The stability and effectiveness of climate coalitions: A comparative analysis of multiple integrated assessment models

Kai Lessmann[*] Ulrike Kornek[*] Valentina Bosetti[†]
Rob Dellink[‡] Johannes Emmerling[§] Johan Eyckmans[¶]
Miyuki Nagashima[||] Hans-Peter Weikard[‡] Zili Yang[**]

October 2, 2013

Abstract: In this paper we report results from a comparison of numerically calibrated game theoretic integrated assessment models that explore stability and performance of international coalitions for climate change mitigation. Specifically, by means of this ensemble of models we are able to identify robust results concerning incentives of nations to commit themselves to a climate agreement, and to estimate what stable agreements can achieve in terms of greenhouse gas mitigation. We also assess the potential of transfers that redistribute the surplus of cooperation in order to foster stability of climate coalitions. In contrast to much of the existing analytical game theoretical literature, we find substantial scope for self-enforcing climate coalitions in most models that close much of the abatement and welfare gap between complete absence of cooperation and full cooperation. This more positive message follows from the use of transfer schemes that are designed to counteract free riding incentives.

Keywords: coalition stability, international environmental agreements, numerical modeling, transfers

---

[*] Potsdam Institute for Climate Impact Research (PIK), PO Box 60 12 03, D-14412 Potsdam, Germany

[†] Università Bocconi and Fondazione Eni Enrico Mattei (FEEM) Milan, Italy

[‡] Environmental Economics and Natural Resources Group, Department of Economics, Wageningen University, The Netherlands

[§] Fondazione Eni Enrico Mattei (FEEM) and Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), Italy

[¶] KU Leuven – Center for Economics and Corporate Sustainability (CEDON), Warmoesberg 26, B-1000 Brussels, Belguim

[||] Research Institute of Innovative Technology for the Earth, Kyoto, Japan

[**] Department of Economics, State University of New York at Binghamton, Binghamton, New York, USA

1 Introduction

1.1 Motivation

This paper reports results from a comparison of models that explore international coalitions for climate policy. Specifically, these models investigate the incentives of nations to commit themselves to a climate agreement, and what makes climate agreements stable in the sense that participation is enforced by the self-interest of its members. Since climate change is a global externality and thus not fully taken into account, national actions are in general not globally efficient. How much a climate coalition improves upon this dilemma depends on the costs and benefits of the individual nations.

All models in this study use a numerical approach to explore climate coalition formation in a cost-benefit structure that reflects real-world regions and their dynamics. Three topics where these numerical models give particularly valuable insights beyond those from their analytical counterparts are (i) the impact of asymmetry, i.e. the regional heterogeneity observed across the world; (ii) quantitative estimates (of the order-of-magnitudes), in particular when trade-offs leave the net effects on, say, coalition stability or free-riding incentives ambiguous; and (iii) the impact of sophisticated (non-orthogonal) reaction functions implict in the models, for example related to the effects of carbon leakage that add to those incentives directly related to the environmental externality. For these topics, both the mechanics and the calibration of the models are of central importance, but the uncertainties both within and between models are large. The strength of our comparison exercise using multiple models is threefold: it will make uncertainty more transparent, help to identify robust results across modeling assumptions and parameterizations, and enable learning from the differences.

Early theoretical investigations resulted in a rather pessimistic assessment of the scope of self-enforcing agreements to cope with international environmental issues. In their seminal paper Carraro and Siniscalco (1993) find stable coalitions generally to be small, and while stable coalitions may be large in the model setup of Barrett (1994), this only holds if the gains from cooperating are small. Much of the ensuing research has investigated this dilemma and ways around it; for a survey of the literature see Finus (2008) and recent advances are discussed in Fuentes-Albero and Rubio (2010).

Transfer payments were identified as likely game-changers, and Barrett (2001) stresses their complementarity to the asymmetry of players: together, transfers and asymmetry may well improve cooperation. But not all transfer schemes have potential to improve the success of environmental agreements: transfers that do not take the strategic implications into account are unlikely to improve an agreement and even may hurt its success (Nagashima et al., 2009; Weikard et al., 2006). In fact, transfers designed specifically to make cooperation more attractive than free-riding greatly improve the success of coalitions (Carraro et al., 2006; Nagashima et al., 2009; Weikard, 2009). Similarly, McGinty (2007) shows the beneficial effect of such transfers within the framework of Barrett (1994).

Recently, studies have focused specifically on the role of asymmetry: Weikard (2009) shows that higher levels of participation under such transfers are spurred by stronger asymmetry among the coalition members, including the grand coalition. This is confirmed by Fuentes-Albero and Rubio (2010) who elaborate that differences in marginal damages (rather than abatement costs or the level of damages) are key to this result. While these studies firmly establish a more optimistic prospect for cooperation and highlight the importance of heterogeneity, the degree of asymmetry often remains a conceptual assumption. Notable exceptions are Carraro et al. (2006) and Nagashima et al. (2009), which rely on integrated assessment models to quantify asymmetries.

This paper is unique in drawing on five (new or updated) projections of the underlying real-world heterogeneity from integrated assessment models of different complexity. These models relax limitations of the more stylized models such as linearity in damages and transferable utility that are common to all studies discussed above. More importantly, their joint assessment allows us to represent the large uncertainties in the estimates of the benefits and costs of climate change mitigation.

Our aim is to explore the implications of real-world asymmetry for coalition stability. To this end, we assess the regional abatement costs and climate change damages in the models, and how this cost-benefit structure translates into the incentive to engage in a climate agreement for specific regions of the world. Our contribution is a better understanding of the well-known cooperation failure, particularly in the heterogeneous setting provided by these numerical models. In addition, we explore the role of transfers and assess their magnitude and direction when used as a tool to enhance cooperation.

1.2 International Climate Agreements

Central to this study is the concept of 'self-enforcing agreements' or 'coalition stability.' A climate coalition is a subset of the world's regions that agree to cooperate on climate change mitigation policies. More specifically, we stipulate that within the coalition climate change is addressed in an efficient manner, i.e. a manner that maximizes coalitional welfare.[1] The coalition adopts a joint climate policy and interacts with the remaining regions as a single player each acting selfishly with respect to the other.

In our comparison exercise we apply the concept of cartel stability (d'Aspremont and Gabszewicz, 1986) to all models. A coalition is considered stable if it is both internally stable, meaning that no member is willing to leave the coalition, and externally stable, meaning that all non-members prefer to remain singletons. Formally, any given coalition S is stable if for the payoff of player i facing coalition S, $\pi_i(S)$, we have

$$\pi_i(S) \geq \pi_i(S \setminus \{i\}) \text{ for all } i \in S, \quad \pi_j(S) \geq \pi_j(S \cup \{j\}) \text{ for all } j \notin S \tag{1}$$

This notion of cartel stability was first applied to international environmental agreements by Hoel (1992), Carraro and Siniscalco (1993), and Barrett (1994).

Building on this concept, we aim at exploring the drivers of cooperation. In particular we also examine the effects of transfer schemes on the prospects for cooperation. We employ the concept of potential internal stability (PIS) introduced in the literature on international environmental agreements for transferable utility by Carraro et al.

---

[1] Most models implement Pareto-efficiency through maximization of the utilitarian sum of individual welfare per region. MICA computes a market-equilibrium with full internalization of the climate change externality.

(2006). A coalition is said to be potentially internally stable (PIS) if there exists a transfer scheme that redistributes payoffs within the coalition such that the coalition is internally stable. Formally, PIS requires the existence of a vector of transfers τ_i such that

$$\pi_i(S) + \tau_i \geq \pi_i(S \setminus \{i\}) \text{ for all } i \in S \tag{2}$$

and the sum of all transfers is zero. Note that the simple addition of transfers in condition (2) is only appropriate in models with transferable utility. For models that do not assume transferable utility but feature a transferable commodity (e.g. consumption), transfers can be implemented at the commodity level. Here, a transfer scheme comprises in a redistribution of the commodity between regions at each time period. For details of the applied procedure see Kornek et al. (2013).

The rest of the paper is structured as follows. An overview of the different integrated assessment models used in this analysis is given in section 2. Section 3 focusses on the role of transfers. Section 4 summarizes results and concludes.

2 Characterization of the Models

All models in this study are built upon Ramsey's dynamic model, four of them include its optimal consumption/savings decision (Ramsey, 1928). This so-called optimal growth framework seems particularly apt for economic models with a long time horizon as required in the assessment of climate change impacts on the economy. An early model to generalize this approach to multiple regions in an application to global warming was the RICE model (Nordhaus and Yang, 1996), which participates in this study in an updated version (Yang, 2008). Closely related is the ClimNeg World Simulation (CWS) model, a modified version of the RICE model, updated to new data in its cost and damage parameters (Bréchet et al., 2011; Eyckmans and Finus, 2006; Eyckmans and Tulkens, 2003). The Model of International Climate Agreements (MICA) follows the same economic framework but also with different assumptions about costs and benefits. Formally, its main distinction is to include international goods markets (Lessmann and Edenhofer, 2011; Lessmann et al., 2009). The fourth model abstracts from the consumption/savings decision: the Stable Coalitions model (STACO) takes as a starting point the balancing of marginal benefits and marginal costs of emission reductions and builds onto this a Ramsey-type dynamic abatement structure (Nagashima et al., 2009, 2011). All four, RICE, CWS, MICA, and STACO, remain relatively stylized in their description of the world economy. The World Induced Technological Change model (WITCH) is a state-of-the-art integrated assessment model of global warming (Bosetti et al., 2006). In contrast to the aforementioned models, which rely on stylized abatement cost functions to model emissions reductions, WITCH incorporates an explicit representation of mitigation options, particularly in the energy system. There is, of course, a trade-off: with increasing computational complexity it becomes less feasible to explore all possible climate coalitions. For WITCH, our study must therefore resort to selected coalitions.

2.1 Non-cooperative and fully cooperative equilibria

The five models in this study represent quite different views of the world economy. This is evident from Table 1, which documents model assumptions and basic behavior in selected numbers.

A key difference of the models is how they value the present against the future. Concerning monetary values such as abatement costs or climate change damages, this is expressed in the models' endogenous interest rate. Simple Ramsey models suggest that this interest rate depends on the pure rate of time preference and, if the intertemporal elasticity of substitution is strictly positive, on consumption growth.[2] Table 1 shows how models differ in the assumed preference parameters. Together with different projections about productivity growth, this results in growth rates of economic output ranging from 1.2 percent to 2.1 percent per year over the first century. The pure rate of time preference is highest for MICA,WITCH, and RICE at 3 percent, and consequently so are the interest rates in these models (around 5 percent).[3] For STACO, the pure rate of time preference is lower (at 2 percent) but the (exogenous) assumption of relatively strong growth in the coming decades leads to a fairly high initial discount rate, especially for emerging economies, which slowly declines over time to values of around 3 percent, and 4.2 percent on average. Finally, in CWS the interest rate is the same as the pure rate of time preference, 1.5 percent, which is the lowest among the models.

In the non-cooperative equilibrium (NC), i.e. where no coalition forms, greenhouse gas (GHG) emissions are of the same order of magnitude in all models, with moderately lower values in MICA and RICE. Non-cooperative emission reductions a of comparable magnitude in most models (about 10 percent of emissions), and about half of that in RICE. In the social optimum, emissions are strongly reduced relative to NC within the first century (again, with the exception of RICE). In the other models, mitigation reduces climate change damages relative to GDP by several percentage points by the end of the century. Especially in STACO and WITCH, the high damages from the non-cooperative scenario are reduced by about four percentage points. In RICE, the formation of the grand coalition leaves climate change damages almost unchanged (we will see later that low damage estimates in RICE are the likely cause for this).

2.2 Cost/benefit information

In this section, we introduce two metrics to characterize abatement costs and the severity of climate change damages in the models, both globally and on the regional scale. Perhaps the most intuitive metric would be to compare marginal cost and marginal damage functions. However, this information is not easy to extract from or compare between all different models. Our metrics are therefore based on model output of scenario runs rather than assumptions regarding functional forms and parameter values.

---

[2] This follows from the Keynes-Ramsey rule $\dot{c}/c = 1/\eta\ (r-\rho)$ with per capita consumption c, elasticity of intertemporal substitution η and pure rate of time preference ρ. At the interest rate r households are indifferent between one unit now or $(1+r)$ units later.

[3] Specifically, the pure rates of time preference are constant in RICE and MICA, but diminish in WITCH from 3 percent initially to 2 percent over the course of a century.

Table 1: Modeling assumptions and key numbers of non-cooperative and fully cooperative model dynamics

Modeling assumptions	MICA	STACO	CWS	WITCH	RICE
Initial year	2005	2011	2000	2005	2000
Time Horizon (years)	190	95	330	145	245
Pure rate of time preference (percent)	3.0	2.0	1.5	3.0	3.0
Intertemp. elast. of subst.	1.0	–	0.0	1.0	1.0
Non-cooperative model behavior	MICA	STACO	CWS	WITCH	RICE
RICE					
Mean GDP growth rate[b]	2.06	1.97	1.54	1.56	1.24
Mean interest rate[c]	5.26	4.17	1.50	5.35	4.98
CO2 emissions (GtC) 2015-2100	1516	1827	1754	1963	1404
Non-cooperative CO2 reductions (percent)[d]	9.8	12.1	10.2	13.0	5.0
Mean CO2 intensity (GtC/tn$)	0.12	0.14	0.13	0.12	0.13
Climate change damage in 2100 (percent)[e]	5.8	7.8	3.2	9.3	1.6
Carbon price 2100: reg. mean ($/tC)	13	89	49	38	8
Cooperative model behavior	MICA	STACO	CWS	WITCH	RICE
CO2 emissions (GtC) 2015-2100	953	984	1094	1122	1242
Climate change damage in 2100 (percent)[e]	3.8	4.0	1.9	4.9	1.5
Carbon price 2100: reg. mean ($/tC)	391	966	529	858	208
Carbon price growth rate to 2100 (percent)	1.90	1.69	0.90	1.02	1.02

a STACO derives the interest rate for discounting payoffs using the Keynes-Ramsey rule to ensure consistency with a logarithmic utility function and a pure rate of 2 per cent

b Using a time horizon of 100 years

c The endogenous rate at which monetary values are discounted in the model, averaged over regions and time

d Emission reduction in the non-cooperative scenario relative to a business-as-usual scenario without climate change damages

e Damages are reported as a share of 2100 economic product

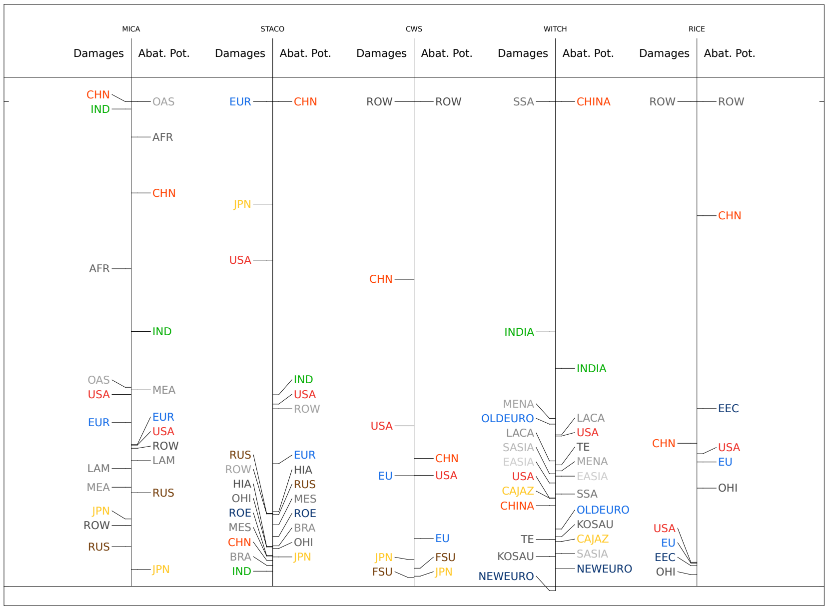

Figure 1: Abatement potential and climate change damages indicators scaled to $[-1, 1]$. Abatement potential was calculated by in every model a common carbon tax trajectory for all regions. The resulting abatement trajectory (measured in tons) was integrated over model's time horizon and scaled relative to the maximum abatement level. The climate change marginal damage indicator for a particular region was calculated by taking the average of the difference in carbon price of the grand coalition and the grand coalition minus the particular region at hand. This indicator was normalized relative to the maximum average difference over all regions. Model regions are specified in Tables 7-11.

Measuring regional abatement costs

Regional abatement costs are inversely related to the regional reductions of emissions at a uniform global carbon price – the higher the emissions reduction at this price, the lower the associated average costs per unit of reduction. We therefore take cumulative regional abatement from a common tax scenario (which results in a uniform global carbon price) as an indicator for a region's abatement potential.[4]

Figure 1 shows for each model the abatement potential indicator, i.e. the abatement undergone by each region in the tax scenario normalized to the maximum abatement

---

[4] For the common tax scenario, all models implement the same global emissions tax trajectory while climate change damages were disabled. The cumulative abatement is the absolute emissions reduction, summed over the models time horizon. We find the global abatement potential to be largest in case of STACO, followed by MICA, CWS, WITCH, and RICE in declining order. The range of costs is large: in WITCH and RICE only about two thirds of the abatement triggered in STACO is achieved. Still, abatement costs are in the same order of magnitude for all five models.

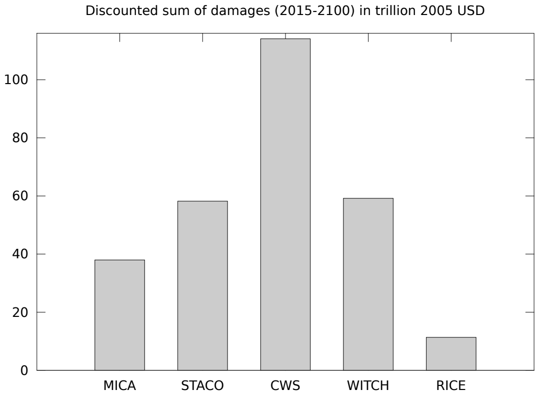

Figure 2: Aggregate total damages 2015–2100 (discounted) in the non-cooperative scenario (NC)

level over all regions. The indicator shows, for example, that China and India always rank high on abatement potential while for Japan the mitigation costs are perceived to be among the largest. We will use the information from this table extensively when discussing the main objectives of this note: incentives of specific regions to join a climate agreement and the characterization of transfers. In general, one can say that the models seem to be in good agreement about their assumptions on the costs of abatement.

Measuring climate change damages

Figure 2 compares aggregate discounted damages (total, not marginal) in the non-cooperative equilibrium across models. Of course, the underlying dynamics of the models are quite different as discussed above: different temperature profiles, economic growth and damage assumptions lead to the dissimilarities in the bars. The figure nevertheless highlights the fact that the damage calibration is low in RICE, resulting to a relatively low carbon price in the cooperative scenario (see Table 1).[5]

In order to compare the marginal damages from climate change between regions for each model, we take a slightly different approach than the total damages shown in Figure 2. Instead, we identify those regions as exhibiting higher marginal damages from climate change that strongly raise the carbon price when joining the grand coalition. The impact on the coalitional carbon price is a good indicator for the way that a joining region suffers from climate externalities because by definition these externalities are now internalized in the grand coalition's carbon price.

To normalize, we take the average of the difference in discounted carbon prices before and after the grand coalition is completed by a joining region. The resulting

---

[5] It should be noted that the STACO model considers only benefits from abatement and results do not rely on the level of damages. Thus, the STACO model is consistent with different damage estimates. The damage value shown here is based on the assumption that for low levels of temperature change, there are net gains from climate change.

measure, normalized to the maximal difference over all regions, is shown in Figure 1 next to the indicator for abatement potential discussed before.

One can see that the assumptions on which players would incur the highest marginal damages from climate change differ greatly between the models. Each model sees a different player to be ranking highest. These differences concerning the marginal damage assumptions will be a main driver of the results of the comparison. We will highlight this point in the separate analyses hereafter.

The characterization of regions in Figure 1 will subsequently be used for describing their behavior in the coalitions.

2.3 Incentives of regions

The different assumptions about model structure and regional characteristics outlined in the previous sections jointly determine the strategic behavior of the regions. This section looks at the incentive of regions to participate in a climate coalition.

The incentive to remain in a coalition is described by the incentive to stay, which is defined as the payoff received as a member of a given coalition minus the payoff outside the coalition (i.e. as a free-riding non-member). For the following discussion, we want to structure the driving forces that determine the incentive to stay for a given region in the following way:

1. First, the benefit of joining the coalition for this region, which is in turn influenced by
    (a) the extent of the externality in this region. When a player joins, this player's externality is henceforth internalized by all coalition members. Thus, the higher the marginal damages in the joining region, the more the coalition abates as a whole. This positive effect is enforced, as a high marginal damage region benefits all the more from additional abatement.
    (b) the reaction functions of the non-members. The free-riding non-members may raise their emissions in reaction to the reduced emissions of the coalition. Such leakage emissions offset the abatement of the coalition and therefore have a negative impact on the benefit of joining.
2. Second, the additional costs incurred by this region upon joining the coalitions. We distinguish the abatement costs of a coalition member and other opportunity costs
    (a) Abatement costs are a result of the distribution of emission reductions which are determined by efficiency in abatement (i.e. the lower the marginal costs, the more a region needs to do), and the overall ambition of the coalition, which depends on the collective marginal climate change damages of all coalition members.
    (b) Other opportunity costs emerge when regions are coupled through more channels than just the externality. For example, when carbon pricing affects the world demand for fossil resources, net exporters of resources will take

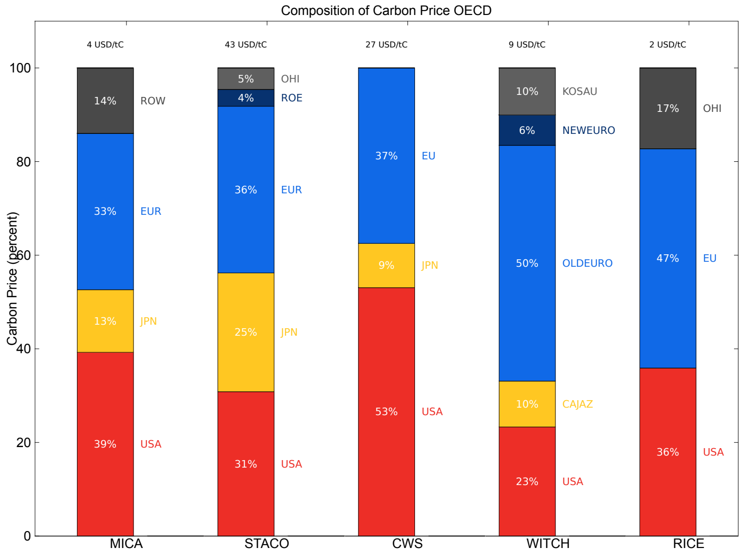

Figure 3: Carbon price in the OECD coalition (as average net present value). Percentages indicate how climate change damages in the member regions contribute to the overall carbon price.

the effect of participating in a coalition on their fossil resource revenues into account.

Some of these drivers of stability were covered above (i.e. in the discussion of models' assumption on costs and benefits) and are summarized in Figure 1. Before we take a look at the incentives, we discuss the distribution of damages, emissions abatement and emission leakage exemplary for the coalition of OECD countries. This coalition is one of the few larger coalitions that can be described in all models. Still, in some cases, we include "mixed" regions of OECD and non-OECD countries. The guiding criterion was whether more than half of the region's economic output was achieved by OECD countries.

Distribution of Damages

The extent to which a region benefits from abatement is measured via the carbon price (see discussion of Figure 1). Figure 3 reports the average net present value carbon price of the first century for each model within the OECD and how individual members contribute to it: when a player leaves this coalition, by what percentage does the price decrease?[6]

---

[6] Since marginal damages are not entirely flat in all models, this procedure is just an approximation of the decomposition of the cooperative carbon price but since the abatement of the OECD coalition is unambitious

The absolute level of the carbon price, given above each bar, shows differences in the ambition in emissions reduction that this coalition will have across models. While for STACO the OECD coalition consists of regions with high damages leading to a very high carbon price, the OECD coalition is much less ambitious e.g. in MICA, and hence free-riding on the OECD is much less attractive in this model.

Second, for the OECD-countries, models agree relatively well who contributes the most to damages: both the USA and Europe score high in every model, and Japan plays a minor role (except in STACO). From the benefit side, these two players therefore gain much from the abatement undergone in the coalition. However, one has additionally to consider their burden from abatement when joining the agreement to determine their net benefit, which we turn to next.

Distribution of Abatement

There are numerous ways to distribute the overall abatement among the members of a climate coalition. Criteria that guide this decision may, for example, be normative or pragmatic criteria, or incentive compatibility considerations. The default distribution of abatement in coalition models is driven by efficiency in the sense of maximizing the social welfare function of the coalition.[7] Efficiency determines first and foremost where how much of the emission reductions ought to be achieved to be efficient. This approach was taken for the models participating in this exercise.

For the OECD coalition, Figure 4 shows that its total abatement over the first centuries is quite different across models (model regions are specified in Tables 7-11). This is in part because not quite the same countries are covered. But since differences turn out to be large even when regions are identical (e.g. in case of single country regions), we conclude that a large part of the differences is due to different cost and benefit assumptions resulting in different carbon prices of the coalition as well the associated abatement potential of its members.

The distribution of abatement is diverse across models, e.g. the share of USA falls anywhere within the range 20-60 percent, for Japan within 1-18 percent, and for Europe within 10-30 percent. The large differences in the cost-benefit assumptions also translate to the efficient burden sharing schemes. All models agree, however, that the largest share of abatement ought to be achieved in the USA, often followed by Europe.[8]

Leakage Emissions

Leakage is the reaction of non-members to the coalition's behavior in terms of increased emissions. As such, we can only discuss leakage for incomplete cooperation, and this section therefore focuses on the OECD coalition.

The sensitivity of the reaction functions depends largely on model features that determine the ways in which non-members are affected by the coalition.

---

and therefore leakage is small, the error is negligible.

[7] Some models use a weighted sum in the social welfare function, see footnote , hence emission choices are Pareto-efficient.

[8] In MICA, the largest share falls onto the rest-of-the-world region, which includes several non-OECD countries and therefore plays a special role.

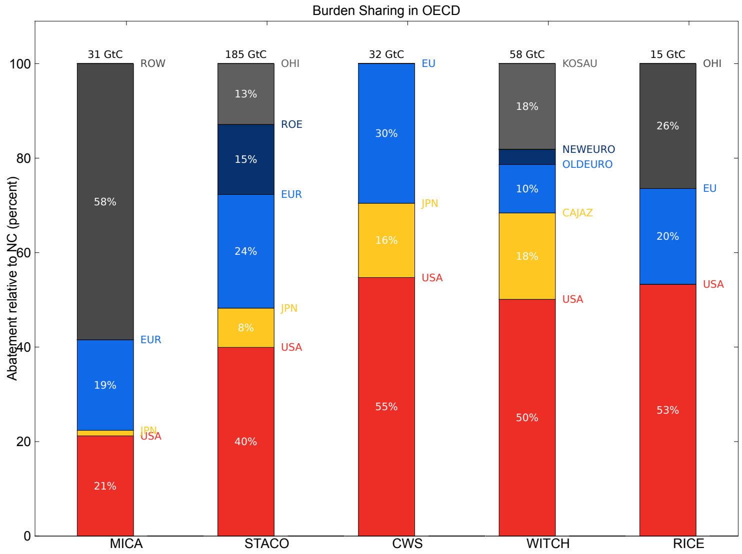

Figure 4: Allocation of emission reductions in the OECD coalition as percentages of overall emissions reduction (time horizon is a century)

On the one extreme, there is zero leakage in STACO. This is a consequence of assuming constant marginal damages, which implies that abatement is chosen independently in the regions. In all other models, the regions react to the abatement decisions of the others.

MICA, CWS, and RICE show only very moderate leakage: for these models, leakage emissions per region are less than one percent of the coalition's abatement.

Regions in WITCH show the strongest free-riding behavior in terms of leakage, with total leakage emissions of 16 percent of the OECD's abatement. In WITCH, the coalition affects non-members through an additional channel, namely energy markets (see Bosetti and De Cian, 2013, for details). A coalition drives down oil prices and free-riders increase their consumption especially of the carbon-intensive oil grades. In addition, climate change damages are especially high in WITCH.

Incentives

The interplay of all the drivers discussed above jointly determines the incentive to join or leave a given coalition. We consider the OECD coalition and the grand coalition in turn.

Figure 5 shows the incentive to stay inside the OECD coalition for all its members. If the incentive to stay was positive for all members, the coalition would be internally stable. Conversely, the figure shows which regions are responsible when the OECD

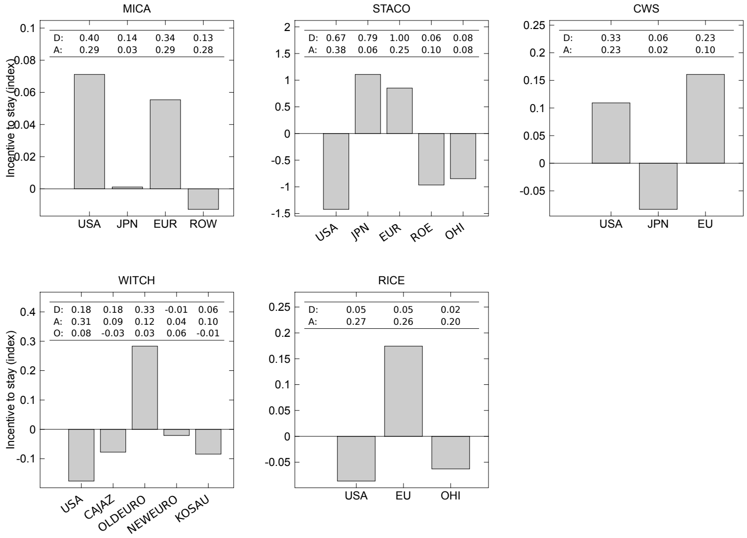

Figure 5: Incentive to stay in the OECD coalition, calculated as the difference of inside and outside payoff and scaled to the gap in global aggregated payoff between no cooperation (=0) and full cooperation (=100). The inset tables list the regional indicators for climate change damages and abatement potential, denoted ‘D’ and ‘A’, respectively.

coalition fails to be internally stable in any of the models. For easy reference, the indicators of abatement potential and climate change damages from Figure 1 are repeated in Figure 5.

While all models agree that the OECD coalition is not stable, they identify different culprits. For example, while the USA player would support the OECD coalition in MICA and CWS, the other three models, STACO, WITCH, and RICE, indicate a strong incentive to leave for this region. In the following, we will use the major players in the OECD coalition (USA, Europe, and Japan) as examples to explain how the drivers discussed above interact to form the incentive to join or leave the coalition.

The USA have a strong motivation to defect from the OECD coalition: their estimated abatement potential is the highest of all OECD players in all models but one where it is second to only one other region (to ROW in MICA). This implies that the USA would carry a large burden of the emission reductions in this coalition and would not join for this reason, which is what three of the models find (STACO, WITCH, and RICE). This is overcompensated by the large gains for the USA in this coalition in the remaining models (MICA and CWS), where the USA incur the highest marginal damages of all coalition members and thus has an incentive to remain in the coalition. In these two models it also helps that the ambition level of the OECD coalition is low, implying a low burden for its members as most high marginal damage regions are outside the OECD coalition. Conversely, none of the models estimates an incentive for the USA to stay in the grand coalition where joint climate change damages imply a much higher burden (not shown).

The incentives of Europe (in the sense of EU-15 countries due to model aggregation, i.e. regions EUR, EU, and OLDEURO) are a relatively simple case: in all models Europe is a typical happy coalition member, characterized by relatively high marginal damages and low abatement potential. Such players have much to gain from cooperation, but pay little as their burden share remains small. The models therefore agree that Europe would want to remain in the OECD coalition.

Japan is modeled as a single country region except for WITCH, where it is part of Canada/Japan/New Zealand (CAJAZ), and in RICE. The models unanimously see little abatement potential in Japan by itself. Thus Japan would carry only a small burden, which makes it better off in MICA and STACO. In CWS, the estimated marginal damages are also very low for Japan. Japan can therefore defect (and save on abatement costs) without substantially lowering the ambition level of the coalition, which turns out to be preferential. The larger aggregate region CAJAZ of WITCH incurs substantial abatement costs, tipping the balance towards defection as marginal damages are only average compared to the other OECD players.

Going from the OECD coalition to the grand coalition, positive net incentives to join become rare (not shown). Of course, this is a consequence of more ambitious emission reductions in this coalition, which places a larger burden on all members. The few exceptions are either of the high damage/low burden type discussed above (e.g. Japan and Europe in STACO, and SSA in WITCH) or very large players (ROW in CWS and RICE), which aggregate a lot of the world's damages and abatement potential due to their sheer size. In WITCH, an additional driver of incentives becomes important for very ambitious coalitions: here, in contrast to other models, net revenues from trade in oil are part of the region's income. Coalitions that strongly abate emissions consume substantially less oil leading to a price drop which increases the outsiders' consumption. Therefore oil-rich regions, while cutting their own oil consumption when joining the coalition, receive large revenues. Interestingly, extraction does not change very much and also the price differences are only minor. However, the pattern of consumption changes between the different grades of oil leading to increased exports of low carbon intensive ones. The top three regions that show the strongest increase in oil revenues (MENA, TE, and SSA) are the three players that prefer to stay in the grand coalition while this effect is negative for CAJAZ and KOSAU.

3 Transfers

3.1 Stable coalitions in IES, IS, and PIS

In this section, we consider the three stability concepts defined in the introduction: internal/external stability (IES), internal stability (IS), and potential internal stability (PIS).[9,10] Table 2 reports the number of stable coalitions for each stability concept, along with maximum coalition size, maximum abatement achieved and maximum welfare achieved. For the latter two, we use the closing the gap indicator to characterize the performance of coalitions, which relates global emission reductions ("environmental effectiveness") and welfare to the gap between the non-cooperative scenario –set to zero– and to full cooperation, set to unity (cf. Eyckmans and Finus, 2007).

Not surprisingly, coalitions that are IES without transfers are small and achieve little, which is in line with the existing literature. RICE and CWS are interesting exceptions: here the best IES coalitions achieve 33 and 77 percent of the global welfare gains of the GC, respectively. For CWS, this is caused by the very large region ROW which enables even a two player coalition to achieve much. The best performing coalition that does not include ROW achieves only a closing the gap indicator for welfare of 21 percent.

When we focus on IS, more coalitions are stable, and the performance improves. We want to highlight two interesting observations: (i) participation remains almost unchanged (with the only exception of an increase from 3 to 4 players in one model, MICA), and (ii) the performance improvement of IS coalitions over IES coalitions is substantial for some models, and negligible in others.

Introducing the transfers that are implicit in PIS has a strong impact: the number of stable coalitions increases by 1-2 orders of magnitude, and the corresponding improvement in the closing the gap indicators is also large. Three of the models find that the grand coalition is PIS, the other two "close the gap" about half. When superadditivity prevails (e.g. in STACO), the PIS coalition generating the highest global welfare is not only IS after receiving the implied transfers but also externally stable (ES) and hence IES (Eyckmans et al., 2010). In other words, the model comparison shows that transfers exist that make it possible to stabilize coalitions that close the welfare gap substantially. This is a considerably more optimistic message than the traditional con-

---

[9] The analysis could be extended to include blocking power (or core stability).

[10] For the models WITCH and RICE, the consumption discount rate was fixed at the one inside the coalition, which leaves the optimization procedure from Kornek et al. (2013) very constraint.

Table 2: Stable coalitions for internal/external stability and potential internal stability.

Model	Concept[a]	Number stable		Max. size	Max. abat.[b]	Max. welf.[c]
MICA	IES	1	(0.05%)	3	0.06	0.09
	IS	54	(2.64%)	4	0.17	0.24
	PIS	480	(23.45%)	6	0.31	0.47
STACO	IES	1	(0.02%)	2	0.03	0.03
	IS	23	(0.56%)	2	0.07	0.07
	PIS	2142	(52.31%)	9	0.59	0.68
CWS	IES	1	(1.59%)	2	0.67	0.77
	IS	5	(7.94%)	2	0.67	0.77
	PIS	61	(96.83%)	6	1.00	1.00
WITCH[d]	IES	1		2	0.03	0.05
	IS	1		2	0.03	0.05
	PIS[e]	5		4	0.17	0.38
RICE	IES	0	(0.00%)			
	IS	3	(4.76%)	2	0.03	0.06
	PIS[e]	7	(11.11%)	2	0.12	0.11

a Stability concepts are abbreviated IES (internal/external stability), IS (internal stability), and PIS (potential internal stability)

b The maximum global abatement achieved by a coalition is measured by the *closing the gap* indicator from 0=no-coopartion to 1=full-cooperation.

c Maximum global welfare is measured by the *closing the gap* indicator.

d In WITCH, only seven selected coalitions were analyzed.

e For the maximization procedure, the discount-rate was held fixed at the level of the PANE-solution.

Table 3: Permit allocation schemes. The permit allocation for a coalition S is determined as follows: each member i of S receives $q_{it} = \omega_{it} \cdot Q_t$ where $Q_t = \sum_{j \in S} e_{jt}$ are the available permits within S. Population, emissions and economic product are abbreviated pop, e, and y.

Scheme	Distribution key
Egalitarian	$\omega_{it} = \mathrm{pop}_{it} / \sum_{j \in S} \mathrm{pop}_{jt}$
Grandfathering	$\omega_{it} = e_{i,t_0} / \sum_{j \in S} e_{j,t_0}$
Quota Nash	$\omega_{it} = e^{NC}_{it} / \sum_{j \in S} e^{NC}_{jt}$
Quota BAU	$\omega_{it} = e^{ND}_{it} / \sum_{j \in S} e^{ND}_{jt}$
Historic responsibility	$\omega_{it} = \left(e^{ND}_{it}\right)^{-1} / \sum_{j \in S} \left(e^{ND}_{jt}\right)^{-1}$
Ability to pay	$\omega_{it} = \left(y_{it}/\mathrm{pop}_{it}\right)^{-1} / \sum_{j \in S} \left(y_{it}/\mathrm{pop}_{it}\right)^{-1}$
Ability to pollute	$\omega_{it} = \left(e_{it}/\mathrm{pop}_{it}\right)^{-1} / \sum_{j \in S} \left(e_{it}/\mathrm{pop}_{it}\right)^{-1}$
Energy efficiency	$\omega_{it} = \left(e_{it}/y_{it}\right)^{-1} / \sum_{j \in S} \left(e_{it}/y_{it}\right)^{-1}$

clusion derived from analytical models so far. In addition, our multi-model approach allows us to conclude that this claim is robust with respect to modeling approaches and parameterizations.

3.2 Transfers and stable coalitions

The previous section already introduced PIS, which implicitly relies on transfers that are designed to make coalitions internally stable wherever this is possible by an ex post reallocation of payoff within the coalition. We complement this "incentive driven" transfer scheme by a list of "conventional" transfers implicitly defined by burden sharing rules (Table 3). In contrast, these allocation rules are designed to be either equitable or pragmatic. The schemes in Table 3 are taken from Altamirano-Cabrera and Finus (2006). To evaluate how burden sharing affects stability of coalitions, we convert permit allocations to monetary transfers using the carbon price of the coalition. The monetary transfers are added either to the consumption streams or payoff (in case of STACO).[11]

How do conventional transfers affect stability?

In a first look at the implications of the conventional transfer schemes, we analyze how a selection of four schemes from Table 3 affects internal stability (IS). Table 4 reports the number of IS coalitions under these transfers and how this number changes relative to the scenario without transfers.

---

[11] In two models, there is no single carbon price within the coalition (WITCH and RICE) because maximization of social welfare for the coalition balances marginal value of emissions in terms of utility but not monetary units. This is different in MICA (where international trade balances marginal utility of consumption) and CWS (which uses a linear utility function). In WITCH and RICE, we use the social cost of carbon for the conversion instead (computed as the marginal utility of carbon inside coalition divided by the average per-capita consumption inside the coalition).

Table 4: How transfers affect coalition stability.

Transfer:	grandfathering		egalitarian		historic responsibility		ability to pay	
Model	coal[a]	Δcoal[b]	coal[a]	Δcoal[b]	coal[a]	Δcoal[b]	coal[a]	Δcoal[b]
MICA	6	-48	4	-50	5	-49	4	-50
STACO	3	-20	9	-14	0	-23	11	-12
CWS	16	11	0	-5	0	-5	3	-2
WITCH[c]	1	0	0	-1	0	-1	1	0
RICE	1	-2	0	-3	0	-3	0	-3

a Number of internally stable coalitions

b Number of internally stable coalitions relative to no-transfers

c Only selected coalition were analyzed in WITCH

The main conclusion is that transfer schemes that were designed without coalition stability in mind have an adverse effect on stability almost unanimously. This is evident from the decrease in the number of IS coalitions (cf. the almost exclusively negative numbers in the column Δcoal). An exception is "grandfathering" in CWS.

Why do conventional transfer schemes fail to induce stability? (A comparison with PIS-transfers)

As we have seen above, the conventional transfer schemes fail to induce much stability in any model. In Table 5 we compare two additional statistics of the transfer schemes to the PIS-transfers in order to track the precise reasons for this in more detail.

The first column of each model displays the share of PIS coalitions where the sign of transfers coincides with PIS transfers, i.e. players that need a positive transfer are receivers, and players with a surplus according to PIS have to pay. Hence by definition, PIS transfers reach the perfect score of 100 percent and other transfers score lower.

We find that most conventional transfer schemes stay well below 100 percent for this indicator, getting the sign of the required transfer right only about half of the time. There are positive exceptions for four of the models (MICA, STACO, CWS, RICE), however the models disagree which of the transfer scheme performs best. This is due to the fact that these models significantly differ in the damage assumptions of players. Therefore the directions of PIS-transfers differ greatly.

The second column displays the average flow of money between the regions across the ensemble. PIS transfers are roughly in the order of magnitude of the "quota bau" and "grandfathering". These two often also perform well for the direction indicator.

Thus, we have identified two problems of the conventional transfer schemes with respect to their negative effect on stability, namely the direction of the induced transfers and their magnitude. In view of these indicators, transfers that are based on business as usual (e.g. BAU Quota and Grandfathering) seem to do better than others. Especially when looking at the magnitude of all transfers listed in the table, there is great agreement that stability-enhancing PIS-transfers only demand relatively low flows of money.

How do PIS transfer schemes depend on the properties of coalition members?

In this section, we characterize the PIS transfers that induce stability. To this end, we consider all coalitions in the ensemble and relate properties of coalition members (as given by the indicators abatement potential and damages from the model characterization) to the frequency with which they receive a positive transfer (Table 6).

Table 6 shows the correlation coefficient of the percentage of coalitions in which a player receives positive transfers to abatement potential, damages, and the ratio of damages to abatement potential.[12]

We find that there is significant agreement among the models that the more damages a region incurs from climate change, the more likely it will be that this region has a surplus to share with other members, i.e. PIS transfers will be negative (column 2).

---

[12] The significance of the one-sided correlation is indicated with a "*" for the p=0.05 level and a "**" for the p=0.01 level.

Table 5: Permit transfer schemes compared to PIS-transfer schemes

	MICA		STACO		CWS		WITCH		RICE	
Transfer	Direction	Magnitude	Direction	Magnitude	Direction	Magnitude	Direction	Magnitude	Direction	Magnitude
PIS-transfer	100	0.019	100	2.662	100	8.334	100	0.239	100	0.002
egalitarian	48	0.232	55	1.529	33	16.574	23	0.383	21	0.044
historic responsibility	71	0.544	56	3.946	71	52.702	54	1.511	64	0.144
grandfathering	50	0.166	40	0.861	72	10.941	54	0.279	50	0.008
quota bau	63	0.025	53	0.194	40	3.212	46	0.096	79	0.001
ability to pay	67	0.409	74	2.755	55	23.075	8	0.542	50	0.103

Table 6: Characterization of PIS transfers with properties of players

	Percentage of positive transfer received		
	Abatement Potential	Damages	Damages/ Abatement
MICA	−0.502	−0.802∗∗	−0.593
STACO	0.117	−0.918∗∗	−0.771∗∗
CWS	−0.857∗	−0.914∗	−0.114
WITCH	0.499	0.273	0.078
RICE	−0.014	−0.186	−0.357

The other two indicators do not give rise to significant correlation although the numbers still indicate the same sign of the relationship in most cases.

4 Summary and Conclusions

In this study, we compared five structurally different models and make the modeling assumptions on the costs and benefits of climate change mitigation comparable through two indicators measuring (a) the regional abatement potential and (b) regional exposure to climate change damages. While the models' estimates for abatement potentials are in agreement for key world regions, we find large differences in the climate change damage estimates that the models prescribe for certain regions. To a large extent, the differences reflect the variations in the literature sources that the model parametrizations are based on and therefore reflect the uncertainty about costs and benefits of climate change mitigation in the literature (cf. Metz et al., 2007).

It is therefore not surprising that the models differ in their assessment whether certain coalitions are stable, and whether certain world regions or nations have an incentive to be members of a given coalition. (A notable exception is the assessment of the EU, for which the models unanimously attest an incentive to support a coalition of OECD countries.) However, when we abstract from the identity of the players and instead consider their cost-benefit characteristics in terms of the two indicators suggested in this study, the models are remarkably consistent in their predictions. We find that the indicators of a region's abatement potential and its exposure to climate change damages capture much of its incentives and allow us to understand the regions membership preference for or against membership in a coalition. When either abatement potential is low (implying a steep marginal abatement cost function) or marginal climate change damages are high in a region, the likelihood for a positive incentive to stay is higher.

In absence of transfers, all models agree that stable coalitions tend to be small and achieve little, due to a lack of internal stability of larger, more ambitious coalitions. This is in accordance with the theoretical literature and therefore not surprising.

Transfers designed to minimize free riding incentives as much as possible achieve much more: the models find that PIS coalitions are substantially larger and achieve about half or more of what full cooperation would achieve both in welfare and GHG abatement terms.

In contrast, conventional transfers do not improve cooperation, they often even undermine existing stable coalitions. The reason is, of course, that conventional transfers are not reflecting incentives: among other things they frequently induce transfers that are (a) too large in their magnitude and (b) transfer wealth in the wrong direction, i.e. regions that need transfers to be convinced to stay in a coalition are made to pay regions that have no incentive to defect from the coalition.

Finally, we examine how the properties of coalition members affect the PIS transfers necessary to stabilize the coalition. We find that players with high damages tend to benefit enough from cooperation such that they can share some of these gains.

The last two findings seem to be robust across the different specifications of the models concerning how an incentive-based transfer scheme should be designed and that its implementation will increase cooperation greatly. On the one hand, its magnitude is comparable to allocation schemes based on historic emissions; the financial flows demanded are therefore comparably small. On the other hand, players with high damages from climate change are eligible for compensating those players with high abatement potential that provide the necessary mitigation.

Acknowledgements

We would like to thank all participants of the two workshops that led to this model comparison (Potsdam 8-9th of February, 2012, and Venice, 24-25th of January, 2013). We also benefited from feedback to presentations of the manuscript at Grantham Institute, LSE, at IGIER, Bocconi, and at the EAERE2013 conference, which is gratefully acknowledged. Ingram Jackard contributed to the review of numerical coalition models an preparation of the initial workshop.

References

Altamirano-Cabrera, J. and Finus, M. (2006). Permit trading and stability of international climate agreements. *Journal of Applied Economics*, 9(1):19–47.

Barrett, S. (1994). Self-enforcing international environmental agreements. *Oxford Economic Papers*, 46:878–894.

Barrett, S. (2001). International cooperation for sale. *European Economic Review*, 45(10):1835–1850.

Bosetti, V., Carraro, C., Galeotti, M., Massetti, E., and Tavoni, M. (2006). WITCH: A World Induced Technical Change Hybrid model. *The Energy Journal*, Special Issue Hybrid Modelling of Energy Environment Policies: Reconciling Bottom-up and Top-down:13–38.

Bosetti, V. and De Cian, E. (2013). A good opening: the key to make the most of unilateral climate action. *Environmental and Resource Economics*, pages 1–22.

Bréchet, T., Gerard, F., and Tulkens, H. (2011). Efficiency vs. stability in climate coalitions: a conceptual and computational appraisal. *Energy Journal*, 32(1):49.

Carraro, C., Eyckmans, J., and Finus, M. (2006). Optimal transfers and participation decisions in international environmental agreements. *Review of International Organizations*, 1(4):379–96.

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

d'Aspremont, C. and Gabszewicz, J. J. (1986). *New Developments in the analysis of market structures*, chapter On the stability of collusion, pages 243–64. Macmillan, New York.

Eyckmans, J. and Finus, M. (2006). Coalition formation in a global warming game: how the design of protocols affects the success of environmental treaty-making. *Natural Resource Modeling*, 19(3):323–358.

Eyckmans, J. and Finus, M. (2007). Measures to enhance the success of global climate treaties. *International Environmental Agreements: Politics, Law and Economics*, 7(1):73–97.

Eyckmans, J., Finus, M., and Mallozzy, L. (2010). A new class of welfare maximizing stable sharing rules for partition function games with externalities. Working Paper. Department of Economics, University of Bath.

Eyckmans, J. and Tulkens, H. (2003). Simulating coalitionally stable burden sharing agreements for the climate change problem. *Resource and Energy Economics*, 25:299–327.

Finus, M. (2008). Game theoretic research on the design of international environmental agreements: Insights, critical remarks, and future challenges. *International Review of Environmental and Resource Economics*, 2:29–67.

Fuentes-Albero, C. and Rubio, S. J. (2010). Can international environmental cooperation be bought? *European Journal of Operational Research*, 202(1):255–264.

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

Kornek, U., Lessmann, K., and Tulkens, H. (2013). NTU implementations of PIS and core stability. Working Paper available upon request.

Lessmann, K. and Edenhofer, O. (2011). Research cooperation and international standards in a model of coalition stability. *Resource and Energy Economics*, 33(1):36–54.

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

McGinty, M. (2007). International environmental agreements among asymmetric nations. *Oxford Economic Papers*, 59(1):45–62.

Metz, B., Davidson, O., Bosch, P., Dave, R., and Meyer, L., editors (2007). *Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Nagashima, M., Dellink, R., Van Ierland, E., and Weikard, H. (2009). Stability of international climate coalitions—a comparison of transfer schemes. *Ecological Economics*, 68(5):1476–1487.

Nagashima, M., Weikard, H., de Bruin, K., and Dellink, R. (2011). International climate agreements under induced technological change. *Metroeconomica*, 62(4):612–634.

Nordhaus, W. D. and Yang, Z. (1996). A regional dynamic general-equilibrium model of alternative climate-change strategies. *The American Economic Review*, 86(4):741–765.

Ramsey, F. (1928). A mathematical theory of saving. *Economic Journal*, 38:543–559.

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

Weikard, H., Finus, M., and Altamirano-Cabrera, J. (2006). The impact of surplus sharing on the stability of international climate agreements. *Oxford Economic Papers*, 58(2):209–232.

Yang, Z. (2008). *Strategic Bargaining and Cooperation in Greenhouse Gas Mitigations: An Integrated Assessment Modeling Approach*. MIT Press.

A Model regions

Table 7: Regions as defined in MICA and corresponding world regions

Region	Countries
AFR	Sub-Saharan Africa without South Africa
CHN	China
EUR	EU-27
IND	India
JPN	Japan
LAM	All American countries except Canada and the United States
MEA	North Africa, Middle Eastern and Arab Gulf countries, resource exporting countries within the former Soviet Union, and Pakistan
OAS	South East Asia, North Korea, South Korea, Mongolia, Nepal, Afghanistan
ROW	Australia, Canada, New Zealand, South Africa and non-EU27 European states except Russia
RUS	Russia
USA	United States of America

Table 8: Regions as defined in STACO and corresponding world regions

Region	Countries
BRA	Brazil
CHN	China
EUR	Europen Union and European Free Trade Association
HIA	High-income Asia, including South Korea and Indonesia
IND	India
JPN	Japan
MES	Middle Eastern countries
OHI	Other high-income countries, including Canada, Australia, New Zealand
ROE	Rest of Europe
ROW	Rest of the world
RUS	Russia
USA	United States of America

Table 9: Regions as defined in CWS and corresponding world regions

Region	Countries
CHN	China
EU	EU-15
FSU	Former Soviet Union
JPN	Japan
ROW	Rest of the world
USA	United States of America

Table 10: Regions as defined in WITCH and corresponding world regions

Region	Countries
CAJAZ	Canada, Japan, New Zealand
CHINA	China, including Taiwan
EASIA	East Asia without China, Japan, Korea
INDIA	India
KOSAU	Korea, South Africa, Australia
LACA	Latin America and Caribbean
MENA	Middle East and Northern Africa
NEWEURO	Recent accessions to the European Union
OLDEURO	EU-15
SASIA	South Asia
SSA	Sub-Saharan Africa
TE	Non-EU East-Europe and Central Asia
USA	United States of America

Table 11: Regions as defined in RICE and corresponding world regions

Region	Countries
CHN	China
EEC	Eastern European countries and the former Soviet Union
EU	European Union
OHI	Other high-income countries
ROW	Rest of the world
USA	United States of America