac{a^2}{2(b + \alpha + n^2 \delta)}$ . Comparing the allocation  $(\hat{e}, \hat{t}, \hat{w})$  with the BAU allocation  $(e_o, t_o, w_o)$  yields  $\hat{w} > w_o$ ,  $\hat{e} < e_o$  and  $\hat{t} > t_o$ . 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.

<sup>11</sup>The best-reply function is derived in the Appendix A.

<sup>12</sup>For the derivation of  $\hat{t}$  we refer to the Appendix A.

**Business as usual: taxes versus caps.** It is clear that the socially optimal allocation is independent of whether full cooperation takes place with cap or tax regulation. In contrast, if all countries act non-cooperatively (BAU) the type of regulation matters. As shown in Eichner and Pethig (2013) the BAU equilibrium caps are  $\bar{e}_o := \frac{a}{b+\alpha+n\delta}$  and BAU welfare is  $\bar{w}_o := \frac{a^2[b+\alpha+(n-2)n\delta]}{2(b+\alpha+n\delta)^2}$  in case of cap regulation. Comparing the BAU allocations with cap and tax regulation yields  $e_o > \bar{e}_o$  and  $w_o < \bar{w}_o$ , where  $e_o[\bar{e}_o]$  denotes the emission and  $w_o[\bar{w}_o]$  denotes the welfare of an individual country in BAU with tax [cap] regulation.

**Result 1.**<sup>13</sup> *In the absence of international cooperation (BAU) emission tax regulation leads to less mitigation and to lower welfare than cap and trade regulation.*

To understand the rationale of Result 1, reconsider the equations (9) and (10). Differentiation yields

$$\frac{\partial e_i^d}{\partial t_i} = -\frac{nb + (n-1)\alpha}{nb(\alpha + b)} < 0 \quad \text{and} \quad \frac{\partial \sum_j e_j}{\partial t_i} = -\frac{1}{(\alpha + b)} < 0$$

which implies  $0 > \frac{\partial \sum_j e_j}{\partial t_i} > \frac{\partial e_i^d}{\partial t_i}$ . Hence a small increase of the tax rate  $t_i$  reduces domestic emissions in country  $i$ ,  $\frac{\partial e_i^d}{\partial t_i} < 0$ , and also reduces total emissions,  $\frac{\partial \sum_j e_j}{\partial t_i} < 0$ , but total emissions shrink less than domestic emissions in country  $i$ . In other words, the reduction of domestic emissions via taxation in country  $i$  induces an increase of emissions in all other countries (so-called carbon leakage). Such a leakage effect is absent in case of cap regulation because all countries take as given the emission caps of all other countries. Since the reduction of one unit of emissions at home reduces total emissions (and climate damage) by less than one unit under tax than under cap regulation, the countries mitigate less under the former than under the latter regulation. That explains why total BAU emissions are higher under tax than under cap regulation.<sup>14</sup>

Another important difference between tax and cap regulation in BAU is that caps are always strategic substitutes (Eichner and Pethig 2013) while we showed above that taxes are strategic complements if and only if the condition  $\alpha(\alpha + b) > n^2b\delta$  is satisfied. Interestingly, the inequalities  $e_o > \bar{e}_o$  and  $w_o < \bar{w}_o$  hold independent of whether tax rates are strategic substitutes or complements. In Section 3.2 below we will take a closer look at the property

---

<sup>13</sup>The non-equivalence of taxes and caps (or cap-and-trade schemes) when non-cooperative governments act strategically, have been derived in various analytical frameworks, e.g. by Ulph (1996), Pizer (2002), Ishikawa and Kiyono (2006), Strand (2010), Karp and Zhang (2012) and Kiyono and Ishikawa (2013).

<sup>14</sup>The identification and interpretation of the asymmetric carbon leakage effect is due to Kiyono and Ishikawa (2013) whose model is similar though different to our BAU model. Result 1 confirms their conclusion that total emissions are larger with taxes than with caps. Their ranking of welfares may differ from that in Result 1 but only if production technologies differ which is excluded in our paper.

of strategies being either substitutes or complements because that distinction will turn out to be important for the stability of coalitions.

### 3 Nash coalitions: taxes versus caps

#### 3.1 Climate coalitions of given size

Suppose next 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  $C := \{1, 2, \dots, m\}$  with  $C$  for coalition, and collect all remaining countries in another group, denoted group  $F := \{m+1, \dots, n\}$  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  $\sum_{j \in C} W^j(t_1, \dots, t_n)$  and who plays Nash against all fringe countries and is therefore called Nash coalition, for short. 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 analytical relief the best reply of the coalition with  $m$  members to the fringe countries' strategies  $t_f$  is given by<sup>15</sup>

$$t_c = \frac{n^3 ab\delta}{N^c} + \frac{(n-m)[\alpha(\alpha+b) - n^2 b\delta]}{N^c} t_f, \quad (14)$$

where  $N^c := n^2 b^2 + (n^2 - m)\alpha^2 + b[(2n^2 - m)\alpha + mn^2\delta] > 0$ . Similarly, we account for  $t_i = t_c$  for  $i \in C$  and  $t_i = t_f$  for  $i \in F$  and convert each fringe country's best-reply function (13) into

$$t_f = \frac{n^3 ab\delta}{N^f} + \frac{m[\alpha(\alpha+b) - n^2 b\delta]}{N^f} t_c, \quad (15)$$

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

Clearly, a coalition-fringe equilibrium is the  $n$ -tuple  $(\underbrace{t_c^*, \dots, t_c^*}_{m\text{-times}}, \underbrace{t_f^*, \dots, t_f^*}_{(n-m)\text{-times}})$  of tax rates solving (14) and (15). The solution turns out to be

$$t_c^* = \frac{mn^3 ab\delta}{N} \quad \text{and} \quad t_f^* = \frac{n^3 ab\delta}{N} \quad (16)$$

with  $N := n^2 b^2 + \alpha^2 H + b[(n^2 + H)\alpha + (n + m(m-1))n^2\delta] > 0$  and  $H := n(n-1) - m(m-1)$ .

---

<sup>15</sup>For the derivation of (14) we refer to the Appendix B.

With the notation  $t_c$  and  $t_f$  we rewrite the welfare (12) of individual countries as  $W^c(t_c, t_f, m) := W^i(\underbrace{t_c, \dots, t_c}_{m\text{-times}}, \underbrace{t_f, \dots, t_f}_{(n-m)\text{-times}})$  for all countries in group  $C$ , as  $W^f(t_c, t_f, m) := W^i(\underbrace{t_c, \dots, t_c}_{m\text{-times}}, \underbrace{t_f, \dots, t_f}_{(n-m)\text{-times}})$  for all countries in group  $F$ , and prove<sup>16</sup> that in the coalition-fringe equilibrium with a coalition of size  $m$  the fringe countries' welfare is higher than the coalition countries' welfare which, in turn, is higher than all countries' welfare in BAU. The coalition countries' emissions are lower and the fringe countries' emissions are higher compared to BAU.

The next step towards analyzing self-enforcing agreements is to investigate how the values of the variables in the coalition-fringe equilibrium vary with exogenous changes in coalition size. To that end we consider *equilibrium* values as functions of  $m \in [2, n]$  and write, in particular

$$\begin{aligned} t_v^* &= \mathcal{T}^v(m), & w_v^* &:= \mathcal{W}^v(m) = W^v[\mathcal{T}^c(m), \mathcal{T}^f(m), m] & \text{for } v = c, f, \\ w^* &:= mw_c^* + (n-m)w_f^* = m\mathcal{W}^c(m) + (n-m)\mathcal{W}^f(m) =: \mathcal{W}(m), \\ e_d^{v*} &= \mathcal{E}^{dv}(m) & \text{for } v = c, f, & e^{s*} = \mathcal{E}^s(m), \\ e^* &:= me_c^{d*} + (n-m)e_f^{d*} = ne^{s*} = n\mathcal{E}^s(m) =: \mathcal{E}(m). \end{aligned}$$

In the Appendix C we provide some general information about the properties of these functions, but the functions are too complex to fully specify their curvature analytically. To make progress, we therefore proceed with a numerical illustration.<sup>17</sup> Although the dependence on coalition size of variables other than welfares and emissions is important, we restrict our subsequent discussion to the curves linking welfares and emissions to coalition sizes in order to keep focused on the stability issue to be addressed in the next section. For our example, denoted Example 1, we choose the parameter values  $\alpha = 1000$ ,  $a = 100$ ,  $b = 20$ ,  $\delta = 1$ ,  $n = 10$  and  $\bar{x} = 12$ . Since Example 1 has also been discussed in Eichner and Pethig (2012) we are able to directly compare the outcome of tax and cap regulation for the same set of parameters.

Figure 1 shows that in case of tax regulation the welfares of coalition and fringe countries (and hence total welfare) are progressively increasing in coalition size and that the free-rider advantage of fringe countries,  $\mathcal{W}^f(m) - \mathcal{W}^c(m) > 0$ , also rises with the coalition

---

<sup>16</sup>See Lemma 2 of the Appendix C.

<sup>17</sup>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 present a single example.

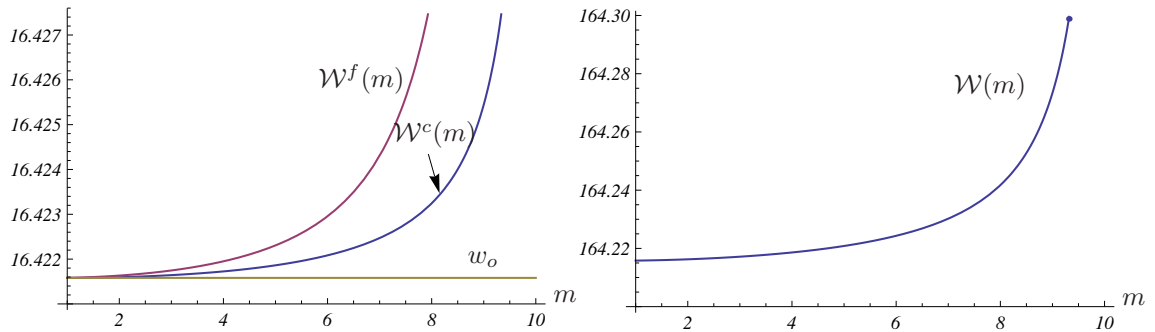


Figure 1: National and total welfare with taxes (Example 1, Nash)

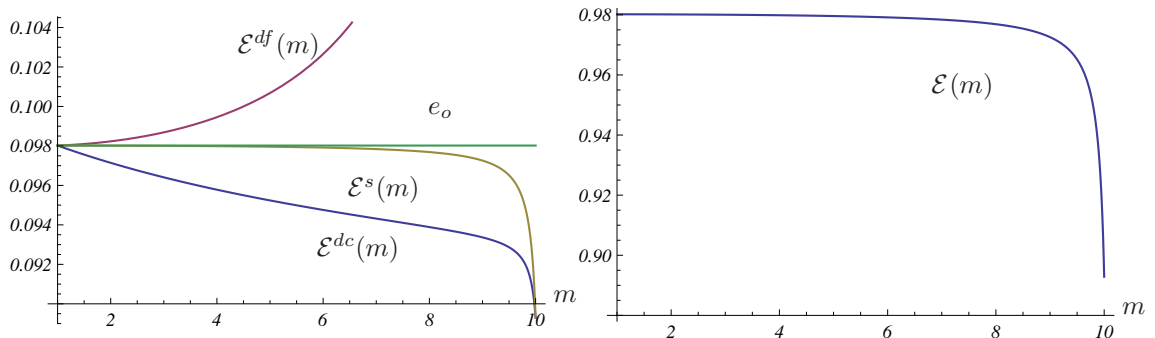


Figure 2: National and total emissions with taxes (Example 1, Nash)

size. In the left panel of Figure 2 the horizontal line ( $e_o$ ) is the fuel supply (= fuel demand = emissions) per country in BAU. In the coalition-fringe equilibrium with coalition of size  $m$ ,  $\mathcal{E}^s(m)$  is the fuel supply per country which is only slightly less than in BAU for small  $m$  and then drops sharply when  $m$  gets very close to  $n = 10$ . The left panel of Figure 2 also shows the divergence of fuel consumption (= emissions) between coalition and fringe countries. The coalition countries' emission levels are always below BAU level and declining in  $m$  because their tax policy is more stringent than in BAU. In contrast, the fringe countries' emissions are above BAU level and increasing in  $m$  which underlines that the fringe countries free ride on the coalition's mitigation effort.

In the Figures 3 and 4 all curves of the Figures 1 and 2 are reproduced with slight changes in scales. In addition, they contain all corresponding curves of Example 1 in the cap-

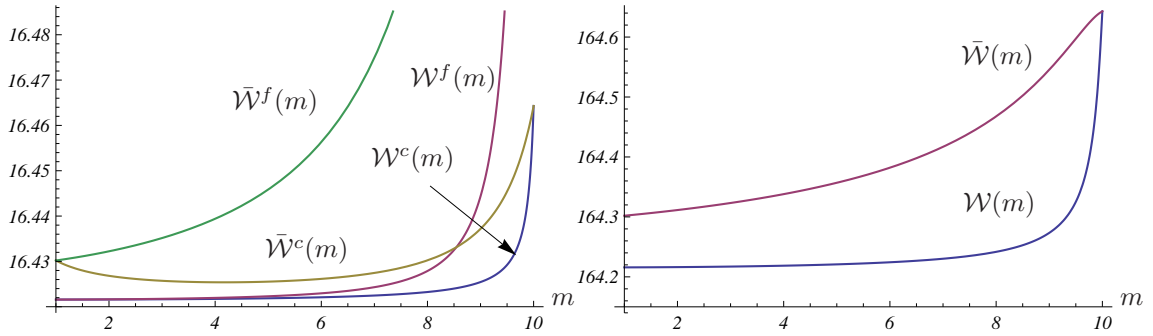


Figure 3: National and total welfare: taxes versus caps (Example 1, Nash)

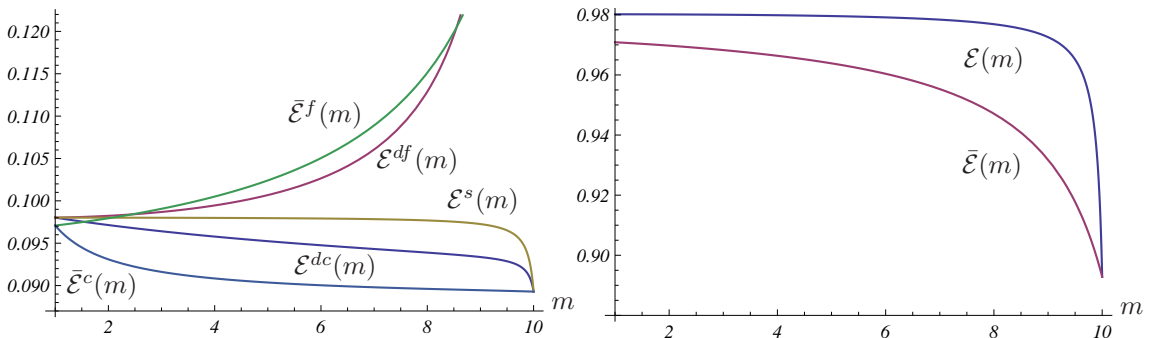


Figure 4: National and total emissions: taxes versus caps (Example 1, Nash)

regulation scenario in Eichner and Pethig (2012). The notation is  $\bar{\mathcal{E}}$ ,  $\bar{\mathcal{W}}$  etc. for functions belonging to the cap-regulation model while functional signs without bars relate to the tax-regulation model, as before. We summarize the information contained in the Figures 3 and 4 in

**Result 2.** *Consider coalition-fringe equilibria with given Nash coalitions of size  $m \in \{2, \dots, n-1\}$ .*

- (i) *Aggregate welfare is monotone increasing in the coalition size with taxes as well as with caps but aggregate welfare is significantly larger under cap than under tax regulation. Furthermore, in case of taxes, aggregate welfare grows only very slowly with the coalition size for all  $m \in \{2, \dots, n-1\}$ .*
- (ii) *Under either regulation, all fringe and coalition countries are better off with caps than with taxes and all fringe countries are better off than coalition countries. The fringe countries' free rider advantage is more pronounced under cap than under tax regulation.*

- (iii) *Aggregate emissions are monotone decreasing in the coalition size with taxes as well as with caps but aggregate emissions are significantly smaller under cap than under tax regulation. Furthermore, in case of taxes, aggregate emissions decline only very slowly with the coalition size for all  $m \in \{2, \dots, n - 1\}$ .*
- (iv) *Under either regulation, all fringe [coalition] countries' emissions are larger [smaller] than in BAU, but the positive difference between the emissions of fringe and coalition countries is larger under cap than tax regulation.*

Interestingly, the mirror-symmetric wording of the Results 2(i) and 2(iii) indicates that under both kinds of regulation the implications regarding aggregate welfare are determined by shifts of aggregate emissions in the opposite direction. Note also that the slow increase of the aggregate welfare for all coalition sizes up to  $m = n - 1$  under tax regulation (Result 2(i)) reflects the coalitions' reluctance for tough taxation due to leakage. That slow increase implies that with the move from the coalition of size  $m = n - 1$  to the grand coalition the increase in aggregate welfare is larger under tax than under cap regulation. This observation will turn out to be important for the subsequent analysis of coalition stability.

## 3.2 Self-enforcing IEAs

In the preceding Section 3.1 we have presupposed the existence of climate coalitions of alternative sizes, and our focus has been on characterizing the coalition-fringe equilibrium and its change with 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).<sup>18</sup> In formal language, an IEA with  $m \in \{2, \dots, n\}$  signatories is said to be self-enforcing or stable if it satisfies the internal stability condition

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

and the external stability condition

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

First we will apply that stability concept to the Example 1, and then we will seek to establish more general stability results. Figure 5 plots the graphs of the functions  $\mathcal{W}^c(m) - \mathcal{W}^f(m - 1)$  and  $\mathcal{W}^f(m) - \mathcal{W}^c(m + 1)$  for Example 1 on the interval  $[1, 4]$ .

---

<sup>18</sup>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).

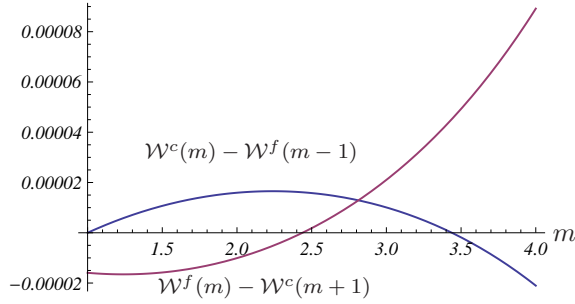


Figure 5: Coalition stability in Example 1

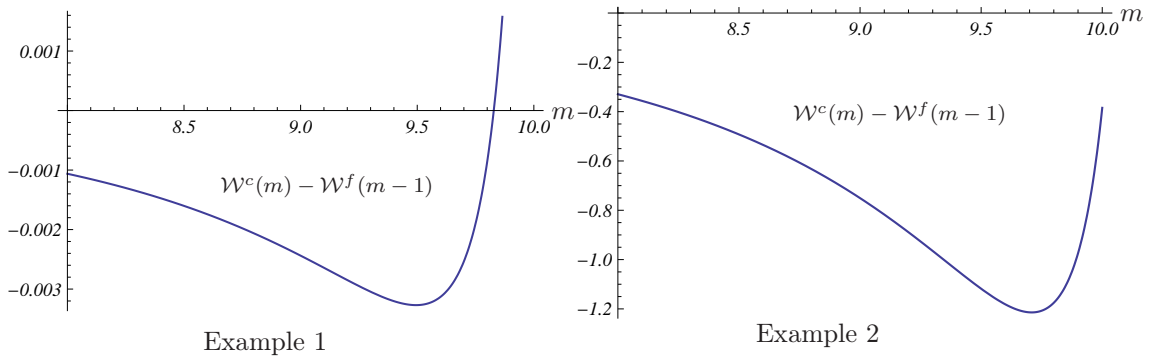


Figure 6: (In)stability of the grand coalition

According to (17) and (18), coalition stability requires both graphs to have non-negative values. Figure 5 exhibits an interval - roughly between 2.5 and 3.5 - where both graphs are positive. The only integer in that interval is  $m = 3$ . Hence the coalition of size  $m = 3$  satisfies the stability conditions (17) and (18). In all examples we ran, the function  $\mathcal{W}^c(m) - \mathcal{W}^f(m - 1)$  was negative valued for all  $m \in \{4, 5, \dots, n - 1\}$ . However, the graph of that function differed significantly across examples in the interval between  $m = n - 1$  and  $m = n$ . As the left panel of Figure 6 shows, in case of Example 1 the graph of  $\mathcal{W}^c(m) - \mathcal{W}^f(m - 1)$  turns from negative to positive at about  $m = 9.98$  so that the grand coalition ( $m = n = 10$ ) stable. Thus in the economy of Example 1 there exist two different stable coalitions, a small one ( $m = 3$ ) and the grand coalition ( $m = 10$ ). That surprising result raises hopes that the grand coalition is stable for all feasible parameter constellations. Unfortunately, this turns out to be incorrect as the right panel of Figure 6 shows for a



counterexample, denoted Example 2, that differs from Example 1 only by substituting the damage parameter  $\delta = 1$  (of Example 1) with  $\delta = 18$ . In Example 2 the grand coalition is unstable because  $\mathcal{W}^c(m) - \mathcal{W}^f(m - 1)$  is negative-valued for all  $m \in \{4, \dots, n\}$ .

The striking finding that the grand coalition is stable under certain conditions and unstable under others calls for analyzing the determinants of its stability in more detail. In the Appendix D we specify analytically the subset of feasible economies in which the grand coalition is stable and provide a rigorous proof of

**Proposition 1.** *Consider the case of tax regulation and Nash coalitions.*

(i) *In the economy  $(a, \alpha, b, \delta, n)$  the grand coalition is stable if and only if*

$$[4(n - 1) + n^2(n - 3)] [\alpha(\alpha + b) - n^2b\delta] \geq n^2(n - 3)(\alpha + b)^2. \quad (19)$$

(ii) *Ceteris paribus, the grand coalition is the more likely stable,*

- *the smaller the total number of countries (the smaller  $n$ );*
- *the less severe the climate damage of emissions (the smaller  $\delta$ );*
- *the higher the fuel extraction costs expressed in terms of the consumption good (the larger  $\alpha$ );*
- *the larger the benefits from fuel consumption (the smaller  $b$ ).*

The inequality (19) in Proposition 1(i) is the necessary and sufficient condition for the stability of the grand coalition. Proposition 1(ii) describes the role of parameters for securing the stability of the grand coalition. Since the *total* climate damage, i.e. the climate damage summed over all countries, is increasing in the parameters  $n$  and  $\delta$ , and is decreasing in  $\alpha$  the stability of the grand coalition appears to be less likely the larger the total climate damage. The link between  $n$ ,  $\delta$  and total climate damage is obvious. To understand the impact of the parameter  $\alpha$  observe that large values of  $\alpha$  imply high opportunity costs of fuel. The more expensive fuel is due to high extraction costs the less fuel is consumed and the lower are total emissions and climate damage. Smaller values of  $b$  increase the utility of consuming fuel, ceteris paribus, and thus further improve the benefit-cost ratio of emissions. Thus, the thrust of Proposition 1 is that the grand coalition is stable in case of tax regulation if the climate damage in BAU is not too severe.

An obvious necessary condition for (19) to be satisfied is that the inequality  $\alpha(\alpha + b) > n^2b\delta$  holds which we have identified in (13) as the condition characterizing tax rates as strategic complements. An equivalent statement is that strategic substitutability is sufficient for the instability of the grand coalition. Note also that although strategic complementarity

is not sufficient for the stability of the grand coalition, the conditions for stable grand coalitions are the same in qualitative terms as the conditions for strategic complementarity.

How do the conditions for stable climate coalitions differ between the scenarios of tax and cap regulation? In case of caps Eichner and Pethig (2012) found either no stable coalition at all or a single stable coalition of size  $m = 2$  for a large subspace of parameter values - no matter how large the total number of countries was chosen. With tax regulation we also identified above cases with small stable coalitions ( $m = 2$  or  $m = 3$ ) as well as cases without a small stable coalition (not documented here).<sup>19</sup> But the crucial difference between both regulation scenarios is that there do not exist parameter constellations in the cap scenario giving rise to stable grand coalitions.<sup>20</sup> That strongly suggests that the following result is general in the present model: *With caps, the grand coalition is never stable, but with taxes the grand coalition is stable under certain parameter constellations specified above.*

In order to improve our intuition for that interesting difference we reconsider the Example 1 with cap and tax regulation and assume that the coalition-fringe equilibrium with the coalition of size  $m = n - 1 = 9$  prevails. Then there exists only a single fringe country, by presupposition, and the question we want to answer is how that country's benefits and costs of joining the coalition differ between cap and tax regulation. To provide the answer in formal language we define the fringe country's equilibrium welfare as  $\mathcal{W}^f(m) = \mathcal{K}^f(m) - \mathcal{D}(m)$ , where  $\mathcal{K}^f(m)$  is the welfare from consuming good  $X$  and fuel, denoted consumption welfare, and where  $(-1) \cdot \mathcal{D}(m)$  is the climate damage multiplied by minus one, denoted climate welfare for short. Hence the fringe country's net benefits of joining the coalition are

$$\mathcal{W}^f(n) - \mathcal{W}^f(n-1) = \underbrace{[\mathcal{D}(n-1) - \mathcal{D}(n)]}_{\text{Benefits of joining}} - \underbrace{[\mathcal{K}^f(n-1) - \mathcal{K}^f(n)]}_{\text{Costs of joining}}.$$

With this notation we establish for the Example 1

**Result 3.** *Let the coalition-fringe equilibrium with the coalition of size  $m = n - 1$  be given and suppose the only fringe country considers joining the coalition. The benefits of joining are larger under tax than under cap regulation and the costs of joining are lower under tax than under cap regulation.*

To understand Result 3, consider first the benefits of joining and recall from the right panel of Figure 4 that total emissions are larger with taxes than with caps for all predetermined coalitions of 2 to 9 members. More specifically, in case of taxes total emissions are  $\mathcal{E}(1) =$

---

<sup>19</sup>It is beyond the scope of the present paper to provide a detailed comparison of the tax and cap scenarios with respect to their small stable coalitions. Such coalitions are of limited interest anyway because the climate damage reduction they achieve compared to BAU is negligible.

<sup>20</sup>A rigorous proof can be obtained from the authors upon request.

0.980 and they decline very slowly to  $\mathcal{E}(9) = 0.973$  which is reflected in the shape of the graph of the function  $\mathcal{E}(m)$  in Figure 4. As  $\mathcal{E}(9)$  is still much higher than the socially optimal level  $\mathcal{E}(10) = 0.897$  the bulk of excessive BAU emissions must and will be mitigated in the transition from the coalition of size  $n - 1$  to the grand coalition. Put differently, when the fringe country joins the coalition it enjoys a massive climate damage reduction,  $\mathcal{D}(n - 1) - \mathcal{D}(n) > 0$ , which represents its benefits of joining. Since for  $m = n - 1$  total emissions are larger with taxes than with caps (right panel of Figure 4) it follows that  $\mathcal{D}(n - 1) - \mathcal{D}(n) > \bar{\mathcal{D}}(n - 1) - \bar{\mathcal{D}}(n)$ , i.e. that the fringe country's benefits of joining the coalition are larger with taxes than with caps (right panel of Figure 7).

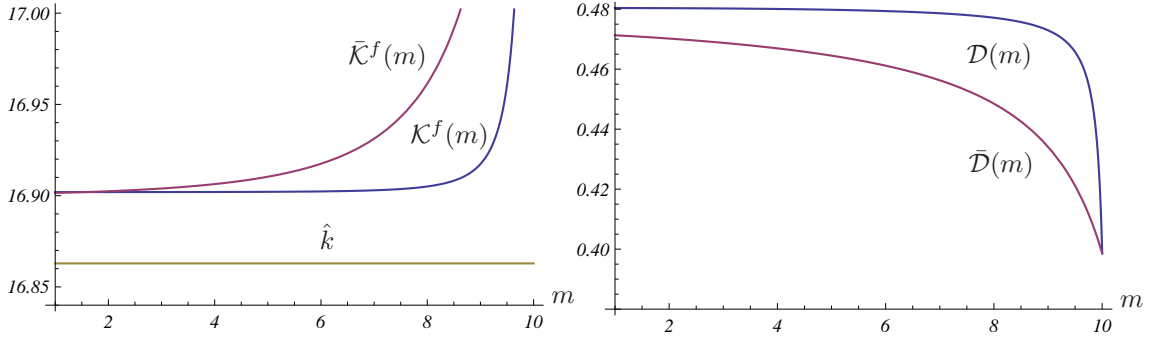


Figure 7: Consumption and climate welfare of fringe countries (Example 1, Nash)

In view of our finding that strategic complementarity is a necessary condition for the stability of the grand coalition (in case of tax regulation) it is interesting to observe that the fringe country's benefits of joining,  $\mathcal{D}(n - 1) - \mathcal{D}(n)$ , are the lower, the stronger the strategic complementarity of taxes. We show that in Appendix E where we first solve the fuel demand (= supply of emissions) from (9) for the tax rates and then insert the latter into (15). After some rearrangements we obtain a derived reaction function of the fringe, say  $(n - m)e_f^d = R(me_c^d, m)$ , which maps the coalition's emissions,  $me_c^d$ , into best-reply emissions of the fringe,  $(n - m)e_f^d$ . The slope of the derived reaction function,  $R_{(me_c^d)}$ , turns out to be always in the interval  $] - 1, 0[$  which means that tax regulation involves leakage at a rate less than one. We also establish in Appendix E that the leakage rate is declining in  $\delta$ ,  $\partial R_{(me_c^d)} / \partial \delta > 0$ . Observe that the lower  $\delta$  the stronger the strategic complementarity. Hence, from  $\partial R_{(me_c^d)} / \partial \delta > 0$  we infer that the stronger the strategic complementarity of tax rates, the smaller the total emission reductions and hence the smaller the benefits of joining the coalition.

Consider now the fringe country's costs of joining the coalition of size  $n - 1$  that consist

of reduced consumption welfare. The left panel of Figure 7 shows that  $\bar{\mathcal{K}}^f(m) - \mathcal{K}^f(m) > \hat{k} > 0$  for all  $m \in \{2, \dots, n-1\}$ , where  $\hat{k} := \bar{\mathcal{K}}^f(n) = \mathcal{K}^f(n)$  is each country's socially optimal consumption welfare. The explanation for the positive difference  $\bar{\mathcal{K}}^f(m) - \mathcal{K}^f(m)$  is that due to caps being strategic substitutes the fringe country's free-rider position is stronger with caps than with taxes, in particular if in the latter case taxes are strategic complements. From  $\bar{\mathcal{K}}^f(m) > \mathcal{K}^f(m) > \hat{k} > 0$  follows  $\bar{\mathcal{K}}^f(n-1) - \hat{k} > \mathcal{K}^f(n-1) - \hat{k} > 0$ , i.e. the fringe country's costs of joining the coalition are lower under tax than under cap regulation.

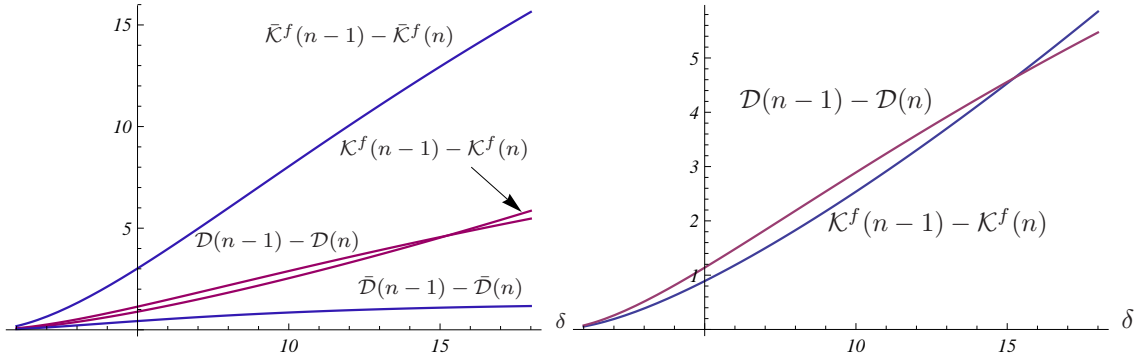


Figure 8: Benefits and costs: taxes versus caps (Example 1, Nash)

Figure 8 illustrates the benefits and costs in dependence of the parameter  $\delta$ . As mentioned before the lower  $\delta$  the stronger the strategic complementarity. The left panel of Figure 8 reconfirms our findings of Result 3 that the benefits of joining are larger under tax than under cap regulation and the costs of joining are lower under tax than under cap regulation. In addition the costs of joining are always larger than the benefits of joining under cap regulation which leaves the grand coalition unstable. Each of the curves is increasing in  $\delta$ , i.e. the stronger the strategic complementarity, the lower both benefits and costs. The right panel of Figure 8 depicts an enlarged version of the cost and benefit curves under tax regulation. It shows that the benefits exceed [fall short of] the costs, if and only if  $\delta$  is smaller [larger] than 15.5. To put it differently, the weaker the strategic complementarity (the larger  $\delta$ ) the more likely is the grand coalition unstable.

## 4 Stackelberg coalitions: taxes versus caps

In the preceding Section 3 we have studied coalition-fringe equilibria with Nash coalitions. Now we turn to an alternative equilibrium concept, the Stackelberg equilibrium in which the

coalition is the Stackelberg leader (Stackelberg coalition, for short) and the fringe countries are Stackelberg followers (Nash players). The assumption of Stackelberg coalitions is applied in various contributions to the literature on IEAs.<sup>21</sup> Eichner and Pethig (2013) analyze the model of Section 2 with Stackelberg coalition and cap regulation. Essentially, they find (i) that the grand coalition is *not* stable, (ii) that in all existing stable coalitions the ratio of coalition members to total number of countries,  $m/n$ , may be up 0.6 no matter how large  $n$  is and (iii) that the reduction of total emissions achieved in equilibria with stable coalitions hardly exceeds the mitigation level in BAU. The result (ii) that relatively large stable coalitions form under certain conditions contrasts significantly with the scenario of cap regulation with Nash coalition (Eichner and Pethig 2012), where stable coalitions are shown to consist of at most two countries. However, owing to result (iii) in the scenario of cap regulation with Stackelberg coalitions all stable coalitions are unattractive, even relatively large ones, because their mitigation performance is very poor. Against this background the interesting question is whether the conditions for large Stackelberg coalitions which are stable as well as effective are more favorable under tax than under tax regulation.

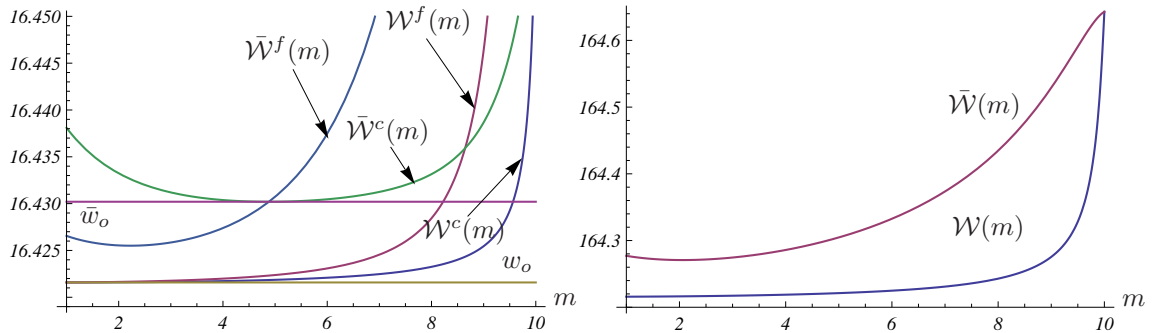


Figure 9: National and total welfare: taxes versus caps (Example 1, Stackelberg)

To answer that question we follow our procedure in Section 3 and look first at how welfares and emissions vary with the size of Stackelberg coalitions before we turn to the stability issue. As informative analytical results cannot be obtained we resort right away to the numerical Example 1 already employed in Section 3. In the Figures 9 and 10 we provide the information for Stackelberg coalitions which is contained in the Figures 3 and 4 for Nash coalitions. As before, the letters with bar, i.e.  $\bar{W}$ ,  $\bar{E}$  etc., belong to the cap and those without bar to the tax scenario. Actually, all graphs of variables with bars in the Figures 9 and 10 are the same as the graphs in the Figures 1 and 2 in Eichner and Pethig (2013)

<sup>21</sup>See e.g. Barrett (1994), Finus (2001), Diamantoudi and Sartzetakis (2006) or Rubio and Ulph (2006). In that literature the reader also finds discussions of the comparative merits of both equilibrium concepts for the issue at hand. Here, we restrict our attention to identifying the differences in outcomes.

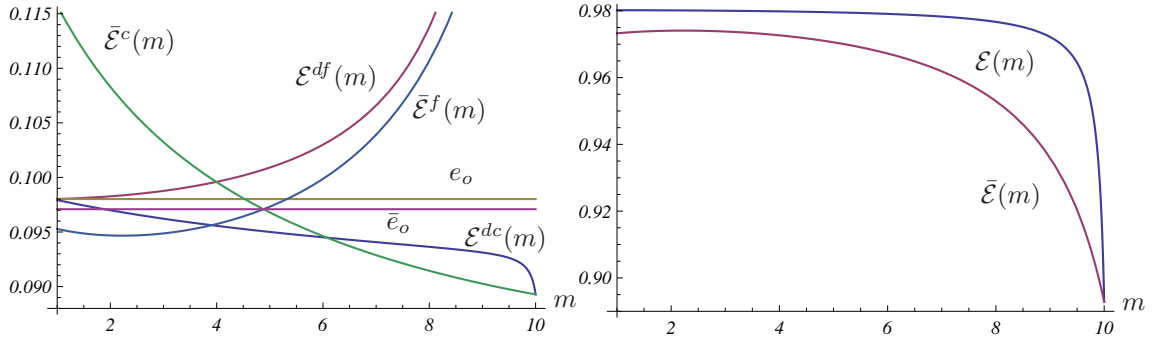


Figure 10: National and total emissions: taxes versus caps (Example 1, Stackelberg)

because in those Figures 1 and 2 the Example 1 is employed as well. We take advantage of that convenient overlap by referring the reader to the discussion of the Figures 1 and 2 in Eichner and Pethig (2013) and immediately proceed to the comparison of both regulation scenarios.

Closer inspection of the right panel of Figure 9 [Figure 10] shows that the shapes of the curves are very similar to the shapes of the curves in the right panel of Figure 3 [Figure 4]. That is, the ranking of the tax and cap scenarios concerning aggregate emissions and aggregate welfares is the same for Nash and Stackelberg coalitions. In particular, under both types of coalitions total emissions are declining in  $m$  in a much more gradual way under cap than under tax regulation such that the decline is much more pronounced with taxes than with caps when  $m$  is very close to  $n$  and approaches  $n$ . As discussed in Section 3.2 that kind of curvatures of the graphs in the right panel of Figure 10 is an indication for the possibility that with taxes the grand coalition might be stable in case of Stackelberg coalitions. In fact, Figure 11 demonstrates, that with tax regulation and Stackelberg coalitions our Example 1 exhibits two stable coalitions, a small one with  $m = 3$  and the grand coalition ( $m = n = 10$ ). Recall from Section 3.2 that for Example 1 with tax regulation and Nash coalitions we also get two stable coalitions, a small one with  $m = 3$  and the grand coalition.

For the case of cap regulation and Stackelberg coalitions Eichner and Pethig (2013) show that the membership of the (only) stable coalition increases with the size of the parameter  $\alpha$  and attains its maximum at  $m = 6$ , if all other parameters of Example 1 remain unchanged. If the same exercise is performed with tax regulation and Stackelberg coalitions the size of the small stable coalition remains constant at  $m = 3$ . Recall, however, that the grand coalition is not stable under cap regulation and that all existing stable coalitions are ineffective with respect to mitigation even if they are relatively large. That is why in our view the crucial and most interesting features of coalition formation are (i) the discrepancy

between the instability of grand coalitions under cap regulation and the possible stability of grand coalitions in case of tax regulation and (ii) the observation that this discrepancy exists in cases of both Nash and Stackelberg coalitions.

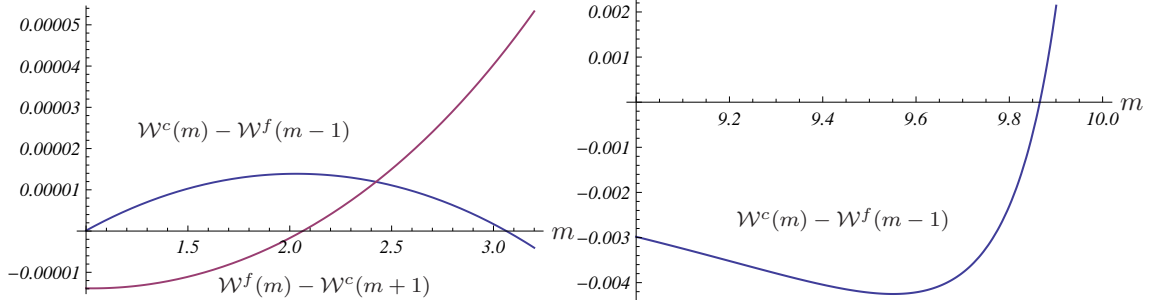


Figure 11: Coalition stability in Example 1 (Stackelberg, tax)

As in Section 3.2, it is desirable to get more general information on the determinants of the stability of grand coalitions. Unfortunately, high analytical complexity prevents us from characterizing the entire subset of feasible economies exhibiting stable grand coalitions analogous to the condition (19) in Proposition 1. In order to demonstrate that the conditions for stable grand coalitions are not exceptional, we construct a large subset of economies with stable grand coalitions, which is smaller than the set of all economies with stable grand coalitions though, and prove in the Appendix D

**Proposition 2.** *Consider the case of tax regulation and Stackelberg coalitions.*

- (i) *There is a parameter space  $\mathcal{M}^+$ , specified in the Appendix D, such that for all economies  $(a, \alpha, b, \delta, n) \in \mathcal{M}^+$  the grand coalition is stable.*
- (ii) *Ceteris paribus, the grand coalition is the more likely stable,*
  - *the smaller the total number of countries (the smaller  $n$ );*
  - *the less severe the climate damage of emissions (the smaller  $\delta$ );*
  - *the higher the fuel extraction costs in terms of the consumption good (the larger  $\alpha$ );*
  - *the larger the benefit from fuel consumption (the smaller  $b$ ).*

The comparison of the Propositions 2(ii) and 1(ii) suggests that the role of parameters for securing stable grand coalitions is the same in qualitative terms in case of Stackelberg and Nash coalitions. There clearly are minor differences between Nash and Stackelberg coalitions

in reaching stable grand coalitions. However, the discussion of such differences is beyond the scope and focus of the present paper. Summing up, the important message of our analysis in Section 4 is

**Result 4.** *Our analysis and extensive numerical calculations suggest that in the model of the world economy of Section 2 the grand coalition is unstable in case of cap regulation whereas it is stable in case of tax regulation if the climate damage is not too severe. This differential effect of cap and tax regulation on the stability of the grand coalition holds independent of whether coalitions are modeled as Nash or Stackelberg coalitions.*

## 5 The role of international trade for coalition formation

In order to investigate the role of international trade for the stability of climate coalitions we compare our stability results of the previous sections with those of the autarky scenario in the otherwise unchanged model. We turn free trade into autarky by shutting down the world markets and replacing them with domestic markets. In autarkic countries, caps and taxes are equivalent for the following reason. If a domestic cap and trade system implements a given cap, the equilibrium permit price can be interpreted as an emission tax which generates emissions exactly equal to the emission cap that was fixed initially. Conversely, if we start from some given emission tax rate that leads to a specific level of emissions we can interpret those emissions as a cap which is then implemented by a permit price exactly equal to the initial tax rate. We take advantage of that equivalence by restricting our subsequent analysis to the autarky model with cap regulation.

Eichner and Pethig (2012) show in their analysis of cap regulation (i) that the autarky model with parametric functions is fully described by the welfare functions,

$$W^i(e_i, \dots, e_n) = \bar{x} + \beta e_i - \frac{\alpha + \gamma}{2} e_i^2 - \frac{\alpha + \gamma}{2} e_i^2 - \frac{\delta}{2} \left( \sum_j e_j \right)^2 \quad i = 1, \dots, n, \quad (20)$$

and (ii) that the autarky model (20) is equivalent to the (parametric version of the) basic model of the coalition formation literature surveyed e.g. by Finus (2003). In other words, the basic model of the literature can be interpreted as the autarky model of the present paper.<sup>22</sup> The convenient implication of both equivalences is that it suffices to investigate the formation of stable coalitions in autarky under cap and tax regulation by simply reporting the results on stable coalitions in the basic model of the extant literature which assumes cap regulation and provides results for both Nash and Stackelberg coalitions.

---

<sup>22</sup>Note also that in our autarky model the BAU equilibrium coincides with the BAU equilibrium under free trade, because non-cooperative identical countries do not gain from trade despite open borders.



For the case of Nash coalitions Finus and Rundshagen (2001), Rubio and Casino (2001) and de Cara and Rotillon (2001) have independently proved that stable coalitions consist of at most two countries in the quadratic basic model of the literature.<sup>23</sup> In case of Stackelberg coalitions Diamantoudi and Sartzetakis (2006) as well as Rubio and Ulph (2006) find that if negative emissions are excluded stable coalitions consist of no more than four countries irrespective of the total number of countries.<sup>24</sup> That leads us to conclude

**Result 5.** *Without international trade there is at most one very small stable coalition in the model of Section 2. This holds for any total number of countries and for coalitions modeled as Nash players or Stackelberg leaders.*

Comparing the Results 4 and 5 reveals an interesting asymmetry between tax and cap regulation when we switch from trade to autarky. In case of cap regulation, all stable coalitions are much smaller than the grand coalition and have a poor mitigation record<sup>25</sup> with and without open borders. So trade does not matter (much) in the cap scenario. With tax regulation, stable coalitions are small in autarky but under free trade the grand coalition may be stable under certain conditions. Hence with taxes trade clearly has a significant impact. To be more specific, consider a world economy with tax regulation, with trade and either Nash or Stackelberg coalition in which the grand coalition is stable. If the borders are then closed, the grand coalition is not stable anymore. The reason for that drastic change is the switch of tax rates from strategic complements in free trade to strategic substitutes in autarky. Regarding the role of strategic substitutes and complements we refer the reader to the corresponding discussion in the previous section.

## 6 An extension: Fossil fuel as intermediate good

We are aware that due to restrictive assumptions made for reasons of tractability the robustness of results is unclear. To make a small step towards more generality we will now examine whether the special structure of the economies of individual countries (Section 2) drives the results. Recall that fossil fuel is modeled as consumption good for households. Although

---

<sup>23</sup>Eichner and Pethig (2012) contribute the additional information that trade liberalization tends to render more unlikely the formation of a two-country stable coalition.

<sup>24</sup>According to Finus (2001) the size of stable coalitions and global welfare are at least as high with Stackelberg than with Nash coalitions. Diamantoudi and Sartzetakis (2006) focus exclusively on subsets of parameters leading to positive equilibrium emissions. Rubio and Ulph (2006) consider a larger parameter space and introduce non-negativity constraints on emissions. For more details see Eichner and Pethig (2013).

<sup>25</sup>Recall that in the free-trade model up to 60% of all countries may be members of stable coalitions, but total emissions in the corresponding coalition-fringe equilibrium are hardly smaller than in BAU.

that setup captures some aspect of the real world, our model of Section 2 fails to account for fossil fuel as an input in the production of some consumption good. This section modifies the model of Section 2 by treating fuel as an input. The aim is to investigate whether that modification changes the results of Section 3, in particular with regard to the conditions for the stability of stable grand coalitions.

**The modified model.** Now the economy of each country is assumed to consist of three sectors. The first is a clean sector that produces the consumer good  $Y$  (quantity  $y_i^s$ ) with the domestic (composite) production factor  $r_{yi}$ , e.g. labor or capital. The second is a dirty sector that generates the consumer good  $X$  (quantity  $x_i^s$ ) with the input fossil fuel  $e_i^d$ . The production functions are

$$y_i^s = Y(r_{yi}) = r_{yi} \quad \text{and} \quad x_i^s = X(e_i^d) = \tilde{\alpha}e_i^d - \frac{\tilde{\beta}}{2}(e_i^d)^2, \quad (21)$$

where  $\tilde{\alpha}$  and  $\tilde{\beta}$  are positive parameters. The production in sector  $X$  causes carbon dioxide emissions in one-to-one relation with fuel input. Fossil fuel is extracted from domestic fossil reserves according to

$$e_i^s = E(r_{ei}) = r_{ei}, \quad (22)$$

where  $r_{ei}$  is the domestic input in extraction. The utility

$$U(y_i^d) + x_i^d - D\left(\sum_j e_j^d\right) \quad (23)$$

of the representative consumer of country  $i$  is additive separable in  $y_i^d$ ,  $x_i^d$  and  $\sum_j e_j^d$ , and linear in  $x_i^d$ . In the sequel we employ the parametric functions  $U(y_i^d) = \tilde{a}y_i^d - \frac{\tilde{b}}{2}(y_i^d)^2$  and  $D\left(\sum_j e_j^d\right) = \frac{\tilde{\delta}}{2}\left(\sum_j e_j^d\right)^2$  with  $\tilde{a} > 0$ ,  $\tilde{b} > 0$  and  $\tilde{\delta} > 0$ .

There are perfectly competitive world markets for good  $X$  (price  $p_x \equiv 1$ ) and for fuel (price  $p$ ), and domestic markets for good  $Y$  (price  $p_{yi}$ ) and for the composite production factor (price  $p_{ri}$ ). The market clearing conditions are

$$\sum_j x_j^s = \sum_j x_j^d, \quad (24)$$

$$\sum_j e_j^s = \sum_j e_j^d, \quad (25)$$

$$y_i^s = y_i^d \quad \text{for } i = 1, \dots, n, \quad (26)$$

$$\tilde{r} = r_{ei} + r_{yi} \quad \text{for } i = 1, \dots, n, \quad (27)$$

where  $\tilde{r}$  is country  $i$ 's exogenous endowment of the production factor. All sectors operate under conditions of perfect competition. The profits of sector  $Y$  and the extraction sector

are

$$\pi_{yi} = p_{yi}Y(r_{yi}) - p_{ri}r_{yi} \quad \text{and} \quad \pi_{ei} = pE(r_{ei}) - p_{ri}r_{ei}. \quad (28)$$

Profit maximization yields

$$p_{yi} = p_{ri} = p, \quad \text{for } i = 1, \dots, n. \quad (29)$$

where  $p$  is the price of fuel in terms of good  $X$  and represents the terms of trade. Due to the simple structure of the model,  $p$  is also the price of good  $Y$  and the price of the domestic production factor (in terms of good  $X$ ).

As we are interested in the differential effect of cap and tax regulation on the formation of stable coalitions we will discuss both modes of regulation beginning with cap regulation.

**Caps.** To implement its emission cap  $e_i$ , government  $i$  issues the amount  $e_i$  of emission permits and auctions them at the permit price  $t_i$ . Sector  $X$  in country  $i$  needs to acquire emission permits to match its purchase of fuel and maximizes its profit

$$\pi_{xi} = X(e_i^d) - (p + t_i)e_i^d. \quad (30)$$

The representative consumer  $i$  takes the climate damage  $D\left(\sum_j e_j^d\right)$  as given and maximizes her (consumption) utility  $U(y_i^d) + x_i^d$  subject to her budget constraint

$$x_i^d + py_i^d = p\bar{r} + \pi_{xi}^* + t_i e_i^d. \quad (31)$$

where the right side of (31) reflects consumer  $i$ 's income (= endowment income plus profit income plus recycled revenues from the permit auction).

Using standard arguments, we show in the Appendix F that profit and utility maximization combined with the market clearing conditions (24) - (27) determine the equilibrium demands for and supplies of all commodities as functions of the politically fixed emission caps  $(e_1, \dots, e_n)$ . Setting  $e_j^d = e_j$  for all  $j$  and inserting the equilibrium demands for the goods  $X$  and  $Y$  in the parametric version of (23) yields the equilibrium welfare of country  $i = 1, \dots, n$ ,

$$\begin{aligned} W^i(e_1, \dots, e_n) = & (\tilde{\alpha} - \tilde{a} + b\tilde{r})e_i - \frac{\tilde{\beta}}{2}e_i^2 + \tilde{a}\tilde{r} - \frac{\tilde{b}}{2}\tilde{r}^2 + \frac{\tilde{b}}{2}\left(\frac{\sum_j e_j}{n}\right)^2 - \tilde{b}e_i\frac{\sum_j e_j}{n} \\ & - \frac{\tilde{\delta}}{2}\left(\sum_j e_j\right)^2. \end{aligned} \quad (32)$$

Next, we compare (32) with the welfare function (12) in Eichner and Pethig (2013) which reads

$$W^i(e_1, \dots, e_n) = ae_i - \frac{b}{2}e_i^2 + \bar{x} + \frac{\alpha}{2}\left(\frac{\sum_j e_j}{n}\right)^2 - \alpha e_i\frac{\sum_j e_j}{n} - \frac{\delta}{2}\left(\sum_j e_j\right)^2. \quad (33)$$

Inspection of the functions (32) and (33) shows that they are isomorphic and coincide, when we set  $a = (\tilde{\alpha} - \tilde{a} + b\tilde{r})$ ,  $b = \tilde{\beta}$ ,  $\bar{x} = \tilde{a}\tilde{r} - \frac{\tilde{b}}{2}\tilde{r}^2$ ,  $\alpha = \tilde{b}$  and  $\delta = \tilde{\delta}$ . Before we will comment on the implication of that unexpected result we want to explore how the equilibrium welfare functions compare in case of tax regulation.

**Taxes.** To analyze the tax regulation scenario, we simply reinterpret  $t_i$  to be the emission tax rate rather than the permit price. Analogous to the procedure in case of cap regulation we show in the Appendix F that profit and utility maximization combined with the market clearing conditions (24) - (27) turn (23) into the equilibrium welfare of country  $i = 1, \dots, n$ ,

$$W^i(t_1, \dots, t_n) = -\frac{t_i^2}{2\tilde{\beta}} + \frac{(\tilde{\alpha} - \tilde{a} + \tilde{r}\tilde{b})n\tilde{\delta}}{(\tilde{b} + \tilde{\beta})^2} \sum_j t_j + \frac{\tilde{b}(\tilde{b} + \tilde{\beta}) - n^2\tilde{\beta}\tilde{\delta}}{2n^2\tilde{\beta}(\tilde{b} + \tilde{\beta})^2} \left( \sum_j t_j \right)^2 + S, \quad (34)$$

where

$$S := \frac{\tilde{a}^2 (\tilde{b} + n^2(-\tilde{\delta}) + \tilde{\beta}) + 2\tilde{a} (\tilde{b} (n^2\tilde{r}\tilde{\delta} + \tilde{r}\tilde{\beta} - \tilde{\alpha}) + n^2\tilde{\alpha}\tilde{\delta} + \tilde{r}\tilde{\beta}^2 - \tilde{\alpha}\tilde{\beta})}{2(\tilde{b} + \tilde{\beta})^2} + \frac{-\tilde{b}^2\tilde{r} (\tilde{r} (n^2\tilde{\delta} + \tilde{\beta}) - 2\tilde{\alpha}) + \tilde{b} (2\tilde{r}\tilde{\alpha} (\tilde{\beta} - n^2\tilde{\delta}) - \tilde{r}^2\tilde{\beta}^2 + \tilde{\alpha}^2) + \tilde{\alpha}^2 (\tilde{\beta} - n^2\tilde{\delta})}{2(\tilde{b} + \tilde{\beta})^2}.$$

Next, we compare (34) with the welfare function (12) of Section 2 which reads, after some rearrangement of terms

$$W^i(t_1, \dots, t_n) = -\frac{t_i^2}{2b} + \frac{an\delta}{(\alpha + b)^2} \sum_j t_j + \frac{\alpha(\alpha + b) - n^2b\delta}{2n^2b(\alpha + b)^2} \left( \sum_j t_j \right)^2 + \bar{x} + \frac{a^2(\alpha + b - n^2\delta)}{2(\alpha + b)^2} \quad (35)$$

Analogous to our comparison of (32) and (33), we conclude that the functions (34) and (35) are isomorphic and coincide if we set  $a = (\tilde{\alpha} - \tilde{a} + b\tilde{r})$ ,  $b = \tilde{\beta}$ ,  $\alpha = \tilde{b}$ ,  $\delta = \tilde{\delta}$  and  $\bar{x} + \frac{a^2(\alpha + b - n^2\delta)}{2(\alpha + b)^2} = S$ .<sup>26</sup>

Summing up, since the functions (32) and (33) as well as the functions (34) and (35) are isomorphic, the straightforward conclusion is

**Result 6.** *The Results 1-5 relating to the model of the Section 2-5 carry over to the model of Section 6.*

Recall that the principal difference between both models is that in the model of the previous sections fossil fuel is passed forward from extraction straight to consumption by households

---

<sup>26</sup>Note that the terms  $\bar{x} + \frac{a^2(\alpha + b - n^2\delta)}{2(\alpha + b)^2}$  and  $S$  are constants that vanish in the first derivative of the respective welfare functions.

whereas in the model of the present section fuel is an intermediate good exclusively used as input in the production of consumer good  $X$ . Result 6 implies that modeling fuel as an intermediate good in production or as a consumption good does not change the conditions for coalition stability substantially. In particular, in both modeling approaches tax regulation and international trade are necessary conditions for the stability of the grand coalition.

## 7 Concluding remarks

In a simple model of identical countries interconnected via trade and global climate damage we have analyzed the conditions for the formation of stable climate coalitions when all countries employ carbon taxes to fight climate change. We have adopted the analytical framework of Eichner and Pethig (2012, 2013) who show for the case of cap regulation that stable coalitions are very small and hence ineffective in reducing climate damage, if coalitions play Nash, and that stable coalitions may be relatively large but also ineffective, if coalitions are modeled as Stackelberg leaders. Previous literature established that taxes and caps are not equivalent in settings with non-cooperative and strategically acting governments. However, to our knowledge that literature does not address coalition formation. To fill that gap we investigate the differences between taxes and caps with respect to their impact on coalition stability, and we also examine the relevance of the alternative assumptions of Nash and Stackelberg coalitions for the impact of taxes.

Rather than summarizing the results in this concluding section systematically, we wish to reemphasize here the rationale of our most striking and unexpected finding that under certain conditions the grand coalition is stable with taxes although it is unstable in case of caps. Recall first that similar as in Kiyono and Ishikawa (2013) it is a carbon leakage effect, which is responsible for the non-equivalence of caps and taxes in the BAU scenario of our model. This leakage effect is absent in case of caps because if cap regulation prevails and an individual country varies its cap, *ceteris paribus*, the change in total emissions is equal to the change in that country's cap. In contrast, if taxes are employed and an individual country increases its tax rate, *ceteris paribus*, that country's domestic emissions shrink, the world fuel price declines, and leakage occurs because all other countries' fuel demands and emissions increase.<sup>27</sup> As a consequence, total emissions decline less than the emissions in the tax-increasing country, and therefore each country's emissions in BAU are larger with taxes than with caps.

---

<sup>27</sup>Note that the decline of the fuel price (terms-of-trade effect) is qualitatively the same under cap and tax regulation. But in case of caps the emissions of the other countries are unchanged by presupposition while in case of taxes the other countries' emissions adjust endogenously.

In all coalition-fringe equilibria with Nash or Stackelberg coalitions of given size ( $m < n$ ) we found the carbon-leakage effect working qualitatively as in BAU. Hence total emissions in those equilibria are always larger under tax than under cap regulation. That observation is of special interest when we consider the coalition-fringe equilibrium with the coalition of size  $m = n - 1$  and examine the only fringe country's decision to join the coalition. Obviously, that country's benefits of joining consist of reduced climate damage. Since climate damage is more severe under taxes than under caps due to the leakage effect the benefit of the (only) fringe country to join the coalition of  $n - 1$  members is higher under tax than under cap regulation. The fringe country's costs of joining that consist of reduced consumption welfare due to increased mitigation turn out to be smaller with tax than with cap regulation. Thus, the incentives of the last remaining fringe country to join the coalition of size  $m = n - 1$  are strictly stronger (on the accounts of both benefits and costs) under tax than under cap regulation. These incentives are always too weak in case of caps, but are strong enough to render the grand coalition stable with tax regulation, if the climate damage is not too severe. Somewhat ironically, it is the relatively poor emission reduction performance of the coalitions under tax regulation which contributes to stabilizing the grand coalition because it implies that the benefits of the (only) fringe country to join the coalition are larger with taxes than with caps.

In summary, the principal messages of the present paper are that - in our framework of analysis - international trade *and* tax regulation are necessary conditions for the stability of the grand coalition, and that in case of international trade and tax regulation the grand coalition is the more likely stable, the less severe the climate damage. Cap regulation is inferior to tax regulation insofar as in case of the former all effective coalitions are unstable, in particular the grand coalition. Further results are that it does not matter (much) for the formation of stable grand coalitions whether coalitions are assumed to be Stackelberg leaders or Nash players or whether fossil fuel is modeled as a consumer good or as an intermediate good. Tractability forced us to employ restrictive assumptions. They do not render uninteresting the results, in our view, but warn against unwarranted generalizations and against offering premature advice on practical climate negotiations. In any case, our analysis strongly suggests directing future research efforts towards the more thorough investigation of the conditions for reaching an effective and encompassing self-enforcing international climate agreement.

## References

- Barrett, S. (1994): Self-enforcing international environmental agreements. *Oxford Economic Papers* 46, 878 – 894.
- Buchner, B., Carraro, C. and I. Cersosimo (2002): Economic consequences of the US withdrawal from the Kyoto/Bonn Protocol. *Climate Policy* 2, 273 – 292.
- Carbone, J.C., Helm, C. and T.F. Rutherford (2009): The case for international emission trade in the absence of cooperative climate policy. *Journal of Environmental Economics and Management* 58, 233 – 263.
- Carraro, C. and D. Siniscalco (1991): Strategies for the international protection of the environment. *CEPR discussion paper* 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.
- De Zeeuw (2008): Dynamic effects on the stability of international environmental agreements. *Journal of Environmental Economics and Management* 55, 163 – 174.
- 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 (2012): Stable climate coalitions (Nash) and international trade. *CESifo working paper* 3915.
- Eichner, T. and R. Pethig (2013): Self-enforcing environmental agreements and international trade. *Journal of Public Economics* 102, 37 – 50.
- Finkelshtain, I. and Y. Kislev (1997): Prices versus quantities: The political perspective. *Journal of Political Economy* 105, 83 – 100.
- Finus, M. (2003): Stability and design of international environmental agreements: The case of transboundary pollution, in: H. Folmer and T. Tietenberg (eds.), *The International Yearbook of Environmental and Resource Economics* 2003/2004, Edward Elgar, Cheltenham, 82 – 158.
- Finus, M. (2001): *Game Theory and International Environmental Cooperation*, Edward Elgar, Cheltenham.
- Finus, M. and D.T.G. Rübbelke (2013): Public good provision and ancillary benefits: The case of climate agreements. *Environmental and Resource Economics* 56, 211 – 226.

- Hoel, M. (1992): International environmental conventions: The case of uniform reductions of emissions. *Environmental and Resource Economics* 2, 141 – 159.
- Hoel, M. and K. Schneider (1997): Incentives to participate in an international environmental agreement. *Environmental and Resource Economics* 9, 153 – 170.
- Hong, F. and L. Karp (2012): International environmental agreements with mixed strategies and investment. *Journal of Public Economics* 96, 685 – 697.
- Ishikawa, J. and K. Kiyono (2006): Greenhouse-gas emission controls in an open economy. *International Economic Review* 47, 431 – 450.
- Karp, L. and J. Zhang (2012): Taxes versus quantities for a stock pollutant with endogenous abatement costs and asymmetric information. *Economic Theory* 49, 371 – 409.
- Kiyono K. and J. Ishikawa (2013): Environmental management policy under international carbon leakage. *International Economic Review* 54, 1057 – 1083.
- Kolstad, C. (2007): Systematic uncertainty in self-enforcing international environmental agreements. *Journal of Environmental Economics and Management* 53, 68 – 78.
- Mansur, E.T. (2013): Prices vs. quantities. *Journal of Regulatory Economics* 44, 80 – 102.
- Montero, J.P. (2002): Prices vs. quantities with incomplete enforcement. *Journal of Public Economics* 85, 435 – 454.
- Na, S.-I. and H.S. Shin (1998): International environmental agreements under uncertainty. *Oxford Economic Papers* 50, 173 – 185.
- Pizer, W. A. (2002): Combining price and quantity controls to mitigate global climate change. *Journal of Public Economics* 85, 409 – 434.
- Rubio, S.J. and A. Ulph (2006): Self-enforcing agreements and international trade in greenhouse emission rights. *Oxford Economic Papers* 58, 233 – 263.
- Strand J. (2010): Taxes versus cap and trade in climate policy when only some importer abate. *CESifo working paper* no. 3233.
- Ulph, A. (1996): Environmental policy instruments and imperfectly competitive international trade. *Environmental and Resource Economics* 7, 333 – 355.
- Weitzman, M. (1974): Prices vs. quantities. *Review of Economic Studies* 41, 477 – 491.



# Appendix

## A: Derivation and comparison of BAU and social optimum

**Business as usual.** Maximizing  $W^i(t_1, \dots, t_n)$  from (12) with respect to  $t_i$  yields the first-order condition

$$W_{t_i}^i = t_i \frac{\partial e_i^d}{\partial t_i} + (e_i^s - e_i^d) \frac{\partial p}{\partial t_i} - D' \left( \sum_j e_j^d \right) \cdot \frac{\partial \left( \sum_j e_j^d \right)}{\partial t_i} = 0, \quad (\text{A1})$$

where

$$\frac{\partial p}{\partial t_i} = \frac{\partial p}{\partial t_j} = -\frac{\alpha}{(\alpha + b)n} < 0, \quad (\text{A2})$$

$$\frac{\partial e_i^d}{\partial t_i} = -\frac{(n-1)nb + \alpha}{(\alpha + b)bn} < 0, \quad \frac{\partial e_i^d}{\partial t_j} = \frac{\alpha}{(\alpha + b)bn} > 0, \quad (\text{A3})$$

$$\frac{\partial \left( \sum_j e_j^d \right)}{\partial t_i} = \frac{\partial \left( \sum_j e_j^d \right)}{\partial t_j} = -\frac{1}{\alpha + b} < 0. \quad (\text{A4})$$

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

$$W_{t_i}^i = -t_i \cdot \frac{\alpha(n-1) + bn}{(\alpha + b)bn} - \left( t_i - \frac{\sum_j t_j}{n} \right) \frac{\alpha}{(\alpha + b)bn} + \frac{\delta}{(\alpha + b)^2} \left( na - \sum_j t_j \right) = 0. \quad (\text{A5})$$

Solving (A5) with respect to  $t_i$  establishes (13).

**Social optimum.** We maximize  $\sum_j W^j(t_1, \dots, t_n)$  with respect to  $t_i$  and obtain

$$\sum_j W_{t_i}^j = \sum_j \left[ t_j \frac{\partial e_j^d}{\partial t_i} + (e_j^s - e_j^d) \frac{\partial p}{\partial t_i} - D' \left( \sum_j e_j^d \right) \cdot \frac{\partial \left( \sum_j e_j^d \right)}{\partial t_i} \right] = 0. \quad (\text{A6})$$

Making use of (A2) - (A4) and the symmetry condition in (A6), we get  $\hat{t} = \frac{n^2 a \delta}{b + \alpha + n^2 \delta}$  which in turn inserted in  $e_i^s$  from (9) yields  $\hat{e} = \frac{a}{b + \alpha + n^2 \delta}$ .

## 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

$$m \left[ V(e_c^s) + T(e_c^s) + p(e_c^s - e_c^d) - D \left( \sum_j e_j^s \right) \right], \quad (\text{B1})$$

where

$$e_c^s = \frac{1}{(\alpha + b)n} [na - (mt_c + (n - m)t_f)], \quad (\text{B2})$$

$$e_c^d = \frac{a}{b} - \frac{t_c}{b} - \frac{\alpha}{(\alpha + b)bn} [na - (mt_c + (n - m)t_f)], \quad (\text{B3})$$

$$p = \frac{\alpha}{(\alpha + b)n} [na - (mt_c + (n - m)t_f)], \quad (\text{B4})$$

$$\sum e_j^s = \frac{1}{(\alpha + b)} [na - (mt_c + (n - m)t_f)] \quad (\text{B5})$$

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

$$t_c \frac{\partial e_c^d}{\partial t_c} + (e_c^s - e_c^d) \frac{\partial p}{\partial t_c} - D' \left( \sum_j e_j^s \right) \cdot \frac{\partial \sum_j e_j^s}{\partial t_c} = 0 \quad (\text{B6})$$

or equivalently

$$-\frac{t_c[nb + (n - m)\alpha]}{nb} - \frac{m(n - m)\alpha(t_c - t_f)}{n^2b} + \frac{m\delta}{(b + \alpha)} [na - (mt_c + (n - m)t_f)] = 0. \quad (\text{B7})$$

Solving (B7) with respect to  $t_c$  yields (14).

## C: Lemmas

### Lemma 1.

- (i) The welfare level of coalition countries,  $\mathcal{W}^c(m)$ , is strictly increasing in  $m$  if  $m > \sqrt{n}$ .
- (ii) The welfare level of fringe countries,  $\mathcal{W}^f(m)$  is strictly increasing in  $m$ .
- (iii) The fossil fuel production,  $\mathcal{E}^s(m)$ , is strictly decreasing in  $m$ .
- (iv) The fossil fuel consumption,  $\mathcal{E}^{df}(m)$ , is strictly increasing in  $m$ .
- (v) The fossil fuel consumption,  $\mathcal{E}^{dc}(m)$ , is strictly decreasing in  $m$ .

**Proof:** Differentiation of  $t_c^* = \mathcal{T}^c(m)$  and  $t_f^* = \mathcal{T}^f(m)$  from (16) with respect to  $m$  yields

$$\mathcal{T}_m^c = \frac{mn^2ab\delta[(2m - 1)\alpha^2 + b[(2n^2 - n + m^2)\alpha + (n - m)^2n^2\delta]]}{N^2}, \quad (\text{C1})$$

$$\mathcal{T}_m^f = \frac{(2m - 1)abn^3\delta[\alpha(\alpha + b) - n^2b\delta]}{N^2}, \quad (\text{C2})$$

We insert (16) in  $W^c(t_c, t_f, m)$ ,  $W^f(t_c, t_f, m)$  and  $e_c^s = e_f^s, e_f^d, e_c^d$  from (8), and differentiate with respect to  $m$  to obtain, after tedious rearrangement of terms:

$$\begin{aligned}\mathcal{W}_m^c &= \frac{n^6 a^2 b \delta [(m-1)n^2 b + (n-m)(n(m-1) + m^2)\alpha^2]}{Z^3}, \\ &\quad + \frac{n^6 a^2 b^2 \delta [n(n-m)(m-1) + n^2(n+1)]\alpha + m(m^2 - n)n^2 \delta}{Z^3}, \\ \mathcal{W}_m^f &= \frac{a^2 b (2m-1)n^6 \delta^2 [n^2 b + (n^2 - 1)\alpha + (2n^2 - 1)b\alpha + n^2 b \delta]}{Z^3} > 0, \\ \mathcal{E}_m^s &= -\frac{abn^4(2m-1)(\alpha + b)\delta}{Z^2} < 0, \\ \mathcal{E}_m^{df} &= \frac{an^3(2m-1)\delta[(n-1)(\alpha + b)\alpha + n^2 b \delta]}{Z^2} > 0, \\ \mathcal{E}_m^{dc} &= -\frac{an^3 \delta [n^2 b^2 + (n-m)^2 \alpha^2 + [(2n(n-m) + m^2)\alpha b + (n^2 - m^2)n^2 b \delta]]}{Z^2} < 0,\end{aligned}$$

where

$$Z := n^2 b + [n(n-1) - m(m-1)]\alpha^2 + b[(n(2n-1) - m(m-1))\alpha + (n + m(m-1))n^2 \delta] > 0.$$

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

- (i)  $\mathcal{W}^f(m) > \mathcal{W}^c(m) > w_o$ ,
- (ii)  $\mathcal{E}^{df}(m) > e_o > \mathcal{E}^s(m) > \mathcal{E}^{dc}(m)$  for all  $m > 2$ .

**Proof:** Comparing  $t_c^*, t_f^*$  yields

$$t_c^* - t_f^* = \frac{(m-1)n^3 ab \delta}{N} > 0. \quad (\text{C3})$$

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

$$W^c(t_c, t_f) - W^f(t_c, t_f) = \frac{t_f^2 - t_c^2}{2b}. \quad (\text{C4})$$

Making use of  $t_f^* < t_c^*$  in (C4) we get  $W^f(t_c^*, t_f^*; m) > W^c(t_c^*, t_f^*; m)$ .

Subtracting  $e_c^d$  and  $e_f^d$  from  $e_c^s$  yields

$$e_c^s - e_c^d = \frac{(n-m)(t_c - t_f)}{bn}, \quad \text{and} \quad e_c^s - e_f^d = -\frac{m(t_c - t_f)}{bn}. \quad (\text{C5})$$

Accounting for  $t_f^* < t_c^*$  in (C4) we get  $\mathcal{E}^{df}(m) > \mathcal{E}^s(m) > \mathcal{E}^{dc}(m)$ . Finally,  $\mathcal{E}^{df}(1) = \mathcal{E}^s(1) = \mathcal{E}^{dc}(1) = e_o$  and Lemma 1(iii)-(v) establish Lemma 2(ii).

## D: Proof of propositions

**Proof of Proposition 1:** Making use of  $\hat{w} = \frac{a^2}{2(b+\alpha+n^2\delta)}$  and inserting  $t_c^*$  and  $t_f^*$  from (16) in  $W^f(t_c, t_f; m)$  yields after rearrangement of terms

$$\begin{aligned} & \hat{w} - W^f(n-1) \\ &= \frac{(n-1)n^4a^2\delta^2F(b, \alpha, \delta, n)}{2(b+\alpha+n^2\delta)[bn^2+2(n-1)\alpha^2+b[(n^2+2(n-1))\alpha+(n^2-2(n-1))\delta]]^2}, \end{aligned} \quad (\text{D1})$$

where

$$F(b, \alpha, \delta, n) := 4(n-1)[\alpha(\alpha+b) - n^2b\delta] - [n^2(n-3)((\alpha+b)b + n^2b\delta)]. \quad (\text{D2})$$

Since the denominator in (D1) is positive, we infer from (D1)

$$\hat{w} \gtrless W^f(n-1) \iff F(b, \alpha, \delta, n) \gtrless 0. \quad (\text{D3})$$

Further rearranging (D3) yields

$$F = [4(n-1) + n^2(n-3)][\alpha(\alpha+b) - n^2b\delta] - n^2(n-3)(\alpha+b)^2. \quad (\text{D4})$$

Proposition 1(i) follows immediately from (D4).

Proposition 1(ii) follows from the properties of the function  $F$  derived below:

- Dependence of  $F$  on  $n$ :

$$\begin{aligned} F(b, \alpha, \delta, 2) &= 4(\alpha+b)^2 > 0, \quad \lim_{n \rightarrow \infty} F(b, \alpha, \delta, n) < 0, \\ F_n &= -3b^2n(n-2) - 3\alpha b(3n^2 - 6n - 4) + 4\alpha^2 - nb\delta[5n^3 - 12n(n-1) + 8], \end{aligned}$$

$F_n < 0$  if  $n$  is sufficiently large;

- Dependence of  $F$  on  $\delta$ :

$$\begin{aligned} F(b, \alpha, 0, n) &= -(\alpha+b)[(n-3)bn^2 - 4(n-1)\alpha], \quad \lim_{\delta \rightarrow \infty} F(b, \alpha, \delta, n) < 0, \\ F_\delta &= -b[n^2(n-3) + 4(n-1)] < 0; \end{aligned}$$

- Dependence of  $F$  on  $\alpha$ :

$$\begin{aligned} F(b, 0, \delta, n) &= -(n-3)n^2b^2 - n^2b\delta[4(n-1) + n^2(n-3)] < 0, \quad \lim_{\alpha \rightarrow \infty} F(b, \alpha, \delta, n) > 0, \\ F_\alpha &= 8(n-1)\alpha - b[n^2(n-3) + 4(n-1)], \end{aligned}$$

$F_\alpha > 0$  if  $\alpha$  is sufficiently large;

- Dependence of  $F$  on  $b$ :

$$\begin{aligned} F(0, \alpha, \delta, n) &= (n-1)4\alpha^2 > 0, \quad \lim_{b \rightarrow \infty} F(b, \alpha, \delta, n) < 0, \\ F_b &= -2b(n-3)n^2 + \alpha[n^3 - 3n^2 + 4(n-1)] - n^2\delta[n^2(n-3) + 4(n-1)], \end{aligned}$$

$F_b < 0$  if  $b$  is sufficiently large.

**Proof of Proposition 2:** The sign of the stability condition  $\mathcal{W}^c(n) - \mathcal{W}^f(n-1)$  is independent of the parameter  $a$ . Figure 12 - 14 provide the proof for the parameter subspace

$$\begin{aligned} M^\alpha &= \{(a, \alpha, b, \delta, n) \mid b = 20, 0 \leq \alpha \leq 5000, \delta = 1, n = 10\}, \\ M^b &= \{(a, \alpha, b, \delta, n) \mid 0 \leq b \leq 100, \alpha = 1000, \delta = 1, n = 10\}, \\ M^\delta &= \{(a, \alpha, b, \delta, n) \mid b = 20, \alpha = 1000, 0 \leq \delta \leq 30, n = 10\}, \\ M^n &= \{(a, \alpha, b, \delta, n) \mid b = 20, \alpha = 1000, \delta = 1, 3 \leq n \leq 25\}. \end{aligned}$$

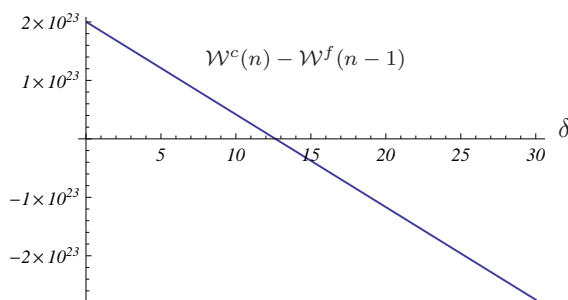


Figure 12: Variations of  $\delta$

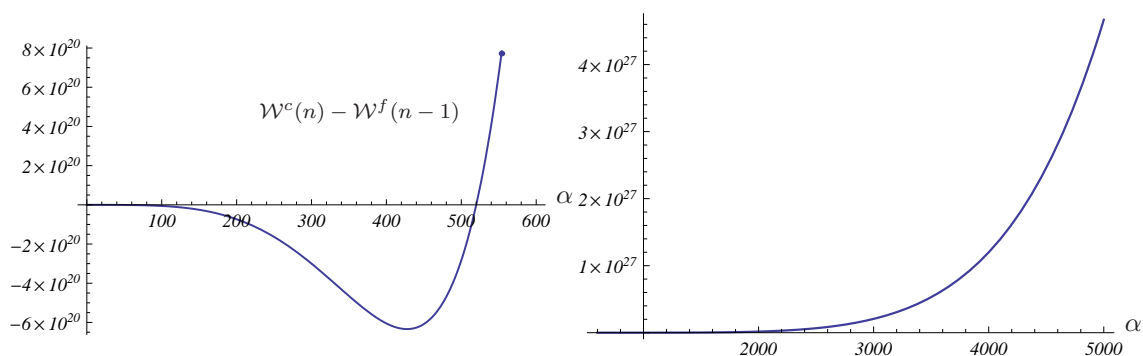


Figure 13: Variations of  $\alpha$

For  $v = \alpha, b, \delta, n$  define  $M^{v+} = \{(a, b, \alpha, \delta, n) \in M^v \mid \mathcal{W}^c(n) \geq \mathcal{W}^f(n-1)\}$  and  $\mathcal{M}^+ := M^{\alpha+} \cup M^{b+} \cup M^{\delta+} \cup M^{n+}$ . From the Figures 12 through 14 it is straightforward that  $\mathcal{M}^+$  is non-empty and quite large.

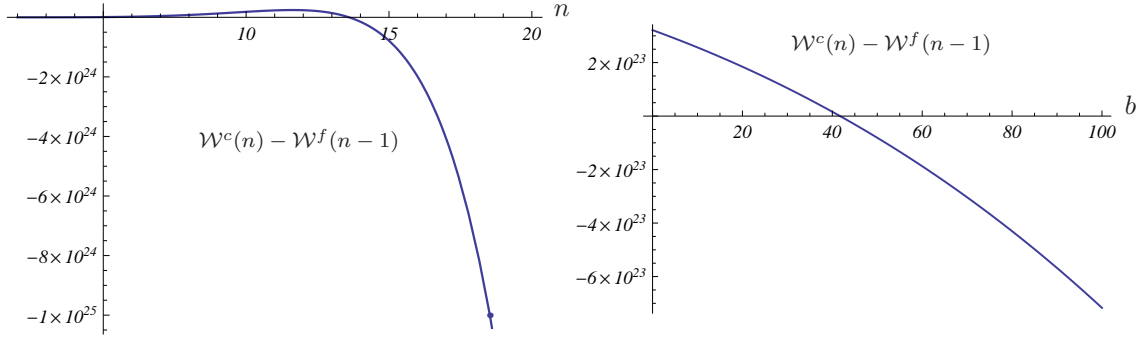


Figure 14: Variations of  $n$  and  $b$

### E: Leakage rates and strategic complementarity

To understand the link between leakage rates and coalition size we solve

$$e_i^d = \frac{a}{b} - \frac{t_i}{b} - \frac{\alpha}{(\alpha + b)bn} \left( na - \sum_j t_j \right) \quad i = 1, \dots, n \quad (\text{E1})$$

from (9) with respect to  $t_1, \dots, t_n$  and get

$$t_i = a - be_i^d - \frac{\alpha}{n} \sum_j e_j^d \quad i = 1, \dots, n. \quad (\text{E2})$$

Introducing the notation  $s_c = me_c^d$  and  $s_f = (n - m)e_f^d$  we consider a coalition country  $i \in C$  and a fringe country  $i \in F$  and rewrite (E2) as

$$t_c = a - \frac{bs_c}{m} - \frac{\alpha}{n}(s_c + s_f), \quad (\text{E3})$$

$$t_f = a - \frac{bs_f}{n - m} - \frac{\alpha}{n}(s_c + s_f). \quad (\text{E4})$$

Inserting (E3) and (E4) and solving with respect to  $s_f$  yields the fringe's reaction function in terms of emissions

$$s_f = R(s_c, m) = \frac{(n - m)[bn + (n - 1)\alpha]an}{A} - \frac{(n - m)[(n - 1)(\alpha + b)\alpha + n^2b\delta]}{A}s_c, \quad (\text{E5})$$

where

$$A := b^2n^2 + (n - m)(n - 1)\alpha^2 + [n(2n - 1) - m(n - 1)]b\alpha + (n - m)n^2b\delta > 0.$$

Differentiation with respect to  $s_c$  and  $m$  yields

$$R_{s_c} = -\frac{(n - m)[(n - 1)(\alpha + b)\alpha + n^2b\delta]}{A} \in [-1, 0], \quad (\text{E6})$$

$$R_{s_c\delta} = \frac{bn^2(b + \alpha)(n - m)(bn^2 + 2(n - 1)\alpha(n - m))}{A^2} > 0 \quad (\text{E7})$$

The leakage rate is expressed by  $|R_{sc}| \in ]0, 1[$ . Since  $R_{sc\delta} > 0$ , the leakage rate is declining in  $\delta$ . The lower  $\delta$ , the stronger the strategic complementarity, the less is the leakage rate declining.

## F: The modified model

**Derivation of equation (32):** Maximizing sector  $X$ 's profit (30) yields the fuel demand

$$e_i^d = \frac{\tilde{\alpha} - p - t_i}{\tilde{\beta}}. \quad (\text{F1})$$

For given  $p$ , the equilibrium permit price is obviously determined by (F1) and

$$e_i^d = e_i \quad \text{for } i = 1, \dots, n. \quad (\text{F2})$$

Combine the first-order condition  $U'(y_i^d) = p$  of maximizing  $U(y_i^d) + x_i^d$  subject to (31) with equations  $y_i^d = y_i^s = r_{yi}$ ,  $e_i^s = r_{ei}$  and (27) to obtain the fuel supply function

$$e_i^s = \frac{p + \tilde{b}\tilde{r} - \tilde{a}}{\tilde{b}}. \quad (\text{F3})$$

The equations (25) and (F2) imply

$$e_i^s = \frac{\sum_j e_j}{n} \quad \text{for } i = 1, \dots, n. \quad (\text{F4})$$

Equating (F3) and (F4) determines the unique fuel price

$$p = \tilde{a} - \tilde{b}\tilde{r} + \frac{\tilde{b}\sum_j e_j}{n}. \quad (\text{F5})$$

Making use of (F2) and (F6) in (F1) yields the unique permit price

$$t_i = \tilde{\alpha} - \tilde{a} + \tilde{b}\tilde{r} - \frac{\tilde{b}\sum_j e_j}{n} - \tilde{\beta}e_i. \quad (\text{F6})$$

Finally, the supply of and the demand for the consumer goods  $X$  and  $Y$  are, respectively,

$$y_i^s = \tilde{r} - \frac{\sum_j e_j}{n} \quad \text{and} \quad x_i^s = X(e_i), \quad (\text{F7})$$

$$y_i^d = \tilde{r} - \frac{\sum_j e_j}{n} \quad \text{and} \quad x_i^d = X(e_i) + p \cdot \left( \frac{\sum_j e_j}{n} - e_i \right). \quad (\text{F8})$$

Inserting  $y_i^d$  and  $x_i^d$  from (F8) in the parametric version of (23) yields (32).

**Derivation of equation (34):** Note that the equations (29), (F1) and (F4) still hold. We consider  $e_i^d$  from (F1) and  $e_i^s$  from (F4) in the market equilibrium condition (25) and solve for the fuel price. Then we insert that price into the functions of fuel supply and demand and obtain

$$p = \frac{\tilde{b}}{(\tilde{b} + \tilde{\beta})n} \left( \frac{n(\tilde{\alpha}\tilde{b} + (\tilde{a} - \tilde{b}\tilde{r})\tilde{\beta})}{\tilde{b}} - \sum_j t_j \right), \quad (\text{F9})$$

$$e_i^s = \frac{\tilde{1}}{(\tilde{b} + \tilde{\beta})n} \left( n(\tilde{\alpha} - \tilde{a} + \tilde{b}\tilde{r}) - \sum_j t_j \right), \quad (\text{F10})$$

$$e_i^d = \frac{\tilde{\alpha}}{\tilde{\beta}} - \frac{t_i}{\tilde{\beta}} - \frac{\tilde{b}}{(\tilde{b} + \tilde{\beta})\tilde{\beta}n} \left( \frac{n(\tilde{\alpha}\tilde{b} + (\tilde{a} - \tilde{b}\tilde{r})\tilde{\beta})}{\tilde{b}} - \sum_j t_j \right), \quad (\text{F11})$$

$$\sum_j e_j^d = \sum_j e_j^s = \frac{\tilde{1}}{(\tilde{b} + \tilde{\beta})} \left( n(\tilde{\alpha} - \tilde{a} + \tilde{b}\tilde{r}) - \sum_j t_j \right). \quad (\text{F12})$$

Accounting for (21), (26), (27),  $r_{ei} = e_i^s$ ,  $x_i^d = X(e_i^d) + p(e_i^s - e_i^d)$  turns the parametric version of (23) into

$$W^i(t_1, \dots, t_n) = \tilde{a}(\tilde{r} - e_i^s) - \frac{\tilde{b}}{2} (\tilde{r} - e_i^s)^2 + \tilde{\alpha}e_i^d - \frac{\tilde{\beta}}{2} (e_i^d)^2 + p(e_i^s - e_i^d) - \frac{\tilde{\delta}}{2} \sum_j e_j^d. \quad (\text{F13})$$

Finally invoke  $p$ ,  $e_i^s$ ,  $e_i^d$  and  $\sum_j e_j^d$  from (F9)-(F12) to convert (F13) into (34).