Cap, Insure & Compensate: Domestic Policies and the Ratification of International Environmental Agreements

Jobst Heitzig

Potsdam Institute for Climate Impact Research, PO Box 601203, 14412 Potsdam, Germany, heitzig@pik-potsdam.de

Ottmar Edenhofer

Potsdam Institute for Climate Impact Research, PO Box 601203, 14412 Potsdam, Germany;
Mercator Research Institute on Global Commons and Climate Change,
Torgauer Straße 12–15, 10829 Berlin, Germany;
Technical University Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany;
Intergoverenmental Panel on Climate Change, Working Group III

EAERE 2013 accepted conference paper, this version 26 April 2013

Using a Putnam's two-level game approach, we discuss the influence of domestic players that can veto the ratification of international environmental agreements on the breadth and depth of such treaties. We first show that in a symmetric Barrett-type payoff model, veto-players can restrict the bargaining set so that either all or only non-grand coalitions become "modest" endogenously, making the grand coalition stable and leading to either "broad but shallow" or "broad and deep" treaties, depending on the exact payoff function of the veto players. We then discuss the possibility of compensating veto-players via domestic policies and the involved timing and commitment problems, and present the novel policy scheme of "cap, insure & compensate" to overcome these problems. In this scheme, the government negotiates an international emissions cap, insures households against climate-related damages and uses the premium to compensate the domestic veto-player for its abatement costs, thus ensuring the ratification of the treaty. We finally analyze the performance of this scheme in a simple two-period model.

Key words: two-level game, veto player, stable coalition, modesty, side-payment, timing

1 Introduction

The 2009 Copenhagen Accord has shown that there is a broad international agreement that global warming should not be allowed to exceed two degrees Celsius above pre-industrial levels, and that this requires vast greenhouse gas emissions reductions. Still, there is no effective global climate treaty yet, and in view of the 2011 Durban and 2012 Doha results, it is unclear whether such a treaty will emerge from the current negotiation process. We argue here that this might be strongly related to the effects that domestic stakeholders of climate policy can have on the strategic interaction between governments on the international level, and that these interactions have not been analysed enough in the existing formal literature.

In the game-theoretically inclined literature on international climate policy, the currently observed low level of international cooperation on climate issues is typically explained along the following line: Mitigating unwanted climate change requires controlling the concentration of greenhouse gases (GHG, mostly CO_2) in the atmosphere by either reducing GHG emissions or increasing the amount of GHG taken out of the atmosphere (e.g., avoiding deforestation or applying technologies such as carbon capture and storage). Lowered GHG concentrations are a public good since GHG mix fast in the atmosphere, hence the provider of the good gets only a small part of the benefits from it. Because of this positive externality, the individually optimal amount of contributions (e.g., emissions reductions or afforestation) to the public good by each relevant agent (e.g., a household or firm) that would maximize that agent's benefit-cost balance is lower than the "socially" optimal amount that would maximize global welfare as measured by some suitable welfare index, a social dilemma very similar to the famous Prisoners' Dilemma (PD). If most relevant agents act in a mainly self-interested way, global welfare and hence most agents' individual welfare will hence be lower than possible. Although in smaller communities such social dilemmas are sometimes attenuated by socio-cultural mechanisms such as moral codes and social pressure that give agents incentives that align their preferences with those of the community, externalities on a national scale are usually dealt with by a domestic regulator (e.g., government) using suitable *policy instruments.* On an international scale, the only viable way seems to be via *treaties* between several sovereign countries. In addition to the obvious *bargaining* problem of selecting a *cost sharing* rule, negotiating a treaty involves at least two additional problems: Without an external enforcement mechanism (e.g., a world government) a treaty's signatories still have an incentive for *non-compliance.* And if more than two countries are relevant, certain countries (e.g., the prospective permit buyers in a cap-and-trade regime) also have an incentive not to sign the treaty in the first place, hoping to benefit from the treaty without sharing the costs, or (for a small number of countries) even believing to benefit from climate change, both of which leads to an additional *non-participation* problem. Possible counter-measures such as the punishment of non-compliance (e.g., via trade sanctions) or of non-participation (e.g., by not ratifying the treaty) are typically harmful for the punishers themselves if they were actually carried out, so if signatories cannot convincingly *commit* to carrying out the punishment, that threat lacks credibility and is void. Accordingly, the prevalent game-theoretical model applied in the literature on International Environmental Agreements (IEAs, see Finus, 2008, for a review) assumes that each country can be represented by one player aiming to maximize that country's welfare and predicts that only small coalitions consisting of two or three countries will form and achieve only little on the global scale.

Upon closer inspection, however, most of the above-mentioned problems might be overcome: Both the existence of so-called tipping elements in the Earth system and the nearing of the corresponding *tipping points* at which irreversible changes can occur rapidly and the argument to rather apply a precautionary principle than an expected utility approach in view of the large uncertainties seem to make the game's strategic structure more similar to a Chicken Game, a Waiting Game, or a Weakest Link Game, in all of which cooperation can be achieved easier than in a PD. The bargaining problem might be alleviated by the identification of *focal points* (e.g., a uniform global carbon price or tax, equal per capita permits or welfare gains, etc.). The compliance problem might be solved using certain *self-enforcing* reciprocal strategies (e.g., Froyn and Hovi, 2008; Heitzig et al., 2011). Finally, the participation problem might be solved by using a *bottom-up* process of possibly hierarchical coalition formation (e.g., Victor, 2008; Heitzig, 2012).

An explanation of the currently observed low level of international cooperation and the slow progress towards a higher one thus requires additional ingredients, two of which we will explore in this paper: the *two-level* character of the game, involving strategic interactions between the international level of treaty negotiations and the domestic level of treaty ratification and domestic cost sharing, and its interrelation to the large *time lag* between costs and benefits of emissions reductions and the corresponding risks. In the following Section, we will provide such an explanation in terms of Putnam's theory of two-level games. As a positive contribution, we will then propose the policy instrument of "Cap, Insure & Compensate" (CIC) as a possible way out in Section

3. In Section 4, we compare CIC with other policy options in a simple two-period model with three domestic players, before concluding in Section 5.

2 Can the two-level approach explain failed cooperation?

While the literature on International Environmental Agreements usually treats each country or closely integrated group of countries (such as the EU) as a single player, real-world treaties are typically first negotiated by representatives of *governments* and then need to be *ratified* in each signatory country. Putnam's (1988) theory of *two-level games* holds that in many countries, both the government and the relevant agents in the ratification process (called the *constituents*) might have different objectives than to maximize the long-term welfare of their country, making the one-player-per-country assumption questionable. Rather, the strategic interaction between the government's behaviour on the *international level* and the behaviour of government and constituents on the *domestic level* must be analysed explicitly. Yet somewhat surprisingly, this concept has not been applied much in game-theoretic models of international climate policy so far.

At first glance, a straightforward explanation of the currently observed low level of international cooperation following the two-level approach could look like this: In at least some of the major emitting countries or world regions (e.g., in the US and Europe), influencial *veto-players* (e.g., fossil energy industry) have the effective power to *block* the ratification of a treaty. A treaty that has a good chance of getting ratified must thus be attractive not only for the negotiating governments but also for all these veto-players. This narrows the set of possible treaty specifications that would get ratified. During negotiations, governments have thus only a *restricted bargaining set* that might contain only treaty specifications that achieve much less than possible from a long-term welfare point of view. Depending on the incentive structure of the veto-players, the bargaining set might even be empty so that no ratifiable treaty can be negotiated at all. But even if it is non-empty but contains only little-achieving treaties, governments might choose to postpone agreement, either because they hope that the incentive structure for the veto-players might change in the future, or because their own incentive structure (e.g., their prospects of getting re-elected) makes it unattractive to sign a little-achieving treaty.

2.1 International environmental agreements as two-level games

The following simplistic model is designed to illustrate the effect of a restricted bargaining set in climate negotiations in a similar way as in Putnam (1988). Assume that two countries, C and U, negotiate a treaty concerning GHG emissions reductions. As in the standard IEA model from Barrett (1994), we assume that each country, i, can reduce their emissions by an arbitrary non-negative amount of abatement, q_i, thus incurring abatement costs which are quadratic in q_i. Both countries' climate-related damages are then reduced in proportion to joint abatement, $Q = q_C + q_U$, showing the positive externality of abatement. We assume that on the international level, each country is represented by a government (denoted by the same letter) aiming to optimize the benefit-cost balance, so that their payoffs can be written as

$$\begin{aligned} \pi_C &= Q - q_C^2/2, \\ \pi_U &= Q - q_U^2/2. \end{aligned} \tag{1}$$

However, in contrast to the standard IEA model, we assume two additional players, U0 and U1, representing two veto-players in country U, whose payoffs are given by

$$\begin{aligned} \pi_{U0} &= (1-\alpha)Q - \alpha\, q_U^2/2, \\ \pi_{U1} &= \alpha Q - (1-\alpha) q_U^2/2 \end{aligned} \tag{2}$$

for some parameter $0 < \alpha < ½$. Players U0 and U1 might be interpreted as two factions in U's parliament, each representing about half of the population thus aiming to maximize the corresponding share of U's payoffs, where U0 is more concerned about climate change (and thus has a larger share of the benefits), while U1 is influenced by the interests of conventional energy producers (and thus has a larger cost share).

The game has now two stages, the first of which represents *negotiations* between government players C and U and results either in a breakdown of negotiations or in signing a treaty that specifies a pair (q^c_C, q^c_U) of agreed "cooperative" abatement levels. The second stage represents *ratification* in which each veto-player, U0 and U1, either accepts or rejects the treaty (if any). If a treaty was signed and both accept it, it is ratified and the agreed levels (q^c_C, q^c_U) are realized, otherwise no treaty comes into effect and instead some "non-cooperative" levels (q^n_C, q^n_U) are realized. For simplicity, let us assume the latter levels correspond to the unique Nash equilibrium of the non-cooperative game between governments C and U, resulting in

$$(q^n_C, q^n_U) = (1,1). \tag{3}$$

We solve the game using backward induction, showing that in the ratification stage, veto-player U*j* will accept the treaty if and only if $\pi_{Uj}(q^c_C, q^c_U) \geq \pi_{Uj}(1,1)$, assuming a preference for ratification in case of equal payoffs. Note that because $\alpha < ½ < 1 - \alpha$, U0 will agree whenever U1 agrees. Thus stage one is equivalent to a classical *bargaining problem* between governments C and U in which the *feasibility set* (here defined as the set of treaties that would get ratified if signed, also called the win set) is defined by the inequality

$$\alpha(q_C + q_U) - (1 - \alpha)q_U^2/2 = \pi_{U1}(q_C, q_U) \;\; \geq \;\; \pi_{U1}(1,1) = (1 + \alpha)/2 \tag{4}$$

and in which the *disagreement point* (also called the "no-agreement outcome") is given by $(q_C, q_U) = (1,1)$. Fig. 1 shows the relevant curves for this bargaining problem in both *issue space* and in *payoff space* (also called "value space"), with the disagreement point in the bottom left corner, the corresponding indifference curves, and the curve of Pareto-efficient points, much similar to Mayer (1992, Figs. 5 and 6). It is easy to see that in issue space the *bargaining set* (here defined as the set of treaties that would get signed and ratified) is simply delimited by the indifference curves of U1 and C, so the question remains which point (q^c_C, q^c_U) fulfilling (4) and the condition

$$q_C + q_U - q_C^2/2 = \pi_C(q_C, q_U) \;\; \geq \;\; \pi_C(1,1) = 3/2 \tag{5}$$

governments will agree on. A commonly assumed bargaining solution concept is the *Nash bargaining solution* (not to be confused with Nash equilibrium) which maximizes the product of gains,

$$(q_C + q_U - q_C^2/2 - 3/2)\,(q_C + q_U - q_U^2/2 - 3/2) \;\; =! \;\; \max \tag{6}$$

inside the bargaining set, i.e., subject to (4) and (5). If the bargaining set contains Pareto-efficient points, the Nash bargaining solution will be Pareto-efficient. For moderately small α (yellow line in Fig. 1), it will thus lie on the top left segment of the Pareto-frontier in payoff space (above the intersection with U1's yellow indifference curve), implying that U's resulting payoff is larger than in the case without veto-players while C's payoff is not much larger than their disagreement payoffs. This confirms the experience that a somewhat restricted bargaining set can be desirable for a negotiator. If α gets even smaller (black and blue lines in Fig. 1), however, the bargaining set no longer contains any Pareto-efficient point and the Nash bargaining solution leads to smaller payoffs for U, also approaching the low disagreement payoffs as α tends to zero. In other words, depending on how asymmetric the veto-players' payoffs are, the bargaining set might only contain little-achieving treaties that may not come into effect for the reasons explained above.

2.2 Endogenous modesty and the stability of the grand coalition

Although, as shown above, the joint payoff of the negotiating countries usually decreases due to the restricted bargaining set, this detrimental effect of domestic veto players may be overcompensated by the more subtle effect they have on the question which countries will enter the negotiating coalition in the first place. We will show here that the "endogenous modesty" implied by the coalition's restricted bargaining set may reduce the incentive to free-ride as a non-member on the efforts of the coalition, so that larger coalitions will be stable than without the presence of veto players. Let us still stick to the open membership single coalition model of Barrett (1994), now with N countries C_i who choose abatement levels q_i and get payoffs

$$\pi_i = Q - q_i^2/2, \quad \text{where} \quad Q = q_1 + \cdots + q_N .$$

In that model, each C_i simultaneously decided whether to enter negotiations or not, and then all who entered negotiations form a coalition that chooses their q_i to maximize their joint payoff, while all others choose their q_i to maximize their individual payoff. Without veto-players, the classical backward induction analysis of this two-stage game quickly shows that in any subgame perfect equilibrium at most three countries form the coalition, regardless of N. The resulting global payoff is then $N^2 + 11N/2 - 12$, which is not much more than the completely non-cooperative Nash equilibrium global payoff of $N^2 - N/2$, as compared to the fully cooperative global payoff of $N^3/2$ which a grand coalition of all N countries would achieve by maximizing global payoff. This low level of participation is due to the ambitious abatement targets of larger coalitions which imply that the members' payoff is smaller than the non-members' payoff of a slightly smaller coalition, so that there is an incentive to leave the large coalition to free-ride on the efforts of the remaining smaller coalition.

Finus and Maus (2008) argue that if larger coalitions were to agree on a more "modest" target instead of the one that maximizes their joint payoff, this free-riding incentive might be so reduced that even the grand coalition of all countries becomes stable. Their proposal, however, leaves the question why an announcement before the participation stage to later on agree on sub-optimal targets should be credible. After all, once the negotiation stage is reached, it is still optimal for the coalition to maximize their joint welfare, so a country deciding not to participate may well expect the participants to still go for the optimal target, no matter what they announced they will do before the participation stage. Nicely, we will see here that the presence of domestic veto players may automatically imply an abatement target modest enough to stabilize the grand coalition. In other words, veto players can be seen as a commitment device.

Broad but shallow agreements for linear veto-player benefits. Assume that in each country C_i there is a veto player V_i who gets payoff

$$\varphi_i = \alpha Q - (1 - \alpha) q_i^2/2,$$

where $0 < \alpha < 1/3$ is a common asymmetry parameter. Consider the three-stage game where first each C_i decides simultaneously whether to participate in negotiations, then those C_i who participate negotiate an agreement over their q_i , and then each V_i whose C_i participates will simultaneously decide whether to ratify the agreement. If all ratify the agreement, the negotiated q_i are implemented, otherwise the participants choose q_i to maximize their individual payoff. Non-participants always choose q_i to maximize their individual payoff. Because of the symmetry of the payoff functions, we can assume that the negotiated agreement assigns the same q_i to all members.

We now solve the game backwards, assuming that a coalition of k countries has formed, and notice that V_i will ratify the agreement iff it makes φ_i, which evaluates as $\alpha[kq_i + (N - k)] - (1 - \alpha)q_i^2/2$, at least as large as in the non-cooperative case, where it takes the value $\varphi_i^0 = \alpha N - (1 - \alpha)/2$. Solving this quadratic inequality for q_i leads to the constraint

$$q_i \leq [\alpha(2k + 1) - 1]/(1 - \alpha) \quad \text{if} \quad k \geq 1/\alpha - 1, \quad \text{and} \quad q_i \leq 1 \quad \text{otherwise.}$$

Since $(\alpha(2k + 1) - 1)/(1 - \alpha) \leq k$, the participants will thus agree on that target if $k \geq 1/\alpha - 1$ and on $q_i = 1$ otherwise, and the agreement will be ratified. Participants then get a payoff of

$$\pi_i^{\text{mem. in } k} = k[\alpha(2k + 1) - 1]/(1 - \alpha) + N - k - [\alpha(2k + 1) - 1]^2/2(1 - \alpha)^2.$$

If a single participant had not decided to participate, it would have gotten instead the non-participant payoff of a coalition of size $k - 1$, which is

$$\pi_i^{\text{non-mem. in } k-1} = (k - 1)[\alpha(2k - 1) - 1]/(1 - \alpha) + N - (k - 1) - ½.$$

The size of the largest stable coalition can now be inferred by equating these two values and solving for k, treating k formally as a continuous variable that is truncated to an integer afterwards. This gives

$$k^* = \lfloor (3 - 3\alpha + [5 - 14\alpha + 9\alpha^2]^{½})/2\alpha \rfloor$$
$$\approx (3 + 5^{½})/2\alpha \approx 2.618/\alpha \quad \text{for small } \alpha$$

where $\lfloor \ldots \rfloor$ denotes truncation to an integer. Each member of this coalition abates

$$q_i^* \approx 2 + (5 - 9\alpha^2)^{½}/(1 - \alpha)^{½},$$

which goes to $2 + 5^{½} \approx 4.236$ for small α. In other words, if the veto players' share of the abatement benefits is small enough and their share of the costs is large enough, arbitrarily large coalitions get stable since the best agreement they can get ratified has only modest per-member targets that are approximately independent of the coalition's size. The coalitional payoff is however only about 4.236 times the non-cooperative one, instead of about k times as in the case of unconstraint joint payoff maximization. Note, however, that also the payoff a thus stabilized grand coalition achieves is only about 4.236 times (instead of N times) the non-cooperative payoff. So with the above payoff structure, one gets "broad but shallow" agreements as in Downs et al. (1998).

Broad and deep agreements for quadratic veto-player benefits. If the shape of the veto player's payoff function differs more from that of the country's, one can also get "broad and deep" agreements. E.g., assume that V_i pays C_i's complete abatement costs but gets only a percentage of C_i's benefits that grows linearly with global abatement from 0% at $Q = 0$ to 50% at the global optimum $Q = N^2$. In other words, assume that

$$\varphi_i = Q^2/2N^2 - q_i^2/2 = (q_{\text{average}}^2 - q_i^2)/2.$$

Then one can show that a coalition of k players who maximize their joint payoff under the constraint that their veto players get at least what they get in the non-cooperative case will choose $q_i = 1$ if $k < N$ and $q_i = N$ if $k = N$. A non-participant to a coalition of size $k < N$ would then get a payoff of $\pi_i = N - ½$, which is strictly smaller than $N^2/2$, the payoff it will get as a member of the grand coalition. Hence no country (and even no group of countries) has an incentive to leave the grand coalition, which is thus stable and achieves the global optimum.

With still other functional forms of φ_i, it can also occur that *all* possible coalitions are stable and the grand coalition still achieves the global optimum, in which case an efficient global treaty can easily emerge through stepwise accession by more and more countries. These stylized examples of payoff functions show that veto players can have quite different effects on both the size and achievements of stable coalitions, depending on how their preferences relate to their governments' preferences.

2.3 Can veto-players be compensated via domestic policies?

For some points in issue space (e.g., for some allocation of emissions caps), the government may be able to link the ratification of the international treaty to the implementation of certain *domestic policies* that would make the combination of both policies profitable for both the veto-player and the government. This would widen the bargaining set and thereby alleviate the ratification problem. In particular, if the domestic policy effectively implements a *side-payment* from the domestic "winners" to the "losers" of the treaty and if the government is indifferent to this side-payment (e.g., if the government maximizes welfare and the side-payment is welfare-neutral), then the resulting bargaining set might indeed be identical to the one that would arise without veto-players. In that case, the original one-player-per-country assumption would be justified ex ante.

In international climate policy, negotiated treaties will likely contain the implementation of either a domestic *carbon tax* or a domestic *emissions cap,* hence the government might use either the tax revenues (via *monetary compensation*) or the domestic emissions permit allocation (via *grandfathering*) to implement a side-payment. E.g., in our first illustrating example above, the government in U could thus implement a side-payment of size $(½ - \alpha)(Q + q_U^2/2)$ from U0 to U1 so that the payoffs of U0 and U1 get redistributed in a way that makes both equal to $Q/2 - q_U^2/4$. Since then both veto-players' payoffs are proportional to government's payoffs, all three players have exactly the same incentive structure, i.e., their preferences are aligned and they can be considered equivalent to one single player representing country U.

Any two-level game explanation of international climate policy thus needs to explain why the possibility of side-payments has not yet resolved the ratification problem. For one thing, compensating fossil energy producers either directly via monetary transfers or through generous permit allocations (grandfathering) might not be acceptable for political reasons. Second, the government might not be indifferent to the side-payment because of its distributional or, again, political side-effects. Also, the asymmetry of the veto-players with respect to costs and benefits of abatement and the fact that benefits accrue later than costs may pose *timing* and/or *commitment* problems for the implementation of side-payments. E.g., in the carbon taxation case, the compensation might only be possible in the form of a stream of transfers paid out of the stream of tax revenues, and the government might not be able to make a credible commitment to actually conduct these payments over a long period of time when it could also use the revenues differently. Finally, when costs accrue immediately and benefits accrue much later, the government might need to *borrow* the money necessary to implement the side-payment for a long time. If it needs to borrow huge amounts from the international financial market which can be expected to apply risk-based pricing, then the loan's interest rate includes a premium for taking the risk of country default. If the maturity and/or the uncertainty about the benefits is too large and thus the default risk is estimated as high, or the country has a bad rating for other reasons, the government might not be able to borrow enough money or the risk-based interest rate exceeds the governments' time discounting rate by so much that borrowing appears infeasible, too.

Because it is so important for a two-level analysis of climate policy whether side-payments are possible, we will focus on this question for the rest of this paper. In particular, we will discuss a novel policy scheme that is designed to overcome the timing, commitment, and risk-related problems of side-payments, thus providing a possible way out of the ratification problem.

3 Cap, Insure & Compensate

There is a policy scheme that would allow a government to implement an early side-payment to a veto-player without having to commit to later transfers and without the need to pay increased interest rates to the financial market. At the same time, it will reduce households' climate-related risk. The idea is to combine a domestic policy instrument leading to emissions reductions (e.g., an emissions cap) with an insurance of households against climate-change related events (e.g., storms, floods, droughts, crop failure, etc.) and an early compensation of veto-players financed out of the collected insurance premiums. More precisely, the "Cap, Insure & Compensate" (CIC) scheme operates as follows:

1. The government offers households insurance against harmful future events the probability and average severity of which are considered to be positively correlated with climate change, for the immediate payment of some insurance premium. The insurance contract is so designed that the insurance comes into effect only after a certain time interval within which the government can decide to withdraw from the contract and pay back the premium with interest.

2. Within that time interval, the government tries to ensure the passing of domestic climate legislation (e.g. the implementation of a domestic cap-and-trade system or the ratification of an international climate treaty) by promising the respective veto-players early compensation to be paid out of the collected insurance premiums.

3. If it does not succeed, it withdraws from the insurance contract, paying back the premium. Otherwise, both the insurance and the climate legislation come into effect and the government pays the promised compensations to veto-players out of the collected premiums and puts the remaining amount into a fund managed by reinsurance industry that can only be used to compensate the insured damages.

Assume that the insured events are sufficiently closely correlated with climate change and the climate legislation is sufficiently welfare-increasing on the aggregate domestic level. Then the premium can be chosen smaller than the net present value of the insured damages the household expects in the case of no climate legislation, and at the same time large enough so that the passed legislation will reduce the probability and severity of the events by so much that the expected compensation of the insured damages after legislation has a smaller net present value than the collected premium minus the paid compensation to the veto-players. In that case, the households will take out the insurance, the legislation will get passed, and the fund will be able to cover the remaining damages with high probability. Note that the government needs only little commitment power since after the establishment of the fund it or a later government can only influence the respective payoffs if it has to but refuses to cover unexpected additional damages exceeding the fund's capital. In collecting the money from the households and placing most of it into a safe fund, the government neither has to pay any additional risk-related interest since in spite of transferring its default risk to the international lender it rather transfers risk from the household to the domestic society (including the next generation though). If households were more risk-averse than the government, the premium could even be reduced further.

4 Analysis in a simple two-period model

Let us study the strategic implications of the proposed CIC mechanism in a stylized model in which there are only a few players: the government, G, a representative household, H, and a veto-player, V. CIC is modelled as a game with two time periods, the first of which has three stages.

- In period one, stage one, G chooses a *premium*, $z \geq 0$.
- In period one, stage two, H chooses whether to take out the insurance at premium z.
- In period one, stage three, G chooses a domestic *abatement* level, $x \geq 0$, and a *compensation*, $y \geq 0$, such that $y \leq z$ (period-one budget constraint).
- In period one, stage four, V chooses whether to agree to abate the amount x for the compensation y. If she agrees and H has taken out the insurance, CIC comes into effect, domestic abatement q equals x, V has abatement *costs*, $C(q) > 0$, where C is some function, and G pays an amount of y to V. Otherwise, CIC does not come into effect, domestic abatement q equals zero and G pays back the premium z to H. Depending on q, global abatement, Q, then equals $L(q) > 0$, where L is another function (interpreted as that Q which G expects to negotiate internationally when promising a domestic q).
- In period two, the insured event happens with a probability $p = P(Q) > 0$, where P is another function. If the event happens, H loses some amount $a > 0$. If it happens and the CIC scheme is in effect, G pays the same amount a to H as compensation. In addition, G has to cover other climate-related damages amounting to $D(Q)$ for some additional function D.

We assume that $C(0) = L(0) = 0$, $C(q)$ and $L(q)$ are strictly increasing, $P(Q)$ and $D(Q)$ are strictly decreasing, $C(q)$ is strictly convex, $P(L(q))$ and $D(L(q))$ are weakly convex, and all four functions are smooth. E.g., a possible simple specification would be this:

Exponential-quadratic example:

$$C(q) = q^2/2, \quad L(q) = lq, \quad P(Q) = b\exp(-cQ), \quad \text{and} \quad D(Q) = d\exp(-cQ), \tag{7}$$

leaving us with only four parameters, a, b, cl, and d. In the following, we analyse the model first with general functions $C, L, P, D,$ and then with this specific form. For simplicity, we assume all players' utility functions are linear in money, hence payoffs can be measured in monetary units.

Household's choice. H's payoff, π_H, equals either 0 (CIC not in effect, event doesn't happen), $-\delta_H a$ (CIC not in effect, event happens), or $-z$ (CIC in effect), where δ_H is H's discount factor per period. Let $\zeta > 0$ be that value of z for which H is indifferent whether CIC comes into effect or not, given damages a. Then

$$\text{H will take out the insurance iff } z \leq \zeta. \tag{8}$$

We call $\zeta(a)$ H's *willingness to pay* here and assume it is strictly increasing and smooth. If H is *risk-neutral,* its expected payoff before the uncertainty about the event happening is resolved, denoted $E\pi_H$, equals either $-\delta_H aP(0)$ (if CIC is not in effect) or $-z$ (if CIC is in effect), hence

$$\text{for risk-neutral or risk-averse H:} \quad \zeta \geq \delta_H aP(0) > 0. \tag{9}$$

Veto-player's choice. V's payoff, π_V, equals either 0 (if CIC is not in effect) or $y - C(x)$ (if CIC is in effect), thus

$$\text{V will agree iff } C(x) \leq y. \tag{10}$$

Hence

$$\text{CIC can come into effect iff } C(x) \leq y \leq z \leq \zeta. \tag{11}$$

The largest possible abatement, ξ, is thus the largest value of x for which $C(x) \leq \zeta$, i.e., $\xi = C^{-1}(\zeta)$, which is at least $C^{-1}(\delta_H aP(0))$ if H is risk-neutral or risk-averse. In the example (7), we get $\xi = (2\delta_H ab)^{½}$ for risk-neutral H.

4.1 Welfare-maximizing government's choice

If G's objective is to maximize welfare and the latter is assumed to equal total domestic payoff, then its payoff, π^w_G, equals either $-\delta_G D(0)$ (if CIC is not in effect and the event doesn't happen), $-\delta_G[a + D(0)]$ (if CIC is not in effect and the event happens), $-C(x) - \delta_G D(L(x))$ (if CIC is in effect and the event doesn't happen), or $-C(x) - \delta_G[a + D(L(x))]$ (if CIC is in effect and the event happens), where δ_G is G's discount factor per period. If G is risk-neutral, its expected payoff before the uncertainty is resolved, $E\pi^w_G$, equals either $-\delta_G[aP(0) + D(0)]$ (if CIC is not in effect) or $-C(x) - \delta_G[aP(L(x)) + D(L(x))]$ (if CIC is in effect), hence

$$\text{for risk-neutral G:} \tag{12}$$
$$\text{CIC is profitable for G iff } \delta_G aP(0) \;\geq\; f(x) \quad \text{with} \quad f(x) = C(x) + \delta_G[aP(L(x)) + D(L(x)) - D(0)].$$

For risk-averse G, the profitability of CIC for G will depend on the specific form of risk-aversion because both compared cases involve uncertainty.

Risk-neutral, equal discounting case. Assume G is risk-neutral, H is risk-neutral or risk-averse, and $\delta_G = \delta_H = \delta$. Then the budget constraint and (8)–(12) imply that CIC can come into effect and is profitable for G iff *x, y, z* can be chosen so that

$$C(x) \;\leq\; y \;\leq\; z \;\leq\; \zeta \;\geq\; \delta aP(0) \;\geq\; f(x), \tag{13}$$

which is trivially fulfilled for $x = y = z = 0$ since $C(0) = f(0) = 0$. Thus CIC will always come into effect and G's optimal choice of x is then given by that value x^{**} of x (the "second-best" abatement level) which minimizes $f(x)$ under the constraint $0 \leq x \leq \xi$, since that maximizes $E\pi^w_G$ under that constraint. Since this objective function $f(x)$ has negative slope at $x = 0$, positive slope for $x \to \infty$, and is strictly convex and smooth, the unconstrained problem has a unique interior solution $x^* > 0$ (the "first-best" abatement level) where the following first-order condition holds:

$$0 \;=\; f'(x^*) \;=\; C'(x^*) + \delta[aP'(L(x^*)) + D'(L(x^*))]L'(x^*). \tag{14}$$

In the example (7), this is equivalent to $x^* = \delta d' cl \exp(-clx^*)$ with $d' = ab + d$, which solves as $x^* = \mathrm{W}(\delta d' c^2 l^2)/cl$, where $\mathrm{W}(s)$ is Lambert's product log function which is strictly increasing and concave and has slope 1 at $s = 0$. Note that still x^* is not strictly increasing in cl but has a maximum at $cl = (e/\delta(ab + d))^{½}$ in this example.

If $x^* > \xi$, then $f(x)$ is strictly decreasing for $x \leq \xi$ and $x^{**} = \xi$, so the constrained problem has the unique solution

$$x^{**} = \min(x^*, \xi). \tag{15}$$

Although G is in principle indifferent to the exact choice of y and z, let us assume it chooses these in favour of H by putting $y = z = C(x)$. Then the choices made will be:

$$x = x^{**},\ y = z = C(x^{**}),\ \text{H takes out the insurance, and also V agrees to CIC.}$$

In particular, the first-best abatement level x^* will be implemented iff $\xi \geq x^*$, otherwise only the second-best abatement, $x^{**} = \xi$. Equivalently:

$$\text{CIC will implement the first-best abatement } x^* \text{ iff } f'(\xi) \geq 0. \tag{16}$$

For risk-neutral or risk-averse H, this condition is at least fulfilled whenever $f'(C^{-1}(\delta_H aP(0))) \geq 0$ since f is convex.

Exponential-quadratic example. For the specification (7), the latter condition is equivalent to

$$(2\delta ab)^{½} \geq \delta d'cl \exp[-cl(2\delta ab)^{½}]. \tag{17}$$

In terms of the "damage coverage ratio"

$$r = aP(0)/(aP(0) + D(0)) = ab/d',$$

this is equivalent to

$$r \geq cl(2\delta ab)^{½} \exp[-cl(2\delta ab)^{½}]/2. \tag{18}$$

In particular, since (i) the right-hand side of (18) is smaller than ½ and since (17) will also be fulfilled for either (ii) $\delta \to 0$, (iii) $ab + d \to 0$ or $ab + d \to \infty$, (iv) $c \to 0$ or $c \to \infty$, or (v) $l \to 0$ or $l \to \infty$, CIC will implement the first-best abatement in our example (7) at least when either (i) the insured event accounts for at least half of the expected climate-related damages, (ii) players are sufficiently impatient, (iii) expected damages are sufficiently small or sufficiently large, (iv) damages decline sufficiently slow or sufficiently fast with abatement, or (v) domestic abatement has sufficiently low or sufficiently high leverage on global abatement.

Fig. 2 (left) shows the dependence of x^* on r for the case $\delta = ½$, $a = b = 1$, and $cl = ¼, ½, 1, 2,$ or 4, for which $\xi = 1$. In all cases, CIC implements the first-best abatement already for fairly small r, as can be seen from the crossing of the x^* curves with the $\xi = 1$ line.

4.2 Effect of an additional expected budget constraint

As the above choices may leave only little capital saved to cover damages, let us assume that G wants to make sure the fund covers at least the *expected* amount of damages, leading to the

$$\text{Expected budget constraint: } z - y \geq \delta_G[aP(L(x)) + D(L(x))]. \tag{19}$$

G now has an incentive to put $y = C(x)$ but $z = \zeta$ instead of $z = C(x) \leq \zeta$. In the risk-neutral, equal discounting case, constraint (19) is then equivalent to $C(x) + \delta[aP(L(x)) + D(L(x))] \leq \zeta$, which is either fulfilled by the first-best solution x^* or by no choice of x. Hence constraint (19) can be fulfilled iff $x^* \leq C^{-1}(\zeta - \delta[aP(L(x^*)) + D(L(x^*))])$, which is smaller than ξ. Consequently,

G will implement CIC either with the first-best choice x^* or not at all.

The choices are then:

$x = x^*$, $y = C(x^*)$, $z = \zeta$, and CIC will come into effect	if $C(x^*) + \delta[aP(L(x^*)) + D(L(x^*))] \leq \zeta$,
CIC will not come into effect	otherwise.

Exponential-quadratic example. In the risk-neutral, equal discounting case with example (7), the condition is

$$r \geq r_{\text{crit.}}, \quad \text{where} \quad r_{\text{crit.}} = \mathrm{W}(\delta d' c^2 l^2)^2 / 2\delta d' c^2 l^2 + \exp(-\mathrm{W}(\delta d' c^2 l^2)).$$

For very small δ, d', c, or l, CIC will not come into effect since then $\exp(-\mathrm{W}(\delta d' c^2 l^2)) \to 1 > r$. On the other hand, for large d', c, or l, with fixed r, CIC will come into effect since then both terms tend to zero. Fig. 2 (right) shows the dependence of $r_{\text{crit.}}$ on r for the case $\delta = ½$, $a = b = 1$, and cl = ¼, ½, 1, 2, or 4. Under the expected budget constraint, CIC will come into effect (and implement the first-best abatement) whenever the dashed diagonal exceeds the solid line, which requires a larger r the smaller cl is.

Numerical example from STACO. Let us consider a more realistic payoff structure derived from the linearized STACO model of Finus et al. (2006) for Europe, with costs

$$C(q) \;=\; 43.1 \times (0.0024\, q^3/3 + 0.01503\, q^2/2)$$

and total discounted damages

$$\delta T(Q) \;=\; \delta[aP(Q) + D(Q)] = d - 37.4 \times 0.236\, Q,$$

where payoffs are in bln$ and abatement is in Gton CO_2. Let us assume that the insured damages are a fixed fraction of total damages, $aP(Q) = rT(Q)$, for some damage coverage ratio r, so that $D(Q) = (1 - r)T(Q)$. Let us also assume that $P(0) = 1$ and that $L(q) = lq$, for some leverage parameter $l \geq 1$. Then $a = rd/\delta$, $\zeta = rd$, and $\xi = \mathrm{C}^{-1}(rd)$. To determine x^*, we solve (14), which reads

$$0 \;=\; 43.1 \times (0.0024\, x^{*2} + 0.01503\, x^*) - 37.4 \times 0.236\, l.$$

Then CIC will come into effect under the expected budget constraint iff

$$(1 - r)d \;\leq\; 37.4 \times 0.236\, lx^* - C(x^*),$$

i.e., if the uninsured part of the business-as-usual damages does not exceed the discounted expected gains from optimal abatement. Fig. 3 shows these values and limits for different values of l and d.

4.3 Effect of veto-player-friendly government

Now let us assume that G's objective is to maximize not welfare but V's payoff, i.e., it has the "corrupt" payoff function $\pi^c_G = \pi_V = y - C(x)$ if CIC comes into effect. If G could leave the fund without capital, this would trivially be maximized by putting $x = 0$ and $y = z = \zeta$, i.e., the maximal amount H is willing to pay for insurance is simply transferred to V without abating or saving for future damages. With the expected budget constraint (19), however, G's problem becomes

$$y - C(x) =! \max, \quad \text{subject to} \quad y \;\leq\; z - \delta_G[aP(L(x)) + D(L(x))] \quad \text{and} \quad z \leq \zeta.$$

So, it will put $z = \zeta$, $y = \zeta - \delta_G[aP(L(x)) + D(L(x))]$, and choose x to minimize $C(x) + \delta_G[aP(L(x)) + D(L(x))]$, the solution of which is again x^*. Hence:

A veto-player-friendly G will implement CIC either with the first-best abatement x^* or not at all.

It will do so iff that is profitable for V, which is the case if and only if $x^* \leq C^{-1}(y)$. Since the latter is at most equal to $C^{-1}(\zeta) = \xi$, we have in particular:

If $\xi < x^*$, a veto-player-friendly G will not implement CIC.

5 Conclusion

Strategic interaction between a government and domestic stakeholders of climate policy can strongly influence the strategic interaction of this government with other governments in international climate policy. In particular, domestic veto-players that can block the ratification of international environmental agreements (IEAs) can affect the outcome various complicated ways. We demonstrated here that Putnam's two-level game approach can be turned into formal game-theoretical models that are suitable to analyse these effects, and that these models can lead to quite different conclusions than classical models of IEAs which ignore the domestic level.

Already a very simple two-country model showed how a domestic veto-player can restrict the international bargaining set in either a profitable or a non-profitable way for its government. The lesson of this model is that a country might profit in international negotiations if it has a domestic situation in which that part of society that benefits less from climate policy and/or is more affected by its costs have a say in the ratification stage but do not have preferences that differ too extremely from society's. Only if a domestic veto-player's preferences are too different from society's, the government might want to seek possibilities to compensate the veto-player.

Our second, many-country model incorporates the aspect of coalition formation and demonstrates additional effects of domestic veto-players on the breadth and depth of IEAs. In particular, veto-players can endogenously lead to "modest" agreements that stabilize larger coalitions than in classical IEA models, leading to either "broad but shallow" or "broad and deep" agreements. E.g., we showed that if the ratio between the veto-player's and society's benefits from abatement grows fast enough with the level of abatement, the equilibrium result can even be a global coalition implementing the first-best abatement level.

The very simple symmetric functional forms chosen as our models' payoff structures are of course not meant to be realistic, hence our results mainly indicate the importance of identifying the domestic veto-players of the major emitters of GHG and their actual payoff functions to be able to assess the direction of their effect on international climate policy.

As the effect of veto-players highly depends on whether they can be compensated via some form of side-payments, we studied this possibility more closely in the second part of the paper. There we demonstrated how the novel policy scheme of "Cap, Insure & Compensate" (CIC) might overcome the timing, commitment, and risk-related problems that make side-payments seem unlikely in the context of abatement. More precisely, we showed how already at the time abatement starts, the government might use the households' willingness to pay for insurance against much later climate-related damages to compensate those paying the price of abatement and thereby ensure the ratification of an international treaty. In a stylized two-period model, we derived conditions under which this scheme would lead to first-best abatement and characterized the otherwise second-best abatement levels. In particular, using for the relevant functions both a simple example specification and a somewhat more realistic specification from the literature, we showed how the "damage coverage ratio" of the scheme's insurance part strongly affects the performance of the scheme, indicating that quite high damage coverage ratios would be necessary under these assumptions on payoffs.

We believe that the models and results presented here also motivate more detailed game-theoretical studies of the various aspects of the interaction between the domestic and the international levels of climate policy that were outlined in this paper.

References

Downs, George W., Rocke, David M., and Barsoom, Peter N. (1998) Managing the Evolution of Multilateralism. *International Organization* 52 (2), 397–419

Finus, Michael, Van Ierland, Ekko, and Dellink, Rob (2006) Stability of Climate Coalitions in a Cartel Formation Game. *Economics of Governance* 7 (3), 271–291

Finus, Michael (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 (1), 29–67

Finus, Michael and Maus, Stefan (2008) Modesty May Pay! *Journal of Public Economic Theory* 10 (5), 801–826

Froyn, Camilla Bretteville and Hovi, John (2008) A climate agreement with full participation. Economics Letters 99, 317–319

Heitzig, Jobst, Lessmann, Kai, and Zou, Yong (2011) Self-enforcing strategies to deter free-riding in the climate change mitigation game and other repeated public good games. *Proceedings of the National Academy of Sciences of the United States of America* 108, 15739–15744

Heitzig, Jobst (2012) Bottom-Up Strategic Linking of Carbon Markets: Which Climate Coalitions Would Farsighted Players Form? *SSRN Climate Change Law & Policy eJournal* 4, 35 (Working Paper)

Mayer, Frederick W. (1992) Managing domestic differences in international negotiations: the strategic use of internal side-payments. *International Organization* 46 (4), 793–818

Putnam, Robert D. (1988) Diplomacy and domestic politics: the logic of two-level games. *International Organization* 42 (3), 427–460

Victor, David G. (2008) Climate accession deals: new strategies for taming growth of greenhouse gases in developing countries. *Harvard Project on International Climate Agreements Discussion Paper* 08-18, 41 pages

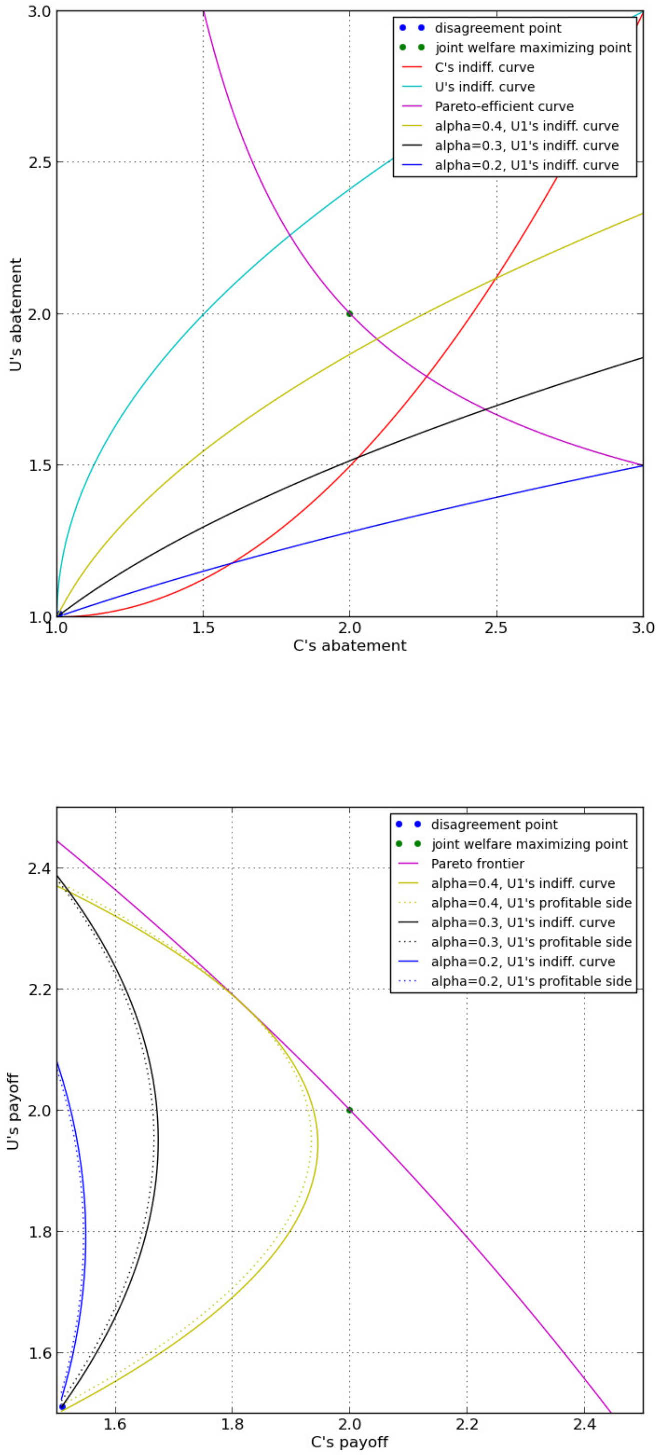

Fig. 1: Example of bargaining set restriction due to domestic veto-players. Countries C, U bargain over abatement levels, where in U there are two veto-players U0, U1. The disagreement point is (1,1) in issue space (top) and (3/2,3/2) in payoff space (bottom). The bargaining set is delimited by the corresponding indifference curves of C (in payoff space this is the vertical axis) and U1. The latter varies with the asymmetry parameter α. For smaller α, typical bargaining solutions give increasing payoffs to U at first but converge to the disagreement point as α approaches 0.

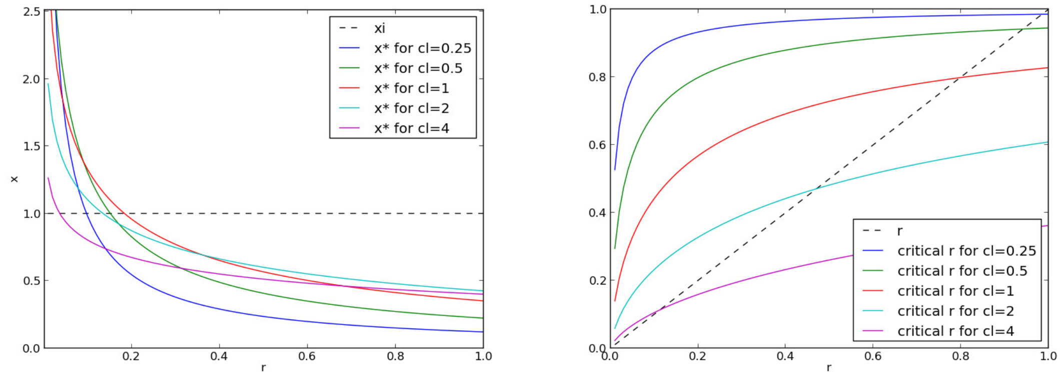

Fig. 2: Example of maximal abatement levels (left) and critical damage coverage ratio (right) with fixed insured damages. Cost, damage, leverage, and probability functions given by (7) with medium discounting of $\delta = ½$, $a = b = 1$, leverage parameter $cl = ¼, ½, 1, 2$, or 4, and varying damage coverage ratio r and uninsured damages $d = 1/r - 1$. Without the expected budget constraint (left), CIC will always come into effect and will implement the first-best abatement level x^* (solid lines) whenever it does not exceed the maximal abatement level ξ which equals 1 here. With the expected budget constraint (right), CIC will only come into effect (and implement the first-best abatement level) if r exceeds the critical r.

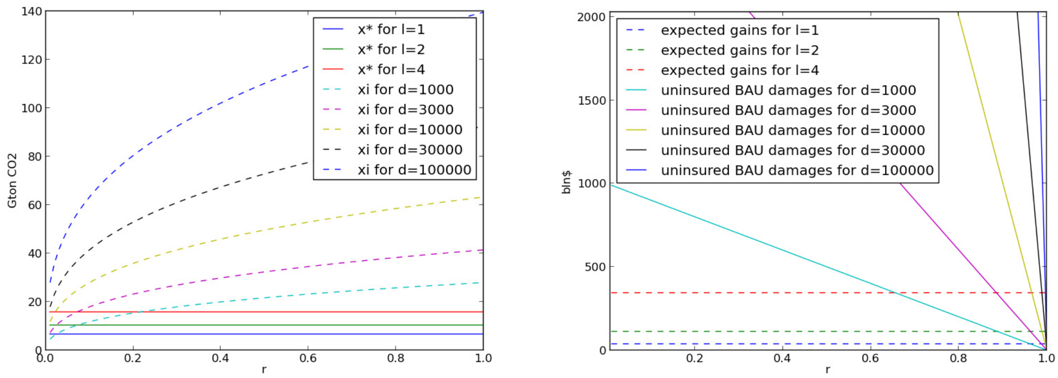

Fig. 3: Rough estimates for Europe. Abatement benefits and costs for Europe from the STACO model, with five different estimates of total business-as-usual (BAU) damages d in Europe, linear leverage with parameter $l = 1, 2$, or 4, and varying damage coverage ratio r. Without the expected budget constraint (left), CIC will always come into effect and will implement the first-best abatement level x^* (solid lines) whenever it does not exceed the maximal abatement level ξ (dashed lines). With the expected budget constraint (right), CIC will only come into effect (and implement the first-best abatement level) if the expected gains (dashed) exceed the uninsured BAU damages (solid). For high leverage and low total damages, this only requires a coverage ratio of about two thirds, while for medium leverage and medium total damages one needs already about 99% damage coverage.