LUISS

Department of Economia e Finanza

Bachelor in Economics and Business

Chair: Games and Strategies

The Stability and Effectiveness of International Environmental Agreements

Prof. Roberto Lucchetti
SUPERVISOR

Lorenzo Comito
CANDIDATE

Academic Year: 2019/2020

CONTENTS

1.INTRODUCTION

International agreements are a means by which countries can commit one another to a particular course of actions in order to improve on the outcome that would occur if they acted unilaterally. To understand the utility of International agreements is necessary to introduce the notion of public goods. Examples of public goods are national defense, street lighting, abatement of carbon emissions and protection of the ozone layer. The fundamental features of these goods is that economic agents cannot be prevented from consuming them (Non-Excludability) and that consumption by one agent does not reduce the good's availability for all other agents (Non-Rivalry). Given these features the provision of a public good that would result from each economic agent maximizing his individual payoff would fall short of the optimal level of provision, the one for which the social marginal benefit of the last unit of public good supplied equals its marginal cost.

However, there is a crucial difference between the examples of public goods that I have provided. In fact, while national defense and street lighting are public goods that are produced and consumed within the boundaries of a state, carbon emissions abatement and the protection of the ozone layer are affected and enjoyed by all countries. They are global public goods. This difference is very relevant because, within the boundaries of a state, there are ways to correct for the under-provision of a public good that would stem by the individual profit-maximizing choices of the economic agents involved. For instance, by means of tax policy the government can collect the funds needed to provide the optimal amount of national defense and street lighting such that their marginal cost will equal their social marginal benefits. Tax policy is effectively a way of coordinating individual behavior to avoid free-riding.

In the case of a domestic public good, it is natural to assume that citizens and firms are the relevant economic agents who choose the provision of the public good, regarding global public goods instead it is usually assumed that countries are the relevant economic agents. Although this assumption is reasonable, perhaps it is a simplification of reality because it implies that countries manage to merge the multitude of individual interests into a coherent and representative set of preferences and they act accordingly. Nonetheless this assumption captures the core of the issue, which is that each state has the right to manage its own affairs and cannot be coerced by foreign intervention.

However states' sovereignty is not absolute, the set of constraints that countries face in conducting their own affairs can be summed up by the obligation of adherence to the 'No harm' principle. This principle is particularly relevant to the provision of international public goods, especially for the ones that involve unidirectional externalities such as pollution of a river that flows downstream from country A to country B. A formal acknowledgement of the 'No harm' principle can be found in the 1972 Stockholm Declaration on the Human Environment. According to Principle 21, "States have, in accordance with the charter of the United Nations and the principles of International Law, the sovereign right to exploit their own resources pursuant to their own environmental policies and the responsibilities to ensure that activities within their jurisdiction or control do not cause damage to the environment of other States or of areas beyond the limits of national jurisdiction''.

It must be pointed out that the boundaries between a country expression of its own sovereignty and its observation of the 'No harm' principle are not clearly drawn. Precisely because the relations among states are not regulated by an overarching international government, there is no clear rule to assess the legitimate conduct of a state. Where the line must be drawn in case of conflict between the two principles is mostly a matter of customs. In international law, customary law refers to the set of informal practices that arose over time out of repeated interactions among states. Because of the lack of formal regulation, deviations from customary law is prevented by other states recognizing deviations to be illicit which in turn damages the deviant who likely must defend itself using diplomatic, legal and material resources

Along with customs, interactions between states are shaped by agreements. Agreements include treaties, conventions, charters, statutes and protocols; they are formal contracts between sovereign states that prescribe binding obligations on a specific matter. Besides, agreements prescribe rules to regulate other parties accession, rules for making amendments, for withdrawing and possibly punishing non-compliance, conditions on the agreement’s entry into force (i.e. a minimum number of participants) and a termination date or a date at which the agreement can be renewed.

This brief discussion on the sources of international law is important to the process of modelling states negotiations. In fact, the structure of formal agreements denote the framework in which state negotiations take place while customary law the sense of fairness that a treaty is supposed to embody. To what extent a treaty imposes fair obligations on its members is not easy to quantify in a model but it is a point that deserves attention, along with the usual focus on how a treaty will affect countries’ payoffs.

In section 2, I will show formally how cooperation can improve on the non-cooperative equilibrium for games that model the provision of a public good. In section 3, I will investigate in more detail some key assumptions in games of treaty negotiations, such as how non-linear cost functions and the introduction of transfers affect cooperation. Finally in section 4 I will consider the branch of the literature that deals with empirical simulations of international agreements designed to reduce carbon emissions.

Section 2 will be mostly based on the book “Environment and Statecraft: The Strategy of Environmental Treaty- Making” by Scott Barrett, whereas section 3 and 4 on academic papers and articles.

2.THE NEGOTIATION GAME

2.1 WHY UNILATERALISM FAILS

	ABATE	POLLUTE
ABATE	2,2	-1,3
POLLUTE	3,-1	0,0

As I have previously introduced, in case of lack of cooperation, a public good will be under-supplied. The reason for this can be better understood highlighting the strategic interactions between players. The provision of a public good in fact can be represented as a Prisoner's dilemma game, in which players decide whether to reduce their emissions (Abate) or not (Pollute). Abatement has public good-like characteristics.

In the matrix above the socially most favorable outcome occurs when both players play 'Abate' because the aggregate payoff will be 4. In case Player 1 plays 'Abate' and Player 2 plays 'Pollute' they will gain respectively a payoff of -1 and 3, because Player 1 will incur the abatement costs but he will benefit only by his share of abatement whereas Player 2 will not incur any abatement costs and nonetheless gain by the abatement supplied by Player 1.

For both players, playing 'Pollute' is the dominant strategy. Which is to say that regardless of Player 2's choice, Player 1 is better off playing 'Pollute'. By the same reasoning Player 2 will play 'Pollute' and therefore (Pollute, Pollute) will be the Nash equilibrium of the game. A Nash equilibrium refers to a set of strategies in which no player can increase his or her payoff by unilaterally changing strategy.

Examining the provision of a public good modeled in this way is tempting to think that the outcome (Abate, Abate) could be reached if only the players managed to coordinate their behavior, for instance by communicating before choosing their abatement level or by imposing a tax on emissions. Considering the latter proposal, suppose that the players agreed to impose a fine equal to 2 in case a player chose to pollute. And that the fine would have to be paid to the other player. Then the payoff matrix would look like this:

	ABATE	POLLUTE
ABATE	2,2	1,1
POLLUTE	1,1	0,0

This matrix shows that if players could commit to pay a fine in case they choose to pollute, the outcome of the game will be the efficient one: (Abate, Abate). However upon little reflection is clear that the problem of cooperation cannot be solved so easily, in fact if a player chose to pollute it would not be in his interests to pay the fine and there is no third party able to enforce the commitment. Both players can easily anticipate that the payment of the fine by the other player is not a credible commitment and hence the actual payoff matrix will be again the first one that I have displayed. In other words, the commitment to pay the fine is not self-enforcing. Before presenting a more realistic way to ensure cooperation is worth clarify in more detail the notion of self-enforcement.

2.2 SELF-ENFORCING AGREEMENTS

The players' agreement on a tax is not a viable solution to the cooperation problem because it is not individually rational for them to maintain their commitment.

Individual rationality is the first condition that a self-enforcing agreement must satisfy and it implies that after an agreement has been negotiated no party of the agreement can gain by withdrawing and no non-party can gain by acceding. This condition can be expressed as:

$\pi_n\left(\alpha - \frac{1}{N}\right) \leq \pi_s(\alpha)$ and $\pi_n(\alpha) \geq \pi_s\left(\alpha + \frac{1}{N}\right)$

Where π_n refers to the payoff of every non-signatory, π_s to the payoff of every signatory and $\alpha \in [0,1]$ to the fraction the N players that join the agreement. Hence at the equilibrium a signatory cannot increase his or her payoff by withdrawing and no signatory can increase his or her payoff by joining the agreement.

The second condition of a self-enforcing agreement is collective rationality. Which implies that the parties of a treaty choose the level of abatement so as to maximize their collective payoff.

If the collective rationality assumption is relaxed, a cooperation can be stable if the countries interact repeatedly. For instance countries can adopt a Grim-strategy which implies that all countries threaten to forsake cooperation if another player would stop supplying the optimal amount of the public good. And this threat is what makes the agreement stable.

Another equilibrium concept which is sometimes used in the literature is the strong Nash equilibrium. This condition requires that no player nor any subgroup of players can increase their payoff by deviating from the equilibrium strategy profile. In turn this implies that the strong Nash equilibrium is Pareto efficient.

2.3 MODELING TREATY PARTICIPATION

Following the model laid out by Scott Barrett in "Environment and Statecraft: The Strategy of Environmental Treaty-Making", I am going to show the determinants of participation in an environmental agreement. Here the Prisoner's dilemma game is extended to $N \geq 2$ symmetric players and the choice of joining the treaty and the abatement decision are modeled as a one-shot, three-stage game. In the first stage players decide whether or not to join the treaty, in the second stage signatories choose their abatement level and in the third stage non-signatories make their abatement decision too. The payoff function is the same for each player:

$$\pi_i = b(Q_{-i} + q_i) - cq_i$$

Where b is the benefit from the reduction in pollution and c the cost of performing the abatement. q_i refers to the discrete abatement choice made by country i, where $q_i = 0$ if country i plays pollute and $q_i = 1$ if it plays abate. Q_{-i} refers to the aggregate abatement provided by all countries other than i.

In order maintain the features of a Prisoner's dilemma game it is assumed that $c > b$ and $Nb > c$ in this way no player will find it individually rational to play abate and full cooperation will be the Pareto efficient outcome. In fact note that if $b = 3$ and $c = 4$ with $N = 2$, the resulting payoff matrix is the same that I have shown at the beginning of the section.

Now the game can be solved by backward induction. It is easy to see that in the third stage non-signatories will play pollute, because it is the individually rational choice. In the second stage signatories will play abate if the resulting aggregate benefits will be higher than the costs, hence:

$q_s = 1$ iff $kb \geq c$ or equivalently $k \geq \frac{c}{b}$, where k is the number of signatories

Hence in the first stage the treaty will be ratified only if signatories expect it to improve on the non-cooperative equilibrium so again if $k \geq \frac{c}{b}$. But no country will enter the agreement if the cutoff point $k = \frac{c}{b}$ has already been reached because it can gain a higher payoff by free-riding on other players abatement. In other words at the equilibrium the number of signatories is just enough to ensure a minimum level of cooperation, not higher because if the cutoff point $k = \frac{c}{b}$ has already been reached an extra signatory will not alter the behavior of other players, since for all players it is individually rational to free-ride on other players abatement. Hence:

$$\frac{c}{b} + 1 \geq k^* \geq \frac{c}{b}$$

The reasoning behind this equilibrium may be clearer plugging numbers into the payoff function, for example if:

$$\pi_i = 3(Q_{-i} + q_i) - 10q_i, \quad N > 4$$

Then $k^* = 4$, each signatory payoff will be $\pi_s = 2$ and each non-signatory payoff will equal $\pi_n = 12$. The outcome is self-enforcing because any signatory is pivotal and if any of them would defect each signatory's payoff would drop to -1 and therefore the non-cooperative outcome $\pi_i = 0$ would ensue.

Note that since all countries are symmetrical each country could be a signatory or a non-signatory, the equilibrium only refers to the number of countries that join the agreement.

The most important conclusion that can be drawn from this participation equilibrium is that full cooperation can rarely be achieved for high values of N and that $k*$ increases in c and decreases in b. The implication for this is that participation in an agreement will be scanter the higher are the potential benefits from cooperation. Denoting by π^c each country payoff in the full cooperation scenario and by π^u each country payoff in the non-cooperative scenario, the gains from cooperation will be:

$$(\pi^c - \pi^u)\, N = (-c + bN)\, N$$

From the above equation is clear that the higher the aggregate potential benefits from full cooperation the lower will be the equilibrium level of signatories.

2.4 ALTERNATIVE SPECIFICATIONS: GRIM TREATY AND THE STRONG NASH EQUILIRIUM

As Scott Barrett (2003) points out the assumption of collective rationality does not allow to reach a higher level of cooperation. If this assumption is dropped, it is easy to see that full participation can be ensured by means of a Grim Treaty agreement. In this setting all countries agree to provide the full abatement level and the treaty stipulates that if any defection would to occur the agreement would collapse and countries would act unilaterally for all the remaining periods of the game.

So the infinitely repeated game is stable if each country is better off by complying rather than cheating. The choice that each country makes depends on what strategy yields the highest expected payoff, discounted by a factor $0 < \delta < 1$. Using the same coefficients as before and assuming say $N = 6$, each country's payoff function is:

$$\pi_i = 3(Q_{-i} + q_i) - 10q_i$$

Each country can expect to gain either π^c or π^n which are the payoffs associated with the strategy 'Comply' and 'Do not comply' respectively.

$\pi^c = 8 + 8\delta + 8\delta^2 + \cdots$ which is equal to $\pi^c = \frac{8}{1-\delta}$

$\pi^{nc} = 15 + 0 + 0 + \cdots$

So for $\delta > \frac{7}{15}$ full participation is ensured. This example can be generalized saying that full cooperation can be sustained as a perfect subgame of the infinitely repeated grim treaty game, for discount factors sufficiently close to 1 (Friedman 1971).

However this outcome is not collectively rational, because even if one defection would occur the other countries would be better off if they ignored the free-rider and carried on cooperation. For this reason a Grim treaty is probably not a realistic model of actual negotiation because its stability is based on the threat to revert to non-cooperation which is not credible (in the collective rationality sense).

If the characterization of negotiations as a Grim treaty allows to reach full-cooperation on the basis of somewhat questionable assumptions, the use of the Strong Nash equilibrium suffers the opposite shortcoming. In fact the latter equilibrium prescribes such strong conditions that it does not allow to ensure any participation in the public good game that I am analyzing. As I have mentioned, a strong Nash equilibrium must be Pareto efficient so considering again the payoff function:

$$\pi_i = 3(Q_{-i} + q_i) - 10q_i, \qquad N = 6$$

The only Pareto efficient outcomes occurs when all six countries provide $q_i = 1$, but obviously this is not a strong Nash equilibrium because any player or subgroup of players can gain from a deviation. In this public good game, the only instances in which a strong Nash equilibrium would exist is for low values of N. In this example for $N \leq 3$, playing $q_i = 0$ would be a strong Nash equilibrium. Note that this result depends on the choice of the benefit and cost parameter; the larger is the ratio of the cost and benefit parameter, the higher is the value of N for which a strong Nash equilibrium exists. In fact a strong Nash equilibrium does exist if $N \leq \frac{b}{c}$. Hence clearly a strong Nash equilibrium is a condition which is excessively difficult to satisfy because it only exists when full cooperation is not profitable.

2.5 CONSENSUS TREATIES

The model that I have presented in the previous sub-section provides quite a discouraging appraisal of the participation level that can be sustained by a self-enforcing treaty. In fact it shows that the equilibrium level of participation is just high enough for the signatories' payoff to be higher than in the non-cooperative scenario. This level of participation is stable because every signatory knows that in case it withdraws it will be worse off since the non-cooperative scenario would ensue. In this sub-section I show how the participation level can be increased thanks to a slight change in the collective rationality assumption. However the increase in participation comes at the expense of the abatement supplied by each signatory.

In the model that I presented in sub-section 2.3, it was assumed that signatories choose the abatement level so as to maximize their collective payoff. This implies that, if the strategy space is restricted to the abatement decision, signatories have no possibility to punish free-riders for not joining the agreement. This model introduces the notion of weak collective rationality (WCR) as opposed to the strong collective rationality (SCR) employed in the first game, precisely to allow signatories to punish free riders.

The features of this model are very similar to the first one, $N > 2$ symmetric countries have to supply a public good (abatement) and are faced with the payoff function:

$$\pi_i = b(Q_{-i} + q_i) - cq_i, \quad with \;\; Nb > c > b$$

However in this model the abatement choice $q_i \epsilon [0,1]$ is continuous and the game is infinitely repeated, the discount factor is set equal to 1. Since the focus of the model is on how to reach a self-enforcing consensus treaty, participation is assumed to be full $k^* = N$.

The treaty specifies an abatement strategy for each player, both for the cooperative phase (when all countries comply) and for the punishment phase (the period following a defection). In the cooperative phase each country earns a payoff π_s, higher or equal to the payoff it would have gotten if all countries played $q_i = 0$ and lower or equal to the payoff it would have gotten if all countries played $q_i = 1$:

$$\pi^{NE} \leq \pi_s \leq \pi^c$$

In case country j fails to comply, a punishment phase would follow. In this phase all other countries would lower their abatement level to q_m^j and country j would have to supply the maximum level of abatement $q_j^j = 1$, as a way to compensate the other players for its willful deviation.

Hence by the weak collective rationality assumption, signatories lower their abatement provision in the punishment phase provided that they are assured to gain a payoff at least as high as their payoff in the cooperative phase. In other words a consensus treaty is considered collectively rational because for every stage of the game it ensures signatories a payoff π_s, which is higher than the Nash equilibrium payoff.

More formally a consensus agreement satisfies two conditions. Both these conditions refer to the punishment phase. The first condition ensures that a unilateral deviation from the agreement makes the deviant worse off, hence if country j deviates from the agreement by choosing $q_j = 0$ the payoff it gets must be lower than what it would gain if it allows cooperation to resume:

$$b(N-1)q_m^j \leq \pi_s, \quad \text{where } m \text{ refers to all players except } j$$

The left-hand side of the equation shows the benefit that country j gets by other players' abatement, where $q_m^j \epsilon [0,1]$ is the level of abatement chosen by each one of the other countries. The right-hand side is the payoff that country j gains allowing a new cooperative phase to become established. Clearly in the period immediately after its deviation (punishment phase) country j would gain a low payoff because it would have to supply the highest level of abatement, but because the discount factor is equal to 1 in the long run this one-time low payoff can be ignored; for this reason the right-hand side shows the cooperative payoff.

The second condition ensures that each one of the $N-1$ countries cannot do better than to punish j:

$$b\left[q_j^j + (N-1)q_m^j\right] - cq_m^j \geq \pi_s, \quad q_j^j = 1$$

Hence the first condition states that no deviation can be profitable and the second condition states that in case a deviation would occur the remaining signatories would be better off punishing the deviant. The second condition rests on the assumption that the deviant would accept to supply the maximum level of abatement ($q_j^j = 1$), because the deviant country knows that it would gain allowing cooperation to resume.

To solve the game it must be found the maximum cooperative payoff (π_s) for which the above conditions hold true.

Setting $q_j^j = 1$

$$\frac{\pi_s - b}{b(N-1)-c} \leq q_m^j \leq \frac{\pi_s}{b(N-1)} \quad \text{or} \quad \pi_s \leq \frac{b^2(N-1)}{c}$$

Hence this model shows that full cooperation can be ensured up to a certain payoff. The rationale behind this result is that the cooperative payoff to maintain the consensus treaty self-enforcing has an upper bound because otherwise the threat to punish defectors would not be credible. Hence the higher is N the lower will be the cooperative payoff with respect to the Pareto efficient outcome. Note that π_s increases in N but at a lower rate than the Pareto efficient outcome.

3.TREATY NEGOTIATION UNDER ALTERNATIVE ASSUMPTIONS AND AN EXPANDED STRATEGY SPACE

In this section I will expand on the games that I have previously presented in order to achieve a more realistic depiction of International Environmental Agreements. Namely I will show how the assumption of quadratic abatement costs affects the equilibrium level of participation and the payoff sustained by a consensus treaty. Moreover I will talk about the role of trade restrictions in International Agreements and I will present a game that allows for money transfers between countries to assess if these measures are beneficial to international cooperation.

3.1 THE EFFECT OF QUADRATIC COSTS ON TREATY PARTICIPATION

In the previous section I have assumed that the costs and benefits of abatement provision grow linearly as more abatement is supplied. Although these assumptions on costs and benefits depend on the nature of the abatement that players are supplying, to assume linear costs and benefits is rarely the most realistic choice. For instance in the case of abatement of carbon emissions, though it is very difficult to understand what is the relationship between emissions and damages, the most widely used estimates assume damages to grow at a growing pace as emissions rise; as a consequence benefits from abatement will grow at a slower rate, as more abatement is supplied

Regarding costs of abatement of carbon emissions instead it is relatively easier to choose a reasonable specification. Most studies assume quadratic costs. The rationale of this assumption is that carbon emissions are released by a large number of sectors (i.e. transportation, industrial production, agriculture etc.) and the costs of reducing emissions varies widely among them. Therefore as more abatement is required there must be found alternative sources of energy also for the sectors for which this transition is more expensive.

In this sub-section I will analyze again the first game that I have presented for $N > 2$ symmetric players, but I assume the following payoff function:

$$\pi_i = b(Q_{-i} + q_i) - c\frac{{q_i}^2}{2}, \qquad Nb > c > b$$

Recall that the game has three stages. In the first stage countries decide whether to sign the agreement, in the second stage signatories make their abatement decision maximizing their collective payoff and in the third stage each non-signatory chooses its abatement level maximizing its individual payoff. Solving by backward induction, in the third stage each non-signatory maximizes:

$$\pi_n = bq_n - c\frac{{q_n}^2}{2}$$

Which is maximized for $q_n^* = \frac{b}{c}$. In the second stage signatories will choose their level of abatement maximizing their collective payoff:

$$\Pi_s = N\alpha b q_s - c\frac{{q_s}^2}{2}$$

Where $\alpha \in [0,1]$ denotes the fraction of signatories. Their aggregate payoff is maximized for $q_s^* = \frac{\alpha N b}{c}$.

Now recall that the conditions that are satisfied by the equilibrium participation level are:

$\pi_n\left(\alpha - \frac{1}{N}\right) \leq \pi_s(\alpha)$ and $\pi_n(\alpha) \geq \pi_s\left(\alpha + \frac{1}{N}\right)$

These conditions ensure that no signatory can gain by unilaterally withdrawing form the agreement and no non-signatory can gain by unilaterally joining the agreement. Plugging the values q_s^* and q_n^* into the above equations it turns out that the two condition holds true respectively for:

$\alpha \geq \frac{2}{N}$ and $\frac{3}{N} \geq \alpha \geq \frac{1}{N}$. Since they must both be satisfied simultaneously: $3 \geq \alpha N \geq 2$.

So the specification of quadratic costs makes impossible to sustain high participation, the highest viable participation level is 3 signatories.

3.2 QUADRATIC COSTS IN THE CONSENSUS TREATY GAME

The quadratic costs assumption yields some interesting results also if applied to the full participation game. So I consider the infinitely repeated game with $N > 2$ symmetric countries that choose their abatement provision to maximize the following payoff function:

$$\pi_i = b(Q_{-i} + q_i) - c\frac{{q_i}^2}{2}, \qquad Nb > c > b$$

The abatement choice is continuous on the interval $q_i \epsilon [0,\ q^{max}]$.The Nash equilibrium level of abatement is $q_i^{NE} = \frac{b}{c}$ and the associated payoff $\pi_i^{NE} = \frac{b^2(2N-1)}{2c}$. While the Pareto efficient abatement level is $q_i^c = \frac{Nb}{c}$ which yields a payoff of $\pi_i^c = \frac{b^2N^2}{2c}$. By assumption the upper bound on q_i is set equal to the full cooperation level of abatement, so $q^{max} = \frac{Nb}{c}$.

Once again to ensure that the treaty is self-enforcing, two conditions must be satisfied. The first one states that in the punishment phase a deviating country j cannot gain a higher payoff than in the cooperative phase, no matter what abatement level (q_j) the deviant chooses:

$$b(Q_{-j}+ q_j) - c\frac{q_j^2}{2} \leq \pi_s^w$$

And the second condition ensures that in the punishment phase signatories gain a payoff at least as high as their cooperative payoff, this is possible because country j must provide a high abatement level q_j^j to make amends of its deviation and to allow cooperation to resume:

$$b\left(Q_{-j} + q_j^j\right) - c\frac{\left[\frac{Q_{-j}}{N-1}\right]^2}{2} \geq \pi_s^w$$

Clearly if these two conditions are met country j will anticipate that it cannot gain by withdrawing from the agreement and the treaty will be self-enforcing.

To find the maximum payoff π_s^w that can be sustained in the cooperative phase consider the two constraints in turn. In the first one country j will choose q_j to maximize its individual payoff, hence it would choose the Nash equilibrium level of abatement $q_j = \frac{c}{b}$. Therefore the first constraint can be re-written as:

$$bQ_{-j} + \frac{b^2}{2c} \leq \pi_s^w$$

In the second constraint, the level of abatement that the treaty obliges country j to provide in the punishment phase is set at the highest possible level, $q_j^j = \frac{Nb}{c}$. Hence:

$$b\left(Q_{-j} + \frac{Nb}{c}\right) - c\,\frac{\left[\frac{Q_{-j}}{N-1}\right]^2}{2} \geq \pi_s^w$$

Rearranging:

$$Q_{-j}\left[b - c\,\frac{Q_{-j}}{2(N-1)}\right] + \frac{b^2N}{c} \geq \pi_s^w$$

The two constraints imply that $Q_{-j} = \frac{Nb-b\sqrt{2N-1}}{c}$, hence the cooperative payoff that can be sustained by a self-enforcing consensus treaty is:

$$\pi_s^w = \frac{b^2\left[1 + 2(N-1)\sqrt{2N-1}\right]}{2c}$$

3.3 THE EFFECT OF TRADE RESTRICTIONS ON TREATY PARTICIPATION

Up to this point the games that I have presented modeled countries' strategic choices only with respect to their abatement decision. However this assumption highly simplifies actual negotiations and may not capture important elements that allow some treaties to reach high level of cooperation. For instance, the Montreal Protocol which is regarded as one of the most successful examples of international cooperation relies on a number of instruments to restructure countries' incentives and foster cooperation. The Montreal Protocol was signed in 1987 and prescribed countries to cut back emissions of chlorofluorocarbons, a class of chemicals responsible for ozone depletion. The protocol underwent several amendments over the years which allowed for greater reductions of emissions and increases in the number of signatories. Now participation is virtually full and so is compliance with the terms of the treaty. Key measures prescribed by the treaty are the trade ban between signatories and non-signatories in the substances controlled by the treaty and the ban on imports of goods containing these substances (i.e. air conditioners).

More generally a treaty can employ trade restrictions to pursue two ends: to punish countries that do not cooperate and to remedy for the higher costs that must be incurred by firms in the cooperative countries. Regarding the latter goal, trade restrictions are especially important if the amount of leakage produced by the agreement is significant. Leakage denotes the mechanism by which pollution-intensive industries shift to countries with looser regulations once their home country implements measures to reduce emissions, increasing the costs of producing pollution-intensive goods. A thorough analysis of leakage is performed by Brian Copeland and Scott Taylor (2000), the authors decompose the best response function of a ‘dirty’-good exporter country to a cutback in emissions by the rest of the world and they identify four effects:

1) The free-rider effect.

2) The producer substitution effect: production of the dirty good rises because of its world price increase.

3) The consumer substitution effect: the increase in the ‘dirty’ good world price means a reduction in the quantity demanded by consumers.

4) Income effect: since environment quality is a normal good, the quantity demanded increases as consumers’ income increases. The increase in income in the ‘dirty’-good exporter country as a result of the higher world price will bring a reduction in emissions.

The identification of this last effect, called ‘bootstrapping’ by the authors, is one of the most valuable contributions of the paper and leads the authors to conclude that a country in response to the rest of the world reducing emissions may even increase its abatement.

To include into a game of international cooperation such a nuanced characterization of the possible trade measures that may be put in place by signatories and the associated response of non-signatories would probably render the model excessively complex, in fact games that include trade considerations into the analysis make a number of simplifying assumptions.

For instance, Scott Barrett in “Environment and Statecraft: The Strategy of Environmental Treaty Making” considers the effect of a trade ban of pollution-intensive goods on non-signatories. The author adds a set of players to the participation game that I have already analyzed in the linear and quadratic specification. They are N imperfectly competitive firms - one for each country – whose cost function increases in the level of abatement that they are required to provide by their home country. Because in this game firm’s profit is an element of country’s payoff function, the author shows that full cooperation can been ensured if signatories impose a trade ban on non-signatories. In fact the trade ban transforms the game into a coordination game with two equilibrium in pure strategy: an equilibrium with no signatories and one with full participation. The full participation equilibrium is the Pareto efficient one and it can be easily achieved if a minimum participation clause is set.

3.4 TRANSFERS AMONG SYMMETRIC COUNTRIES

Another common instrument to increase treaty participation is the use of side payments. A side payment is a transfer of money from one country to another, conditional on the receiving country’s accession to the agreement and its provision of the public good. The following analysis is based on the paper Scott Barrett (2001) in which the author shows that, using the usual internal/external stability concept, under certain assumptions side payments can be beneficial to treaty participation.

The rationale of side payments is that countries’ aggregate payoff increases as treaty participation is widened, hence signatories may manage to counteract the free-riding incentive of non-signatories by offering them a side payment and still be better off thanks to the extra abatement supplied. First of all, I consider the linear specification of the one-shot participation game that I have presented in sub-section 2.3, to check if side payments can be beneficial in this setting. Consider once again the payoff function that I have used previously as an example:

$$\pi_i = 3(Q_{-i} + q_i) - 10q_i \quad \text{with say} \quad N = 6$$

Where the equilibrium number of signatories is $k^* = 4$, signatories payoff is $\pi_s = 2$ and non-signatories payoff $\pi_n = 12$. A money transfer m from signatories to a non-signatory should satisfy these conditions:

$$\pi_s(k^* + 1) - \frac{m}{k^*} > \pi_s(k^*)$$

$$\pi_s(k^* + 1) + m > \pi_n(k^*)$$

The first condition ensures that all signatories find it rational to pay the side payment m to increase participation, note that the cost of m is equally shared among signatories. And the second condition ensures that the non-signatory who is offered the side payment is better off accepting it.

For the coefficients that I am using in this example, $m = 8$ satisfies these constraints. In this way after the side payment is paid the original four signatories will have a payoff of $\pi_s = 3$, the newly convinced signatory will have a payoff of $\pi_s' = 13$ and the only remaining free-rider will have a payoff of $\pi_n = 15$. However it is easy to see that this outcome is not stable because each of the four original signatories can gain by withdrawing. Indeed if any of the four original signatories would withdraw the treaty participation level would bounce back to the equilibrium $k^* = 4$ and the withdrawing country would be better off since it would enjoy the non-signatory payoff $\pi_n = 12$.

More generally a side payment cannot increase participation because the payoff of a signatory that contributes to the payment of m is strictly less than what it could gain by withdrawing:

$$\pi_s(k^* + 1) - \frac{m}{k^*} < \pi_n(k^*)$$

3.5 MODELLING TREATY PARTICIPATION WITH ASYMMETRIC COUNTRIES

The above analysis cannot be conclusive because all players are assumed to be symmetric. Real countries obviously are not symmetric and in treaties that prescribe side payments there are always notable differences between givers and recipients. To account for countries' asymmetries, consider two types of countries N_1 and N_2 . Each country is faced with the binary choice 'Abate' or 'Pollute' which provides the following payoffs:

$$\pi_A^i = \alpha_i(b_1 z_1 + b_2 z_2) - c \qquad \pi_P^i = \alpha_i(b_1 z_1 + b_2 z_2)\ , \qquad i = 1,2$$

Where z_1 and z_2 are the type 1 and type 2 countries that play 'Abate'. Then assume $b_2 > b_1$, $\alpha_2 = 1$ and $\alpha_1 \epsilon [0,1]$. Type 2 countries can be seen as the developed countries, in fact if a type 2 country abates its emissions this has a larger effect than the abatement performed by a type 1 country $(b_2 > b_1)$ because developed countries generally pollute more. Moreover type 2 countries are assumed to benefit more from an equal amount of abatement than type 1 countries $(\alpha_2 > \alpha_1)$. Then assume the following conditions to make the game a prisoners' dilemma game:

$c > b_2$ and $\alpha_1 N_1 + N_2 > \frac{c}{b_1}$

So first of all consider the three-stage game without side payments. In the first one, all countries choose simultaneously to be a signatory or a non-signatory. In the second stage, signatories choose jointly whether to play ‘Pollute’ or ‘Abate’. And in the third stage, non-signatories simultaneously and independently make their abatement choice.

Solving the game by backward induction, it is easy to see that in the third stage non-signatories will play pollute.

To find the equilibrium of the second stage consider the aggregate payoff of signatories, because by assumption signatories maximize their collective payoff:

$$\Pi_s = [(\alpha_1 k_1 + \alpha_2 k_2)b_1 - c]z_1 + [(\alpha_1 k_1 + \alpha_2 k_2)b_2 - c]z_2$$

Note that k_i refers to the number of countries that sign the treaty to play ‘Abate’ while z_i denote the number of countries which actually play ‘Abate’, but since non-signatories will play ‘Pollute’ and signatories sign the treaty only if they anticipate that playing ‘Abate’ will be profitable, these values always coincide. Maximizing Π^s with respect to z_i gives the following solution:

$z_{\ i}^{*} = k_i$ if $\alpha_1 k_1 + \alpha_2 k_2 > \frac{c}{b_i}$ or $z_i^* = 0$ if $\alpha_1 k_1 + \alpha_2 k_2 < \frac{c}{b_i}$

The solution requires that signatories of the same type act in the same way, so three equilibria can occur:

$z_1^* = z_2^* = 0$ or $z_1^* = 0,\ z_2^* = k_2$ or $z_1^* = k_1$ and $z_2^* = k_2$. Note that it cannot occur the equilibrium in which only type 1 countries abate, because the benefit they get from abatement is strictly lower than what type 2 countries get; therefore if type 1 countries find it rational to abate, it will be necessarily rational to do so for type 2 countries too.

Anyways it is clear that countries will sign the agreement only if they anticipate that is profitable to provide the abatement, hence the equilibrium of the first stage is the critical one. In this stage three equilibria $\{k_1^*, k_2^*\}$ are possible:

(1) If $\alpha_1 N_1 > \frac{c}{b_1}$ then $\frac{c}{b_1} + \alpha_1 > \alpha_1 k_1^* > \frac{c}{b_1}$ and $k_2^* = 0$

(2) If $\alpha_2 N_2 > \frac{c}{b_2}$ then $k_1^* = 0$ and $\frac{c}{b_2} + \alpha_2 > \alpha_2 k_2^* > \frac{c}{b_2}$

(3) $k_1^*\ and\ k_2^*\ that\ satisfy\ these\ two\ conditions$:
$\frac{c}{b_2} + \alpha_1 > \alpha_1 k_1^* + \alpha_2 k_2^* > \frac{c}{b_1}$ and $\alpha_1 k_1^* + \alpha_1 k_2^* \frac{b_2}{b_1} > \frac{c}{b_1}$

As in the equilibrium with symmetric countries in these equilibria the number of signatories is just high enough to make cooperation profitable, in fact if a signatory would withdraw all other signatories would play pollute.

As an example consider the following values $N_1 = N_2 = 50,\ \alpha_2 = 1,\ \alpha_1 = 0.5, b_2 = 6, b_1 = 3,\ c = 100$ there exists only one equilibrium of type (2), namely $\{0{,}17\}$.

Note that more than one type of equilibria can exists. For values $N_1 = N_2 = 50,\ \alpha_2 = 1,\ \alpha_1 = 0.75$, $b_2 = 6$, $b_1 = 3$, $c = 100$ there exists two equilibria, of type (1) and (2). Namely $\{45{,}0\}$ and $\{0{,}17\}$.

The third equilibrium exists only when the countries are quite symmetric, in fact note that to satisfy the first condition it must be that $\frac{c}{b_2} + \alpha_1 > \frac{c}{b_1}$. For instance for values $N_1 = N_2 = 50, \alpha_2 = 1,\ \alpha_1 = 0.9$, $b_2 = 6,\ \ b_1 = 5.9$, $c = 100$ there exists six equilibria $\{19{,}0\}$, $\{18{,}1\}$, $\{17{,}2\}$, $\{16{,}3\}$, $\{15{,}4\}\ and$ $\{0{,}17\}$.

From these examples it can be seen that when the two types of countries are strongly asymmetric, like in the first example, only type 2 countries will play ‘Abate’ because type 1 countries do not find it collectively rational to do so. However the benefit that type 2 countries would get from type 1 countries’ abatement exceeds the costs of supplying it. This observation opens the possibility that the use of transfers may increase countries’ aggregate payoff.

3.6 TRANSFERS AMONG ASYMMETRIC COUNTRIES

Given the above consideration now consider the case in which the two types of countries are strongly asymmetric, hence only equilibrium (2) exists. To incorporate side payments into the game consider a one-shot four-stage game. In stage one, every type 2 countries chooses whether to sign the agreement. In stage two, type 2 countries collectively choose their abatement and whether to offer a side payment m to every type 1 country that agrees to join the agreement and play ‘Abate’. In stage three, every type 1 country decides whether to join the agreement. In stage four, every non-signatory of both types makes its abatement decision.

Solving the game by backward induction, as usual non-signatories will play ‘Pollute’ in the fourth stage. In the third stage, a type 1 country will join the agreement if it finds individually rational to do so. Hence if the following condition is satisfied:

$m \geq c - \alpha_1 b_1$

So assuming that side payments are not rationed, all type 1 countries will act in the same way: either $k_1^* = N_1$, if the above condition is satisfied or $k_1^* = 0$ if it is not satisfied.

In stage two, type 2 countries have to make both the abatement and the side payment decision. They will collectively choose to play ‘Abate’ if the usual condition $k_2^* > \frac{c}{b_2}$ is satisfied, note that α_2 was dropped because it equals 1 by assumption. So, if the above condition is satisfied and no side payments are offered, the collective payoff of type 2 country equals:

$$\Pi^s = k_2^*\,(b_2 k_2^* - c)$$

If instead a side payment equal to $m = c - \alpha_1 b_1$ is offered to every type 1 country, to the above payoff it must be added the benefit of the extra abatement and subtracted the cost of the side payments, so:

$$\Pi^s = k_2^*\,(b_2 k_2^* - c) + k_2^* b_1 N_1 - N_1(c - \alpha_1 b_1)$$

Clearly type 2 countries will agree to offer the side payment if the second term in the above equation is larger than the third term, hence:

$$m = c - \alpha_1 b_1 \quad \text{if} \quad k_2^* \geq \frac{c}{b_1} - \alpha_1 \qquad \text{otherwise} \quad m = 0$$

Given the assumption that countries are strongly asymmetric only equilibrium (2) would hold in this setting if side payments were not employed, allowing for side payments there are two possible equilibria in stage 1:

(2a) If $\frac{c}{b_1} - \alpha_1 > N_2 > \frac{c}{b_2}$ then $k_1^* = 0$ and $\frac{c}{b_2} + \alpha_2 > \alpha_2 k_2^* > \frac{c}{b_2}$

(2b) If $\alpha_2 N_2 > \frac{c}{b_1} - \alpha_1$ then $k_1^* = N_1$ and $\frac{c}{b_1} + \alpha_2 - \alpha_1 > \alpha_2 k_2^* > \frac{c}{b_1} - \alpha_1$

Equilibrium (2a) is the same as equilibrium (2) in the previous sub-section and so it refers to the case in which no side payments are offered. Equilibrium (2b) is the new one which occurs if side payments are offered and accepted. Hence $k_1^* = N_1$ because all type 1 countries will accept a side payment $m = c - \alpha_1 b_1$. $k_2^* > \frac{c}{b_1} - \alpha_1$ because otherwise type 2 countries would not have offered the side payment. And $k_2^* < \frac{c}{b_1} + \alpha_2 - \alpha_1$ is the usual upper bound above which joining an agreement is not individually rational.

Simulations show that thanks to the assumption of asymmetric countries, side payments can be beneficial to cooperation. For instance consider the first example presented in the previous sub-section, where $N_1 = N_2 = 50$, $\alpha_2 = 1$, $\alpha_1 = 0.5$, $b_2 = 6$, $b_1 = 3$, $c = 100$ which leads to an equilibrium $\{0,17\}$. Allowing for side payments instead, it can be seen that the values satisfy the conditions of (2b) and $\{50,33\}$ is the equilibrium.

So transfers have significantly increased the number of signatories and the aggregate welfare, that was 5950 without transfers and became 17800 with transfers.

4. EMPIRICAL STUDIES OF INTERNATIONAL ENVIRONMENTAL AGREEMENTS

In this section I will focus on a critical strand of the literature on international agreements which deals with simulations of international environmental agreements to assess the stability of an agreement concerning the provision of a particular public good and the magnitude of the effects that it will bring about. Specifically I will present simulations of international agreements that aim at reducing carbon emissions made on the basis of Integrated Assessment models. First of all I will briefly describe the purpose and the functioning of Integrated Assessment Models and then I will focus on two papers that make use of them to study the stability of a climate change agreement. This section will allow me to give an overview of this strand of the literature, to compare the conclusions of two representative works and to judge if the conclusions offered by the papers presented in the previous sections of this thesis are confirmed by empirical studies.

Because in this section I deal with empirical studies clearly the subject matter of the international agreement must be clearly defined, as opposed to the previous sections in which the theoretical findings can be applied to the provision of international public goods in general. I decided to focus on climate change agreements because arguably for no other environmental problem the need for cooperation is so essential, not only for the magnitude of the changes required but most importantly because reduction of carbon emissions is a global and intergenerational public good which therefore requires coordination among many widely heterogeneous countries. The difficulty of reaching a successful agreement and the extent of the benefits that can be achieved from cooperation make it the ideal public good problem to focus on.

4.1 INTEGRATED ASSESSMENT MODELS

Firstly a brief description Integrated Assessment Models (IAM). Integrated assessment models integrate features of the world economy and of the natural world to assess the interplay and the feedback effects between economic growth, the climate and the stock of earth's natural resources. Interactions between these variables are modeled over long time spans (usually more than a century) so that the model's forecasting can be used as a reference for policies that have long-term effects. Naturally a number of simplified assumptions and estimates are needed to accomplish such task, even more so if the goal is to predict how a region's future welfare responds to different climate policies to draw conclusions on its interests in joining an agreement. For instance an important choice of the modeler is in how many regions the world economy must be divided, in fact no IAM accounts for all countries as single agents but rather they aggregate them in vast geographical regions.

The second paper that I will present (Nordhaus 2015) is based on Coalition-DICE model, a IAM designed specifically to find a stable coalition in a climate agreement. Countries are aggregated into 15 regions that maximize their individual welfare in a one-shot game. The model includes several variables for each region: population, exogenous output, baseline carbon emissions and a baseline trade matrix. Carbon emissions are defined as an externality of production. The damages from emissions are defined as the social cost of carbon (SCC), which corresponds to the damages (in monetary terms) of a ton of carbon emissions. The SCC is constant by assumption and since estimates about it vary widely, simulations are made using several values (from 12.5 to 100 dollars per ton). National social costs of carbon are even more difficult to estimate, so they are assumed to be proportional to the region's share of global GDP. Abatement costs are quadratic in the quantity of emission reduction and they are region-specific, they depend mostly on the region's carbon-intensity of production (the ratio of emissions to output) and on technological and sectoral differences. An important trademark of the model is that it accounts for the effect of international trade and tariffs on regions' economic welfare, where the effect of tariffs is estimated using a tariff-impact function. Another peculiarity of the model is that there is no discount rate because it is implied in the social cost of carbon and even if the discount rate influences also investment and consumption decisions, these second-order effects of the discount rate are not deemed essential to be accounted for since they do not significantly affect the conclusions of the simulation.

The other paper that I will focus on employs a similar model the: CLIMNEG Integrated Assessment model.

4.2 NEW ROADS TO INTERNATIONAL ENVIRONMENTAL AGREEMENTS: THE CASE OF GLOBAL WARMING

This paper (Eyckmans and Finus 2003), based on the CLIMNEG model, analyses the stability of international environmental agreements under a number of different assumptions. Namely the authors consider open versus exclusive membership, transfers versus no-transfers scenarios and single versus multiple coalition formation. The usual internal-external stability concept is adopted.

By the single coalition formation assumption, only one coalition can form and players can either join or act as singletons, instead the multiple coalition formation assumption allows players to arrange themselves in more than one non-trivial coalitions. It is not easy to judge which coalition formation assumption better depicts actual international agreements, because on the one hand it is true that each public good problem is usually tackled with a single agreement (i.e. the Montreal protocol for the ozone layer depletion or the Paris agreement to reduce carbon emissions) but on the other hand these agreements prescribe different provisions for different groups of countries (i.e. developed versus developing countries) so the multiple coalition formation assumption might yield a more realistic analysis.

The open/exclusive membership assumption concerns the ability of a newcomer to join a coalition: if membership is open every player is free to join any coalition, instead if membership is exclusive the newcomer must be allowed to join by either all members of the coalition or by at least the majority of them (both specifications of the exclusive membership are considered).

The CLIMNEG model aggregates countries into six regions (US, EU, Japan, China, Former Soviet Union and Rest of the World) that maximize discounted consumption over 35 ten-year periods. In each period gross production $(Y_{i,t})$ is allocated to consumption $(Z_{i,t})$, investments $(I_{i,t})$, investments in abatement $(Y_{i,t}C_i(\mu_{i,t}))$ and damages from climate change $(Y_{i,t}D_i(\Delta T_i))$. Where $\mu_{i,t} \in [0,1]$ refers to the fraction of emission abatement with respect to the Business-as-usual scenario.

$$Y_{i,t} = Z_{i,t} + I_{i,t} + Y_{i,t}C_i(\mu_i) + Y_{i,t}D_i(\Delta T_i)$$

Since damages from climate change and abatement costs are included in gross production, $Y_{i,t}$ must be interpreted as potential output: what could have been produced in the absence of climate change. The cost and damage functions are increasing and convex functions of emission abatement and temperature change respectively.

Treaty negotiation is modeled as a two-stage game. In the first one each region decides whether to join a coalition and in the second stage regions choose their economic and abatement strategy maximizing the collective payoff of their coalition and acting non-cooperatively towards outsiders. In the first stage each region $I\{1, \dots, N\}$ chooses an address $\sigma = \{\sigma_1, \dots, \sigma_N\}$ which is mapped into the resulting coalition structure $c = \{c^1, \dots, c^M\}$ via a coalition function $\varphi(\sigma)$. The outcome of the first stage depends on the membership assumption adopted, for the single membership game the choice of the address is restricted to $\sigma = \{0,1\}$ so:

$$c^i = \begin{cases} \{i\}\ if\ \sigma_i = 0 \\ \{j \backslash \sigma_j = 1\}\ if\ \sigma_i = 1 \end{cases}$$

Player i can act as a singleton by choosing $\sigma_i = 0$ or if he chooses $\sigma_i = 1$ he will join the single non-trivial coalition with any other player that has chosen to do so.

In the multiple membership agreements, the coalition function φ maps the membership decision $\sigma = \{0, \dots, N\}$ into the coalition structure in the following way:

$$c^i = \{i\} \cup \{j \backslash \sigma_i = \sigma_j\}$$

So two players are part of the same coalition if they make the same announcement.

In the second stage the economic decision of each coalition depends on the valuation function which is a composition of two functions $v = w \circ \varepsilon$, where ε is a function that maps the coalition structure into economic strategies and w is a function that maps economic strategies into welfare levels. So a coalition's welfare $w(s)$ depends on a vector of all other coalitions' economic strategies. An economic strategy refers to emission abatement and capital investment decisions. The authors further assume that does exist a vector of economic strategies s^* which is a unique interior equilibrium.

Because for every coalition structure there exists a unique vector of payoffs v^*, the equilibrium of the game can be expressed in terms of the equilibrium announcements $\sigma^* = \{0, \dots, N\}$ such that the resulting vector of payoffs satisfies the conditions of internal and external stability.

For the single coalition game, the stability conditions are very similar to the ones employed in the games that I have presented so far, the only difference refers to the external stability condition in case the assumption of exclusive membership is adopted. In the single coalition game let I^{NC} be the set of players announcing $\sigma_j^* = 0$ and I^C the set of players announcing $\sigma_i^* = 1$ then consider the following stability conditions:

a) Internal stability: $\forall i \in I^C \quad v_i(\sigma_i^* = 1, \sigma_{-i}^*) > v_i(\sigma_i = 0, \sigma_{-i}^*)$
b) External stability 1: $\forall j \in I^{NC} \quad v_j(\sigma_j^* = 0, \sigma_{-j}^*) > v_j(\sigma_j^* = 1, \sigma_{-j}^*)$
c) External stability 2: $v_j(\sigma_j^* = 0, \sigma_{-j}^*) < v_j(\sigma_j^* = 1, \sigma_{-j}^*)$ and

c.1) $\exists S \epsilon I^C,\ S \geq \frac{I^C}{2} \quad v_i(\sigma_j^* = 0, \sigma_{-j}^*) > v_i(\sigma_j = 1, \sigma_{-j}^*)$

c.2) $\exists i \epsilon I^C \quad v_i(\sigma_j^* = 0,\ \sigma_{-j}^*) > v_i(\sigma_j = 1,\ \sigma_{-j}^*)$

So membership can either be open, exclusive requiring a majority vote or exclusive requiring a unanimous vote. In all three cases a stable coalition structure must satisfy the internal stability condition, meaning that parties of the non-trivial coalition cannot improve their welfare by leaving unilaterally. If membership is open also condition (b) must be satisfied, so no player who is acting as a singleton can increase his payoff by joining. In case of exclusive majority voting membership either condition (b) or condition (c.1) must hold, the latter meaning that if a player who is acting as a singleton would be better off by joining the coalition, then the coalition structure is stable if every player of a subgroup S of players belonging to the non-trivial coalition gain by not letting the newcomer in, where S must comprise at least half of all the players who belong to the non-trivial coalition. Finally if membership is exclusive requiring an unanimous vote either condition (b) or condition (c.2) must hold, the latter meaning that if a newcomer wants to join, all players belonging to the non-trivial coalition must benefit from his accession.

The conditions for the stability of a multiple coalition game are fairly similar. Regardless of the membership assumption a stable coalition structure must satisfy the internal stability condition. The external stability condition is supplanted by the intracoalitional condition which is satisfied when no player can increase his payoff by joining another coalition. As in the previous case under the exclusive membership assumption if a player would benefit from joining another coalition, either the majority or all the players of the target coalition must be better off by the newcomer's accession.

Another possibility tested by the authors are monetary transfers among members of a coalition. So in case of transfers, countries' valuations must be corrected for the amount paid or received:

$$v'_i(c) = v_i(c) + t_i$$

And transfers are calculated using the following equation:

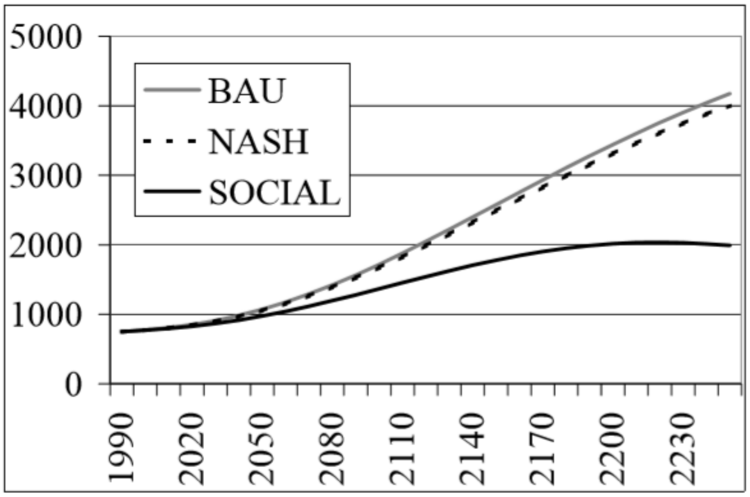
$$t_i = [v_i(c^N) - v_i(c)] + \alpha_i \sum_{i \epsilon c_k} [v_i(c) - v_i(c^N)]$$

Where α_i is the share of discounted marginal damages from climate change of country i over the sum of discounted marginal damages of all members of coalition c^k. The first term of the transfer equation sets every player back to what they would earn in the singleton coalition structure (c^N) and the second term redistributes the difference between the aggregate valuation of coalition's members and the aggregate valuation of coalition's members in the singleton coalition structure to each player weighted by α_i.

The authors point out that in their dataset the superadditivity property is satisfied, so as a newcomer joins a coalition the accession always increases aggregate welfare of insiders. As a newcomer joins a coalition two effects occur: 1) Insiders increase their abatement as they internalize a larger fraction of total emissions 2) Outsiders decrease their abatement level. Since the superadditivity property holds it means that the first effect outweighs the second. The implication of this is that if transfers are allowed, in any coalition structure coalition members are better off with respect to the singleton coalition structure. Note that this does not imply that the grand coalition is stable because even if a larger coalition makes insiders better off it obviously increases the incentives to free ride, too.

To have some benchmarks for the abatement achieved by the stable coalitions the authors show the concentration of CO_2 (in gigatons) resulting under different scenarios: Business-as-usual, Nash equilibrium (singleton coalition structure) and socially optimal outcome (Grand Coalition).

Figure (1)



It can be seen that the grand coalition achieves a significant reduction in emissions with respect to the other two scenarios. However this difference in abatement does not translate itself into a walloping difference in regions' welfare, in fact aggregate world discounted consumption in the socially optimal scenario is only 0.52% higher than in the Nash equilibrium. The reason is that the model does not assign to climate damages a high value with respect to total production and besides much of the benefit from abatement occurs in the distant future, hence it is heavily discounted.

In the following tables the authors show which among the 203 possible coalition structure are stable, in the transfer and in the no-transfer case. The blue rows at the top and at the bottom of the tables denote particular coalition structures: coalition structure No.203 is the grand coalition, the No.1 is the singleton coalition structure and the No.196 represents the signatories of the Kyoto Protocol. Coalition structures are ranked in terms of aggregate welfare produced. The columns (3), (4), (5) refer to the open, exclusive majority voting and exclusive unanimity voting membership assumption respectively, each column in sub-divided into an S and an M-column referring to whether the single or multiple coalition formation assumption is applied. In the grid formed by these columns each 'y' and 'n' denotes whether that coalition structure is stable or not stable respectively. Then columns (6), (8) and (10) show total welfare, concentration of CO_2 (in gigatons) and total emissions respectively. Columns (7), (9) and (11) show for each coalition structure the difference in welfare, CO_2 concentration and emissions with respect to the grand coalition, expressed in percentages. For instance looking at column (6) it can be seen that the singleton coalition structure yields only 0.52% less aggregate welfare than the grand coalition whereas the percentage difference regarding CO_2 concentration and emissions is much higher.

Because the six regions differ widely in the relevant parameters (abatement costs and climate damages) it is interesting to see if similarities among regions are conducive to the formation of a coalition. The authors show that in the no-transfer case, regions with similar climate damage and abatement cost parameters tend to be in the same coalitions. Instead if transfers are allowed regions with different interests can manage to cooperate. For instance it can be seen that Rest of the World often forms a coalition with Japan or the US, this is because ROW suffers high climate damages and bears low abatement costs whereas Japan suffers low damages and bears high costs and the US suffers moderate damages and bears moderate costs, these coalitions are stable because ROW contributes much to joint abatement and enjoys much of the benefit while for Japan and the US the opposite is true.

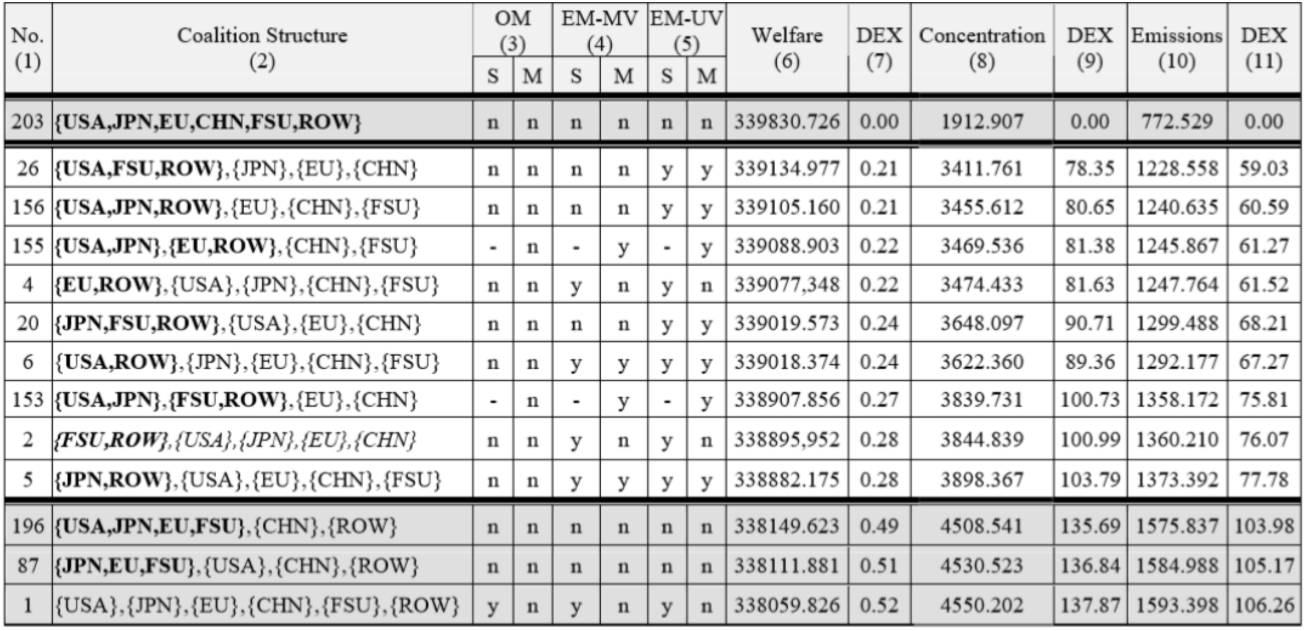
Note that the grand coalition is never stable and that single coalitions in stable coalition structures never have more than three members, confirming the usual finding that stable coalitions tend to be small.

Looking more closely at the diferent specifications regarding the membership assumption, it can be seen that exclusive membership significantly increases the stability of coalition structures especially in the no-transfers scenario. In fact without transfers no coalition is stable if membership is open.

Regarding single versus multiple coalition formation it can be seen that especially with transfers letting regions strike multiple agreements is beneficial, in fact in the transfer case the four coalition structures which gain the highest welfare are all composed of multiple coalitions.

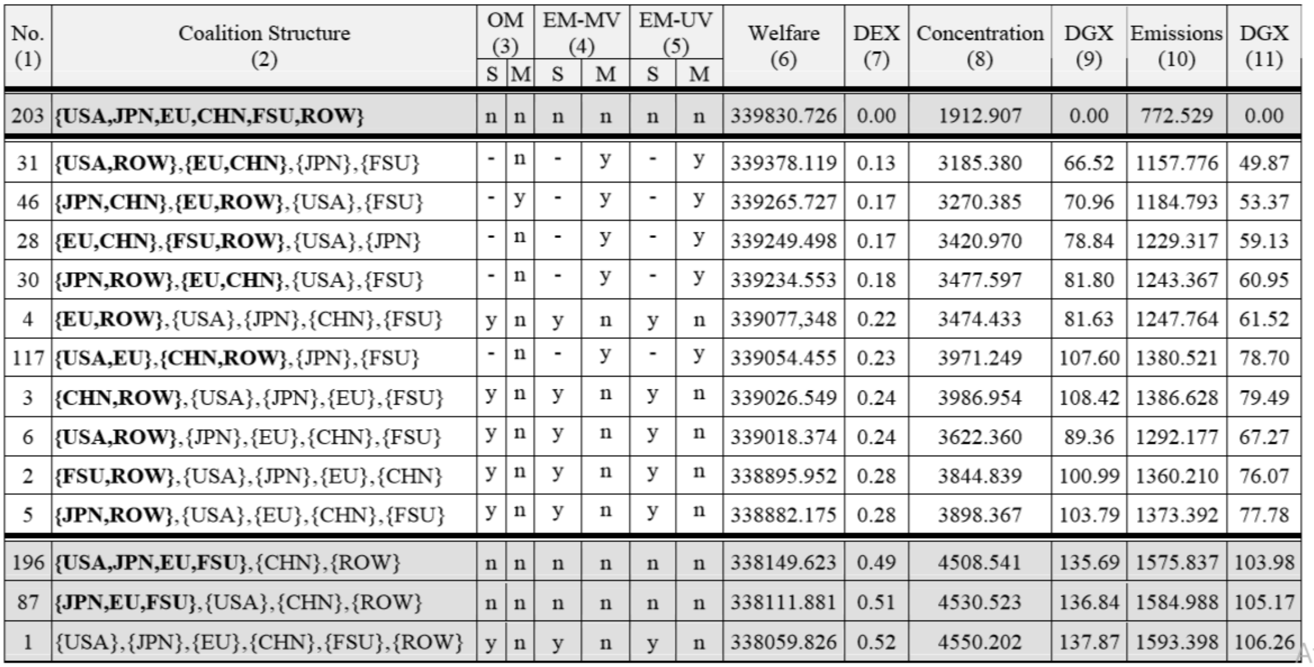
Transfers are clearly welfare-improving, as can be seen comparing the DEX index (column 7) of the most profitable coalition structures in the transfer and no-transfer case.

Figure (2) Stable coalitions without transfers



No. (1)	Coalition Structure (2)	OM (3)		EM-MV (4)		EM-UV (5)		Welfare (6)	DEX (7)	Concentration (8)	DEX (9)	Emissions (10)	DEX (11)
		S	M	S	M	S	M						
203	**{USA,JPN,EU,CHN,FSU,ROW}**	n	n	n	n	n	n	339830.726	0.00	1912.907	0.00	772.529	0.00
26	**{USA,FSU,ROW}**,{JPN},{EU},{CHN}	n	n	n	n	y	y	339134.977	0.21	3411.761	78.35	1228.558	59.03
156	**{USA,JPN,ROW}**,{EU},{CHN},{FSU}	n	n	n	n	y	y	339105.160	0.21	3455.612	80.65	1240.635	60.59
155	**{USA,JPN}**,**{EU,ROW}**,{CHN},{FSU}	-	n	-	y	-	y	339088.903	0.22	3469.536	81.38	1245.867	61.27
4	**{EU,ROW}**,{USA},{JPN},{CHN},{FSU}	n	n	y	n	y	n	339077,348	0.22	3474.433	81.63	1247.764	61.52
20	**{JPN,FSU,ROW}**,{USA},{EU},{CHN}	n	n	n	n	y	y	339019.573	0.24	3648.097	90.71	1299.488	68.21
6	{**USA,ROW**},{JPN},{EU},{CHN},{FSU}	n	n	y	y	y	y	339018.374	0.24	3622.360	89.36	1292.177	67.27
153	**{USA,JPN}**,**{FSU,ROW}**,{EU},{CHN}	-	n	-	y	-	y	338907.856	0.27	3839.731	100.73	1358.172	75.81
2	***{FSU,ROW}***,*{USA},{JPN},{EU},{CHN}*	n	n	y	n	y	n	338895,952	0.28	3844.839	100.99	1360.210	76.07
5	**{JPN,ROW}**,{USA},{EU},{CHN},{FSU}	n	n	y	y	y	y	338882.175	0.28	3898.367	103.79	1373.392	77.78
196	**{USA,JPN,EU,FSU}**,{CHN},{ROW}	n	n	n	n	n	n	338149.623	0.49	4508.541	135.69	1575.837	103.98
87	**{JPN,EU,FSU}**,{USA},{CHN},{ROW}	n	n	n	n	n	n	338111.881	0.51	4530.523	136.84	1584.988	105.17
1	{USA},{JPN},{EU},{CHN},{FSU},{ROW}	y	n	y	n	y	n	338059.826	0.52	4550.202	137.87	1593.398	106.26

Figure (3) Stable coalitions with transfers



No. (1)	Coalition Structure (2)	OM (3)		EM-MV (4)		EM-UV (5)		Welfare (6)	DEX (7)	Concentration (8)	DGX (9)	Emissions (10)	DGX (11)
		S	M	S	M	S	M						
203	**{USA,JPN,EU,CHN,FSU,ROW}**	n	n	n	n	n	n	339830.726	0.00	1912.907	0.00	772.529	0.00
31	**{USA,ROW}**,**{EU,CHN}**,{JPN},{FSU}	-	n	-	y	-	y	339378.119	0.13	3185.380	66.52	1157.776	49.87
46	**{JPN,CHN}**,**{EU,ROW}**,{USA},{FSU}	-	y	-	y	-	y	339265.727	0.17	3270.385	70.96	1184.793	53.37
28	**{EU,CHN}**,**{FSU,ROW}**,{USA},{JPN}	-	n	-	y	-	y	339249.498	0.17	3420.970	78.84	1229.317	59.13
30	**{JPN,ROW}**,**{EU,CHN}**,{USA},{FSU}	-	n	-	y	-	y	339234.553	0.18	3477.597	81.80	1243.367	60.95
4	**{EU,ROW}**,{USA},{JPN},{CHN},{FSU}	y	n	y	n	y	n	339077,348	0.22	3474.433	81.63	1247.764	61.52
117	**{USA,EU}**,**{CHN,ROW}**,{JPN},{FSU}	-	n	-	y	-	y	339054.455	0.23	3971.249	107.60	1380.521	78.70
3	**{CHN,ROW}**,{USA},{JPN},{EU},{FSU}	y	n	y	n	y	n	339026.549	0.24	3986.954	108.42	1386.628	79.49
6	**{USA,ROW}**,{JPN},{EU},{CHN},{FSU}	y	n	y	n	y	n	339018.374	0.24	3622.360	89.36	1292.177	67.27
2	**{FSU,ROW}**,{USA},{JPN},{EU},{CHN}	y	n	y	n	y	n	338895.952	0.28	3844.839	100.99	1360.210	76.07
5	**{JPN,ROW}**,{USA},{EU},{CHN},{FSU}	y	n	y	n	y	n	338882.175	0.28	3898.367	103.79	1373.392	77.78
196	**{USA,JPN,EU,FSU}**,{CHN},{ROW}	n	n	n	n	n	n	338149.623	0.49	4508.541	135.69	1575.837	103.98
87	**{JPN,EU,FSU}**,{USA},{CHN},{ROW}	n	n	n	n	n	n	338111.881	0.51	4530.523	136.84	1584.988	105.17
1	{USA},{JPN},{EU},{CHN},{FSU},{ROW}	y	n	y	n	y	n	338059.826	0.52	4550.202	137.87	1593.398	106.26

These simulations confirm the general conclusions of the previous sections, namely that stable coalitions tend to be small and that transfers are welfare improving if players are heterogeneous. Moreover this paper shows that multiple agreements often lead to better results than single ones and that exclusive membership can significantly improve the stability and the profitability of an agreement.

4.3 CLIMATE CLUBS: OVERCOMING FREE-RIDING IN INTERNATIONAL CLIMATE POLICY

The study of international cooperation presented in this paper (Nordhaus 2015) differs in some major ways with respect to the previous works, namely regarding the stability concept employed and the design of the treaty. The stability concept adopted is the strong Nash equilibrium (or coalition Nash equilibrium) and the major difference in the design of the treaty is the use of external penalties on non-signatories, namely trade tariffs. Besides, negotiations are modeled as a one-shot game and the formation of a single coalition is assumed.

Simulations are based on the integrated assessment model C-DICE where the players of the negotiations game are the 15 regions in which countries are aggregated (US, EU, China, India, Former Soviet Union, Japan, Canada, South Africa, Tropical Africa, Mideast and North Africa, Eurasia, Latin America, Brazil, Middle-income Asia and Rest of the World).

Another difference with respect to the previous works consists in the top-down approach of the climate agreement. The author distinguishes between agreements based on a bottom-up approach and those based on a top-down approach. All the agreements analyzed so far belong to the first category because they assume that the resulting coalition will choose the level of abatement so as to maximize its aggregate payoff, on the other hand in a top-down agreement the level of abatement that signatories must provide is determined from the onset. Namely signatories must set a cost of carbon at least equal to the global cost of carbon, which is defined as the monetary damage caused by one ton of carbon emissions.

As a side note, regarding how signatories may actually ‘set’ the cost of carbon emissions there are two theoretically equivalent ways: a carbon tax or a cap and trade system that allocates carbon emission allowances (or a hybrid version of the two). Although both methods can yield the same outcome, in practice they both have advantages and shortcomings: with a carbon tax it may be difficult to foresee the resulting quantity of emissions because the demand for carbon emissions must be accurately estimated, in a cap-and-trade system the quantity of emissions is fixed but the price of emission allowances may be subject to high volatility if there are shifts in the demand curve. Also in a cap-and-trade system it may be more difficult to agree on the ideal quantity of allowances, because in climate negotiations countries would have to agree both on the total quantity of emissions and on the per-country share and since every country will have incentives to claim a high quantity of the latter, the resulting aggregate quantity of emissions will likely exceed the optimal quantity. Instead if countries were to agree on just one variable (a carbon tax) such problem would not be there (see Nordhaus 2013 for a more thorough discussion).

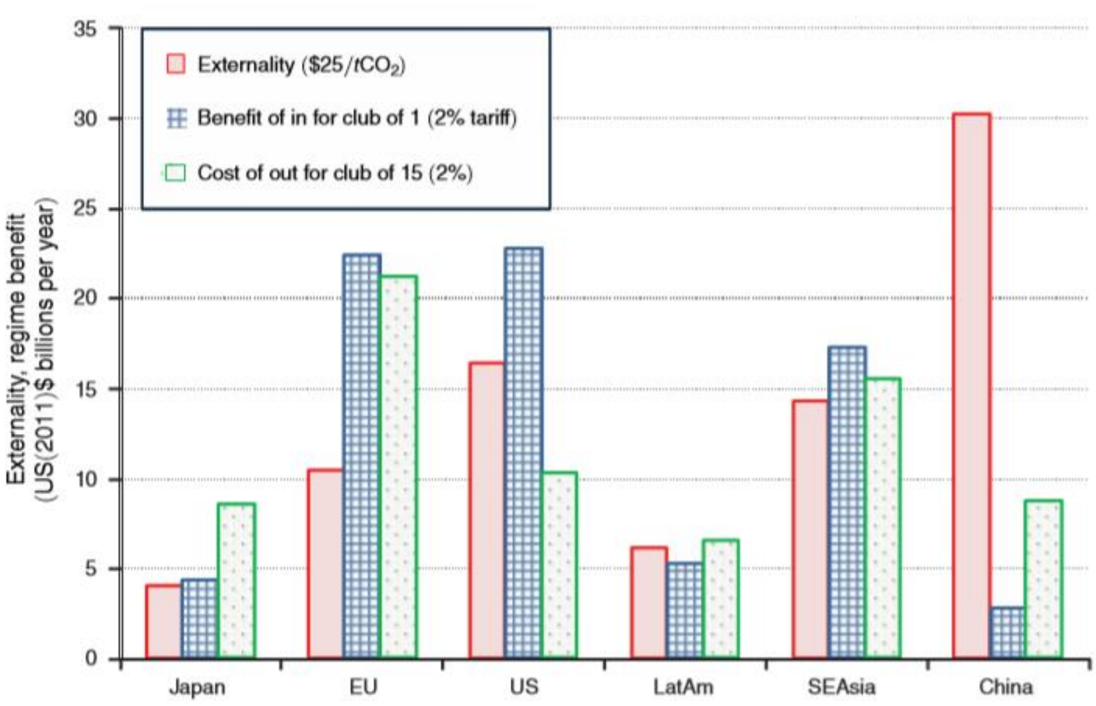
The author shows that assuming that climate damages are proportional to a region’s share of world output then a region’s cost of carbon (τ_i^{NC}) in the non-cooperative scenario is equal to the global cost of carbon (SCC) weighted by the region’s share of output (θ_i):

$$\tau_i^{NC} = \theta_i \, SCC$$

However the above derivation is a simplification with respect to the one performed by the integrated assessment model because countries differ in carbon intensities and damage parameters.

To ensure that countries have the right incentives to cooperate, the agreement prescribes a uniform ad valorem trade tariff on all exports from non-signatories to signatories. The trade tariff consists in an external penalty and it is particularly suitabe to enforce cooperation because it harms the receiver and benefits the sender. However this reasoning is based on a fundamental assumption which is that amendments to international Law would allow for trade tariffs as a means to enforce a climate agreement and retaliation would be prohibited. Granted that this is a major assumption, the author shows that a trade tariff of 2% would produce a cost fairly similar to the negative externality that the non-signatory is imposing on the other regions. Where the externality is calculated as the global cost of carbon minus the region’s cost of carbon times the region’s emissions in the Nash equilibrium minus the region’s emission at the global cost of carbon. The following picture assumes a global cost of carbon of 25 dollar per ton of emissions and it shows that the cost of being a non-signatory when all other 14 regions are part of the club (cost of out) is in the same order of magnitude of the externality produced by the non-signatory (externality), so the the trade tariff is a fairly well-targeted instrument to incentivize cooperation. Besides the picture shows the benefit of being the only signatory with a tariff of 2% (benefit of in).

Figure (4)



Because of the high uncertainty around climate damage, four alternatives for the social cost of carbon have been picked by the author $\{12.5, 25, 50, 100\}$, in each case the social carbon price will be chosen as the price that signatories are required to set (target price). Also several values for the tariff rate have been considered, from 0 to 10%. Combinations of these parameters give rise to 44 regimes. For each one of them the author looked for a stable coalition (according to the strong Nash equilibrium) using an evolutionary algorithm.

Only 6 regimes have been found to not sustain a stable coalition, because they were combinations of high social cost of carbon and low tariff rates. In figure (5) the coloured bars represent the 11 tariff rates, it can be seen that most regimes sustain the maximum number of participants especially for low social costs of carbon and moderate tariff rates. Note that a regime where the tariff rate is zero (leftmost bar) always fails to support a coalition. Figure (6) shows the global average carbon price weighted by each region carbon emissions for each regime. Clearly when participation is full the average and the target price coincide whereas as participation declines the resulting free-riding brings the average carbon price down.

Figure (5)

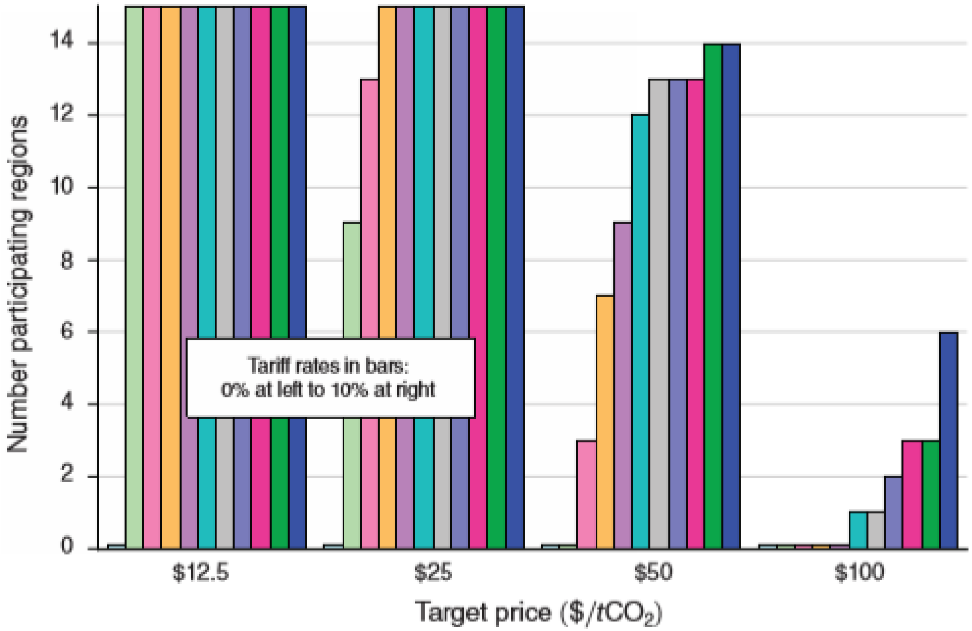
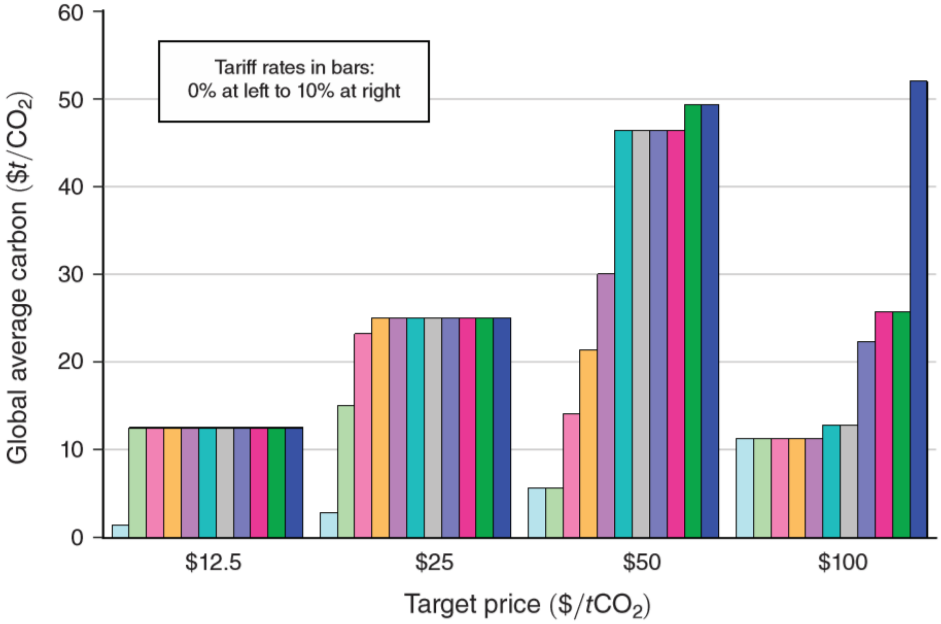


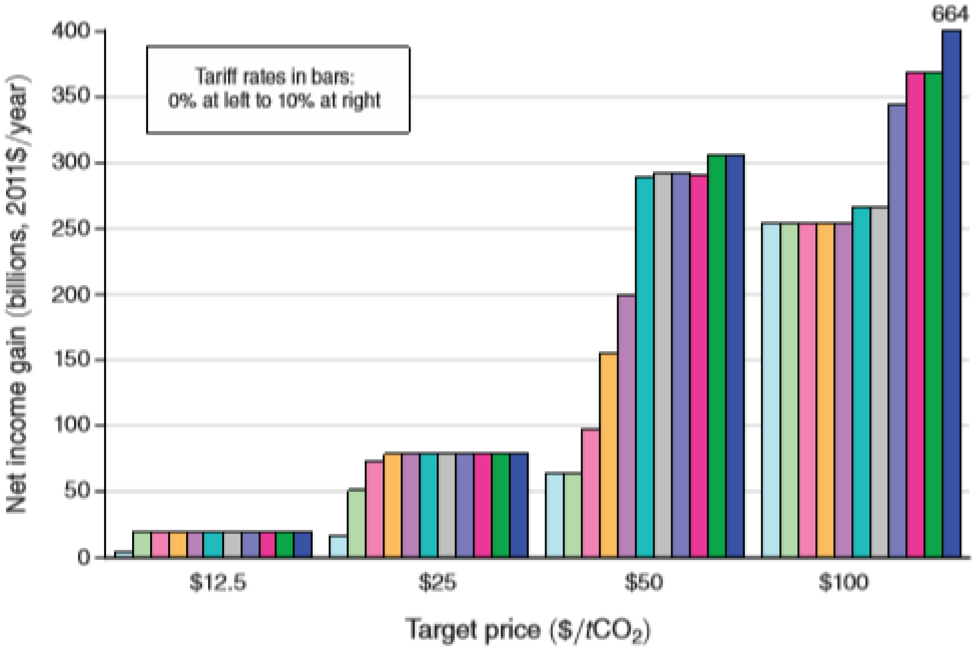
Figure (6)



So it is clear that an external penalty is necessary to provide the right incentives for coordination and this design makes the agreement very successful except for high target price for which evidently the cost of the abatement required (net of the avoided climate damage) exceeds the cost of trade tarrifs.

Regarding the pattern of gains and losses for each region not surprisingly the most important determinants are the region' carbon-intensity of production (i.e. South Africa and Eurasia) and its trade openess, regions that have high values of these parameters must either bear high abatement costs if they join the agreement or they suffer high trade costs otherwise. On the contrary regions that suffer high climate damage (i.e. India) gain in all regimes. Anyways the magnitude of losses is small compared to the gains, indeed no regime produces aggregate losses as can be seen in figure (7) which shows the gain in every regime with respect to zero abatement.

Figure (7)



This paper presents in normative terms an ideal of international cooperation. Indeed the astounding achievements of the agreement envisioned by the author require major changes to international law to allow for trade tariff to stimulate participation. Anyway this proposal is based on a sound argument, which is that a trade tariff is an appropriate instrument to recoup the externality cost that non-signatories impose on all the other regions. The results of the simulations show that trade tariffs are essential and they make it possible to reach full cooperation for most estimates of the social cost of carbon even considering the strict stability concept adopted.

5. CONCLUSIONS

In this thesis I have explored the inherent difficulties that prevent countries from cooperating in the provision of an international public good and what is the best way to design an international agreement in order to overcome such difficulties and harness the great benefits produced by cooperation.

In section 2, I have showed that many istances of environvental problems can only be solved by providing a public good and because countries act strategically, the natural outcome is a Nash equilibrium in which the quantity provided of the public good falls short of the optimal level. Given this starting point I showed that an international agreement can improve on this outcome, formally this has been demostrated by modeling countries decisions as a multi-stage game in which countries had the opportunity to form a coalition and subsequently make their abatement decision. I then stressed how this result relies on the assumptions made, namely the symmetry of the players, the stability concept employed and the assumption about individual and collective rationality. I especially focused on this latter assumption showing how a modification in the collective rationality assumption can make it possible to achieve a full-participation agreement. However to make an agreement of this type self-enforcing, the amount of public good provided by each signatory had to be within an interval so as to reduce the incetives to free-ride and make the threat of punishment for a defector credible.

In section 3, I considered alternative assumptions to offer a more realist analysis of international cooperation. I considered again the models presented in the previous section under the assumption of quadratic abatement costs. This new specification offered interesting results especially in the strong collective rationality model, under this assumption the scope for cooperation was severely undermined in fact it was shown that only agreements of maximum three signatories were stable at the equilibrium. Besides I considered the effect of an extended strategy space on cooperation, namely money transfers to induce participation and I showed that they have a very positive effect on treaty participation provided that countries are asymmetric.

In the last section I turned my attention to empirical studies of international cooperation which made use of Integrated Assessment Models to model countries' consumption, investment and abatement decisions and to highlight the interrelatedness between economic growth and the climate. The two papers that I presented confirmed the general finding of the previous sections, that is that only small coalitions tend to be stable and transfers are beneficial to cooperation. Moreover these papers also offered new indications regarding what design features can enhance the effectiveness of an agreement, the first one showed that exclusive membership agreenments fare significantly better than open membership ones, while the second paper showed that the use of external penalties (trade tariffs) is highly effective in fostering cooperation.

6. BIBLIOGRAPHY

- **Barrett, Scott**. 1994. *"Self-Enforcing International Environmental Agreements."* Oxford Economic Papers

- **Barrett, Scott**. 2001. *"International Cooperation for sale."* European Economic Review

- **Barrett, Scott.** 2002. *" Consensus Treaties."* Journal of institutions and Theoretical Economics

- **Barrett, Scott**. 2003.*"Environment and Statecraft: The Strategy of Environmental Treaty-Making*." Oxford and New York: Oxford University Press.

- **Copeland, B. and Taylor, S**. 2000. *" Free Trade and Global Warming: a Trade Theory View of the Kioto Protocol. "* Mimeo, Department of Economics, University of British Columbia

- **Eyckmans, J. and Finus, M.** 2003. *"New Roads to International Environmental Agreements: The Case of Global Warming"* Center For Economic Studies

- **Friedman, J.** 1971 *" A Noncooperative equilibrium for Supergames."* Review of Economic Studies

- **Nordhaus, W and Stztorc, P.** 2013 *" Dice 2013R Introduction and User's Manual"*

- **Nordhaus, W**. 2015 *"Climate Clubs: Overcoming Free Riding in International Climate Policy."* American Economic Review

- **Weitzman, M**. 2014. "*Can Negotiating a Uniform Carbon Price Help to Internalize the Global Warming Externality?*" Journal of the Association of Environmental and Resource Economists