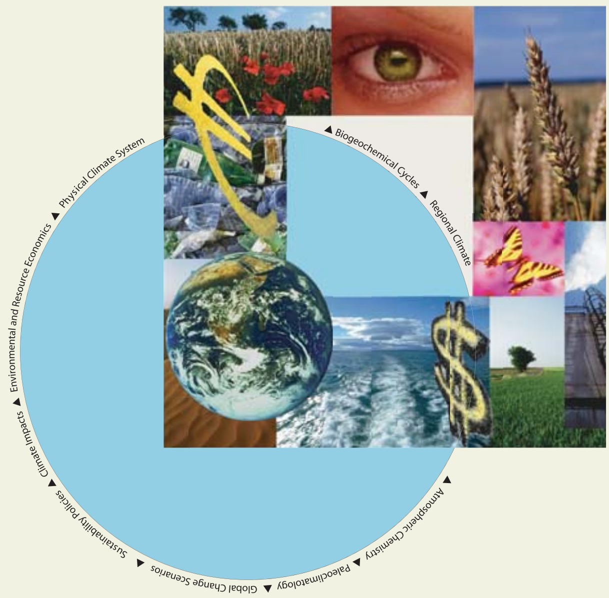
International Max Planck Research School on Earth System Modelling

Stability and Burden Sharing Emissions in International Environmental Agreements A Game-Theoretic Approach

Dritan Osmani

PhD Thesis prepared within the International Max Planck Research School on Earth System Modelling



Outline of Thesis

The body of literature on International Environmental Agreements (IEA) has two conflicting views. One is based on cooperative game theory and concludes that the grand coalition is stable, by assuming transferable utility, then using the γ-core concept and implementing transfers to solve the heterogeneity of the countries involved (Chander & Tulkens 1997, 2006, Eyckmans & Tulkens 2003, Chander 2007). This represents an optimistic view of the possibility of international cooperation on solving global environmental problems. The other view is rooted in the non-cooperative game theory and became the dominant path in the literature (Barrett 1994, 2003, Botteon & Carraro 2001, Osmani & Tol 2005, Finus.etal 2006, Rubio & Ulph 2006, McGinty 2007). The usual approach of non-cooperative game theory to stable IEAs is based on the idea developed for cartel stability (d'Aspremont et al. (1983)) and requires so-called internal and external stability. Internal stability means that a country does not have an incentive to leave the coalition, while external stability means that a country does not have an incentive to join the coalition. This part of the literature reaches the conclusion that the size of a stable coalition is typically very small, thus representing a pessimistic view of global environmental goods[1].
Papers from one to five perform a game theoretic approach, while paper six use welfare analysis.
Firstly, in our game theoretic approach, a link between the economic activity and the physical environment is established which is generated from the integrated assessment model (except first paper which uses stylized social welfare functions). The social welfare function captures the difference between the profit from pollution and the environmental damage. Following this approach, players (countries) play a two stage-game. In the first stage, each country decides whether to join the coalition and become a signatory (or coalition member) or stay singleton and non-signatory (*membership game*). In the second stage, every country decides on emissions (*strategic game*). Within the coalition, players play cooperatively (by maximizing their joint welfare) while the coalition and single countries compete in a non cooperative way (by maximizing their own welfare).

In first paper we use non-cooperative game theory in order to develop further a model from Barrett (1994a). The main assumptions of the model are:

- all countries are identical,
- each country's net benefit function is known and known to be known etc. by all countries
- pollution abatement is the only policy instrument,
- costs are independent of one another.

Being aware of the work on coalition theory by Ray and Vohra (1994), Yi and Shin (1995) and Bloch (1996, 1997) we think that modelling two self-enforcing IEA (employing the stability concept of d'Aspremont et al. (1983)) can bring a better understanding of improving capacity of IEA's. We are less concerned with developing a general theory of coalition formation. Rather, we present and apply a method for computing the maximum size of two coalition. The loss in generality is compensated by a gain in practically. The main contribution of this paper is the discussion on the *possibility of improving capability* (size and emission reduction) of two self-enforcing IEA compared to one self-enforcing IEA by modelling the IEA as a one-shot game. Another contribution is *a different formulation (as nonlinear optimization problem)* of finding α ($\alpha N =$ the number of signatories) in extended Barrett's model. Although our work is less general than that of Yi and Shin, Bloch etc. we are actually able to compute the coalition sizes and optimal abatement levels. We would like to stress that we reinforce the conclusions of Asheim et. al

[1]The references for this paragraph are found in the second paper. Other references can be found in the bibliography of paper which they introduce.

(2006) and Carraro (2000) by following a different method, that is nonlinear optimization. Our first paper considers stylized cost-benefit functions and identical countries which are the main drawbacks of this research. These drawbacks are overcome in the second paper where cost-benefit functions from FUND (developed by Richard Tol) are used, and different world regions are considered (thus not identical).

In the second paper, the IEAs are analyzed by using the farsighted stability concept within the framework of mixed non-cooperative and cooperative game theory. We investigate what outcomes are stable, which implies that they cannot be replaced by any coalition of rational, farsighted and selfish countries. The selfishness of players shapes the aspects of *non-cooperative approach.* The idea of farsightedness means that one should check for multi-step stability by comparing the profits of a coalition member after a series of deviations has come to an end. The deviation is possible only if players (regions) display *cooperate attitude (by forming a coalition)* to each-other in order to increase their welfare.
Farsighted stability developed further the notation of stable sets of von Neumann & Morgenstern (1947). Stable sets are defined to be self-consistent. The notion is characterized by internal and external stability. Internal stability guarantees that the solution set is free from inner contradictions, that is, any two outcomes in the solution set cannot dominate each other and external stability guarantees that every outcome excluded from the solution is accounted for, that is, it is dominated by some outcome inside the solution. Harsanyi (1974) criticizes the von Neumann and Morgenstern solution also for its failing to incorporate foresight. He introduced the concept of indirect dominance to capture foresight. An outcome indirectly dominates another, if there exists a sequence of outcomes starting from the dominated outcome and leading to the dominating one, and at each stage of the sequence the group of players required to enact the inducement prefers the final outcome to its status quo. His criticism inspired a series of works on abstract environments including among others those of Chwe (1994), Mariotti (1997) and Xue (1998). Chwe (1994) introduces the notation of farsighted stability which is applied to the problem of IEAs by Diamantoudi & Sartzetakis (2002) and by Eyckmans (2003). Diamantoudi & Sartzetakis (2002) consider identical countries while asymmetric countries are taken into account in our model. Eyckmans (2003) studies only single farsightedly stable coalitions while we allow multiple farsightedly stable coalitions. In addition, a more systematic way of finding farsightedly stable coalitions is introduced in our approach (as we have 16 different world regions, Eyckmans consider only 5 world regions).
The welfare functions of sixteen world regions are taken from the Climate Framework for Uncertainty, Negotiation and Distribution model FUND (see Section 2.2).
Solving farsighted stability by using combinatorial algorithms, we find all farsightedly stable coalitions. We show that by applying FUND, two farsightedly stable coalitions can be formed. The number of regions in both coalitions is around two thirds of all regions, and they improve welfare and abatement levels significantly. This is one of the few relatively optimistic results of non-cooperative game theory. Another contribution of this paper is the further extension of the farsighted stability concept to preferred farsighted stability. The preferred farsightedly stable coalition is a farsightedly stable coalition where the majority of country members reach higher profits in comparison to any other farsightedly stable coalition[2].

The third paper presents a detailed discussion and comparison of D'Aspremont stability and farsighted stability. We show that the D'Aspremont stable coalitions are often sub-coalitions of farsightedly stable coalitions. Besides, farsightedly stable coalitions can be frequently the largest size coalition. Moreover, they create always the biggest improvement in environmental and welfare that game theory without side payments can realize. Similarly to preferred farsightedly stable coalitions, we introduce preferred D'Aspremont stable coalitions. All D'Aspremont stable coalitions are found and multiple D'Aspremont coalitions are compared with multiple farsighted ones.

The fourth paper develops further our previous work on farsightedly stable coalitions and preferred farsightedly stable coalitions. In our second paper, we show that multiple preferred farsightedly stable coalitions include two thirds of countries and improve significantly the welfare and environment which are optimistic results. Here we extend the discussion on the issue, which farsightedly stable coalitions are more likely to form in different time horizons and how much improvements they bring to welfare and environment. We raise the question if the farsightedly stable coalitions can be maintained for a long-term period. The improvements in welfare and abatement levels of full non-cooperative behavior (atom structure) and grand coalition are considered also.

[2]We consider only *economic incentives* that a region has to join a coalition for environmental protection. Other factors like commitment to cooperation are not taken into account.

The main contribution of the fifth paper is the discussion on the assumption of *joint welfare maximization.* As the members of coalition *play cooperatively* by maximizing their joint welfare, we compare the joint welfare maximization with classical cooperative game theory value such as Shapley Value, Nash bargaining solution and Consensus Value.

The sixth paper addresses the distributional issue which is the cumbersome point of benefit-cost analysis. As per capita income is lower in poor countries, then willingness-to-pay based estimates of damages in poor countries are lower than in the developed countries even though the impact is identical in human, physical or ecological terms. One way of managing this would be to use a normative approach by introducing weight factors based on the different marginal value of money in the different regions of the world. This would give higher weight to costs in the poor countries. Environmental equity can be understood as assuming a new decision criterium that requires that the value of lost lives (also any environmental goods) in rich and poor countries has to be weighted differently. The normative approach is not realistic as it does not respect the WTP estimates which are the base of economic valuation. However, we would like to experiment in order to test the Kyoto emissions reduction targets requiring that the value of life is identical in poor and rich countries.
There are different views in favor and against of using weight factors. We do not plan to review this discussion. We simply assume that weight factors are considered appropriate from a normative point of view, and then examine if the Kyoto emissions reduction targets are consistent with the requirement of valuing the life in poor and developed countries by weighting them differently.
Following Ray (1984) and Stenman (2000) we obtain the equity weights by totally differentiating the social welfare function. The social welfare function depends on three parameters, which are the incomes per capita, the elasticity of marginal utility e and the inequality aversion parameter γ. The incomes per capita depend on GDP, population and costs and benefits from pollution abatement which are obtained from two different integrated assessment models namely the Climate Framework for Uncertainty, Negotiation and Distribution (FUND) model developed by Richard Tol, and the MITs Emissions Prediction and Policy Analysis (EPPA) model developed at MIT. The essential point of our paper is the development of a relation between e and γ by equalizing the value of life in poor and rich regions following Fankhauser et al. (1997) and Stenman (2000), but allowing for larger range of parameters values that relates e and γ. As e and γ take their values in specific intervals, there is a significant advantage to have a relation between them, as it is possible to restrict the intervals for e and γ when the world regions are approximated in only two types, namely poor and rich. That is, when the world social welfare for different e and γ is maximized and pollution abatements levels are found, one focuses on smaller intervals for e and γ which are going to give smaller variation for abatement levels. Finally it is possible to test if the abatement targets of Kyoto protocol respect the condition that the value of life is identical in poor and rich countries when equity weights are used. The costs and benefits from pollution abatements are calculated for the year 2010, which is considered as the representative year of the first commitment period of Kyoto protocol that includes the years 2008 through 2012.

Papers are introduced consecutively, first paper first chapter, second paper second chapter and so on. The last chapter provides the conclusions.