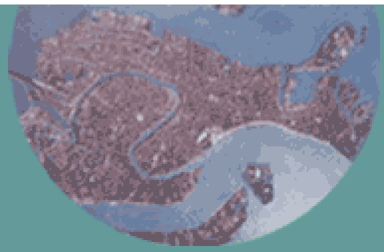
EAERE FEEM VIU
European Summer School
In Resources and Environmental Economics
POLITICAL ECONOMY OF THE ENVIRONMENT
September, 1st - 6th, 2003



Emission trading and the stability of environmental agreements

Author:

Gergely Ujhelyi

Harvard University

USA

Emission Trading and the Stability of Environmental Agreements[1]

Gergely Ujhelyi
Department of Economics, Harvard University
ujhelyi@fas.harvard.edu

April 29, 2003

[1]The first version of this paper was my MA Thesis at the Department of Economics, Central European University. I would like to thank participants of the Harvard Theory Lunch and the EEPHU student seminar for comments.

Abstract

This paper is a first step in analysing how a market for tradeable permits created by an environmental agreement affects the stability of the agreement. I present a model in which the stability implications of (a) the market structure of the permit market and (b) the principle of initial allocation of permits can be analysed. I show that with identical players and an allocation giving every player an equal amount of permits, the market structure has no effect on the stability of the agreement, while for a given market structure this allocation makes the agreement more stable than any other allocation.

1 Introduction

Environmental agreements and increasing reliance on artificial markets are two striking features of modern environmental policy. In the twentieth century, more than 150 international environmental conventions, protocols, treaties and amendments were signed, almost all of them after 1950 [4], [16]. Artificial property rights markets are implemented in a number of domestic regulatory systems, and an international market for emission permits is currently being developed under the 1997 *Kyoto Protocol* of the UN's *Framework Convention on Climate Change.* To a certain extent, national environmental policy can also be viewed as the outcome of agreements between the various interest groups making up society.[1] Artificial markets are established by agreements: creating them requires some cooperation between future trading partners.

In order for environmental policy to be viable, the agreement creating it must be stable, e.g., countries must not have the incentive to withdraw from an international treaty. Such incentives are of course shaped by the characteristics of the market itself. Many observers noted the crucial role the incorporation of emission trading played in the acceptance of the Kyoto Protocol [9], [12].[2] The payoffs countries receive from the Protocol, and thereby the incentives for new members to join and existing members to leave it will also depend on the specific characteristics of the market, such as the method of initial allocation of permits, the trading regime, the market structure or the liability rule. Similarly, there is a considerable amount of evidence on how various characteristics of domestic permit trading systems (initial allocation of permits, above all) influence the political acceptability of such systems. (See [10] on the RECLAIM program in the Los Angeles basin, [11] and [13] on sulfur trading under the Clean Air Act, and [19] on the failure of emission trading in the UK.) In setting up such markets, attention must therefore be paid to their stability implications. C. Carraro and D. Siniscalco write: "The various economic instruments needed to implement an agreement...must also be designed to promote the stability of the agreements." ([3], p327). This requires an understanding of how various economic instruments affect the stability of the agreement. The present study is a first step in this direction.

[1] Such a political economy perspective on the development of domestic environmental regulation is emphasised in many studies. See [1], [8], [11], [13], and [14] for different approaches.

[2] One observer notes that the U.S. probably signed the agreement *because* it included emission trading, while many of the details of the system were left to be worked out at a later stage in order to achieve a compromise with China and India, who initially thtreatened to veto any agreement *with* trading [9].

This paper investigates how various characteristics of a market for emission permits affect the stability of the environmental agreement establishing the market. I develop a model to examine how (a) the market structure, and (b) the principle of initial allocation affects the stability of an agreement. The permit market is modeled using a cooperative bargaining game to allocate the permits and a social planner to decide on their optimal total quantity [22]. Interaction of signatories and non-signatories follows the Partial Agreement Nash Equilibrium framework [5], in which all the players choose their emissions simultaneously, with signatories acting as a group (represented by the social planner) and non-signatories setting their pollution levels as singletons. Stability is defined in a shortsighted manner [2], [3], which assumes that if a player leaves of joins a particular agreement, the new agreement remains formed. Stability implications of market structure and initial allocation are investigated in a special case of the model, where players are assumed to be identical. I show that in this model, once an initial allocation equalising players' profits is fixed, varying the market structure (perfect competition or market with transaction costs) has no effect on the stability of the agreement. In contrast, if one takes market structure as given and varies the principle of allocation, the agreement becomes more or less stable. Specifically, the allocation in which permits are distributed equally yields agreements that are more stable than any other agreement, both for a perfect market, and for a market with transaction costs.

The remainder of the paper is organised as follows. Section 2 presents the model. I first describe the standard way of modeling environmental agreements, then take a specific model, due to S. Barrett [2], to illustrate how this standard model works and where it can be modified. In Section 2.3, I present the modified general model, which includes emission trading. This model is solved in Section 2.4. Section 3 contains the analysis of stability. I first show how market structure affects stability, then turn to the stability consequences of varying the allocation principle. Section 4 concludes.

2 Modeling Environmental Agreements

In this section I first present the standard model of an environmental agreement. I determine which part of the standard approach is extended by the present paper. Section 2.2 presents an example illustrating the standard model and its extension. Finally, I describe and solve the modified model.

2.1 The Standard Model

The standard model of an environmental agreement involves n players (e.g., firms or countries), who emit a global pollutant. The quantity emitted by player i is denoted q_i. Each player incurs damages from total emissions, and costs (such as abatement costs) that are related to his own emissions only. The damage facing player i is $d_i(Q)$, where $Q = \sum_{i=1}^{n} q_i$ with d_i strictly increasing and strictly convex. The cost function, $c_i(q_i)$, is assumed to be strictly decreasing and strictly convex: less pollution (more abatement) costs increasingly more. It is also assumed that $d_i'' + c_i'' > 0$ over the relevant range. In the absence of agreement, players behave according to the non-cooperative Nash model. That is, they maximise their profit, π_i, taking emission of everyone else as given:

$$\max_{q_i} \{\pi_i = -d_i(Q) - c_i(q_i)\} \tag{1}$$

The necessary and sufficient first order condition for player i's maximum is:

$$\frac{\partial d_i(Q)}{\partial q_i} = -\frac{\partial c_i(q_i)}{\partial q_i}. \tag{2}$$

The n equations for $i = 1, ..., n$ determine the Nash equilibrium emission levels. These unregulated equilibrium emissions will be denoted q_i^u. Players face the standard externality problem. They could collectively be made better off by choosing $q_1^{opt}, ..., q_n^{opt}$ so as to solve:

$$\max_{\{q_j\}_{j=1}^{n}} \sum_j \left[-d_j(Q) - c_j(q_j)\right], \tag{3}$$

which would yield

$$\sum_j \frac{\partial d_j(Q^{opt})}{\partial q_i} = -\frac{\partial c_i(q_i^{opt})}{\partial q_i} \tag{4}$$

for all i. Clearly $q_i^{opt} < q_i^u$: unregulated pollution is socially excessive. Moreover, players have an incentive to free-ride: given that everyone else chooses q_i^{opt}, player i' is better off polluting more.

To turn the above set-up into a public good problem, the key variable of the model is sometimes taken to be the level of abatement, y_i, rather than the level of pollution. A status quo level of pollution is then assumed implicitly, and we have $y_i \equiv \mathit{status\ quo} - q_i$. When I use this setup, the status quo will conveniently be chosen to equal the non-cooperative equilibrium pollution q_i^u. Total abatement is then denoted by $Y = \sum_{i=1}^{n} y_i$. In this approach a comparison of the first order conditions equivalent to (2) and (4) would show that the Nash equilibrium in abatement levels yields socially suboptimal abatement, and players have an incentive to free-ride on others' abatement efforts.

Although full cooperation as in (3) would be socially optimal, as will be shown below, partial cooperation can also make society better off compared to the Nash outcome. With partial cooperation, a subset of the n players agree to coordinate their actions. I will denote by S (for "Signatories") the set of cooperating players, which I call an *agreement*, and N (for "Non-signatories") the set of all remaining players. The crucial requirement is that this agreement be such that no player has an incentive to deviate from it. We have seen above that, in this sense, full cooperation is not *stable*. Specifically, denoting player i's profit by π_i^S if he cooperates and π_i^N if he does not, stability requires that

$$\pi_i^S(\mathbf{x}) \geq \pi_i^N(\mathbf{x}) \text{ for all } i \in S \tag{5}$$

and

$$\pi_i^N(\mathbf{x}) \geq \pi_i^S(\mathbf{x}) \text{ for all } i \in N, \tag{6}$$

where $\mathbf{x}$ is a vector of variables - exogenous to the model above - which affect players' profits (these need not be the same for all players: what variables are taken into account depends on the form of the functions π_i). Equation (5) requires that every signatory should be at least as well off as she would be if she left the agreement. This is sometimes called *internal stability.* Similarly, equation (6) requires that every non-signatory be as least as well off as she would be if she joined the agreement. This is referred to as *external stability.* The actual stability concept used depends on the specification of the $\pi_i(\mathbf{x})$-s. A key variable in $\mathbf{x}$ is S itself: the identity (or at least the number) of cooperating players. For example, when $|S| = n$, players profits are determined from (3), whereas with $S = \{\emptyset\}$, profits are calculated from (1). In addition, $\pi_i(\mathbf{x})$ depends on at least three things: the dynamic consequences of players' actions that are taken into account, the interaction of S and N, and the nature of cooperation within S.

1. *Dynamic consequences.* When deciding whether to join S or remain in N, players have to consider what the consequences of such a move will be. Carraro and Siniscalco [3] assume that if a member of S exits the coalition, the remaining players continue to cooperate. Chwe [6] uses the concept of "farsighted coalitional stability", and argues that players should take into account all the dynamic consequences of their action. Thus, if player i's exiting S gives all other players an incentive to exit as well, then player i's relevant π_i^N corresponds to his Nash equilibrium profit.

2. *Interaction.* One has to specify how players in S take into account actions of players in N, and vice-versa, when actions are determined. Barrett [2] assumes a Stackelberg-type game, in which the coalition S is a leader and moves before N. Chander and Tulkens [5] in contrast take a Cournot setup, in which S and N are constrained to determine their pollution levels simultaneously. In both studies only one agreement, S, can form, and N remains disintegrated. The alternative approach is to assume that multiple coalitions can form, as in [18].

3. *The nature of cooperation.* The third issue which needs to be specified is what exactly is meant by "cooperation", i.e. how the pollution levels q_i (or abatement levels y_i) are determined for $i \in S$. In this respect the literature is surprisingly unified: cooperation is always assumed to mean that individual actions are chosen so as to maximise the sum of signatories' profits. Signatories thus take collective action by solving (3) with $j \in S$.

In contrast to previous literature, the focus of the present paper is the third issue above, the nature of cooperation. In practice pollution levels that maximise the sum of signatories' profits seem hard to determine directly. Emission trading is an indirect means of implementing this collective optimum.

2.2 An Example

To illustrate the standard model and motivate its modification, in this section I consider the example proposed by S. Barrett [2] on the decision of countries whether or not to join an environmental agreement. I first describe the original example, then extend it to include some of the concepts put forward in the present paper.

There are n identical countries, with quadratic benefit and cost functions in which abatement (y) rather than pollution is the key variable. Profit of country i is:

$$\pi_i = \frac{b}{n}\left(aY - \frac{Y^2}{2}\right) - \frac{c}{2}y_i^2,$$

where a, b and c are positive parameters. Since in Barrett's model countries are identical, the coalition structure is conveniently described with the help of a parameter α denoting the fraction of the n countries who decide to sign an environmental agreement. Thus, we have $|S| = \alpha n$ and $|N| = (1-\alpha)n$. As discussed in the previous section, solving the model requires three sets of assumptions. In Barrett's model, these are as follows.

The nature of cooperation. Signatories are assumed to maximise the sum of their profits:

$$\max_{\{y_i\}_{i\in S}} \sum_{i\in S}\left[\frac{b}{n}\left(aY - \frac{Y^2}{2}\right) - \frac{c}{2}y_i^2\right] \tag{7}$$

Since countries are identical, the optimum yields $y_i = y^S$ for $\forall i \in S$, and we can write $Y^S = \alpha n y^S$ to denote total abatement by signatories.

Interaction. A Stackelberg-type equilibrium is considered in which countries in S are allowed to move before the remaining $(1-\alpha)n$ countries. Non-signatories act as singletons, maximising their individual profits as in the non-cooperative case:

$$\max_{y_i} b\left(aY - \frac{Y^2}{2}\right) - \frac{c}{2}y_i^2 \tag{8}$$

for $i \in N$. Problem (8) yields a reaction function for each non-signatory. Since countries are identical, in equilibrium we will have $y_i = y^N$ for all $i \in N$. Thus non-signatories' aggregate abatement is $Y^N = (1-\alpha)ny^N$. Using the reaction functions from the solution of problem (8), we can solve the signatories' collective problem (7). The equilibrium values of Y^S and Y^N are obtained in the standard Stackelberg fashion, yielding the profits $\pi^S(\alpha)$ and $\pi^N(\alpha)$ for signatories and non-signatories respectively. Table 1 displays simulation results for $n = 10$, $a = 100$, $b = 1$ and $c = 0.25$ for different values of α. Notice that increasing the share of cooperating countries steadily increases total abatement.

Dynamic consequences. The last ingredient needed to get a concept of stability determining α is a specification of what dynamic consequences are taken into account by the player who decides whether or not to cooperate. Barrett chooses the simple shortsighted concept. Every country i is assumed to make a comparison between its profit given the current number of signatories - $\pi^S(\alpha)$ for a signatory ($\pi^N(\alpha)$ for a non-signatory) -, and what it would

α	Y^S	Y^N	Y	π^S	π^N
0	-	80.0	80.0	-	472.0
0.1	1.9	76.8	78.7	476.8	468.1
0.2	8.3	69.9	78.2	474.0	466.6
0.3	20.0	59.0	78.9	472.3	468.9
0.4	**35.6**	**45.4**	**81.1**	**472.2**	**474.9**
0.5	52.6	31.6	84.2	473.7	482.5
0.6	68.1	19.7	87.7	476.4	489.4
0.7	80.2	10.8	91.0	479.5	494.3
0.8	88.8	5.0	93.8	482.7	497.3
0.9	94.3	1.6	95.9	485.4	498.8
1	97.6	-	97.6	487.8	-

Source: [2], Table 1, p883.

Table 1: Equilibrium abatement levels and profits

get if it moved in the other group, i.e. withdrew from (joined) the agreement, given that the coalition $S \setminus \{i\}$ ($S \cup \{i\}$) continues to cooperate. Country i's new profit would be $\pi^N(\alpha - 1/n)$ if the country left and $\pi^S(\alpha + 1/n)$ if the country joined the agreement. Stability, as in (5) and (6), requires that α be such that no country has an incentive to make this change, i.e.

$$\begin{aligned} \pi^S(\alpha) &\geq \pi^N(\alpha - 1/n) \\ \pi^N(\alpha) &\geq \pi^S(\alpha + 1/n). \end{aligned}$$

For the simulation in Table 1, these conditions are only satisfied for $\alpha = 0.4$. An agreement consisting of 4 countries out of 10 is therefore the only stable coalition in this example. Barrett's important conclusion is that cooperation in international environmental policy may indeed be incomplete, involving only a subset of all countries.

To start the type of analysis performed in the present paper, I modify the assumptions regarding the nature of cooperation. Suppose that signatories of the agreement do not simply allocate pollution abatement efforts among themselves. Instead, once the optimal total quantity Y^S has been specified, pollution permits are given to the players, who are free to trade these on a market involving the signatory countries. Let y_i^0 denote the abatement level required by the initial endowment of permits allocated to country $i \in S$. Assuming perfect compliance, a country has to buy permits if its abatement level is below what is required. Its profit therefore becomes

$$\pi_i = \frac{b}{n}\left(aY - \frac{Y^2}{2}\right) - \frac{c}{2}y_i^2 - p\left(y_i^0 - y_i\right),$$

where p is the market price of permits. Assuming that the permit market is perfectly competitive, a well known result states that abatements y_i will be unaffected by the initial allocation [17]. The market price p is therefore also unaffected, and the initial allocations simply represent transfers between the signatory countries.[3] Those getting $y_i^0 > y_i$ subsidise those who receive $y_i^0 < y_i$.

Barrett's example can be interpreted as corresponding to the special case when $y_i^0 = y_i$ for all i. As countries are assumed to be identical, this allocation gives equal profits to every signatory. An allocation with this property is called Equal Gains Allocation (EGAL) ([22], see a precise definition below). Notice that here EGAL coincides with most bargaining solutions - with identical countries the Nash solution or the Kalai-Smorodinsky solution would also result in $y_i^0 = y_i$.

Suppose however that for some reason initial allocation does not follow the EGAL principle. For example, consider the case when whoever is last in joining the agreement is allocated a quantity of permits which requires him to abate twice as much as the other members of the coalition. Although this allocation is quite arbitrary, it will serve to illustrate the basic point made in this paper: stability of environmental agreements is affected by the nature of cooperation created by these agreements. In Table 2 the last, disadvantaged member of the coalition is indexed by d (parameter values were chosen as before). When a non-signatory is considering whether to join the agreement, $\pi_d^S(\alpha + 1/n)$ is the profit he takes into account. On the other hand, both disadvantaged signatories (with profit $\pi_d^S(\alpha)$) and other signatories (with profits $\pi^S(\alpha)$) have to consider whether to exit from the coalition, and earn a profit equal to $\pi^S(\alpha - 1/n)$. As can be seen from the table, in such a framework $\alpha = 0.3$ yields the only stable agreement. Because of the unequal treatment he would receive relative to others, the fourth player has no incentive to join S. The size of the stable agreement has decreased by one player compared to the case when permits were allocated following the EGAL principle.

As this simple example illustrates, stability of the agreements is influenced by the institutions these agreements create. With an international permit trading system, different rules of initial allocation might lead to stable agreements of different size. This point will be further elaborated upon later in this paper.

Another point worth mentioning in relation to Barrett's model concerns the interaction of signatories and non-signatories. Barrett assumes that sig-

[3] Notice that the optimal total quantity is also independent from the initial allocation. This is the Coase theorem (cf. [7]).

α	p	y^0_{id}	y^0_i	π^S_d	π^S	π^N
0	-	-	-	-	-	472.0
0.1	0.5	-	1.9	-	476.8	468.1
0.2	1.0	5.5	2.8	472.6	475.5	466.6
0.3	**1.7**	**10.0**	**5.0**	**466.8**	**475.0**	**468.9**
0.4	2.2	14.3	7.1	460.3	476.1	474.9
0.5	2.6	17.5	8.8	455.2	478.3	482.5
0.6	2.8	19.4	9.7	453.4	481.0	489.4
0.7	2.9	20.0	10.0	454.9	483.6	494.3
0.8	2.8	19.7	9.9	458.7	486.1	497.3
0.9	2.6	18.9	9.4	463.5	488.2	498.8
1	2.4	17.7	8.9	468.3	490.0	-

Table 2: Stability with a non-EGAL initial allocation of permits

natories can behave in a Stackelberg fashion, which is not necessarily justified.[4] Chander and Tulkens [5] propose another approach, and this is also the one I will use. In the concept they call Partial Agreement Nash Equilibrium (PANE), the Stackelberg assumption is replaced by a Cournot framework. Signatories and non-signatories are assumed to choose abatement levels simultaneously, with non-signatories acting as singletons and signatories so as to maximise their collective profit. To see how this would modify Barrett's example, consider Table 3, for the case of no permit trading (or, alternatively, for the case of EGAL initial allocation of permit). The last two columns reveal that in this case no stable coalition exists.

This last point illustrates a weakness of the game theoretical literature on environmental agreements. There is no single model of the game being played, and different assumptions regarding the dynamic consequences of players' actions or the interaction of cooperating and non-cooperating players may well yield different results. Recognising this weakness, the present paper will nevertheless stick to a particular equilibrium concept (PANE), and investigate the nature of cooperation rather than dynamic consequences or interaction in greater detail. To do so, I first modify the standard model described in Section 2.1 along the lines suggested in the preceding example. Results obtained using this modified model are presented in Section 3.

[4]The Stackelberg assumption is especially odd for small coalitions. In Table 1, Barrett uses the this approach also for $\alpha = 0.1$. This amounts to allowing a single player to act as a leader just because it decides to do so.

α	Y^S	Y^N	Y	π^S	π^N
0	-	80.0	80.0	-	472.0
0.1	8.0	72.0	80.0	472.0	472.0
0.2	27.6	55.2	82.8	461.4	479.2
0.3	48.6	37.8	86.5	458.0	487.2
0.4	65.3	24.5	89.8	461.5	492.7
0.5	76.9	15.4	92.3	467.5	495.9
0.6	84.7	9.4	94.1	473.4	497.6
0.7	89.9	5.5	95.4	478.3	498.5
0.8	93.4	2.9	96.4	482.3	499.1
0.9	95.9	1.2	97.0	485.4	499.4
1	97.6	-	97.6	487.8	-

Table 3: Stability with PANE

2.3 Modifying the Model: Emission Trading

This section extends the model presented in Section 2.1 to include emission trading between members of S. Setting up a market for tradeable permits involves three steps. First the total quantity of permits is determined, second, this quantity is allocated to participants of the market, and finally a market structure is specified in which trading takes place. The modified model is described in three steps, proceeding backwards from the market structure.

The Market for Tradeable Permits

In the *competitive* market for tradeable permits players face an exogenous total quantity of permits, $\bar{q}$, a vector of initial allocations $\mathbf{q}^0 \doteq [q_i^0]_{i\in S}$, and a market price p of permits. Players buy permits for emissions exceeding their initial allocation, and sell permits they do not use at the market price.[5] Thus player i chooses his emission q_i so as to solve

$$\max_{q_i} \left\{ \pi_i = -d_i(Q) - c_i(q_i) - p(q_i - q_i^0) \right\} . \tag{9}$$

Necessary and sufficient first order conditions are

$$-\frac{\partial d_i(Q)}{\partial q_i} - \frac{\partial c_i(q_i)}{\partial q_i} = p \tag{10}$$

for $i \in S$ (cf. Equation (2)). The market price p is determined from (10) and the market clearing condition $\bar{q} = Q^S$. Thus individual pollutions depend on

[5]The possibility of noncompliance is ignored.

the market price, which in turn depends on the total quantity of permits, $\bar{q}$. As is well known, pollutions are independent of initial allocations: a perfect market achieves the emission target $\bar{q}$ cost effectively, regardless of the initial distribution of permits [17].

To investigate how market structure affects the stability of agreements, I will consider one alternative structure: a market *with transaction costs*. This model is based on Stavins [20]. Players are assumed to face transaction costs which depend on the amount of permits they sell or purchase. Letting $t_i \doteq q_i - q_i^0$, transaction costs are given by the twice continuously differentiable, convex function $T(t_i)$, which satisfies $T(t_i) \equiv T(-t_i)$ (and therefore also $T'(0) = 0$) and $T'(t_i)t > 0$ if $t \neq 0$.[6] Players thus face identical transaction cost functions, which are increasing in the level of transaction, both for buying and selling.

With transaction costs, player i's problem becomes

$$\max_{q_i} \left\{ \pi_i = -d_i(Q) - c_i(q_i) - pt_i - T(t_i) \right\}. \tag{11}$$

The necessary and sufficient first order condition for this problem is

$$-\frac{\partial d_i(Q)}{\partial q_i} - \frac{\partial c_i(q_i)}{\partial q_i} - \frac{\partial T(t_i)}{\partial t_i} = p. \tag{12}$$

Again, market equilibrium is determined by Equation (12) for $i \in S$ and the market clearing condition. Here q_i depends on the initial allocation q_i^0 and the price p, which in turn depends on the vector of initial allocations $\mathbf{q}^0$ as well as on the total quantity of permits. Therefore, with transaction costs it will no longer be true that a market for tradeable permits leads to a cost effective realisation of the emission target regardless of the initial allocation of permits.

Next, I turn to the part of the model describing the initial allocation of permits.

The Initial Allocation of Permits

Initial allocation of permits is modeled using the Endowment Game proposed in [22], where many details of the model can be found. In this approach, distribution of permits takes place in a cooperative bargaining game, where player i's payoff is the profit (9) or (11) he realises in the market. That is, players in the Endowment Game look at their subsequent market profits as a function of their initial endowment, and distribute these endowments according to a cooperative bargaining solution (e.g., the Nash bargaining solution).

[6] In [22] I call this type of transaction cost "monotonic".

In this bargaining game, payoffs, denoted by $\pi_i(\mathbf{q}^0)$, correspond to the value functions associated to problems (9) or (11) for $i \in S$. Reference points (i.e. what players would get if no agreement was reached concerning the initial allocation of permits) are naturally chosen to equal the Nash equilibrium payoffs corresponding to problem (1). If players cannot agree on an initial allocation of permits, no cooperation is possible, and S disintegrates. This reference payoff will be denoted by D_i for player i: $D_i \doteq -d_i\left(\sum_j q_j^u\right) - c_i(q_i^u)$. The net gain of a player is defined as

$$R_i(\mathbf{q}^0) \doteq \pi_i(\mathbf{q}^0) - D_i.$$

How, then, is the initial allocation of permits determined? In [22] I argue that three principles are of particular relevance, partly because of their theoretical characteristics, partly because the criteria they embody have a potential connection to principles put forward in international climate change negotiations (see [21] on the principles endorsed by the Framework Convention on Climate Change and the Kyoto Protocol). These principles, or solution concepts are defined as follows.

Definition 1 *An allocation* $\hat{\mathbf{q}}^0$ *is a Utilitarian Allocation, if it maximises the sum of the players' payoff:*

$$\hat{\mathbf{q}}^0 \in \arg\max_{\mathbf{q}^0} \sum_{i \in S} \pi_i(\mathbf{q}^0)$$

Definition 2 *An allocation* $\hat{\mathbf{q}}^0$ *is a Nash Allocation, if the corresponding payoffs give the Nash solution of the Endowment Game.*

Definition 3 *An allocation* $\hat{\mathbf{q}}^0$ *is an Equal Gains Allocation (EGAL), if it equates the players' net gains, i.e.* $R_i(\hat{\mathbf{q}}^0) = R_j(\hat{\mathbf{q}}^0)$ *for all* i, $j \in S$.

Some of the main characteristics of these principles are the following (see [22], Section 3 for the proofs):

- With a competitive market, every allocation is a Utilitarian Allocation, while the Nash Allocation is unique. Furthermore, if an EGAL exists, then it is also unique, and corresponds to the Nash Allocation.
- With transaction costs, there is exactly one Utilitarian Allocation. It is characterised by $q_i^0 = q_i$ for all i, and in this case pollution levels are the same as in the competitive market. It is also possible to show that with transaction costs, it cannot be the case that an allocation

satisfies exactly two of the three principles above. Either there exists an allocation which satisfies all three criteria,[7] or an allocation satisfies exactly one, or none of them.

Total Quantity of Permits

The last step in setting up our market for tradeable permits is to specify how signatories set the total quantity of permits, $\bar{q}$, which in equilibrium will equal their total emissions Q^S. In doing so, signatories (or a social planner acting on their behalf) take into account players' profits, as determined by the allocation principle which will be adhered to, and the market structure in which permits are to be traded. A natural choice is to assume that the social planner picks this total quantity so as to maximise the signatories' collective profit, i.e. he solves:

$$\max_{\bar{q}} \sum_{i \in S} \pi_i(p(\mathbf{q}^0, \bar{q}), \mathbf{q}^0, \bar{q}) \tag{13}$$

Notice that problem (13) is quite different from the standard problem in the literature, such as (7), in which signatories directly choose $\{q_i\}_{i \in S}$ so as to maximise the sum of their profits. Here individual pollutions are determined in three stages, through initial allocation and a permit market, and explicitly depend on permit market institutions. Problem (13) and a standard problem like (7) will only result in the same pollution levels for specific allocation principles, or a competitive market. Comparing different allocation principles and market structures from the point of view of optimal total quantity yields the following results (see [22] Section 4 for the proofs):

- In any market, (13) yields the socially optimal total quantity if and only if the allocation $\mathbf{q}^0$ is Utilitarian.
- For non-Utilitarian allocations, the social planner chooses different $\bar{q}$ total quantities for different market structures.

2.4 Solving the Model

In what follows, I will consider the special case of the model given in the previous section, in which every player is identical ($d_i = d$ and $c_i = c$ for all i). Although this is clearly unrealistic, it will help in focusing on some of the

[7] This is the special case when parameters of the model are such that in the Utilitarian Allocation with $q_i^0 = q_i$ we also have $R_i(\mathbf{q}^0) = R_j(\mathbf{q}^0)$ for all $i, j \in S$. One important case in which this holds is when all the players are identical.

stability issues that arise. Identical players are also assumed in [2], [3], and [18].

To solve the model, one has to make specific assumptions regarding the interaction of cooperating and non-cooperating players and the dynamic consequences players take into account. In particular, I will look for a Partial Agreement Nash Equilibrium (PANE) between S and N [5], i.e. assume that cooperating and non-cooperating players choose emissions simultaneously, the former setting the total quantity of permits, the latter choosing their emission levels as singletons. I also assume that players calculate their profits in the shortsighted manner, as in [2] and [3]: they assume that the agreement they join or leave remains formed. Specifically, every non-signatory $i \in N$ takes emission by everyone else as given, and solves the following problem:

$$\max_{q_i} - d\left(Q^S + \sum_{j \in N} q_j\right) - c(q_i). \tag{14}$$

Signatories (or a social planner acting on their behalf) on the other hand, take the market structure and the principle of allocation (as well as non-signatories' pollution) as given, and choose the total quantity of permits, $\bar{q}$, so as to maximise their joint profit. Take first the case of a competitive market. As shown in Section 2.3, individual pollution levels are determined by the market equilibrium conditions

$$-d'(Q) - c'(q_i) = p \text{ for all } i \in S, \text{ and } \sum_{j \in S} q_j = \bar{q}. \tag{15}$$

Since players are identical, with a competitive permit market in equilibrium we have $q_i = q^S$ for all $i \in S$. The problem of the social planner is then:

$$\max_{\bar{q}} \alpha n \left[-d\left(\alpha n q^S(p(\bar{q})) + Q^N\right) - c(q^S(p(\bar{q})))\right], \tag{16}$$

where $q^S(p(\bar{q}))$ denotes equilibrium pollutions given by (15).

A Partial Agreement Nash Equilibrium is a vector of pollution levels $[q_i]_{i \in N}$ and a quantity $\bar{q} = Q^S$ which solve problems (14) and (16). Since players are identical, in PANE we have $q_i = q^N$ for all $i \in N$. Thus PANE is characterised by the first order conditions:

$$-d'\left(Q^S + Q^N\right) - c'(q^N) = 0 \tag{17}$$

and

$$-\alpha n d'(Q^S + Q^N) - c'(q^S) = 0. \tag{18}$$

As discussed in Section 2.3, the remaining variable of the model, the vector of initial allocations, $\mathbf{q}^0$, can be chosen arbitrarily, this will not affect the choice of the other variables.

With transaction costs, equilibrium in the permit market is determined by the following conditions:

$$-d'(Q) - c'(q_i) - T'(t_i) = p \text{ for all } i \in S, \text{ and } \sum_{j \in S} q_j = \bar{q}. \tag{19}$$

Here, market equilibrium is no longer independent of $\mathbf{q}^0$. The social planner solves

$$\max_{\bar{q}} \sum_{i \in S} \left[-d\left(\bar{q} + Q^N\right) - c(q_i(p(\mathbf{q}^0, \bar{q}), q_i^0)) - T(q_i(p(\mathbf{q}^0, \bar{q}), q_i^0) - q_i^0) \right].$$

Thus, in PANE, condition (18) is replaced by

$$-(\alpha N - 1)d' + \sum_{i \in S} T'(q_i(p(\mathbf{q}^0, \bar{q}), q_i^0) - q_i^0) \frac{dq_i^0}{d\bar{q}} + p(\mathbf{q}^0, \bar{q}) = 0. \tag{20}$$

where I have made use of conditions (19).

For a given agreement S, players' equilibrium profits, $\pi_i^S(S)$ and $\pi^N(S)$, determine the players' incentives to become part of S or N.[8,9] The importance of having S as large as possible can be shown if we recall from Section 2.1 the fact that social optimum would require having all players sign the agreement. As shown in Appendix A for the case of a competitive permit market, increasing the number of signatories steadily pushes PANE towards this social optimum, i.e. total pollution diminishes in a monotonic way. See the third column of Table 3 for an illustration.

The factors which determine the stability of an agreement are investigated next.

[8] To simplify the notation, I will denote by $\pi_i^S(S)$ the profits of a signatory regardless of the market structure. Whether $\pi_i^S(S)$ is determined by (9) (competitive market) or (11) (transaction costs) will always be clear from the context.

[9] Since players are identical, $\pi_i^N(S) \equiv \pi^N(S)$ for every nonsignatory.

3 Stability

With the modification of the standard model of environmental agreements presented in the previous section, it is possible to investigate how the nature of cooperation (the characteristics of a system of tradeable permits) affect the stability of the agreement.

I am using the short-sighted stability concept, as in [2], [3]. Stability of an agreement S thus requires that

$$\pi_i^S(S) \geq \pi^N(S\backslash\{i\}) \text{ for all } i \in S, \quad (21)$$

$$\pi^N(S) \geq \pi_{i'}^S(S \cup \{i'\}) \text{ for all } i' \in N. \quad (22)$$

Internal stability, equation (21), assumes that the coalition $S\backslash\{i\}$ remains formed, and $\pi^N(S\backslash\{i\})$ is determined in the corresponding new PANE. Similarly, external stability, equation (22), assumes that joining players only look "one agreement ahead", and profits are determined in the PANE corresponding to the agreement $S \cup \{i'\}$. Whether a specific agreement is stable or not is not the primary focus of the analysis below, however. This can only be determined for specific functional forms, and even there numerical examples might be necessary (as in Barrett's example in Section 2.2). Instead, I am interested in *comparing* the stability of agreements with given emission trading institutions. An agreement S will be called *more stable* than S', if the stability of S' implies the stability of S, but not vice-versa. Thus, S is more stable than S' if moving from the stable S' to S (e.g., expanding a stable coalition), we know that the resulting agreement has to be stable, but we cannot say whether a change in the opposite direction, moving from S to S' (e.g., reducing a stable coalition), results in a stable or unstable agreement.[10]

Below I first investigate how changing the market structure affects stability for a given principle of allocation. I then take market structure as fixed, and ask how stability is affected if the initial allocation is changed.

3.1 Stability and Market Structure

I begin by investigating the stability effects of market structure. Specifically, I take a given principle of allocation, and compare an agreement with a competitive market to one with transaction costs. I show that for the three

[10]Notice that the definition does not require any of the agreements to be stable (at this level of generality of the model, stability cannot be established). An agreement can be more stable than another one, and both can still be unstable.

principles of interest, stability of the agreement is unaffected by the market structure.

First the three allocation principles defined in Section 2.3 are characterised. The first lemma establishes a basic result for the case of identical players, already noted in relation to Barrett's example in Section 2.2.

Lemma 4 *With identical players and a competitive market, $q_i^0 = q_i$ for all $i \in S$ gives the EGAL and the Nash Allocation.*

Proof. The Nash Allocation and the EGAL always coincide (see Section 2.3), and the EGAL obviously satisfies the equalities. ■

The next Lemma establishes an important equivalence result for the case of transaction costs.

Lemma 5 *In a market with transaction costs and identical players the Utilitarian Allocation is also the only Nash Allocation and the only EGAL.*

Proof. As shown in [22], in a market with transaction costs, if the Utilitarian Allocation is an EGAL, then it is also a Nash allocation (see Section 2.3). Since the profits corresponding to a Nash Allocation are always uniquely determined, I only need to show that with identical players the Utilitarian Allocation is the only EGAL. Suppose not. Then there is an allocation $\hat{\mathbf{q}}^0$ different from the Utilitarian Allocation, in which for any i and j we have

$$\pi_i(\hat{q}_i^0) = \pi_j(\hat{q}_j^0). \tag{23}$$

But since players are identical, $\pi_i \equiv \pi_j \doteq \pi$ for all i, j. We know that (23) is satisfied for the Utilitarian allocation, so if our assumption is correct, then the function π must not be monotonous. However, from (11), we have

$$\frac{d\pi(q_i^0)}{dq_i^0} = T' + p,$$

using the envelope theorem. This is clearly strictly positive, therefore (23) cannot hold for an allocation different from the Utilitarian Allocation. ■

Since $q_i^0 = q_i$ for all i yield the same pollutions in both market structures, from Lemma 4 and Lemma 5 we have.

Corollary 6 *With identical players, the Nash Allocation / EGAL in a competitive market is identical to the EGAL / Nash / Utilitarian Allocation in a market with transaction costs.*

Using these results, it is straightforward to show that with identical players, stability of the agreement is unaffected by the market structure.

Proposition 7 *With identical players and EGAL / Nash Allocation, stability of the agreement is unaffected by the market structure.*

Proof. Since the allocation, the optimal pollution levels and the optimal total quantity are identical in the two market structures considered (Corollary 6, Section 2.3), and since in an EGAL equilibrium transaction costs are zero, equations (21) and (22) are unaffected. ■

Quite intuitively, Proposition 7 states that if in equilibrium players do not trade on the market, then introducing transaction costs does not affect any of the relevant variables of the model. Although this paper uses the specific concept of stability in (21-22), since Proposition 7 states that profits of the players do not change, this result holds for all meaningful definitions of stability.

A few remarks are in order concerning the way in which the model is assumed to react to the various changes assumed in the discussion. The first such change is when a player leaves or joins the agreement. In specifying the dynamic consequences of such moves, I assumed that the new agreement is completely reorganised: a new optimal total quantity is chosen in the new PANE, and permits are reallocated according to the given principle among the new parties. In practice, this might take some time, this scenario can therefore be considered a "long run" reaction. The alternative, "short run" scenario would be to assume that in response to a player's leaving (joining) the agreement, the total quantity of permits is reduced (increased) by a fixed amount. One possibility that naturally arises, is to assume that an exiting player takes away q_i^0, while an entering player gets $q_i^0 = q_i$. Although in general this might alter some of the implications of the model, it clearly does not affect the results presented above, where identical players and EGAL were considered. Those results hold for the "short run" as well as the "long run" analysis.

The second change which the model had to react to was the change in the market structure. Here again, the "long run" adjustment assumes that in the new market structure the agreement is completely reorganised. A "medium run" scenario could be if $\bar{q}$ was held fixed, but the initial allocation would be allowed to adjust according to the given principle, e.g., an agreement where transaction costs appear has to construct a new Nash Allocation from the total quantity chosen in the previous, competitive market structure. In the "short run" both the allocation $\mathbf{q}^0$ and the total quantity of permits $\bar{q}$ might be held fixed. In general, these three scenarios might lead to very different results. In the analysis above, however, since neither the initial allocation satisfying a given principle nor optimal total quantity was affected by the market structure, the results obtained hold for all three "runs".

3.2 Stability and Initial Allocation

Next, I take the market structure as given, and investigate how changing the principle of allocation affects the stability of the agreement. I show that the EGAL (Utilitarian and Nash Allocation) yields agreements that are more stable than others, for both competition and transaction costs.

As the following proposition shows, with identical players and a competitive market the EGAL / Nash Allocation is particularly attractive from a stability point of view.

Proposition 8 *With identical players and a competitive market, the EGAL / Nash Allocation makes an agreement more stable than any other allocation.*

Proof. Stability of an agreement S^* with a given initial allocation requires that

$$\begin{aligned}\left[-d(Q)-c(q_i)-p\left(q_i-q_i^0\right)\right]_{S=S^*} &\geq \pi^N(S^*\backslash\{i\}) \\ \pi^N(S^*) &\geq \left[-d(Q)-c(q_{i'})-p\left(q_{i'}-q_{i'}^0\right)\right]_{S=S^*\cup\{i'\}}\end{aligned}$$

for all $i \in S$ and all $i' \in N$. Adding these inequalities up for $i \in S$ and $i' \in N$ respectively, and using the fact that $\sum_j (q_j - q_j^0) = 0$, we get

$$\begin{aligned}|S^*|\cdot[-d(Q)-c(q_i)]_{S=S^*} &\geq |S^*|\cdot\pi^N(S^*\backslash\{i\}) \\ |N|\cdot\pi^N(S^*) &\geq |N|\cdot[-d(Q(S))-c(q_{i'})]_{S=S^*\cup\{i'\}}\,.\end{aligned}$$

Dividing by $|S^*|$ and $|N|$ respectively, and noticing that $\pi^N(S^*\backslash\{i\})$ and $\pi^N(S^*\cup\{i'\})$ are not affected by the allocation principle, these are indeed the conditions for EGAL to insure stability. Thus, if a given agreement is stable with a specific allocation, then it is also stable with EGAL. Clearly the reverse is not necessarily true. ■

Next, I turn to transaction costs. I show that if we slightly modify the EGAL (also Utilitarian and Nash Allocation), this reduces the stability of the agreement.

The problem in this case with transaction costs is that total quantity of permits, $\bar{q}$, depends on the allocation of permits (see equation (20)). Thus, if we modify the allocation, the optimal quantity of permits changes, *depending on the relationship between total quantity and allocation*. Since in this model all three of the allocation principles derived from the Endowment Game lead to the same endowments, there is no obvious way to represent this relationship between total quantity and initial allocation. One way to model this relationship, which has some merit, is the following. Let us assign a constant β_i to every player i, and suppose that the initial allocation takes the form

$$q_i^0 = \frac{\beta_i}{\sum_{j\in S}\beta_j}\bar{q}.$$

Suppose also that initially we have $\sum_{j\in S}\beta_j = 1$. This setup defines a method of initial allocation which requires that player i should always be allocated a given fraction of the permits, relative to other signatories. (Notice that here q_i^0 depends on $\bar{q}$ as well as S.) EGAL corresponds to the special case in which $\beta_i = \frac{1}{\alpha n}$ for all $i \in S$.

Consider then, what happens if we slightly perturb the initial allocation, i.e. change player i's β_i by a small amount. From (19) and (20) we know that any such change must satisfy three conditions:

$$-d''\frac{d\bar{q}}{d\beta_i} - c''(q_j)\frac{dq_j}{d\beta_i} - T'(t_j)\left(\frac{dq_j}{d\beta_i} - \frac{dq_j^0}{d\beta_i}\right) = \frac{dp}{d\beta_i} \tag{24}$$

for all $j \in S$,

$$\sum_{j\in S}\frac{dq_j}{d\beta_i} = \frac{d\bar{q}}{d\beta_i}, \tag{25}$$

and finally

$$-(\alpha N - 1)d''\frac{d\bar{q}}{\beta_i} + \sum_{j\in S}T''(t_j)\left(\frac{dq_j}{d\beta_i} - \frac{dq_j^0}{d\beta_i}\right)\beta_j + T'(t_i) + \frac{dp}{d\beta_i} = 0, \tag{26}$$

where the last equation uses the fact that $\frac{dq_j^0}{d\bar{q}} = \beta_j$. Solving this system for the EGAL yields

Lemma 9 *Moving slightly away from the EGAL leaves optimal total quantity of permits unchanged, i.e.* $\frac{d\bar{q}}{d\beta_i} = 0$.

Proof. See Appendix C. ■

Using Lemma 9 we can show that moving slightly away from the EGAL reduces the stability of the agreement.

Proposition 10 *With transaction costs and identical players, the EGAL (Utilitarian and Nash Allocation) makes an agreement "locally" more stable than any other allocation (more stable than any allocation in the neighbourhood of the EGAL).*

Proof. From the envelope theorem,

$$\frac{d\pi_i^S}{dq_i^0} = -(q_i - q_i^0)\frac{dp}{dp_i^0} + p + T'(q_i - q_i^0).$$

Thus, in the EGAL $\frac{d\pi_i^S}{dq_i^0} = p$. For any agreement, a move *towards* the EGAL equalises profits by reducing the profits of the players whose endowment decreases, and increasing the profits of those whose endowment increases. Thus, for players with the lowest profits, $\pi_i^S(S)$ increases, and for players with the highest potential profit in N, $\pi_{i'}^S(S \cup \{i'\})$ decreases. On the other hand, since according to Lemma 9 total quantity is unaffected by a small change, $\pi^N(S)$ and $\pi^N(S \backslash \{i\})$ remain constant. Hence, if (21-22) held for a non-EGAL allocation, they hold for EGAL too. Stability of an agreement with an allocation in the neighbourhood of the EGAL therefore implies the stability of the agreement with EGAL. By similar reasoning, one can see that the reverse does not hold, which completes the proof. ■

Let us, again, briefly consider how the model is assumed to react to changing conditions. A crucial assumption upon which the validity of Proposition 8 rests is that once a player leaves or joins the cooperating players, the agreement is re-optimised, i.e. a new optimal total quantity is determined and reallocated between the members of the new agreement. As a result, $\pi^N(S^* \backslash \{i\})$ and $\pi^N(S^* \cup \{i'\})$ are not affected by the allocation principle, and stability conditions for any two allocation are easily comparable. This is also the "long run" approach, mentioned in the previous section, in which it is assumed that after conditions change, every variable of the agreement is optimally readjusted. As discussed above, the alternative ("short run") assumption is to let exiting (joining) players decrease (increase) total quantity by a given amount. To see why this would lead to very different results even in this case of identical players, consider what happens if exiting players simply take away their initial endowment q_i^0 from the agreement. In this case, the potential profit of an exiting player, $\pi^N(S^* \backslash \{i\})$, changes as his initial endowment is modified. In particular, allocating slightly more permits to player i increases (!) his motivation to leave the agreement. This is so because a higher q_i^0 means that the exiting player can diminish the total pollution of the signatories to a larger extent. To see this, notice first that

$$\frac{d\pi^N(S\backslash\{i\})}{dq_i^0} = \frac{d\pi^N(S\backslash\{i\})}{dQ^S(S\backslash\{i\})} \cdot \frac{dQ^S(S\backslash\{i\})}{dq_i^0}.$$

According to our assumption that exiting players take away their initial endowments, the second term in the multiplication is simply -1. Thus

$$\frac{d\pi^N(S\backslash\{i\})}{dq_i^0} = -\frac{d\pi^N(S\backslash\{i\})}{dQ^S(S\backslash\{i\})},$$

and the sign of $\frac{d\pi^N(S\backslash\{i\})}{dQ^S(S\backslash\{i\})}$ can be established by standard comparative statics calculations (see Appendix B): it is negative. Therefore $\frac{d\pi^N(S\backslash\{i\})}{dq_i^0} > 0$, and increasing the endowment of a player increases his motivation to exit the agreement. Similar remarks are valid in relation to Proposition 10.

Similarly, a "short run" and a "long run" adjustment can be distinguished in how the model reacts to a change in the principle of allocation. In the "long run", a new optimal total quantity is determined, while the "short run" analysis would be to fix $\bar{q}$, and look at the consequences of simply reallocating it among the players according to the new principle. Since in a competitive market optimal total quantity is independent of the initial allocation, and since in the proof of Proposition 10 this is shown to hold locally for transaction costs, both Proposition 8 and Proposition 10 stay valid in the "short run" analysis.

4 Conclusion

Artificial markets exhibit theoretically interesting, distinctive features, when compared to the standard markets studied in economics. Being artificial, market institutions are not given 'by nature', but develop as a result of conscious action, typically involving some cooperation between future competitors. The international market for tradeable permits, for example, is an exchange economy[11] in which endowments of the consumers are determined entirely by the consumers themselves. As a result, once one allows coalitions to form, one has to worry about stability in two distinct levels. First, there is stability of the market itself, given the endowments of the players, as in standard equilibrium analysis. One can define stability in the sense of the *core*, and examine whether a coalition of market participants have an incentive to form a smaller market, and trade their endowments among themselves. For example, the standard result that the competitive equilibrium is in the core still holds.[12] Second, we must take a step back, and see if players have an incentive to create the market in the first place. These incentives depend on what players expect to gain from the market, given its various characteristics. The present paper is an attempt at this second type of stability analysis.

This paper made a first step in analysing how a market for emission permits created by an environmental agreement influences the stability of the agreement. A model was described, in which the stability implications of (a) the structure of the permit market and (b) the initial allocation of permits could be investigated. Emission levels of signatories were determined in three steps, through the decision of a social planner on the total quantity of permits, a cooperative bargaining game allocating the initial endowments, and finally a permit market. Signatories and non-signatories were assumed to interact according to the Partial Agreement Nash Equilibrium model, and players always supposed that the agreement they joined or left remained formed. In the special case of the model involving identical players, three results were established:

- In the case of an equal allocation of permits (conforming to the Utilitarian, Nash or EGAL principle) the market structure does not affect the stability of the agreement.

[11] Under the Kyoto Protocol, "production" might be be possible, however, if the Clean Development Mechanism - according to which a country can acquire aditional permits by reducing pollution in countries that are not parties to the Protocol - is implemented.

[12] See for example [15], Chapter 18.

- With a competitive market, equal allocation yields more stable agreements than any other allocation.
- With transaction costs, equal allocation yields more stable agreements than any other allocation in its neighbourhood.

These results illustrate the type of insight which might be gained from such models. Of course, various extensions and generalisations of the model might be desirable in order to get results with strong practical implications. The most striking simplification is the assumption of identical players. It is not clear, however, whether general results can be obtained without this assumption. Considering specific functional forms might be necessary, as in the example of Section 2.2. Other ways of modeling the interaction of signatories and non-signatories and the dynamic consequences of players' actions were mentioned in Section 2.1. Varying these assumptions might be fruitful. Finally, the nature of cooperation, i.e. the market for tradeable permits could be extended to include other principles of allocation in the Endowment Game and other market structures, such as the case of market power. Examining how other characteristics of a permit market - such as the compliance regime or the possibility of banking permits - affect the stability of the agreement might also be useful.

References

[1] Amacher, G.S and A.S. Malik, Instrument Choice When Regulators and Firms Bargain, *Journal of Environmental Economics and Management* 30, 233-53 (1998).

[2] Barrett, S., Self-enforcing International Environmental Agreements, *Oxford Economic Papers* 46, 878-94 (1994).

[3] Carraro, C. and D. Siniscalco, Strategies for the international protection of the environment, *Journal of Public Economics* 52, 309-28 (1993).

[4] Center For International Earth Science Information Network, *Environmental Treaties and Resource Indicators*, Palisades, NY. Available at http://sedac.ciesin.org/entri

[5] Chander, P. and H. Tulkens, The Core of an Economy with Multilateral Environmental Externalities, *International Journal of Game Theory* 26, 379-401 (1997).

[6] Chwe, M. S., Farsighted Coalitional Stability, *Journal of Economic Theory* 63, 299-325 (1994).

[7] Coase, R., The Problem of Social Cost, *Journal of Law and Economics* 3, 1-44 (1960).

[8] Congleton, R.D., Political Institutions and Pollution Control, *The Review of Economics and Statistics* 74, 412-21 (1992).

[9] Garner, R., *Environmental Politics*, Macmillan, London, UK (2000).

[10] Harrison, D. Jr, Turning theory into practice for emissions trading in the Los Angeles air basin, *in*: S. Sorrell and J. Skea (eds.): *Pollution for Sale. Emissions Trading and Joint Implementation*, Edward Elgar, Cheltenham, UK (1999).

[11] Hausker, K., The Politics and Economics of Auction Design in the Market for Sulfur Dioxide Pollution, *Journal of Policy Analysis and Management* 11(4), 553-72 (1992).

[12] International Institute for Sustainable Development, *Earth Negotiations Bulletin*, Vol. 12, various issues. Available at http://www.iisd.ca.

[13] Joskow, P.L. and R. Schmalensee, The Political Economy of Market-Based Environmental Policy: The U.S. Acid Rain Program, *Journal of Law and Economics* 41, 89-135 (1998).

[14] Laffont, J-J. and M. Boyer, Toward a Theory of the Emergence of Environmental Incentive Regulation, *Rand Journal of Economics* 30(1), 137-57 (1999).

[15] Mas-Collel, A., M. Whinston and J. Green, *Microeconomic Theory*, Harvard University Press, Cambridge, MA (1995).

[16] McCormick, J., *The Global Environmental Movement*, Wiley, Chichester, UK (1995).

[17] Montgomery, W. K., Markets in licenses and efficient pollution control programs, *Journal of Economic Theory* 5, 395-418 (1972).

[18] Ray, D. and R. Vohra, Coalitional Power and Public Goods, *Journal of Political Economy* 109(6), 1355-84 (2001).

[19] Sorrell, S., Why sulphur trading failed in the UK?, *in*: S. Sorrell and J. Skea (eds.): *Pollution for Sale. Emissions Trading and Joint Implementation*, Edward Elgar, Cheltenham, UK, (1999).

[20] Stavins, R. N., Transaction Costs and Tradeable Permits, *Journal of Environmental Economics and Management* 29, 133-48 (1995).

[21] Tóth, F.L., *Fair Weather? Equity Concerns in Climate Change*, Earthscan, London, UK (1999).

[22] Ujhelyi, G., Endowment Game and Optimal Total Quantity in the Market for Tradeable Permits, CEU-Economics Working Paper 1/2002.

A Appendix

The Partial Agreement Nash Equilibrium pollution levels Q^N and Q^S (and total pollution $Q = Q^N + Q^S$ are determined by the following two equations:

$$-d'(Q) - c'\left(\frac{Q^N}{(1-\alpha)n}\right) = 0 \tag{A1}$$

and

$$-\alpha n d'(Q) - c'\left(\frac{Q^S}{\alpha n}\right) = 0, \tag{A2}$$

where $\alpha n = |S|$ (and $(1-\alpha)n = |N|$). Notice that $\alpha \geq \frac{1}{n}$ and the convexity assumption on $c()$ imply

$$q^N = \frac{Q^N}{(1-\alpha)n} \geq \frac{Q^S}{\alpha n} = q^S, \tag{A3}$$

i.e. in PANE non-signatories' individual emissions are always higher than signatories'. Taking derivatives of (A1-A2) with respect to α yields

$$\mathbf{H}\begin{bmatrix} \frac{\partial Q^S}{\partial \alpha} \\ \frac{\partial Q^N}{\partial \alpha} \end{bmatrix} = \begin{bmatrix} c''\left(\frac{Q^N}{(1-\alpha)n}\right)\frac{Q^N}{(1-\alpha)^2 n} \\ nd'(Q) - c''\left(\frac{Q^S}{\alpha n}\right)\frac{Q^S}{\alpha^2 n} \end{bmatrix},$$

where $\mathbf{H}$ denotes the Hessian of system (A1-A2). We can use Cramer's rule to calculate $\frac{\partial Q^S}{\partial \alpha}$ and $\frac{\partial Q^N}{\partial \alpha}$. Adding these up we get

$$\begin{aligned} \frac{\partial Q}{\partial \alpha} &= \frac{\partial Q^S}{\partial \alpha} + \frac{\partial Q^N}{\partial \alpha} = \\ &= \frac{c''\left(\frac{Q^N}{(1-\alpha)n}\right)}{\det \mathbf{H}}\left[\frac{c''\left(\frac{Q^S}{\alpha n}\right)}{(1-\alpha)\alpha n^2}\left(\frac{Q^N}{1-\alpha} - \frac{Q^S}{\alpha}\right) + \frac{d'(Q)}{1-\alpha}\right]. \end{aligned}$$

From (A3), the term in brackets is positive, while $\det \mathbf{H} < 0$ from second order conditions. Thus $\frac{\partial Q}{\partial \alpha} < 0$: increasing the number of signatories steadily reduces world pollution.

B Appendix

In the Partial Agreement Nash Equilibrium, non-signatories pollution levels satisfied the following first order condition (equation (17)):

$$-\frac{\partial d\left(Q^S + (1-\alpha)nq^N\right)}{\partial q_i} - \frac{\partial c(q^N)}{\partial q_i} = 0.$$

Taking the derivative with respect to Q^S and rearranging, one gets:

$$\frac{dq^N}{dQ^S} = -\frac{d''}{c''} \cdot \left(1 + \frac{dQ^N}{dQ^S}\right), \tag{B1}$$

where $Q^N = (1-\alpha)nq^N$. Multiplying by $(1-\alpha)n$ and rearranging, we have

$$\frac{dQ^N}{dQ^S} = -\frac{(1-\alpha)n\frac{d''}{c''}}{(1-\alpha)n\frac{d''}{c''} + 1}.$$

Substituting back into (B1):

$$\frac{dq^N}{dQ^S} = -\frac{\frac{d''}{c''}}{(1-\alpha)n\frac{d''}{c''} + 1}. \tag{B2}$$

To find how the equilibrium profit of a non-signatory changes in response to a change in Q^S, notice that

$$\frac{d\pi^N}{dQ^S} = -d' \cdot \left(1 + \sum_{\substack{j \neq i \\ j \in N}} \frac{dq_j}{dQ^S}\right)$$

from the envelope theorem. Using (B2),

$$\frac{d\pi^N}{dQ^S} = \frac{-d'}{(1-\alpha)n\frac{d''}{c''} + 1}\left(\frac{d''}{c''} + 1\right),$$

which is negative.

C Appendix

Solving (24) for $\frac{dq_j}{d\beta_i}$ and substituting into (25) and (26) one gets a system of two linear equations with two unknowns, $\frac{d\bar{q}}{d\beta_i}$ and $\frac{dp}{d\beta_i}$:

$$\frac{d\bar{q}}{d\beta_i} = \left(-d''\frac{d\bar{q}}{d\beta_i} - \frac{dp}{d\beta_i}\right)\sum_{j\in S}\frac{1}{c''(q_j)+T''(t_j)} + \sum_{j\in S}\frac{T''(t_j)\frac{dq_j^0}{d\beta_i}}{c''(q_j)+T''(t_j)} \tag{C1}$$

and

$$-(\alpha N-1)d''\frac{d\bar{q}}{\beta_i} + \frac{dp}{d\beta_i} + \sum_{j\in S}T''(t_j)\frac{-d''\frac{d\bar{q}}{d\beta_i} - \frac{dp}{d\beta_i} - c''(q_j)\frac{dq_j^0}{d\beta_i}}{c''(q_j)+T''(t_j)}\beta_j + T'(t_i) = 0. \tag{C2}$$

Recalling that with the EGAL $q_j = q_j^0 = q^S$ and $\beta_j = \frac{1}{\alpha n}$ for all $j \in S$, equations (C1) and (C2) simplify to

$$\frac{d\bar{q}}{d\beta_i} = \left(-d''\frac{d\bar{q}}{d\beta_i} - \frac{dp}{d\beta_i}\right)\frac{\alpha n}{c''+T''} + \frac{d\bar{q}}{d\beta_i}\frac{T''}{c''+T''}$$

and

$$-(\alpha N-1)d''\frac{d\bar{q}}{\beta_i} + \frac{dp}{d\beta_i} + T''\frac{-d''\frac{d\bar{q}}{d\beta_i} - \frac{dp}{d\beta_i} - c''\frac{1}{\alpha n}\frac{d\bar{q}}{d\beta_i}}{c''+T''} = 0$$

respectively. Solving, we find $\frac{d\bar{q}}{d\beta_i} = \frac{dp}{d\beta_i} = 0$, as stated.