Dynamic core-theoretic cooperation in a two-dimensional international environmental model

M. Germain, H. Tulkens and A. Magnus

Discussion Paper 2009-15

Institut de Recherches Économiques et Sociales
de l'Université catholique de Louvain

UCL

Dynamic core-theoretic cooperation in a two-dimensional international environmental model*

Marc Germain[†], Henry Tulkens[‡] and Alphonse Magnus[§]

March 2009

Abstract

This article deals with cooperation issues in international pollution problems in a two dimensional dynamic framework implied by the accumulation of the pollutant and of the capital goods. Assuming that countries do reevaluate at each period the advantages to cooperate or not given the current stocks of pollutant and capital, and under the assumption that damage cost functions are linear, we define at each period of time a transfer scheme between countries, which makes cooperation better for each of them than non-cooperation. This transfer scheme is also strategically stable in the sense that it discourages partial coalitions.

Keywords : stock pollutant, capital accumulation, international environmental agreements, dynamic core solution.

JEL Classification : Q54, Q58, F42, F53, O21.

*The authors wish to thank their colleagues Guillaume Rocheteau, Gérard Hamiache, Aart the Zeeuw, Russell Cooper and Parkash Chander for valuables discussions and, for their hospitality and support, the Chair Lhoist Berghmans at CORE as well as the Department of Economics at the National University of Singapore.

[†]Laboratoire EQUIPPE, Université de Lille-3, France and Department of Economics, Université catholique de Louvain, Belgium. Corresponding author : marc.germain@univ-lille3.fr, marc.germain@uclouvain.be .

[‡]CORE, Université catholique de Louvain, Belgium.

[§]Institut de mathématique, Université catholique de Louvain, Belgium.

Introduction

This paper extends to a two dimensional setting - pollutant emissions and investment in capital goods - the dynamic game theoretic results established for the one dimensional model of international environmental agreements by Germain Tulkens and de Zeeuw 1998 (closed loop, linear case) and by Germain, Toint, Tulkens and de Zeeuw 2003 (closed loop, non linear case) [1]

Economically, a serious limitation of these contributions is that in the model they use, pollution abatement is independent of investment and of capital accumulation, so that growth of the economies is either absent or exogenous. There is indeed only one control variable for each country, namely the level of its emissions and, for the economy taken as a whole, only one state variable. The economic model is one of partial equilibrium cost minimization. The aim of this paper is to broaden the perspective by considering investment in each country as a second control variable as well as the country's accumulated capital as a second state variable. The economic model thus moves towards one of general equilibrium utility maximization.

In this context, the central issue of these two papers is taken up again here, namely whether stable and efficient cooperation among countries in the core-theoretic sense of a cooperative game at each period can be established, possibly using appropriately designed transfers. As in the two previous papers, a positive answer is obtained in terms of a sequence of self enforcing cooperative international agreements, that we call a coalitionally (or strategically) stable path of the economic-ecological system.[2] Here as in Germain et al. 1998, damage cost functions are assumed linear. Thus far, the non linear case of Germain et al. 2003 remains open for a similar extension.[3]

The issues raised by the necessity of cooperation amongst the countries involved in an transfrontier pollution problem if a social optimum is to be achieved, have been addressed in the literature in terms of game theory concepts starting with static (one shot) games (see Mäler (1989), followed by Chander and Tulkens (1992), Kaitala, Mäler and Tulkens (1995)). These are only suitable for flow pollution models. Stock pollutant problems introduce an intertemporal dimension, for which dynamic game theory is a more appropriate tool of analysis, as is done in e.g. van der Ploeg and de Zeeuw (1992), Kaitala, Pohjola and Tahvonen (1992), Hoel (1992), Tahvonen (1993), Petrosjan and Zaccour (1995).

Except for the last one, the other four contributions leave aside the issue of the *voluntary* implementation of the international optimum. This is an important drawback since no supranational authority can be called upon to impose the optimum in a context where the countries' interest in cooperation diverges strongly between one another, and especially if some countries loose when the social optimum is implemented. In view of ensuring such implementation, it was already shown in the static case by Tulkens (1979) that there exists feasible transfers between the countries involved that would provide incentives towards cooperation. This property, understood later as inducing strategies belonging to the core of a cooperative game, has in effect been demonstrated for a static game by Chander and Tulkens (1995, 1997), who propose a specific transfer scheme achieving the

[1]Hereafter, Germain et al. 1998 and Germain et al. 2003, respectively.

[2]Another concept of coalitional stability, called "internal and external stability" and due to d'Aspremont and Jaskold Gabszewicz 1986, was introduced in an international environmental dynamic game by Rubio and Ulph 2008, leading to quite different results than those obtained here as to the formation of the grand coalition.

[3]The non linear case was treated in an open loop setting by Eyckmans and Tulkens 2003. Coalitional stability was also established there, albeit of a substantially different kind due to the open loop assumption.

stated purpose. This result has been extended by Germain, Toint and Tulkens (1998) to the larger context of open-loop dynamic games, and thereafter to closed-loop (or feedback) dynamic games in the two papers mentioned at the outset. For a similar approach where transfers guarantee dynamic individual rationality understood as time-consistency and agreeability (with an application to a international pollution problem), see Jorgensen, Martin-Herran and Zaccour (2003).

The structure of the present paper is as follows. Section 1 presents the international stock pollutant model with capital accumulation in each country. In section 2, transfers are formulated so as to ensure that each country is not worse off when it participates in an international agreement, compared to a no agreement situation described as a Nash equilibrium. Section 3 computes the main variables of the model, including the transfers, at each period. Section 4 presents a theorem establishing that a specific form of the transfers achieves coalitional stability in the sense of the core of a cooperative game.

1 Preliminaries

1.1 Components of the model

Our economic model is written in discrete time. Consider n countries indexed by $i \in \mathcal{N} = \{1, 2, ..., n\}$ and some planning period $\mathcal{T} = \{1, 2, ..., T\}$ (T, the planning horizon, is a positive integer, possibly infinite). In the following, all variables (production, investment, capital stock, pollutant emissions, pollution stock) are positive, the first three of which are economic while the last two are ecological. We therefore deal with an economic-ecological model.

Each country i is characterised by an aggregate production function :

$$y_{it} = F_i(k_{it}, e_{it}) \tag{1}$$

where y_{it}, k_{it}, e_{it} are respectively the production, the stock of capital and the consumption of energy of country i at time t. We assume throughout that F_i satisfies the standard properties of a neo-classical production function, as reported in the following assumption :

Assumption 1: The function $F_i : R_+^2 \longrightarrow R_+$ satisfies the following conditions:
(i) $f(0,0) = 0$.
(ii) f is twice continuously differentiable.
(iii) $f_k(k,e) > 0$ and $f_e(k,e) > 0$, for all k, e.
(iv) $f_e(k,0) = +\infty$, for all $k > 0$.
(v) f is differentiably strictly concave jointly in (k,e), i.e.

$$f_{kk} < 0, f_{ee} < 0, \text{ and } f_{kk}f_{ee} - f_{ke}^2 > 0.$$

All components of this assumption are standard in production theory and have well known economic interpretations.

The capital stock is assumed to accumulate following :

$$k_{it} = [1 - \delta_i]k_{i,t-1} + h_{it}, \; i \in \mathcal{N} \tag{2}$$

where h_{it} is investment of country i at date t and δ_i is the rate of depreciation of capital of country i ($0 < \delta_i < 1$).

Pollution emitted by a country is assumed to be proportional to its energy consumption. For the sake of simplicity and without loss of generality, we denote the two quantities by the same symbol.Thus $\mathbf{e_t} = (e_{1t}, ..., e_{nt})'$ is the vector of the different countries' emissions of a certain pollutant at time t. These emissions spread uniformly in the atmosphere and contribute to a stock of pollutant z according to the equation

$$z_t = [1-\gamma]z_{t-1} + \sum_{i=1}^{n} e_{it} \tag{3}$$

where the initial stock of pollutant z_0 is given and where γ is the pollutant's natural rate of degradation $(0 < \gamma < 1)$. As described in (3), the current stock is a linear function of the inherited stock z_{t-1} and of the current emissions $\mathbf{e}_t$.This model describes, for example, the basics of the climate change problem where the flows of emissions of greenhouse gases accumulate into a stock, which only gradually assimilates and which is the cause of the climate change.

The stock of pollutant causes damages to each country's environment. For country i $(i \in \mathcal{N})$, these damages during period t are measured in monetary terms by the function $D_i(z_t)$, where D_i is supposed to be a linear function of the current stock z_t :

Assumption 2: The function $D_i : R_+ \to R_+$ *is of the form:*

$$D_i(z_t) = \pi_i z_t \tag{4}$$

where π_i is positive number.

Finally, let for each $i \in \mathcal{N}$

$$W_i = \sum_{t=1}^{T} \rho^{t-1} \left[F_i(k_{i,t}, e_{i,t}) - h_{i,t} - \pi_i z_t\right] \tag{5}$$

denote the discounted sum of the stream of collective consumption enjoyed by the population of country i over the planning period $\mathcal{T}$, where $0 < \rho < 1$ is a discount factor.

Bringing together all the above, we call the resulting model an "economic-ecological system". For such a system, any $(3n+1)T$ dimensional vector $\{(e_{i,t}, h_{i,t}, k_{i,t}, z_t), i \in \mathcal{N}, t \in \mathcal{T}\}$ satisfying (2) and (3) constitutes a feasible path for the planning period $\mathcal{T}$, yielding each country i a collective consumption (5).

1.2 Non cooperative vs cooperative behaviors and the issue of coalitional stability

For the so described international economy with environmental interactions, we deal in this paper with two planning issues, each of which corresponds to an alternative behavioral assumption on the countries' policies regarding pollutant emissions. In the first case we assume that each country i simply pursues its own interest in its choice of both emission and investment policies, $(e_{i,t}, t \in \mathcal{T})$and $(h_{i,t}, t \in \mathcal{T})$ respectively, ignoring the external effects induced by the former on the other countries through the stock accumulation process described by equation (3). Such behavior is formally expressed, for each country, by the solution of the dynamic optimization problem consisting in maximizing the value of (5) with respect to the control variables e_{it} and h_{it}, subject to the constraints (2) and (3) whereby the state variables $(k_{i,t}, t \in \mathcal{T}, i \in \mathcal{N})$ and $(z_t, t \in \mathcal{T})$, respectively, are defined.

Internationally, the outcome of such parallel behaviors is in the nature of a Nash equilibrium of a non cooperative game in which the players are the countries and their strategies are the policies just mentioned. We call this outcome the Non Cooperative Nash Equilibrium (NCNE) and denote it $(\overline{e}_{i,t}, i \in \mathcal{N}, t \in \mathcal{T})$, $(\overline{h}_{i,t}, i \in \mathcal{N}, t \in \mathcal{T})$ and $(\overline{z}_t, t \in \mathcal{T})$

In the second case, the assumption is made that while still seeking their interest as expressed by (5), all countries also do take into account the external effects induced on the other countries by their emission decisions. This is formalized by the solution of the alternative, and single, dynamic optimization problem consisting in maximizing the sum over all i's of (5) with respect to all control variables e_{it} and h_{it}, $i = 1, 2, ..., n$, subject as before to (2) and (3) whereby the state variables are defined. Internationally, the outcome of such coordinated behaviors is in the nature of Pareto efficient strategies of the cooperative version of the game defined above. We call this outcome the International Optimum (IO) and denote it $(e^{\mathcal{N}}_{i,t}, i \in \mathcal{N}, t \in \mathcal{T})$, $(h^{\mathcal{N}}_{i,t}, i \in \mathcal{N}, t \in \mathcal{T})$ and $(z^{\mathcal{N}}_t, t \in \mathcal{T})$.

The well known difference between the Nash equilibrium and Pareto efficient solutions just stated prompts the following normative question: can a policy instrument be devised whereby the countries would be induced to move away from the Nash equilibrium trajectory, which is inefficient, and adopt instead the Pareto efficient one, while being assured that they would not loose from this move, neither individually nor by forming coalitions? This is exactly what is achieved by the Chander-Tulkens transfer scheme mentioned above whose dynamic extension by Germain et al. 1998 was formulated for the one-dimensional case only. We now turn to the two-dimensional economic setting introduced in the previous subsection.

2 Transfers to stabilize the international optimum

2.1 The transfers at the final time T

The matter is handled by backward induction.

2.1.1 The partial Nash equilibrium w.r.t. a coalition U at time T

If a coalition forms, there first should be specified which state of the system is going to prevail, that is, which paths are going to be followed by the countries. For that purpose, we transpose to the present dynamic context a concept introduced by Chander and Tulkens (1997) and define the partial Nash equilibrium w.r.t. a coalition U at time T (PANE_U$_T$) as follows. Let U be a subset of $\mathcal{N}$ formed by countries that behave cooperatively *between themselves*. The other countries are assumed to behave individually. Formally, for given $\mathbf{k}_{\mathcal{N},T-1} =_{def} \{k_{i,T-1}, i \in \mathcal{N}\}$ and z_{T-1} :
(i) the members of coalition U maximize jointly the sum of their utilities :

$$\max_{\mathbf{h}_{UT}, \mathbf{e}_{UT}} \sum_{i \in U} [F_i(k_{iT}, e_{iT}) - h_{iT} - \pi_i z_T] \tag{6}$$

where $\mathbf{h}_{U,T} =_{def} \{h_{iT}, i \in U\}$, $\mathbf{e}_{U,T} =_{def} \{e_{iT}, i \in U\}$, under constraints (2) and (3) and taking $e_{j,T}, h_{j,T}$ $(\forall j \notin U)$ as given.
(ii) each country outside the coalition solves :

$$\max_{h_{iT}, e_{iT}} F_i(k_{iT}, e_{iT}) - h_{iT} - \pi_i z_T, \ i \notin U \tag{7}$$

under constraints (2) and (3), and taking $e_{j,T}, h_{j,T}$ $(\forall j \neq i)$ as given.

Given (2), the previous objectives can be equivalently maximized w.r.t. $\mathbf{k}_T$ and $\mathbf{e}_T$. The simultaneous solution of these problems leads to a set of FOC that fully characterize the PANE_U$_T$:
(i) for the members of coalition U :

$$\frac{\partial F_i}{\partial k_{iT}} = 1 \tag{8}$$

$$\frac{\partial F_i}{\partial e_{iT}} = \pi_U =_{def} \sum_{i \in U} \pi_i \tag{9}$$

(ii) for a country outside the coalition :

$$\frac{\partial F_i}{\partial k_{iT}} = 1 \tag{10}$$

$$\frac{\partial F_i}{\partial e_{iT}} = \pi_i \tag{11}$$

Let k_{iT}^U, e_{iT}^U $(i \in \mathcal{N})$ be the solution of the FOC (8) to (11) and

$$z_T^U = [1-\gamma]z_{T-1} + \sum_{i=1}^{n} e_{iT}^U \tag{12}$$

The payoffs of the countries are given by the following value functions :

$$w_{iT}^U(k_{i,T-1}, z_{T-1}) = F_i(k_{iT}^U, e_{iT}^U) - k_{iT}^U + [1-\delta_i]k_{i,T-1} - \pi_i z_T^U,\ i \in \mathcal{N} \tag{13}$$

so that the payoff of coalition U is given by the value function :

$$W_T^U(\mathbf{k}_{U,T-1}, z_{T-1}) = \sum_{i \in U} w_{iT}^U(k_{i,T-1}, z_{T-1}) = \sum_{i \in U} \left[F_i(k_{iT}^U, e_{iT}^U) - k_{iT}^U + [1-\delta_i]k_{i,T-1} - \pi_i z_T^U\right] \tag{14}$$

where $\mathbf{k}_{U,T-1} =_{def} \{k_{i,T-1}, i \in U\}$.

2.1.2 Two particular cases

The PANE_U$_T$ generalizes the two important particular outcomes mentioned earlier, namely the NCNE and the IO. Indeed on the one hand, the non-cooperative Nash equilibrium at time T (NCNE$_T$) is obtained when the coalition is a singleton ($U = v =_{def} \{i\}$ for any i), i.e. all countries are assumed to behave individually. Then the payoffs of the countries are given by the following value functions:

$$w_{iT}^v(k_{i,T-1}, z_{T-1}) = F_i(k_{iT}^v, e_{iT}^v) - k_{iT}^v + [1-\delta_i]k_{i,T-1} - \pi_i z_T^v,\ i \in \mathcal{N} \tag{15}$$

where k_{iT}^v, e_{iT}^v are the solution of the FOC (8) to (11) when $U = v$, and

$$z_T^v = [1-\gamma]z_{T-1} + \sum_{i=1}^{n} e_{iT}^v \tag{16}$$

On the other hand, the international optimum at date T (IO$_T$) obtains when the coalition contains all countries ($U = \mathcal{N}$) and all members are supposed to maximize jointly the sum of their utilities. The payoff of country i at the optimum is given by the following value function :

$$w_{iT}^{\mathcal{N}}(k_{i,T-1}, z_{T-1}) = F_i(k_{iT}^{\mathcal{N}}, e_{iT}^{\mathcal{N}}) - k_{iT}^{\mathcal{N}} + [1-\delta_i]k_{i,T-1} - \pi_i z_T^{\mathcal{N}} \tag{17}$$

so that the global payoff of all countries is given by the value function :

$$W_T^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},T-1}, z_{T-1}) = \sum_{i=1}^{n} w_{iT}^{\mathcal{N}}(k_{i,T-1}, z_{T-1}) = \sum_{i\in\mathcal{N}} \left[F_i(k_{iT}^{\mathcal{N}}, e_{iT}^{\mathcal{N}}) - k_{iT}^{\mathcal{N}} + [1-\delta_i]k_{i,T-1} - \pi_i z_T^{\mathcal{N}}\right] \tag{18}$$

where $\mathbf{k}_{\mathcal{N},T-1} =_{def} \{k_{i,T-1}, i \in \mathcal{N}\}$, $k_{iT}^{\mathcal{N}}, e_{iT}^{\mathcal{N}}$ $(i \in \mathcal{N})$ are the solution of the FOC (8) to (11) when $U = \mathcal{N}$, and

$$z_T^{\mathcal{N}} = [1-\gamma]z_{T-1} + \sum_{i=1}^{n} e_{iT}^{\mathcal{N}} \tag{19}$$

By definition of the optimum, one verifies that

$$W_T^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},T-1}, z_{T-1}) \geq \sum_{i=1}^{n} w_{iT}^{U}(k_{i,T-1}, z_{T-1}),\ \forall U \subset \mathcal{N} \tag{20}$$

Thus, from a collective point of view, the IO$_T$ is preferable to any solution of the partial agreement type, the least preferred being the NCNE$_T$. The difference $W_T^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},T-1}, z_{T-1}) - \sum_{i=1}^{n} w_{iT}^{v}(k_{i,T-1}, z_{T-1})$ is called by Chander and Tulkens (1997) the *ecological surplus* resulting from international cooperation.

However (20) may not be sufficient to ensure cooperation[4]. Indeed if $\exists U \subset \mathcal{N}$ such that $\sum_{i\in U} w_{iT}^{\mathcal{N}}(k_{i,T-1}, z_{T-1}) < W_T^{U}(\mathbf{k}_{U,T-1}, z_{T-1})$, then coalition U will not cooperate without compensation for the lower payoff it obtains.

2.1.3 Transfers

Since the horizon of time is limited to the single period $t = T$, one can use the transfers formula proposed by Chander and Tulkens (1997) in a static framework. The idea is that the required compensation can be achieved by an appropriate sharing of the ecological surplus. Let

$$\begin{aligned}\theta_{iT}(\mathbf{k}_{\mathcal{N},T-1}, z_{T-1}) \quad = \quad & -[w_{iT}^{\mathcal{N}}(k_{i,T-1}, z_{T-1}) - w_{iT}^{v}(k_{i,T-1}, z_{T-1})] \\ & + \frac{\pi_i}{\pi_{\mathcal{N}}}[W_T^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},T-1}, z_{T-1}) - \sum_{i=1}^{n} w_{iT}^{v}(k_{i,T-1}, z_{T-1})]\end{aligned} \tag{21}$$

be the transfer (> 0 if received, < 0 if paid) to country i at time T, where $\pi_{\mathcal{N}} =_{def} \sum_{i\in\mathcal{N}} \pi_i$. By construction, the budget of the transfers defined by (21) is balanced, i.e. :

$$\sum_{i=1}^{n} \theta_{iT}(\mathbf{k}_{T-1}, z_{T-1}) = 0 \tag{22}$$

[4]This is why we refrain from calling the IO an international *cooperative* optimum, as is often, and mistakenly, asserted in the literature. Cooperation results from individual and coalitional rationality; efficiency, i.e. the fulfilment of (18), is not a sufficient condition for cooperation.

Then at the international optimum country i's payoff *including* transfers becomes :

$$\begin{aligned}\widetilde{w}_{iT}^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},T-1}, z_{T-1}) &= w_{iT}^{\mathcal{N}}(k_{i,T-1}, z_{T-1}) + \theta_{iT}(\mathbf{k}_{\mathcal{N},T-1}, z_{T-1}) \\ &= w_{iT}^{v}(k_{i,T-1}, z_{T-1}) + \frac{\pi_i}{\pi_{\mathcal{N}}}[W_T^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},T-1}, z_{T-1}) - \sum_{i=1}^{n} w_{iT}^{v}(k_{i,T-1}, z_{T-1})]\end{aligned} \tag{23}$$

Since the ecological surplus is positive, one has :

$$\widetilde{w}_{iT}^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},T-1}, z_{T-1}) - w_{iT}^{v}(k_{i,T-1}, z_{T-1}) = \frac{\pi_i}{\pi_{\mathcal{N}}}[W_T^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},T-1}, z_{T-1}) - \sum_{i=1}^{n} w_{iT}^{v}(k_{i,T-1}, z_{T-1})] \geq 0, \ \forall i \in \mathcal{N} \tag{24}$$

Thus cooperation with the transfers (21) is *individually rational* at time T, in the sense that each country is better off at the IO *with these transfers* than at the non cooperative Nash equilibrium,

Following Chander and Tulkens (1997), one also has the following stronger result[5] :

$$\sum_{i \in U} \widetilde{w}_{iT}^{\mathcal{N}}(k_{i,T-1}, z_{T-1}) \geq W_T^{U}(\mathbf{k}_{U,T-1}, z_{T-1}), \ \forall U \subset \mathcal{N} \tag{25}$$

i.e. cooperation with transfers is *coalitionally rational, or self enforcing* at time T, in the sense that any coalition would enjoy only a lower payoff than the one it would obtain at the IO_T *with transfers,* if it were to defect from this IO_T and revert to the PANE_T w.r.t. itself. In terms of game theory, such a property is summarized by saying that the vector $(\widetilde{w}_{iT}^{\mathcal{N}}, i \in N)$ is an imputation that belongs to the core of a cooperative game associated with the economic model. Full details are given in the paper just quoted.

2.2 The transfers at earlier periods

Countries know that, whatever they do previously to T, transfers exist (defined by (21)) that make the IO_T preferable for each of them with respect to all other solutions. The problem we wish to consider now is whether transfers can be designed that make all countries interested in cooperating for periods previous to T as well.

Let us suppose that there exists a sequence of transfers that makes the IO_τ's preferable for all countries at times $\{\tau = t+1, ..., T\}$. We make two important behavioral assumptions :
(i) these transfers induce effectively cooperation from $t+1$ onwards[6], and
(ii) *all* countries therefore (rationally) expect at t that they will cooperate in all subsequent periods $\{t+1, ..., T\}$.

Given these assumptions, we show in the following that it is indeed possible to define transfers that make the IO_t preferable for all countries at time t.

[5]The proof of their Theorem 1 carries over directly to the present case because the variables $k_{i,T-1}$ and z_{T-1} only play a parametrical role here.The expression (25) holds whatever the inherited stock of pollutant z_{T-1} and the vector of capital stocks $\mathbf{k}_{\mathcal{N},T-1}$.

[6]Note that, following Chander and Tulkens (1997, section 5), one could indeed obtain the cooperative optimum with transfers as an equilibrium, called *ratio-equilibrium.*

2.2.1 The partial fallback equilibrium w.r.t. a coalition U at time t

Let U be a subset of $\mathcal{N}$ formed by countries that behave cooperatively *between themselves at* t (while expecting international cooperation sustained by transfers from $t+1$ onwards) and take the other countries' investment and energy consumption as given. This coalition will thus hold only for period t. The other countries are assumed to behave individually. Formally, for given $\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}$:
(i) the members of coalition U maximize jointly the sum of their utilities :

$$\max_{\mathbf{h}_{U,t},\mathbf{e}_{U,t}} \sum_{i\in U} \left[F_i(k_{i,t}, e_{i,t}) - h_{i,t} - \pi_i z_t + \rho \widetilde{w}_{it+1}^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t}, z_t)\right] \tag{26}$$

where $\mathbf{h}_{U,t} =_{def} \{h_{i,t}, i \in U\}, \mathbf{e}_{U,t} =_{def} \{e_{i,t}, i \in U\}$, under constraints (2) and (3), taking $e_{j,t}, h_{j,t}$ $(\forall j \notin U)$ as given and $\widetilde{w}_{it+1}^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t}, z_t)$ is defined as in (23) where t is substituted for T. $\widetilde{w}_{it+1}^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t}, z_t)$ is the future payoff obtained by country i at the IO's during $\{t+1, ..., T\}$.
(ii) each country outside the coalition solves :

$$\max_{h_{i,t},e_{i,t}} F_i(k_{i,t}, e_{i,t}) - h_{i,t} - \pi_i z_t + \rho \widetilde{w}_{it+1}^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t}, z_t),\ i \notin U \tag{27}$$

under constraints (2) and (3), taking $e_{j,t}, h_{j,t}$ $(\forall j \neq i)$ as given.

As before, given (2), the previous objectives can be equivalently maximized w.r.t. $\mathbf{k}_t$ and $\mathbf{e}_t$. Solving simultaneously these problems leads to a set of FOC that fully characterize the partial fallback equilibrium w.r.t. coalition U at time t (PAFE_U$_t$) :
(i) for the members of coalition U :

$$\frac{\partial F_i}{\partial k_{i,t}} = 1 - \rho \sum_{i\in U} \frac{\partial \widetilde{w}_{it+1}^{\mathcal{N}}}{\partial k_{i,t}} \tag{28}$$

$$\frac{\partial F_i}{\partial e_{i,t}} = \pi_U - \rho \sum_{i\in U} \frac{\partial \widetilde{w}_{it+1}^{\mathcal{N}}}{\partial z_t} \tag{29}$$

(ii) for a country outside the coalition :

$$\frac{\partial F_i}{\partial k_{i,t}} = 1 - \rho \frac{\partial \widetilde{w}_{it+1}^{\mathcal{N}}}{\partial k_{i,t}}, \tag{30}$$

$$\frac{\partial F_i}{\partial e_{i,t}} = \pi_i - \rho \frac{\partial \widetilde{w}_{it+1}^{\mathcal{N}}}{\partial z_t} \tag{31}$$

Let $k_{i,t}^U, e_{i,t}^U$ $(i \in \mathcal{N})$ be the solutions of the FOC (28) to (31), and $z_t^U = [1-\gamma]z_{t-1} + \sum_{i=1}^n e_{i,t}^U$. Then the payoffs of the countries are given by the following value functions :

$$w_{i,t}^U(k_{i,t-1}, z_{t-1}) = F_i(k_{i,t}^U, e_{i,t}^U) - k_{i,t}^U + [1-\delta_i]k_{i,t-1} - \pi_i z_t^U + \rho \widetilde{w}_{it+1}^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t}^U, z_t^U),\ i \in \mathcal{N} \tag{32}$$

so that the payoff of coalition U is given by the value function :

$$\begin{aligned} W_t^U(\mathbf{k}_{U,t-1}, z_{t-1}) &= \sum_{i\in U} w_{i,t}^U(k_{i,t-1}, z_{t-1}) \\ &= \sum_{i\in U} \left[\begin{array}{c} F_i(k_{i,t}^U, e_{i,t}^U) - k_{i,t}^U + [1-\delta_i]k_{i,t-1} - \pi_i z_t^U \\ +\rho \widetilde{w}_{it+1}^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t}^U, z_t^U) \end{array} \right] \end{aligned}$$

where $\mathbf{k}_{U,t-1} =_{def} \{k_{i,t-1}, i \in U\}$. Notice that all the above is defined for $t < T$.

2.2.2 Two particular cases

As before, the PAFE_U$_t, t < T$ generalizes two important particular cases : the fallback non-cooperative equilibrium and the international optimum. These are similar but not identical to the parallel cases defined above for $t = T$.

The fallback non-cooperative equilibrium at time t (FNCE$_t$): The fallback non-cooperative equilibrium is obtained when the coalition is a singleton ($U = v = \{i\}$ for any $i \in \mathcal{N}$), i.e. all countries are assumed to behave in an individualistic manner at time t (and cooperatively afterwards). Then the payoffs of the countries are given by the following value functions :

$$w_{i,t}^{v}(k_{i,t-1}, z_{t-1}) = F_i(k_{i,t}^{v}, e_{i,t}^{v}) - k_{i,t}^{v} + [1-\delta_i]k_{i,t-1} - \pi_i z_t^{v} + \rho\widetilde{w}_{it+1}^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t}^{v}, z_t^{v}),\ i \in \mathcal{N} \tag{33}$$

where $k_{i,t}^{v}, e_{i,t}^{v}$ are solution of the FOC (28) to (31) when $U = v$, and

$$z_t^{v} = [1-\gamma]z_{t-1} + \sum_{i=1}^{n} e_{i,t}^{v} \tag{34}$$

For every $t < T$, the key difference between the FNCE$_t$ and the NCNE$_T$ lies in the presence of the terms $\rho\widetilde{w}_{it+1}^{\mathcal{N}}$ in (33) compared to (15) : this is where our behavioral assumption on expectations comes into play.

The international optimum at time t (IO$_t$): The international optimum is obtained when the coalition contains all countries ($U = \mathcal{N}$). Thus all countries are supposed to behave cooperatively (i.e. they maximize jointly the sum of their utilities) at time t as well as in the whole future. The payoff of country i at the IO is given by the following value function :

$$w_{i,t}^{\mathcal{N}}(k_{i,t-1}, z_{t-1}) = F_i(k_{i,t}^{\mathcal{N}}, e_{i,t}^{\mathcal{N}}) - k_{i,t}^{\mathcal{N}} + [1-\delta_i]k_{i,t-1} - \pi_i z_t^{\mathcal{N}} + \rho\widetilde{w}_{it+1}^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t}^{\mathcal{N}}, z_t^{\mathcal{N}}),\ i \in \mathcal{N} \tag{35}$$

where $\mathbf{k}_{\mathcal{N},t} =_{def} \{k_{i,t}, i \in \mathcal{N}\}$, $k_{it}^{\mathcal{N}}, e_{it}^{\mathcal{N}}$ $(i \in \mathcal{N})$ are solution of the FOC (28) to (31) when $U = \mathcal{N}$, and

$$z_t^{\mathcal{N}} = [1-\gamma]z_{t-1} + \sum_{i=1}^{n} e_{it}^{\mathcal{N}} \tag{36}$$

Thus the global payoff of all countries is given by the value function :

$$\begin{aligned} W_t^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) &= \sum_{i=1}^{n} w_{i,t}^{\mathcal{N}}(k_{i,t-1}, z_{t-1}) \qquad (37)\\ &= \sum_{i\in\mathcal{N}} \left[\begin{array}{c} F_i(k_{i,t}^{\mathcal{N}}, e_{i,t}^{\mathcal{N}}) - k_{i,t}^{\mathcal{N}} + [1-\delta_i]k_{i,t-1} - \pi_i z_t^{\mathcal{N}} \\ +\rho\widetilde{w}_{it+1}^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t}^{\mathcal{N}}, z_t^{\mathcal{N}}) \end{array} \right] \end{aligned}$$

By definition of the IO, one verifies that

$$W_t^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) \geq \sum_{i=1}^{n} w_{i,t}^{U}(k_{i,t-1}, z_{t-1}),\ \forall U \subset \mathcal{N} \tag{38}$$

From a collective point of view, the IO$_t$ is preferable to any solution ot the partial agreement type.

However (38) may not be sufficient to ensure cooperation. Indeed if $\exists U \subset \mathcal{N}$ such that $\sum_{i\in U} w_{it}^{\mathcal{N}}(k_{i,t-1}, z_{t-1}) < W_t^{U}(\mathbf{k}_{U,t-1}, z_{t-1})$, then coalition U will not cooperate without compensation for the lower payoff it obtains.

2.2.3 Transfers at time t

To induce all countries to cooperate to the achievement of the international optimum at time t, we proceed as in period T (see subsection 2.1.3). Let

$$\begin{aligned} \theta_{i,t}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) &= -[w^{\mathcal{N}}_{i,t}(k_{i,t-1}, z_{t-1}) - w^{v}_{i,t}(k_{i,t-1}, z_{t-1})] \\ &\quad + \frac{\pi_i}{\pi_{\mathcal{N}}}[W^{\mathcal{N}}_t(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) - \sum_{i=1}^{n} w^{v}_{i,t}(k_{i,t-1}, z_{t-1})] \end{aligned} \tag{39}$$

be the transfer (> 0 if received, < 0 if paid) to country i at time t, where as before $\pi_{\mathcal{N}} = \sum_{i\in\mathcal{N}} \pi_i$. By construction, the budget of the transfers defined by (39) is balanced, i.e. :

$$\sum_{i=1}^{n} \theta_{i,t}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) = 0 \tag{40}$$

The difference $W^{\mathcal{N}}_t(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) - \sum_{i=1}^{n} w^{v}_{i,t}(k_{i,t-1}, z_{t-1})$ is the ecological surplus resulting from the *extension* of international cooperation to period t. Then at the international optimum country i's payoff *including* transfers becomes

$$\begin{aligned} \widetilde{w}^{\mathcal{N}}_{i,t}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) &= w^{\mathcal{N}}_{i,t}(k_{i,t-1}, z_{t-1}) + \theta_{i,t}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) \\ &= w^{v}_{i,t}(k_{i,t-1}, z_{t-1}) \\ &\quad + \frac{\pi_i}{\pi_{\mathcal{N}}}[W^{\mathcal{N}}_t(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) - \sum_{i=1}^{n} w^{v}_{i,t}(k_{i,t-1}, z_{t-1})] \end{aligned} \tag{41}$$

Since the ecological surplus is positive, one has :

$$\begin{aligned} & \widetilde{w}^{\mathcal{N}}_{i,t}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) - w^{v}_{i,t}(k_{i,t-1}, z_{t-1}) \\ = \; & \frac{\pi_i}{\pi_{\mathcal{N}}}[W^{\mathcal{N}}_t(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) - \sum_{i=1}^{n} w^{v}_{i,t}(k_{i,t-1}, z_{t-1})] \geq 0, \; \forall i \in \mathcal{N} \end{aligned}$$

Thus cooperation with transfers is *individually rational* at time t, in the sense that each country is better off at the IO_t *with transfers* than at the FNCE_t, whatever the inherited stock of pollutant z_{t-1} and the vector of capital stocks $\mathbf{k}_{\mathcal{N},t-1}$.

In section 4 below, we generalize Chander and Tulkens (1997) result to our dynamic setting and show that :

$$\sum_{i\in U} \widetilde{w}^{\mathcal{N}}_{i,t}(k_{i,t-1}, z_{t-1}) \geq W^{U}_t(\mathbf{k}_{U,t-1}, z_{t-1}), \; \forall U \subset \mathcal{N}, \; \forall t \in \mathcal{T} \tag{42}$$

i.e. cooperation with transfers is *rational in the sense of coalitions* at time t, in the sense that a coalition would enjoy a lower payoff than the one it would obtain at the IO_t *with transfers,* whatever the inherited stock of pollutant z_{t-1} and the vector of capital stocks $\mathbf{k}_{\mathcal{N},t-1}$. This is similar to the core property mentioned above with (25) for $t = T$. It is here extended to all $t < T$.

We have thus shown that :

- at time T, the IO_T with transfers is preferable for all countries to what they would obtain either as a member of a coalition or individually (in the sense of the PANE_U_T defined above);

- if $\forall \tau = t+1, ..., T$, the IO$_\tau$ with transfers is preferable for all countries to what they would obtain either as a member of a coalition or individually (in the sense of the PAFE_U$_t$ defined above), then the IO$_t$ with transfers is also preferable at time t.

Proceeding backwards from $t = T$ to $t = 1$, the final result is that all countries cooperate in each period (no coalition will ever form). This determines the emission and the capital stock levels in each period and also the trajectory of the stock of pollutant, given its initial value z_0. In turn this trajectory determines the values of the functions w_i^v, $w_i^{\mathcal{N}}$ and $\widetilde{w}_i^{\mathcal{N}}$, and therefore also the values of the transfers θ_i.

In the infinite horizon case, the backward reasoning considered above applies no more. However, we can consider the stationary solution by taking advantage of the fact that (i) the production functions F_i and damage functions D_i do not depend directly on time as well as (ii) the sharing parameters $\pi_i/\pi_{\mathcal{N}}$ not to depend directly on time either. The functional forms of the solutions thus only vary in time through the varying stocks $\mathbf{k}$ and z. The structure of the problem is then the same as in the finite horizon case.

3 Solution

We now compute the trajectories of the emissions, the capital stocks, the pollutant stock, as well as the values of the payoffs and of the transfers.

3.1 The system of equations

For each[7] $U \subseteq \mathcal{N}$ and each $t \in \mathcal{T}$, the PAFE_U$_t$ is characterized by the following payoffs :

$$W_t^U(\mathbf{k}_{U,t-1}, z_{t-1}) = \max_{\mathbf{k}_{U,t}, \mathbf{e}_{U,t}} \left\{ \sum_{i \in U} \left[\begin{array}{c} F_i(k_{i,t}, e_{i,t}) - k_{i,t} + [1-\delta_i]k_{i,t-1} - \pi_i z_t \\ +\rho w_{i,t+1}^v (k_{i,t}, z_t) \\ +\rho \frac{\pi_i}{\pi_{\mathcal{N}}} \left[W_{t+1}^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t}, z_t) - \sum_{j \in \mathcal{N}} w_{j,t+1}^v(k_{j,t}, z_t) \right] \end{array} \right] \right\} \tag{43}$$

$$w_{i,t}^U(k_{i,t-1}, z_{t-1}) = \max_{k_{i,t}, e_{i,t}} \left\{ \begin{array}{c} F_i(k_{i,t}, e_{i,t}) - k_{i,t} + [1-\delta_i]k_{i,t-1} - \pi_i z_t \\ +\rho w_{i,t+1}^v (k_{i,t}, z_t) \\ +\rho \frac{\pi_i}{\pi_{\mathcal{N}}} \left[W_{t+1}^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t}, z_t) - \sum_{j \in \mathcal{N}} w_{j,t+1}^v(k_{j,t}, z_t) \right] \end{array} \right\}, \ i \notin U \tag{44}$$

under constraint (3), with terminal values $W_{T+1}^U(\mathbf{k}_{U,T}, z_T) = w_{i,T+1}^U(k_{i,T}, z_T) = 0$. We observe that $W_t^U(\mathbf{k}_{U,t-1}, z_{t-1})$ and $w_{i,t}^U(k_{i,t-1}, z_{t-1})$ are linear function of their arguments. Let

$$W_t^U(\mathbf{k}_{U,t-1}, z_{t-1}) = P_t^U + \sum_{i \in U} q_{i,t}^U k_{i,t-1} + R_t^U z_{t-1}$$

and

$$w_{i,t}^U(k_{i,t-1}, z_{t-1}) = p_{i,t}^U + q_{i,t}^U k_{i,t-1} + r_{i,t}^U z_{t-1}, \ \forall i \notin U$$

where the parameters $P_t^U, q_{i,t}^U, R_t^U$ $(i \in U, t \in \mathcal{T})$ and $p_{i,t}^U, q_{i,t}^U, r_{i,t}^U$ $(i \notin U, \ t \in \mathcal{T})$ are to be identified.

[7] Clearly, for $U = \mathcal{N}$, the expression (44) disappears as well as those derived from it in the sequel.

Assume that $\mathbf{k}^U_{\mathcal{N},t}$ and $\mathbf{e}^U_{\mathcal{N},t}$ are the solution of the FOC characterising the PAFE_U$_t$. From (43) and (44), it follows that :

$$P^U_t + \sum_{i\in U} q^U_{i,t} k_{i,t-1} + R^U_t z_{t-1} = \sum_{i\in U} \left[\begin{array}{c} F_i(k^U_{i,t}, e^U_{i,t}) - k^U_{i,t} + [1-\delta_i] k_{i,t-1} - \pi_i z^U_t \\ +\rho \left[p^v_{i,t+1} + q^v_{i,t+1} k^U_{i,t} + r^v_{i,t+1} z^U_t \right] \\ +\rho \frac{\pi_i}{\pi_\mathcal{N}} \left[\begin{array}{c} P^\mathcal{N}_{t+1} + \sum_{j\in\mathcal{N}} q^\mathcal{N}_{j,t+1} k^U_{j,t} + R^\mathcal{N}_{t+1} z^U_t \\ - \sum_{j\in\mathcal{N}} \left[p^v_{j,t+1} + q^v_{j,t+1} k^U_{j,t} + r^v_{j,t+1} z^U_t \right] \end{array} \right] \end{array} \right] \tag{45}$$

and

$$p^U_{i,t} + q^U_{i,t} k_{i,t-1} + r^U_{i,t} z_{t-1} = \left[\begin{array}{c} F_i(k^U_{i,t}, e^U_{i,t}) - k^U_{i,t} + [1-\delta_i] k_{i,t-1} - \pi_i z^U_t \\ +\rho \left[p^v_{i,t+1} + q^v_{i,t+1} k^U_{i,t} + r^v_{i,t+1} z^U_t \right] \\ +\rho \frac{\pi_i}{\pi_\mathcal{N}} \left[\begin{array}{c} P^\mathcal{N}_{t+1} + \sum_{i\in\mathcal{N}} q^\mathcal{N}_{i,t+1} k^U_{i,t} + R^\mathcal{N}_{t+1} z^U_t \\ - \sum_{j\in\mathcal{N}} \left[p^v_{j,t+1} + q^v_{j,t+1} k^U_{i,t} + r^v_{j,t+1} z^U_t \right] \end{array} \right] \end{array} \right], \ i \notin U \tag{46}$$

where $z^U_t = [1-\gamma] z_{t-1} + E^U_t$ and $E^U_t =_{def} \sum_{i=1}^n e^U_{i,t}$. The parameter identification in the two members of (45) gives :

$$\begin{aligned} P^U_t &= \sum_{i\in U} \left[\begin{array}{c} F_i(k^U_{i,t}, e^U_{i,t}) - k^U_{i,t} - \pi_i E^U_t \\ +\rho \left[p^v_{i,t+1} + q^v_{i,t+1} k^U_{i,t} + r^v_{i,t+1} E^U_t \right] \\ +\rho \frac{\pi_i}{\pi_\mathcal{N}} \left[\begin{array}{c} P^\mathcal{N}_{t+1} + \sum_{j\in\mathcal{N}} q^\mathcal{N}_{j,t+1} k^U_{j,t} + R^\mathcal{N}_{t+1} E^U_t \\ - \sum_{j\in\mathcal{N}} \left[p^v_{j,t+1} + q^v_{j,t+1} k^U_{j,t} + r^v_{j,t+1} E^U_t \right] \end{array} \right] \end{array} \right] &(47) \\ q^U_{i,t} &= 1 - \delta_i, \ i \in U &(48) \\ R^U_t &= [1-\gamma] \sum_{i\in U} \left[-\pi_i + \rho r^v_{i,t+1} + \rho \frac{\pi_i}{\pi_\mathcal{N}} \left[R^\mathcal{N}_{t+1} - \sum_{j\in\mathcal{N}} r^v_{j,t+1} \right] \right] &(49) \end{aligned}$$

Similarly, the identification in the two members of (46) gives :

$$\begin{aligned} p^U_{it} &= \left[\begin{array}{c} F_i(k^U_{i,t}, e^U_{i,t}) - k^U_{i,t} - \pi_i E^U_t \\ +\rho \left[p^v_{i,t+1} + q^v_{i,t+1} k^U_{i,t} + r^v_{i,t+1} E^U_t \right] \\ +\rho \frac{\pi_i}{\pi_\mathcal{N}} \left[\begin{array}{c} P^\mathcal{N}_{t+1} + \sum_{j\in\mathcal{N}} q^\mathcal{N}_{j,t+1} k^U_{j,t} + R^\mathcal{N}_{t+1} E^U_t \\ - \sum_{j\in\mathcal{N}} \left[p^v_{j,t+1} + q^v_{j,t+1} k^U_{j,t} + r^v_{j,t+1} E^U_t \right] \end{array} \right] \end{array} \right], \ i \notin U &(50) \\ q^U_{i,t} &= 1 - \delta_i, \ i \notin U &(51) \\ r^U_{it} &= [1-\gamma] \left[-\pi_i + \rho r^v_{i,t+1} + \rho \frac{\pi_i}{\pi_\mathcal{N}} \left[R^\mathcal{N}_{t+1} - \sum_{j\in\mathcal{N}} r^v_{j,t+1} \right] \right], \ i \notin U &(52) \end{aligned}$$

Finally, the FOC derived from (43) and (44) lead to the following equations :

$$\begin{aligned} \frac{\partial F_i}{\partial k_{i,t}}(k^U_{i,t}, e^U_{i,t}) &= 1 - \rho q^v_{i,t+1} - \rho \left[q^\mathcal{N}_{i,t+1} - q^v_{i,t+1} \right] \sum_{j\in U} \frac{\pi_j}{\pi_\mathcal{N}}, \ i \in U &(53) \\ \frac{\partial F_i}{\partial k_{i,t}}(k^U_{i,t}, e^U_{i,t}) &= 1 - \rho q^v_{i,t+1} - \rho \left[q^\mathcal{N}_{i,t+1} - q^v_{i,t+1} \right] \frac{\pi_i}{\pi_\mathcal{N}}, \ i \notin U &(54) \\ \frac{\partial F_i}{\partial e_{i,t}}(k^U_{i,t}, e^U_{i,t}) &= \sum_{j\in U} \left[\pi_j - \rho r^v_{j,t+1} + \rho \frac{\pi_j}{\pi_\mathcal{N}} \left[R^\mathcal{N}_{t+1} - \sum_{l\in\mathcal{N}} r^v_{l,t+1} \right] \right], \ i \in U &(55) \\ \frac{\partial F_i}{\partial e_{i,t}}(k^U_{i,t}, e^U_{i,t}) &= \pi_i - \rho r^v_{i,t+1} + \rho \frac{\pi_i}{\pi_\mathcal{N}} \left[R^\mathcal{N}_{t+1} - \sum_{j\in\mathcal{N}} r^v_{j,t+1} \right], \ i \notin U &(56) \end{aligned}$$

Equations (47) to (56) form a system of equations whose unknowns are $\{(P_t^U, q_{i,t}^U\ (i \in U), R_t^U)$, $(p_{it}^U, q_{i,t}^U, r_{it}^U; i \notin U)$, $(k_{i,t}^U, e_{i,t}^U; i \in \mathcal{N}); t \in \mathcal{T}\}$ where all variables are nil when $t = T+1$.

3.2 Solving the system

3.2.1 $q_{i,t}^U$, $i \in \mathcal{N}, t \in \mathcal{T}$

These parameters are immediately known via (48) and (51).

3.2.2 R_t^U and r_{it}^U $(i \notin U)$, $t \in \mathcal{T}$

First one observes that in the particular case $U = \mathcal{N}$, (49) leads to $R_t^{\mathcal{N}} = [1-\gamma]\left[-\pi_{\mathcal{N}} + \rho R_{t+1}^{\mathcal{N}}\right]$ (recall that $\pi_{\mathcal{N}} = \sum_{j\in\mathcal{N}} \pi_j$). On the other hand, when $U = v$,
$r_{i,t}^v = [1-\gamma]\left[-\pi_i + \rho r_{i,t+1}^v + \rho \frac{\pi_i}{\pi_{\mathcal{N}}}\left[R_{t+1}^{\mathcal{N}} - \sum_{j\in\mathcal{N}} r_{i,t+1}^v\right]\right] \Rightarrow \sum_{i\in U} r_{i,t}^v = R_t^U$, $\forall U \subset \mathcal{N}$. This is particular true for $U = \mathcal{N}$. Then (49) reduces to :
(i) coalition U :

$$R_t^U = [1-\gamma]\left[-\pi_U + \rho R_{t+1}^U\right]$$

where $\pi_U =_{def} \sum_{i\in U} \pi_i$. Given that $R_{T+1}^U = 0$, it follows that :

$$R_t^U = -\pi_U[1-\gamma]\frac{1-\rho^{T+1-t}[1-\gamma]^{T+1-t}}{1-\rho[1-\gamma]},\ U \subset \mathcal{N} \tag{57}$$

When $U = v$,

$$r_{i,t}^v = -\pi_i[1-\gamma]\frac{1-\rho^{T+1-t}[1-\gamma]^{T+1-t}}{1-\rho[1-\gamma]},\ i \in \mathcal{N} \tag{58}$$

(ii) For non-members of the coalition, it follows that $r_{it}^U = [1-\gamma]\left[-\pi_i + \rho r_{i,t+1}^v\right] \Rightarrow r_{it}^U = r_{i,t}^v\ (i \notin U)$.

3.2.3 $k_{i,t}^U$ and $e_{i,t}^U$, $i \in \mathcal{N}, t \in \mathcal{T}$

FOC (53) to (56) reduce to :

$$\frac{\partial F_i}{\partial k_{i,t}}(k_{i,t}^U, e_{i,t}^U) = 1 - \rho\left[1-\delta_i\right],\ i \in \mathcal{N} \tag{59}$$

$$\frac{\partial F_i}{\partial e_{i,t}}(k_{i,t}^U, e_{i,t}^U) = \pi_U - \rho R_t^U,\ i \in U \tag{60}$$

$$\frac{\partial F_i}{\partial e_{i,t}}(k_{i,t}^U, e_{i,t}^U) = \pi_i - \rho r_{i,t}^v,\ i \notin U \tag{61}$$

Given that R_t^U and $r_{i,t}^v$ are known via (57) and (58), this system enables to compute the solutions $\mathbf{k}_{\mathcal{N},t}^U$ and $\mathbf{e}_{\mathcal{N},t}^U$.

For example, if F_i is the familiar Cobb-Douglas production function defined by $y_i = F_i(k,e) = k^{\alpha_i}e^{\beta_i}$ (with $0 < \alpha_i, \beta_i < \alpha_i + \beta_i < 1$), then :

$$
\begin{aligned}
k_{i,t}^U &= \left[\frac{1-\rho[1-\delta_i]}{\alpha_i}\right]^{[1-\beta_i]/[\alpha_i+\beta_i-1]} \left[\frac{\pi_U\{1-[\rho[1-\gamma]]^{T+1-t}\}}{\beta_i[1-\rho[1-\gamma]]}\right]^{\beta_i/[\alpha_i+\beta_i-1]} \\
e_{i,t}^U &= \left[\frac{1-\rho[1-\delta_i]}{\alpha_i}\right]^{\alpha_i/[\alpha_i+\beta_i-1]} \left[\frac{\pi_U\{1-[\rho[1-\gamma]]^{T+1-t}\}}{\beta_i[1-\rho[1-\gamma]]}\right]^{[1-\alpha_i]/[\alpha_i+\beta_i-1]} \\
y_{i,t}^U &= (k_{i,t}^U)^{\alpha_i}(e_{i,t}^U)^{\beta_i} = \left[\frac{1-\rho[1-\delta_i]}{\alpha_i}\right]^{\alpha_i/[\alpha_i+\beta_i-1]} \left[\frac{\pi_U\{1-[\rho[1-\gamma]]^{T+1-t}\}}{\beta_i[1-\rho[1-\gamma]]}\right]^{\beta_i/[\alpha_i+\beta_i-1]}.
\end{aligned}
$$

3.2.4 P_t^U and p_{it}^U $(i \notin U), t \in \mathcal{T}$

(47) reduces to :

$$
P_t^U = \sum_{i\in U}\left[F_i(k_{i,t}^U, e_{i,t}^U) - k_{i,t}^U - \pi_i E_t^U + \rho\left[p_{i,t+1}^v + [1-\delta_i]\,k_{i,t}^U + r_{i,t+1}^v E_t^U + \frac{\pi_i}{\pi_{\mathcal{N}}}\left[P_{t+1}^{\mathcal{N}} - \sum_{j\in\mathcal{N}} p_{j,t+1}^v\right]\right]\right]
$$

Thus

$$
\begin{aligned}
P_t^{\mathcal{N}} &= \sum_{i\in N}\left[F_i(k_{i,t}^{\mathcal{N}}, e_{i,t}^{\mathcal{N}}) - k_{i,t}^{\mathcal{N}}\right] - \pi_N E_t^{\mathcal{N}} + \rho\left[P_{t+1}^{\mathcal{N}} + \sum_{i\in\mathcal{N}}[1-\delta_i]\,k_{i,t}^{\mathcal{N}} + R_{t+1}^{\mathcal{N}} E_t^{\mathcal{N}}\right] \\
&= \sum_{\tau=t}^{T}\rho^{\tau-t}\left[\sum_{i\in N}\left[F_i(k_{i,\tau}^{\mathcal{N}}, e_{i,\tau}^{\mathcal{N}}) - k_{i,\tau}^{\mathcal{N}} + \rho\,[1-\delta_i]\,k_{i,\tau}^{\mathcal{N}} + \rho R_{\tau+1}^{\mathcal{N}} E_\tau^{\mathcal{N}}\right] - \pi_{\mathcal{N}} E_\tau^{\mathcal{N}}\right] \qquad (62)
\end{aligned}
$$

When $U = v$,

$$
p_{it}^v = F_i(k_{i,t}^v, e_{i,t}^v) - k_{i,t}^v - \pi_i E_t^v + \rho\left[p_{i,t+1}^v + [1-\delta_i]\,k_{i,t}^v + r_{i,t+1}^v E_t^v + \frac{\pi_i}{\pi_{\mathcal{N}}}\left[P_{t+1}^{\mathcal{N}} - \sum_{j\in\mathcal{N}} p_{i,t+1}^v\right]\right] \qquad (63)
$$

for all $i \in \mathcal{N}$. Summing on i gives :

$$
\sum_{i\in\mathcal{N}} p_{it}^v = \sum_{i\in\mathcal{N}}\left[F_i(k_{i,t}^v, e_{i,t}^v) - k_{i,t}^v\right] - \pi_{\mathcal{N}} E_t^v + \rho\left[P_{t+1}^{\mathcal{N}} + \sum_{i\in\mathcal{N}}[1-\delta_i]\,k_{i,t}^v + \sum_{i\in\mathcal{N}} r_{i,t+1}^v E_t^v\right] \qquad (64)
$$

(62) and (64) give :

$$
P_t^{\mathcal{N}} - \sum_{j\in\mathcal{N}} p_{j,t}^v = \sum_{i=1}^{n}[\Delta F_i - \Delta k_{i,t}] - \pi_{\mathcal{N}}\Delta E_t + \rho\left[\sum_{i=1}^{n}[1-\delta_i]\,\Delta k_{i,t} - R_{t+1}^{\mathcal{N}}\Delta E_t\right] \qquad (65)
$$

where by definition $\Delta x_t = x_t^{\mathcal{N}} - x_t^v$ is the difference between the values of x_t at the IO_t (where $U = \mathcal{N}$) and at the fallback non-cooperative equilibrium (where $U = v$). From the previous computations, all the quantities of the RHS are known so that $P_t^{\mathcal{N}} - \sum_{j\in\mathcal{N}} p_{j,t}^v$ is computable. Then $p_{i,t}$ $(\forall i)$ follows as well from the integration of (63) :

$$
p_{it}^v = \sum_{\tau=t}^{T}\rho^{\tau-t}\left[F_i(k_{i,\tau}^v, e_{i,\tau}^v) - k_{i,\tau}^v - \pi_i E_\tau^v + \rho\left[[1-\delta_i]\,k_{i,\tau}^v + r_{i,\tau+1}^v E_t^v + \frac{\pi_i}{\pi_{\mathcal{N}}}\left[P_{\tau+1}^{\mathcal{N}} - \sum_{j\in\mathcal{N}} p_{\tau,t+1}^v\right]\right]\right]
$$

and the whole system is solved.

4 Coalitional rationality

Assume that at date t there exists a sequence of future transfers that makes international cooperation rational in the sense of coalitions from $t+1$ onwards. Let $\mathbf{k}_t^{\mathcal{N}}$ and $z_t^{\mathcal{N}}$ be the vector of capital stocks and the stock of pollution at the international optimum at date t. Let $\mathbf{k}_t^v$ and z_t^v be the vector of capital stocks and the stock of pollution at the fallback non-cooperative equilibrium at date t. Let $(w_{1t}^{\mathcal{N}}, ..., w_{nt}^{\mathcal{N}})$ and $(w_{1t}^v, ..., w_{nt}^v)$ be the vectors of the countries'payoffs at the IO *without current transfers* and at the fallback non-cooperative equilibrium at date t respectively. Let $W_t^{\mathcal{N}}$ and W_t^v be the sum of the components of these vectors. Let the vector of countries'payoffs *with current transfers* $(\widetilde{w}_{1,t}^{\mathcal{N}}, ..., \widetilde{w}_{n,t}^{\mathcal{N}})$ be defined by

$$\widetilde{w}_{i,t}^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) = w_{i,t}^{\mathcal{N}}(k_{i,t-1}, z_{t-1}) + \theta_{i,t}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}),\ i \in \mathcal{N}$$

where the vector of current transfers $(\theta_{1t}, ..., \theta_{nt})$ is such that

$$\theta_{i,t}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) = -[w_{i,t}^{\mathcal{N}}(k_{i,t-1}, z_{t-1}) - w_{i,t}^v(k_{i,t-1}, z_{t-1})] + \frac{\pi_i}{\pi_{\mathcal{N}}}[W_t^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) - W_t^v(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1})]. \tag{66}$$

Theorem : With the transfers (66), one has for every coalition $U \subset \mathcal{N}$:

$$\sum_{i \in U} \widetilde{w}_{i,t}^{\mathcal{N}}(k_{i,t-1}, z_{t-1}) \geq \sum_{i \in U} w_{i,t}^U(k_{i,t-1}, z_{t-1}) = W_t^U(\mathbf{k}_{U,t-1}, z_{t-1}) \tag{67}$$

that is, the vector $(\widetilde{w}_{i,t}^{\mathcal{N}}, i \in \mathcal{N})$ has the core property (42). Thus the IO_t with transfers (66) make international cooperation rational in the sense of coalitions at every date t as announced at the end of section 2.

Proof of the theorem : The proof makes use of the 3 following lemmas.

Lemma 0 : The transfers (66) can be rewritten :

$$\begin{aligned}\theta_{it}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) \quad &= \quad -\left[y_{i,t}^{\mathcal{N}} - y_{i,t}^v - [1 - \rho\,[1-\delta_i]]\left[k_{i,t}^{\mathcal{N}} - k_{i,t}^v\right]\right] \\ &\quad + \frac{\pi_i}{\pi_{\mathcal{N}}} \sum_{j=1}^n \left[y_{j,t}^{\mathcal{N}} - y_{j,t}^v - [1 - \rho\,[1-\delta_j]]\left[k_{j,t}^{\mathcal{N}} - k_{j,t}^v\right]\right]\end{aligned}$$

Proof : Recall that $W_t^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) = P_t^{\mathcal{N}} + \sum_{i\in\mathcal{N}} q_{i,t}^{\mathcal{N}} k_{i,t-1} + R_t^{\mathcal{N}} z_{t-1}$ and $W_t^v(\mathbf{k}_{t-1}, z_{t-1}) = \sum_{i=1}^n w_{i,t}^v(k_{i,t-1}, z_{t-1}) = \sum_{i=1}^n \left[p_{i,t}^v + q_{i,t}^v k_{i,t-1} + r_{i,t}^v z_{t-1}\right]$. Recall that $q_{i,t}^{\mathcal{N}} = q_{i,t}^v$ and $R_t^{\mathcal{N}} = \sum_{i=1}^n r_{i,t}^v$ (see section 3). It follows that

$$\theta_{i,t}(\mathbf{k}_{t-1}, z_{t-1}) = -[w_{i,t}^{\mathcal{N}}(k_{i,t-1}, z_{t-1}) - w_{i,t}^v(k_{i,t-1}, z_{t-1})] + \frac{\pi_i}{\pi_{\mathcal{N}}}\left[P_t^{\mathcal{N}} - \sum_{i=1}^n p_{i,t}^v\right] \tag{68}$$

Now :

$$\begin{aligned} & w_{i,t}^{\mathcal{N}}(k_{i,t-1}, z_{t-1}) \\ = \quad & y_{i,t}^{\mathcal{N}} - k_{i,t}^{\mathcal{N}} + [1-\delta_i]k_{i,t-1} - \pi_i z_t^{\mathcal{N}} + \rho[w_{i,t+1}^v(k_{i,t}^{\mathcal{N}}, z_t^{\mathcal{N}}) + \frac{\pi_i}{\pi_{\mathcal{N}}}\left[W_{t+1}^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t}^{\mathcal{N}}, z_t^{\mathcal{N}}) - W_{t+1}^v(\mathbf{k}_{\mathcal{N},t}^{\mathcal{N}}, z_t^{\mathcal{N}})\right] \\ = \quad & y_{i,t}^{\mathcal{N}} - k_{i,t}^{\mathcal{N}} + [1-\delta_i]k_{i,t-1} - \pi_i z_t^{\mathcal{N}} + \rho\left[p_{i,t+1}^v + q_{i,t+1}^v k_{i,t}^{\mathcal{N}} + r_{i,t+1}^v z_t^{\mathcal{N}} + \frac{\pi_i}{\pi_{\mathcal{N}}}\left[P_{t+1}^{\mathcal{N}} - \sum_{i=1}^n p_{i,t+1}^v\right]\right]\end{aligned}$$

Similarly,

$$\begin{aligned}
& w_{i,t}^{v}(k_{i,t-1}, z_{t-1}) \\
= \ & y_{i,t}^{v} - k_{i,t}^{v} + [1-\delta_i]k_{i,t-1} - \pi_i z_t^v + \rho[w_{i,t+1}^{v}(k_{i,t}^{v}, z_t^v) + \frac{\pi_i}{\pi_{\mathcal{N}}}\left[W_{t+1}^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t}^{v}, z_t^v) - W_{t+1}^{v}(\mathbf{k}_{\mathcal{N},t}^{v}, z_t^v)\right] \\
= \ & y_{i,t}^{v} - k_{i,t}^{v} + [1-\delta_i]k_{i,t-1} - \pi_i z_t^v + \rho\left[p_{i,t+1}^{v} + q_{i,t+1}^{v}k_{i,t}^{v} + r_{i,t+1}^{v}z_t^v + \frac{\pi_i}{\pi_{\mathcal{N}}}\left[P_{t+1}^{\mathcal{N}} - \sum_{i=1}^{n} p_{i,t+1}^{v}\right]\right]
\end{aligned}$$

Thus, making use of the solution for $q_{i,t+1}^{v}$ and $r_{i,t+1}^{v}$ (recall (48), (51) and (58)), of $z_t^v = [1-\gamma]z_{t-1} + E_t^v$ and of $z_t^{\mathcal{N}} = [1-\gamma]z_{t-1} + E_t^{\mathcal{N}}$,

$$\begin{aligned}
w_{i,t}^{\mathcal{N}}(k_{i,t-1}, z_{t-1}) - w_{i,t}^{v}(k_{i,t-1}, z_{t-1}) &= y_{i,t}^{\mathcal{N}} - y_{i,t}^{v} - [1-\rho\,[1-\delta_i]]\left[k_{i,t}^{\mathcal{N}} - k_{i,t}^{v}\right] + \pi_i\left[E_t^{\mathcal{N}} - E_t^{v}\right] \\
& \quad -\pi_i[1-\gamma]\frac{1-\rho^{T-t}[1-\gamma]^{T-t}}{1-\rho[1-\gamma]}\left[E_t^{\mathcal{N}} - E_t^{v}\right]
\end{aligned}$$

On the other hand, given (65) :

$$\begin{aligned}
P_{t+1}^{\mathcal{N}} - \sum_{i=1}^{n} p_{i,t+1}^{v} &= \sum_{i=1}^{n}\left[y_{i,t}^{\mathcal{N}} - y_{i,t}^{v} - \left[k_{i,t}^{\mathcal{N}} - k_{i,t}^{v}\right] - \pi_i\left[E_t^{\mathcal{N}} - E_t^{v}\right]\right] \\
& \quad +\rho\left[\sum_{i=1}^{n}[1-\delta_i]\left[k_{i,t}^{\mathcal{N}} - k_{i,t}^{v}\right] - \pi_{\mathcal{N}}[1-\gamma]\frac{1-\rho^{T-t}[1-\gamma]^{T-t}}{1-\rho[1-\gamma]}\left[E_t^{\mathcal{N}} - E_t^{v}\right]\right]
\end{aligned}$$

Putting these two last expressions in (68) completes the proof.■

Lemma 1 : At date t, comparing the fallback non-cooperative equilibrium and the PAFE_U, one has

$$\text{(a) } y_{i,t}^{U} - [1-\rho\,[1-\delta_i]]\,k_{i,t}^{U} \ > \ y_{i,t}^{v} - [1-\rho\,[1-\delta_i]]\,k_{i,t}^{v}, \ \ i \in U \qquad (69)$$

$$\text{(b) } y_{i,t}^{U} - [1-\rho\,[1-\delta_i]]\,k_{i,t}^{U} \ = \ y_{i,t}^{v} - [1-\rho\,[1-\delta_i]]\,k_{i,t}^{v}, \ \ i \notin U \qquad (70)$$

Proof : (i) The FOC characterizing the PAFE_U$_t$ are :

$$\frac{\partial F_i}{\partial k_{i,t}}(k_{i,t}^{U}, e_{i,t}^{U}) \ = \ 1-\rho\,[1-\delta_i] =_{def} \mu_i \qquad (71)$$

$$\frac{\partial F_i}{\partial e_{i,t}}(k_{i,t}^{U}, e_{i,t}^{U}) \ = \ \pi\left[1+\rho[1-\gamma]\frac{1-\rho^{T-t}[1-\gamma]^{T-t}}{1-\rho[1-\gamma]}\right] =_{def} \pi\eta_t \qquad (72)$$

with

$$\begin{aligned}
\pi &= \pi_U \text{ if } i \in U \\
\pi &= \pi_i \text{ if } i \notin U
\end{aligned}$$

One observes that the FOC are the same at the FNCE$_t$ and at the PAFE_U$_t$ for $i \notin U$, so that $k_{i,t}^{U} = k_{i,t}^{v}$ and $e_{i,t}^{U} = e_{i,t}^{v} \Rightarrow y_{i,t}^{U} = y_{i,t}^{v}$ $(i \notin U)$. Thus (b) is proved.

(ii) Let us differentiate totally (71) and (72) w.r.t. π :

$$\begin{aligned}\frac{\partial^2 F_i}{\partial k_{i,t}^2}dk_{i,t}+\frac{\partial^2 F_i}{\partial k_{i,t}\partial e_{i,t}}de_{i,t} &= 0 &(73)\\ \frac{\partial^2 F_i}{\partial k_{i,t}\partial e_{i,t}}dk_{i,t}+\frac{\partial^2 F_i}{\partial e_{i,t}^2}de_{i,t} &= \eta_t d\pi &(74)\end{aligned}$$

Solving the system leads to :

$$de_{i,t}=\frac{\frac{\partial^2 F_i}{\partial k_{i,t}^2}}{\frac{\partial^2 F_i}{\partial k_{i,t}^2}\frac{\partial^2 F_i}{\partial e_{i,t}^2}-\left[\frac{\partial^2 F_i}{\partial k_{i,t}\partial e_{i,t}}\right]^2}d\pi =_{def} \Delta_i d\pi \tag{75}$$

$\Delta_i < 0$ because of the assumptions on F_i (see section 1). On the other hand, $y_{i,t}=F_i(k_{i,t},e_{i,t}) \Rightarrow$

$$dy_{i,t}=\frac{\partial F_i}{\partial k_{i,t}}dk_{i,t}+\frac{\partial F_i}{\partial e_{i,t}}de_{i,t} \tag{76}$$

Given (71) and (75), one has at the PAFE_U$_t$:

$$dy_{i,t}-\mu_i dk_{i,t}=\frac{\partial F_i}{\partial e_{i,t}}\eta_t d\pi < 0 \text{ if } d\pi>0 \tag{77}$$

because of the assumptions on F_i (see section 1) and $\eta_t > 0$.

For coalition members, moving from the FNCE$_t$ to the PAFE_U$_t$ can be seen as a succession of small increments $d\pi$, starting at π_i an finishing at $\pi_U=\sum_{i\in U}\pi_i$ $(>\pi_i)$. So that point (a) follows.■

Lemma 2 : if $\exists\, U\subset\mathcal{N}$ such that :

$$W_t^U(\mathbf{k}_{U,t-1},z_{t-1}) > \sum_{i\in U}\widetilde{w}_{it}^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t-1},z_{t-1}) \tag{78}$$

then the vector $(\widehat{w}_{1t},...,\widehat{w}_{nt})$ defined by

$$\begin{aligned}\widehat{w}_{it}(\mathbf{k}_{\mathcal{N},t-1},z_{t-1}) &= w_{it}^{\mathcal{N}}(k_{i,t-1},z_{t-1})+\widehat{\theta}_{it}(\mathbf{k}_{\mathcal{N},t-1},z_{t-1}),\ i\in\mathcal{N} &(79)\\ \text{where } \widehat{\theta}_{it}(\mathbf{k}_{\mathcal{N},t-1},z_{t-1}) &= -[w_{it}^{\mathcal{N}}(k_{i,t-1},z_{t-1})-w_{it}^{U}(k_{i,t-1},z_{t-1})] &(80)\\ &\quad +\frac{\pi_i}{\pi_{\mathcal{N}}}\left[W_t^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t-1},z_{t-1})-\sum_{j=1}^{n}w_{it}^{U}(k_{i,t-1},z_{t-1})\right]\\ \text{where } w_{it}^{U}(k_{i,t-1},z_{t-1}) &= F_i(k_{i,t}^U,e_{i,t}^U)-k_{i,t}^U+[1-\delta_i]k_{i,t-1}-\pi_i[[1-\gamma]z_{t-1}+E_t^U] &(81)\\ &\quad +\rho w_{i,t+1}^{v}\left(k_{i,t}^U,[1-\gamma]z_{t-1}+E_t^U\right)\\ &\quad +\rho\frac{\pi_i}{\pi_{\mathcal{N}}}\left[W_{t+1}^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t}^U,[1-\gamma]z_{t-1}+E_t^U)-W_{t+1}^{v}(\mathbf{k}_{\mathcal{N},t}^U,[1-\gamma]z_{t-1}+E_t^U)\right]\end{aligned}$$

dominates $(\widetilde{w}_{1t}^{\mathcal{N}},...,\widetilde{w}_{nt}^{\mathcal{N}})$ in the sense that

$$\begin{aligned}\text{(i) }\sum_{i\in U}\widehat{w}_{it}(\mathbf{k}_{\mathcal{N},t-1},z_{t-1}) &> \sum_{i\in U}\widetilde{w}_{it}^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t-1},z_{t-1}) &(82)\\ \text{(ii) }\widehat{w}_{it}(\mathbf{k}_{\mathcal{N},t-1},z_{t-1}) &\geq \widetilde{w}_{it}^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t-1},z_{t-1}),\ i\notin U &(83)\end{aligned}$$

Proof : (i) (79) et (80) $\Rightarrow$

$$\begin{aligned}\widehat{w}_{it}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) &= w_{it}^U(k_{i,t-1}, z_{t-1}) + \frac{\pi_i}{\pi_{\mathcal{N}}}\left[W_t^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) - \sum_{j=1}^{n} w_{it}^U(k_{i,t-1}, z_{t-1})\right] \\ &\geq w_{it}^U(k_{i,t-1}, z_{t-1})] \end{aligned} \tag{84}$$

because the term between brackets is necessarily positive by definition of the IO. Then

$$\sum_{i\in U}\widehat{w}_{it}(\mathbf{k}_{t-1}, z_{t-1}) \geq \sum_{i\in U} w_{it}^U(k_{i,t-1}, z_{t-1}) = W_t^U(\mathbf{k}_{U,t-1}, z_{t-1}) > \sum_{i\in U}\widetilde{w}_{it}^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1})$$

where the first inequality results from (84) and the second follows from assumption (78). Thus (82) is verified.

(ii) (83) can be rewritten

$$\begin{aligned}\widehat{w}_{it}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) &= w_{it}^{\mathcal{N}}(k_{i,t-1}, z_{t-1}) + \widehat{\theta}_{it}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) \\ &\geq \widetilde{w}_{it}^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) = w_{it}^{\mathcal{N}}(k_{i,t-1}, z_{t-1}) + \theta_{it}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}),\ i\notin U\end{aligned}$$

Thus verifying (83) is equivalent to verity that

$$\widehat{\theta}_{it}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) \geq \theta_{it}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}),\ i\notin U$$

Now, by Lemma 0,

$$\begin{aligned}\theta_{it}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) &= -\left[y_{i,t}^{\mathcal{N}} - y_{i,t}^{v} - [1-\rho[1-\delta_i]]\left[k_{i,t}^{\mathcal{N}} - k_{i,t}^{v}\right]\right] \\ &\quad + \frac{\pi_i}{\pi_{\mathcal{N}}}\sum_{j=1}^{n}\left[y_{j,t}^{\mathcal{N}} - y_{j,t}^{v} - [1-\rho[1-\delta_j]]\left[k_{j,t}^{\mathcal{N}} - k_{j,t}^{v}\right]\right]\end{aligned}$$

By a similar reasoning,

$$\begin{aligned}\widehat{\theta}_{it}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) &= -\left[y_{i,t}^{\mathcal{N}} - y_{i,t}^{U} - [1-\rho[1-\delta_i]]\left[k_{i,t}^{\mathcal{N}} - k_{i,t}^{U}\right]\right] \\ &\quad + \frac{\pi_i}{\pi_{\mathcal{N}}}\sum_{j=1}^{n}\left[y_{j,t}^{\mathcal{N}} - y_{j,t}^{U} - [1-\rho[1-\delta_j]]\left[k_{j,t}^{\mathcal{N}} - k_{j,t}^{U}\right]\right]\end{aligned}$$

Then $\widehat{\theta}_{it}(\mathbf{k}_{t-1}, z_{t-1}) \geq \theta_{it}(\mathbf{k}_{t-1}, z_{t-1})$ $(i\notin U) \Rightarrow$

$$-\left[y_{i,t}^{U} - y_{i,t}^{v} - [1-\rho[1-\delta_i]]\left[k_{i,t}^{U} - k_{i,t}^{v}\right]\right] + \frac{\pi_i}{\pi_{\mathcal{N}}}\sum_{j=1}^{n}\left[y_{j,t}^{U} - y_{j,t}^{v} - [1-\rho[1-\delta_j]]\left[k_{j,t}^{U} - k_{j,t}^{v}\right]\right] \geq 0,\ i\notin U$$

This is indeed true because of Lemma 1, so (83) is verified.■

Proof of the theorem : (82) and (83) $\Rightarrow$

$$\sum_{i=1}^{n}\widehat{w}_{it}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) > \sum_{i=1}^{n}\widetilde{w}_{it}^{\mathcal{N}}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1})$$

$$\Rightarrow \sum_{i=1}^{n} \left[w_{it}^{\mathcal{N}}(k_{i,t-1}, z_{t-1}) + \widehat{\theta}_{it}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) \right] > \sum_{i=1}^{n} \left[w_{it}^{\mathcal{N}}(k_{i,t-1}, z_{t-1}) + \theta_{it}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) \right] \tag{85}$$

$$\Rightarrow \sum_{i=1}^{n} \widehat{\theta}_{it}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) > \sum_{i=1}^{n} \theta_{it}(\mathbf{k}_{\mathcal{N},t-1}, z_{t-1}) \tag{86}$$

Now this last result is impossible because these two sums are by construction equal to 0. This contradicts the thesis of Lemma 2, so the theorem is demonstrated.■

5 Bibliography

Chander, P. and H. Tulkens, 1992 "Theoretical foundations of negotiations and cost sharing in transfrontier pollution problems", European Economic Review, 36 (2/3), 288-299.

Chander, P. and H. Tulkens, 1995, A core-theoretic solution for the design of cooperative agreements on transfrontier pollution, International Tax and Public Finance 2, 279-293.

Chander, P. and H. Tulkens, 1997, The core of an economy with multilateral environmental externalities, International Journal of Game Theory 26, 379-401.

d'Aspremont, C. and J. Jaskold Gabszewicz, 1986, "On the stability of collusion", chapter 8 (pp. 243-264) of J.E. Stiglitz and G.F. Mathewson (eds), New developments in the analysis of market structure, Cambridge (Mass.), The MIT Press.

Eyckmans, J. and H. Tulkens, 2003, "Simulating coalitionally stable burden sharing agreements for the climate change problem", Resource and Energy Economics 25, 299-327.

Germain M., Ph. Toint and H. Tulkens, 1998, Financial transfers to ensure cooperative international optimality in stock pollutant abatement, chapter 11 in Faucheux, S., Gowdy, J. and Nicolaï, I. eds, Sustainability and Firms: Technological Change and the Changing Regulatory Environment, Edward Elgar, Cheltenham.

Germain, M., H. Tulkens and A. de Zeeuw, 1998, Stabilité stratégique en matière de pollution internationale avec effet de stock : le cas linéaire, Revue Economique 49(6), 1435-1454.

Germain, M., P. Toint, H. Tulkens and A. de Zeeuw, 2003, "Transfers to Sustain Dynamic Core-Theoretic Cooperation in International Stock Pollutant Control", Journal of Economic Dynamics and Control 28, 79-99.

Hoel, M., 1992, Emission taxes in a dynamic international game of CO$_{2}$ emissions, in: R. Pethig, ed., Conflicts and Cooperation in Managing Environmental Resources, Microeconomic Studies (Springer Verlag, Berlin).

Jorgensen S., G. Martin-Herran and G. Zaccour, 2003, Agreeability and time consistency in linear-state differential games, Journal of Optimization Theory and Applications, 119(1), p. 49-63.

Kaitala, V. K.G. Mäler and H. Tulkens 1995, "The acid rain game as a resource allocation process, with application to negotiations between Finland, Russia and Estonia", The Scandinavian Journal of Economics, 97(2) 325-343. Reprinted in Carlo Carraro, Ed., Governing the Global Environment, The Globalization of the World Economy, An Elgar Reference Collection, Cheltenham 2003.

Kaitala, V., M. Pohjola and O. Tahvonen, 1992, Transboundary air pollution and soil acidification: A dynamic analysis of an acid rain game between Finland and the USSR, Environmental and Resource Economics 2, 161-181.

Mäler, K.-G., 1989, The acid rain game, in: H. Folmer and E. van Ierland, eds., Valuation Methods and Policy Making in Environmental Economics (Elsevier, Amsterdam).

Petrosjan, L. and G. Zaccour, 1995, A multistage supergame of downstream pollution, G-95-14, GERAD, Ecole des Hautes Etudes Commerciales, Université de Montréal.

van der Ploeg, F. and A. de Zeeuw, 1992, International aspects of pollution control, Environmental and Resource Economics 2, 117-139.

Rubio, S.J. and A. Ulph 2007, An Infinite-horizon Model of Dynamic Membership of International Environmental Agreements, mimeo .

Tahvonen, O., 1996, Carbon dioxyde abatement as a differential game, European Journal of Political Economy 10(4), 685-705.

Tulkens, H. (1979). An Economic Model of International Negotiations Relating to Transfrontier Pollution, in K. Krippendor (ed.), Communication and Control in Society, Gordon and Breach, New York, 199-212.

Institut de Recherches Économiques et Sociales
Université catholique de Louvain

Place Montesquieu, 3
1348 Louvain-la-Neuve, Belgique

ISSN 1379-244X D/2009/3082/015