Department of Economics
University of Southampton
Southampton SO17 1BJ
UK

Discussion Papers in Economics and Econometrics

AN INFINITE-HORIZON MODEL OF DYNAMIC MEMBERSHIP OF INTERNATIONAL ENVIRONMENTAL AGREEMENTS

Santiago J Rubio
Alistair Ulph

No. 0210

1. Introduction.

There is now an extensive literature on international environmental agreements (see Barrett (2002) and Finus (2001) for excellent recent books summarising this literature). Yet, with a few exceptions discussed below, this literature works with simple static models of pollution despite the fact that many of the important problems (climate change, ozone depletion, acid rain) on which this literature seeks to shed light involve stock pollutants. In this paper we introduce a simple infinite-horizon model of a stock pollutant in which membership of an IEA changes over time as the stock of pollution varies.

There are a small number of papers which consider the formation of IEAs to deal with a stock pollutant. Rubio and Casino (2001) use the concept of self-enforcing IEAs familiar from the work of Carraro and Siniscalco (1993) and Barrett (1994) but extend these models to allow for a stock pollutant. It is well-know that the static model generates pessimistic results, in which when a large number of countries join an IEA, the gains from the IEA are small. As in the static model, the dynamic model is analysed as a two-stage game. Countries first of all decide whether or not to join an IEA. Then countries choose their paths of emissions. These emission paths are calculated by solving a differential game in either open-loop or feedback strategies assuming that the IEA signatories act to maximise their joint welfare while non-signatories just maximise individual welfare. Having solved for the emission paths and hence evaluated payoffs to signatories and non-signatories, Rubio and Casino then ask how many countries will want to join the IEA, using the same kind of stability analysis as in the static model. But in this model the dynamics of the stock pollutant affect only emissions strategies, not IEA membership. Many of the attempts to model empirically how many countries might join an IEA to deal with climate change have the same feature that countries are assumed to make a once-for-all decision whether to join an IEA, with the dynamics of the stock pollutant affecting only emissions paths and hence present-value payoffs (see, e.g. Eyckmans (2001)).

Germain, Toint, Tulkens and de Zeeuw [GTTZ] (2002) extend the framework of Chander and Tulkens (1995) to a dynamic model of a stock pollutant. As is now well understood the Chander and Tulkens approach, based on core concepts, is able to obtain the more optimistic conclusion that the grand coalition will be formed, because they assume that if one country defects from the grand coalition all countries will revert to non-cooperative behaviour. This punishment is sufficiently severe to deter defections. By contrast the stability analysis of Barrett and others assumes that if one country leaves an IEA, the remaining members will act to optimise their joint interests. With asymmetric countries it is necessary to use income transfers to ensure the stability of the grand coalition. In an important paper, GTTZ (2002) extend this analysis by showing that it is possible to devise dynamic transfers to ensure stability of the grand coalition when there is a stock pollutant. The membership of the grand coalition is thus maintained over time, although, importantly, in this case this is achieved by appropriate design of the transfers, rather than just by assuming that membership decisions are taken once-and-for-all.

Both Rubio and Casino and GTTZ analyse models in which an IEA operates over the whole life of a stock pollutant. Karp and Sacheti (1997) consider a two-period model of a stock pollutant but assume that an IEA will only form in one of the periods – either just in the first, because the IEA will fall apart at the end of one period, or in the second, because there may be substantial delays in forming an IEA. They then assess how the incentives to join an IEA[1] are affected by the dynamics of the stock pollutant, as well as by differences in the extent to which the pollutant is local or global and the extent to which planners discount the future. In Karp and Sacheti membership of the IEA varies sharply over time, but in a way that is exogenously imposed.

In an earlier paper (Rubio and Ulph (2002)) we extended the model of self-enforcing IEAs found in Carraro and Siniscalco (1993), Barrett (1994) to a two-period model of a stock pollutant. We studied two ways in which IEA membership might be decided. In the *fixed* membership model we followed Rubio and Casino and assumed that countries made a once-for-all decision at the outset whether or not to join an IEA. In the *variable* membership model we assumed that countries decided each period whether or not to join an IEA. We showed that in the variable membership case the

[1] They also use a rather different approach to modeling an IEA as a "modest" perturbation on a non-cooperative Nash equilibrium.

number of countries would rise over time, and that for a wide range of parameter values, variable membership gave higher global welfare than fixed membership. So allowing membership to vary over time matters. In one version of the variable membership model there was an extreme case of the rising membership model in which no country would join in the first period, but some countries would join in the second period. This provided an explanation for one of the patterns of membership imposed exogenously by Karp and Sacheti.

However, the two-period model has the undesirable characteristic that, even if unit damage costs are an increasing function of the stock of pollution, emissions generated in the second period are less damaging than emissions generated in the first period simply because their effects are felt for a shorter time horizon. It was this which drove the result that membership would rise over time. In this paper we extend the model of Rubio and Ulph (2002) to an infinite-horizon model, but we consider only the variable IEA membership case. The model retains a number of the special simplifying features of Rubio and Ulph (2002), in particular the assumption that countries are identical and that in each period countries have to make a discrete choice of emissions (pollute or abate). We first analyse the outcomes when all countries act non-co-operatively and when all countries act cooperatively. We then analyse IEAs where membership varies over time. We show that there exists a steady-state stock of pollution, which lies between the cooperative and non-cooperative steady-states, with a corresponding steady-state IEA membership. We show that membership is a decreasing function of the stock of pollution, so that if the initial stock of pollution lies below (above) the steady-state, then the membership will decline (rise) as the stock moves towards steady-state. The crucial parameters in this model are those which determine the level of unit damage costs and how quickly these costs rise with the stock of pollution. We show that as these parameters increase, both the initial and steady-state memberships decline, until, in the limit, membership is 2 in every time period. So for high values of these parameters membership dynamics disappear. Not surprisingly, it is for these parameter values that the potential gains from cooperation are greatest. These results are just the dynamic generalisation of the pessimistic static results.

2. The Model

There are N identical countries. Consider a typical period (we ignore time subscripts). Each country can choose a level of emissions q = 0,1, which we interpret to mean it can either abate or pollute. The total of emissions by all other countries is denoted by Q, so total emissions in the period are $q + Q$. Suppose at the start of the period the cumulative stock of emissions is z. Then the cumulative stock of emissions at the start of the next period is $\rho z + q + Q$ where ρ $(0 < \rho < 1)$ is the *decay factor* per period. We denote the initial stock of emissions by $z_0 \geq 0$.

If a country pollutes in the period it derives a unit of benefit which we shall normalise to1 and assume is constant over time. On the other hand, each unit of the stock of pollution at the start of the period generates for each country a unit of damage costs, $\gamma(z)$, which is a strictly increasing function of the stock of pollution at the start of the period. Thus the flow of net benefits to a country in the period is given by: $\pi(q,Q,z) = q - \gamma(z)z$. Finally, the *discount factor* per period is δ, $(0 < \delta < 1)$.

In the rest of this section we consider what happens when all countries act non-cooperatively, and when all countries act cooperatively. In the next section we analyse international environmental agreements.

2.1 Non-Cooperative Equilibrium.

Let $U(z)$ be the present value of current and all future net benefits to a country when the stock of emissions at the start of the current period is z and in each period each country selects its optimal non-cooperative emission strategy. Then, in the current period a typical country takes as given the total emissions of all other countries, Q, and chooses its emission strategy q =0, 1 to maximise present value payoff:

$$\Pi(q,Q,z) = \pi(q,Q,z) + \delta U(\rho z + q + Q) = q - \gamma(z)z + \delta U(\rho z + q + Q).$$

Then it is straightforward to see that, for given Q and z, it pays a country to pollute if

$$1 \geq \delta[U(\rho z + Q) - U(\rho z + Q + 1)] \qquad (1)$$

We shall show shortly that *U(z)* is a strictly decreasing and concave function of *z*. So (1) has the usual interpretation that it will pay a country to pollute if the instantaneous gain to it from a unit of emissions is greater than the reduction in present value future net benefits it will suffer from a unit increase in the stock of emissions. Given the properties of *U(z)* a sufficient condition for each country to pollute no matter what decisions other countries make is that (1) holds for *N*-1. We shall show shortly that, with $Q = N - 1$, as $z \to \infty$, (1) cannot hold. However, we shall shortly make an assumption on parameter values, Assumption A, which ensures that (1) holds for $Q = N - 1$ and $0 \leq z \leq \bar{z}$, where $\bar{z}$ is an upper bound on the stock of emissions sufficiently large that the stock of emissions never reaches $\bar{z}$. It then follows trivially that:

Claim 1. *Given Assumption A, for all relevant values of z, namely,* $0 \leq z \leq \bar{z}$ *the optimal non-cooperative strategy is for every country to pollute in every period.*

Then, the value function *U* is defined by the recursive equation:

$$U(z) = 1 - \gamma(z)z + \delta U(\rho z + N). \qquad (2)$$

Define $\tilde{z}$ as the *steady-state* stock of pollution when countries act non-cooperatively. Then $\tilde{z} = \frac{N}{1-\rho}$, and, as we shall show in Assumption A, $\tilde{z} < \bar{z}$. From (2) we derive:

$$U(\tilde{z}) = \frac{1 - \gamma(\tilde{z})\tilde{z}}{1-\delta}$$

so we can characterise the value function in steady-state, and clearly $U'(\tilde{z}) < 0$, so the value function must be decreasing in *z* when the stock of emissions reaches its steady-state value. Moreover it is clear that if $z_0 < (>)\tilde{z}$, then the stock will rise (fall) monotonically to its steady-state value.

To make further progress, we shall consider particular functional forms so that we can derive the value function explicitly. We assume that $\gamma(z) = \lambda + \mu z, \quad \lambda > 0, \mu > 0$ and that *U(z)* takes the form $U(z) = a - bz - cz^2$. Substituting in (2) we get:

$$\begin{aligned} a - bz - cz^2 &= 1 - (\lambda + \mu z)z + \delta[a - b(\rho z + N) - c(\rho z + N)^2] \\ &= [1 + \delta a - \delta bN - \delta cN^2] - [\lambda + \delta\rho b + 2\delta\rho cN]z - [\mu + \delta\rho^2 c]z^2 \end{aligned}$$

Thus *U(z)* is indeed quadratic, and equating coefficients we get:

Result 1. *For the infinite horizon non-cooperative equilibrium with unit damage cost* $\gamma(z) = \lambda + \mu z$*, the value function takes the form* $U(z) = a - bz - cz^2$ *where:*

$$c = \frac{\mu}{1 - \delta\rho^2} > 0; \quad b = \frac{\lambda + 2\delta\rho Nc}{(1 - \delta\rho)} > 0; \quad a = \frac{[1 - \delta Nb - \delta N^2 c]}{1 - \delta}$$

Thus $U'(z) < 0, U''(z) < 0$. Finally, using the specific functional forms, we can rewrite (1) as:

$$1 \geq \delta[b + c + 2c(\rho z + N - 1)] \qquad (1')$$

As noted earlier, it is straightforward to see that, since $c > 0$, the RHS of (1') increases linearly in *z*, and so must fail to hold for large enough values of *z*. We define $\bar{z} = \phi\tilde{z} = \phi\frac{N}{1 - \rho}$ where $\phi > \max\, [1, \frac{z_0}{\tilde{z}}]$. If we ensure that condition (1') holds for $\bar{z}$, then in the non-cooperative model it will always pay countries to pollute, since, by definition of $\bar{z}$ and the dynamics of the non-cooperative model, the stock of emissions can never lie above $\bar{z}$. Substituting for values of *b, c* and defining: $\vartheta \equiv \frac{\delta}{1 - \delta\rho}; \quad \zeta \equiv \frac{\delta}{1 - \delta\rho^2}, \quad \varphi \equiv \frac{\phi\rho}{(1 - \rho)} + \frac{1}{(1 - \delta\rho)} > 1$

Assumption A *For the infinite horizon non-cooperative equilibrium with unit damage cost* $\gamma(z) = \lambda + \mu z$*, we assume that parameters satisfy the condition:*

$$1 \geq \vartheta\lambda + \zeta(2N\varphi - 1)\mu \qquad \text{(A1)}$$

To summarise, in the non-cooperative equilibrium with unit damage costs $\lambda + \mu z$, where λ, μ satisfy Assumption A, the optimal non-cooperative strategy is for all countries to pollute in all periods. Starting from the initial stock of emissions, z_0, the stock will rise (fall) monotonically to the steady-state value $\tilde{z}$ assuming that $z_0 < (>) \tilde{z}$. The present value of net benefits for each country is given by $U(z_0) = a - bz_0 - cz_0^2$ where *a, b, c* are given in Result 1.

2.2 Cooperative Equilibrium

Now suppose that all countries cooperate and let *V(z)* be the present value of current and future net benefits each country will receive if the stock of emissions at the start of the current period is *z* and in each period the countries collectively choose their optimal cooperative strategy. To determine this optimal strategy, suppose total current emissions of all other countries is *Q*. Then the optimal current period strategy for a particular country to maximise collective present value of current and future net benefits will be to abate if:

$$1 \leq \delta N[V(\rho z + Q) - V(\rho z + Q + 1)] \qquad (3)$$

We shall show shortly that *V'* < 0, so (3) has the obvious interpretation that it will be optimal for a country to abate if the current benefit it derives from emitting a unit of pollution is less than the present value loss imposed on all countries by increasing the stock of emissions by one unit. We shall show shortly, that provided parameters satisfy Assumption B, then (3) holds for all $Q > 0$, $0 \leq z \leq \bar{z}$. Then we get, trivially,

Claim 2 *For all values of z,* $0 \leq z \leq \bar{z}$, *the optimal cooperative strategy is for all countries to abate pollution in each period.*

Given this optimal strategy, it is clear that the stock of emissions will decline to its steady-state value, 0, and that the value function is given by the recursive relation:

$$V(z) = -\gamma(z)z + \delta V(\rho z) \qquad (4)$$

To get further we again assume the specific functional form for the unit damage cost $\gamma(z) = \lambda + \mu z$, and that the value function takes the quadratic form $V(z) = \alpha - \beta z - \chi z^2$. Then (4) becomes:

$$\alpha - \beta z - \chi z^2 = -(\lambda + \mu z)z + \delta\alpha - \delta\beta\rho z - \delta\chi\rho^2 z^2$$

Thus the value function is indeed quadratic, and equating coefficients we have:

Result 2 *For the infinite-horizon cooperative equilibrium with unit damage cost* $\gamma(z) = \lambda + \mu z$ *the value function takes the form* $\alpha - \beta z - \chi z^2$ *where* $\alpha = 0, \quad \beta = \dfrac{\lambda}{1 - \delta\rho}, \quad \chi = \dfrac{\mu}{1 - \delta\rho^2}$.

For these specific functional forms (3) becomes:

$$1 \le \delta N[\beta + \chi + 2\chi(\rho z + Q)] \qquad (3')$$

For this inequality to hold for all non-negative Q and z it is sufficient that it holds for $Q = z = 0$. Substituting for β, χ we need the following assumption on parameters to guarantee that (3') is satisfied.

Assumption B. *For the cooperative equilibrium with unit damage cost* $\lambda + \mu z, \quad \lambda, \mu$*, satisfy the condition:*

$$1 \le N\vartheta\lambda + N\zeta\mu \qquad \text{(A2)}$$

To summarise, in the cooperative equilibrium with unit damage costs $\lambda + \mu z$, where λ, μ satisfy Assumption B, the optimal cooperative strategy is for all countries to abate in all periods. Starting from initial stock of emissions, z_0, the stock will fall monotonically to the steady-state value 0. The present value of net benefits for each country is given by $V(z_0) = \alpha - \beta z_0 - \chi z_0^2$ where α, β, χ are given in Result 2.

Finally, to ensure that the cooperative and non-cooperative equilibria have the properties set out above, we need to check that we can find parameter values that satisfy both Assumptions A and B, i.e. $\vartheta\lambda + \zeta(2N\varphi - 1)\mu \le 1 \le N\vartheta\lambda + N\zeta\mu$. It is straightforward to see that if $\frac{1}{N} < \vartheta\gamma < 1; \quad 0 < \mu < \frac{1-\vartheta\lambda}{\zeta(2N\varphi - 1)}$ then both Assumptions A and B are satisfied. For future reference, the way we shall select parameters λ, μ is to select θ_1, θ_2 as any two numbers lying strictly between 0 and 1. Then we set:

$$\lambda = \frac{1}{\vartheta}[\theta_1 + \frac{1-\theta_1}{N}]; \quad \overline{\mu}(\theta_1) \equiv \frac{1-\vartheta\lambda}{\zeta(2N\varphi - 1)}; \quad \mu = \theta_2 \overline{\mu}(\theta_1) \qquad \text{(A3)}$$

3. International Environmental Agreements with Dynamic Membership

We now consider the formation of a sequence of International Environmental Agreements (IEAs), in which, in each period, countries are free to join or leave an IEA. The model of IEA formation in each period is a dynamic version of the model of self-enforcing or stable IEAs introduced in the work of Carraro and Siniscalco (1993) and Barrett (1994), and well surveyed in Finus (2001). The model can be viewed as a two-stage game, in which countries first decide whether or not to join an IEA and then determine their emissions. In the second-stage emission game, non-signatory countries (whom we denote by the symbol f to mean free-rider or fringe country) choose emissions in a non-cooperative fashion similar to section 2.1; signatory countries (whom we denote by the symbol s) act in a cooperative fashion similar to 2.2, choosing emissions to maximise the total net benefits of all signatory countries, taking as given the behaviour of non-signatories. In the first-stage membership game we look for a stable IEA in which no individual signatory country would wish to switch to be a non-signatory, and no non-signatory would wish to switch to being a signatory. This is equivalent to a Nash equilibrium of the membership game. Before getting into detailed analysis we make three general remarks about our modelling strategy.

Firstly, the key difference between our model and the work of Barrett, Carraro and Sinsicalco *et al* is that we allow the membership to vary over time as the stock of emissions varies. Following Rubio and Ulph (2002), in which we analyse a two-period version of this model, we call this a *variable membership* model. As we noted in the introduction, this contrasts with the way in which IEAs for a stock pollutant are modelled by, for example, Rubio and Casino (2001) where it is assumed that countries make a once-for-all decision at the outset whether or not to join an IEA. In Rubio and Ulph (2002) we called this a *fixed membership* model, and for the two-period model we were able to contrast the outcomes with fixed and variable membership models. In this paper we consider only the variable membership model.

Secondly, we shall show that the optimal strategy for signatories in each period is to abate pollution. However, we do not provide any analysis of why signatories find it in their interests to abide by this strategy rather than free-ride. Rather, like much of the work on stable IEAs, the analysis of the stability of IEA membership is conducted independently of any analysis of why signatories abide by their agreement. Since the analysis of commitment by IEA signatories usually rests on repeated game arguments, (see, for example Barrett (1997)) it can be argued that the failure to use the dynamic nature of our model to analyse both what determines the stable size of IEA each period and what makes countries stick to their agreements is a serious limitation of our analysis. We recognise this issue and hope to rectify it in future research.

Thirdly, since we have assumed all countries are identical all we can determine is the size of the stable IEA in each period. We cannot say which countries become signatories in any period. This leads to an important modelling assumption. For in conducting stability analysis we need to be able to assess how a decision by a country to change its status in one period (i.e. to switch from being a signatory to a non-signatory or vice versa) affects its payoff not just in that current period but in future periods. The way a status change affects a country's future net benefits depends on two issues. First, a change of status in the current period will affect the stock of emissions in future periods, which will in turn affect the size of future IEA membership. We will take account of this through the dependency of the value function and the size of IEA membership on the stock of pollution. Second, future net benefits will also depend on whether the country contemplating a change of status in the current period will be a signatory or not in future periods. But, as just noted, we have no way of determining *which* countries will be signatories or non-signatories. To get round this, we shall assume that each country believes that in each period there is a random process for determining which countries become signatories, such that the probability of any country being a signatory in that period is simply the size of the stable IEA in that period divided by the total number of countries. This probability of being a signatory is clearly the same for all countries, and is independent of whether a country was a signatory or non-signatory in previous periods. Since the size of membership varies over time, obviously the probability of being a signatory in any period varies over time. So each country has the same *expected* present value of future net benefits, which will depend on the stock of emissions at the start of next period. This corresponds to what, in Rubio and Ulph (2002), we called the *Random Assignment Rule*. Of course a more realistic treatment might be to assume that current status has an important impact on future status, and in the two-period model of Rubio and Ulph (2002) we contrasted the outcome using the Random Assignment Rule with an alternative model (the *Status Quo Assignment Rule*) in which current status is an important determinant of future status. However, while that is tractable in a two-period model, it becomes much harder to analyse in an infinite horizon model.

With this preamble, we define *W(z)* as the *expected* present value of current and future net benefits to a country when (i) the stock of emissions at the start of the current period is z; (ii) in each period the size of membership of an IEA is that which constitutes a unique stable IEA for that period; (iii) in each period, signatories and non-signatories choose their optimal emission strategies; and (iv) in each period there is a random process for determining which countries become signatories, with each country having an equal probability of being a signatory. We now turn to the analysis of the second stage emission game and then the analysis of the stable IEA.

3.1 Second-stage Emission Game

Suppose that in a period with initial stock of emissions z, the outcome of the first-stage membership game is that there are n signatories to an IEA. We now need to derive the optimal emission strategies for non-signatories and signatories.

Non-Signatories

For a typical non-signatory country, suppose the total of current emissions of all other countries is Q. Then it will pay that non-signatory to pollute if

$$1 \geq \delta[W(\rho z + Q) - W(\rho z + Q + 1)] \qquad (5)$$

Not surprisingly this has the same form and interpretation as (1), except that there is a different value function.

Signatories

A typical signatory country will choose its emissions so as to maximise the total net benefits of the n signatories, recognising that all non-signatories will pollute.

Assuming all other signatories abate, it will pay any one signatory country to abate as long as:

$$1 \le \delta n[W(\rho z + N - n) - W(\rho z + N - n + 1)] \equiv \omega(n, z) \qquad (6)$$

Again, not surprisingly, (6) has the same form and interpretation as (3) except that there are only *n* countries who cooperate and there is a different value function.

We need to say more about the properties of $\omega(n, z)$. Clearly $\omega(0, z) = 0$. By comparing (5) and (6) it is readily seen that $\omega(1, z) < 1$. It will be useful to approximate $\omega(n, z) = -\delta n W'(\rho z + N - n)$. Then we have: $\omega_n = -\delta W' + \delta n W''$; $\omega_z = -\delta \rho n W''$. We shall show later that $W' < 0, W'' < 0$. So $\omega_z > 0$. We assume that $|W''|$ is sufficiently small that $\omega_n > 0$.

Define $n(z)$ as the value of *n* for which $\omega(n(z), z) = 1$. Totally differentiating we obtain: $\frac{dn}{dz} = -\frac{\omega_z}{\omega_n} < 0$. Hence:

Lemma 1 *For any z there is a unique positive value of n(z)* > 1*; moreover n(z) is a decreasing function of z.*

Define *m(z)* as the smallest integer no less than *n(z)*. Clearly *m(z)* ≥ 2. *m(z)* is the critical minimum size of IEA membership at which it just pays signatories to abate pollution. Then we have:

Result 3

(i) For all $n \ge m(z)$*, the optimal strategies are for non-signatories to pollute and signatories to abate, and the resulting payoffs to signatories and non-signatories are:*

$$W^s(n, z) = -\gamma(z)z + \delta W(\rho z + N - n); \quad W^f(n, z) = W^s(n, z) + 1$$

(ii) For all n < m(z), the optimal strategies are for both non-signatories and signatories to pollute and the payoffs to signatories and non-signatories are:

$$W^s(n, z) = W^f(n, z) = 1 - \gamma(z)z + \delta W(\rho z + N)$$

Thus Result 3 tells us for any *n, z* what the optimal emission strategies and payoffs are for signatories and non-signatories. We can now go back to the first-stage game and determine the size of membership which constitutes a stable IEA.

3.2 First-stage Membership Game

In a period with initial stock of emissions *z* we define an IEA of size $\hat{n}$ as stable if it satisfies the 2 properties:

Internal Stability

$$W^s(\hat{n}, z) \geq W^f(\hat{n}-1, z)$$

so no signatory has any incentive switch to being a non-signatory.

External Stability

$$W^f(\hat{n}, z) > W^s(\hat{n}+1, z)$$

so no non-signatory has any incentive to switch to being a signatory.

Then we have:

Result 4 *If the initial emissions at the start of a period is z, then the unique stable IEA in that period has membership m(z).*

Proof:

We first show that *m(z)* is stable.

Internal Stability

From the definition of internal stability and Result 3 we require:

$$-\gamma(z)z + \delta W(\rho z + N - m(z)) \geq 1 - \gamma(z)z + \delta W(\rho z + N)$$
$$\Leftrightarrow 1 \leq \delta[W(\rho z + N - m(z)) - W(\rho z + N)]$$

From the definition of *m(z)*

$$1 \leq \delta m(z)[W(\rho z + N - m(z)) - W(\rho z + N - m(z) + 1)]$$

But if $W' < 0, W'' < 0$ then $y[W(X) - W(X+1)] < [W(X) - W(X+y)] \quad \forall X, y > 0$

So internal stability is satisfied.

External Stability

From the definition of external stability and Result 3 we require:

$$1-\gamma(z)z+\delta W(\rho z+N-m(z))>-\gamma(z)z+\delta W(\rho z+N-m(z)-1)$$
$$\Leftrightarrow 1>\delta[W(\rho z+N-m(z)-1)-W(\rho z+N-m(z))]$$

which is satisfied by (5).

Finally, proof of the External Stability condition shows that no $n > m(z)$ can be internally stable. The fact that, from Result 3, for all $n < m(z)$ payoffs to signatories and non-signatories are identical and independent of n means that no $n < m(z)$ is externally stable. *QED*

The intuition is simply that, as long as countries know that current membership is strictly greater than *m(z)* then defection by one country will mean that remaining signatories will continue to abate, so the incentive to defect is exactly the same as the incentive for a non-signatory to pollute. So it must pay a country to leave any IEA larger than *m(z)*. But once membership has reached *m(z)* any further defection by a single country would cause remaining signatories to pollute, and, by the definition of *m(z),* the cost of this outweighs any gain a single IEA member country would get from defecting and polluting. Finally for membership below *m(z)* countries are indifferent between joining or not joining, and our definition of external stability is that if they are indifferent they will join.

Now that we have derived the size of the stable IEA and the optimal strategies for signatories and non-signatories, we can determine the payoffs to signatory and non-signatory countries at the start of a period in which the initial stock of emissions is z as follows:

$$W^s(z)=-\gamma(z)z+\delta W(\rho z+N-m(z)); \quad W^f(z)=W^s(z)+1$$

Then assuming that each country has the same probability *m(z)/N* of being selected as a signatory, the expected present-value of current and future net benefits when the initial stock of emissions is given by:

$$W(z) = (\frac{m(z)}{N})W^s(z) + (1 - \frac{m(z)}{N})W^f(z)$$
$$= 1 - \frac{m(z)}{N} - \gamma(z)z + \delta W(\rho z + N - m(z)) \quad (7)$$

We can also define the steady-state stock of emissions by: $\hat{z} = \frac{N - m(\hat{z})}{1-\rho} < \tilde{z}$. We define $m_0 \equiv m(z_0)$ as the initial membership and $\hat{m} \equiv m(\hat{z})$ as the steady-state membership. But we know from Lemma 1 that *if* the stock is rising, then membership must be falling.

Note that the result that the size of membership of the stable IEA falls as the stock of pollution rises is just the dynamic version of the result in Barrett (1994) that as damage costs rise (relative to the benefits of emissions) so the gains from cooperation rise, but conversely the number of countries joining the IEA falls. In Barrett's case that was a comparative statics result across different pollution problems. In our model it occurs endogenously as the accumulation of pollution drives up damage costs. The intuition for why this result occurs is that, as we have noted, free-riding means that the size of the stable IEA is the *minimum* size of IEA at which it just pays IEA members to abate pollution; as the damage costs from pollution rise, the minimum size of IEA at which it pays signatories to abate falls.

3.3 Particular Functional Forms

In order to be able to use (7) to solve explicitly for the value function *W*, we again resort to special functional forms, though in this case, as we shall see, we shall need to make a number of approximations. So again we assume that $\gamma(z) = \lambda + \mu z$ and suppose that we can approximate *W* by the quadratic function $W(z) = A - Bz - Cz^2$.

Then the condition for determining *n(z)* becomes:

$$1 = \delta n(z)[(B + C) + 2C(\rho z + N - n(z))]$$
$$= n(z)[\delta(B + C) + 2\delta CN + 2\delta\rho Cz] - 2\delta C(n(z))^2 \quad (8)$$

Define: $\tau \equiv \delta(B+C)+2\delta CN; \sigma \equiv 2\delta\rho C; \xi(z) \equiv \tau + \sigma z; \psi(z) \equiv \sqrt{\xi^2 - 8\delta C} < \xi(z)$.

Then we can solve (8) to get[2]:

$$n(z) = \frac{\xi(z)-\psi(z)}{4\delta C}; \quad n'(z) = \frac{\sigma}{4\delta C}(1-\frac{\xi}{\psi}) < 0; \quad n''(z) = 2\sigma^2\psi^{-3} > 0. \quad (9)$$

The first approximation we make is to ignore the fact that membership must be an integer and work with *n(z)* not *m(z)*. Then steady-state membership is defined as :

$$\hat{z} \cong \frac{N-n(\hat{z})}{1-\rho} \qquad (10)$$

Straightforward manipulations allow us to solve (10), which is quadratic, to yield:

$$\hat{z} = \frac{\omega + \sqrt{\omega^2 + 4\upsilon\eta}}{2\upsilon} \qquad (11)$$

where:

$$\upsilon \equiv 2(1-\rho)\delta C; \quad \omega \equiv \{2N\delta C - (1-\rho)\delta(B+C)\};$$
$$\eta \equiv \{N\delta(B+C)-1\} > 0. \qquad (12)$$

Since η, $\upsilon > 0$, we have chosen the upper root of the quadratic in (10) to ensure that $\hat{z} > 0$. This also ensures that $\hat{z}$ is unique.

The second approximation we make is to approximate *n(z)* by a quadratic expression: $\bar{n}(z) \equiv f - gz + hz^2$ where *f, g, h* are all positive. We choose *f, g h* so as to fit $\bar{n}(z)$ to *n(z)* as follows: $\bar{n}(z_0) = n(z_0) \equiv n_0; \bar{n}(\hat{z}) = n(\hat{z}) \equiv \hat{n}; \bar{n}'(\hat{z}) = n'(\hat{z}) \equiv \hat{n}' < 0$; i.e. the value of $\bar{n}$ should coincide with the true value at the initial stock, z_0, and at steady-state stock, $\hat{z}$; moreover the slope of $\bar{n}$ should coincide with the true slope at steady-state. The values of *f, g, h* which satisfy these requirements are:

$$h = \frac{(n_0 - \hat{n}) + \hat{n}'(\hat{z}-z_0)}{(\hat{z}-z_0)^2} > 0; \; f = (\hat{n} - \hat{n}'\hat{z}) + h\hat{z}^2 \quad g = 2h\hat{z} - \hat{n}' > 0 \qquad (13)$$

[2] It is straightforward to see that (8) has two positive roots. We take the lower root because the upper root lies above *N* as long as $N\delta(B+C) > 1$, which is a necessary condition for (6) to hold.

Note that by the properties of *n(z)* given in (9), $0 < -\hat{n}' < \frac{n_0 - \hat{n}}{\hat{z} - z_0}$, guaranteeing that *h* is indeed positive.

We can now rewrite (7) as:

$$W(z) = 1 - \frac{\overline{n}(z)}{N} - (\lambda + \mu z)z + \delta W(\rho z + N - \overline{n}(z))$$

or:

$$A - Bz - Cz^2 = 1 - \frac{f - gz + hz^2}{N} - (\lambda + \mu z)z + \delta[A - BX - CX^2]$$

where $X \equiv (N - f) + (\rho + g)z - hz^2$

Substituting for *X,* collecting terms which have the same power of *z*, but ignoring terms which involve higher than quadratic powers of *z* (our third approximation) and equating the remaining coefficients yields:

$$\begin{aligned} A &= 1 - \frac{f}{N} + \delta A - \delta B(N - f) - \delta C(N - f)^2; \\ B &= \lambda - \frac{g}{N} + \delta B(\rho + g) + 2\delta C(\rho + g)(N - f) \qquad (14) \\ C &= \frac{h}{N} + \mu - \delta Bh + \delta C(\rho + g)^2 - 2\delta Ch(N - f) \end{aligned}$$

Note that (14) is not a simple set of linear equations in *A, B, C* which can be solved in the way we did in section 2, because *f, g, h* are complicated non-linear functions of *A, B, C* through (9), (11), (12) and (13).

To make progress, we have resorted to numerical methods to solve (14). Starting from some initial set of values for *A, B, C* (which we take to be *A* = *a, B* = *b, C* = *c*, i.e. we start from the non-cooperative value function), we use (9) to solve for $n_0 = n(z_0)$. Then from (11) and (12) we solve for $\hat{z}$, and from (9) we solve for $\hat{n} = n(\hat{z}), \hat{n}' = n'(\hat{z})$. This allows us to use (13) to solve for *f, g, h*. Finally we use (14) to determine a new set of values for *A, B, C*. We iterate until differences between old and new values of *A, B, C* are negligible. We report the results in the next section.

3.4 Dynamics of the Stock and Membership.

To complete the description of the model with dynamic IEA membership, we describe how the stock of the pollutant, and hence membership, changes over time. The equation of motion for the stock is:

$$z_{t+1} - z_t = -(1-\rho)z_t + N - m(z_t) \cong -(1-\rho)z_t + N - n(z_t) \qquad (15)$$

Now we know the following: (i) from (11) there is a unique steady-state stock, $\hat{z}$; (ii) from (9) and footnote 3 $n(z) < N$ for all $z \geq 0$; (iii) $n'(z_t) < 0$, but *n(z)* is bounded below by 1. So the relationship between $N - n(z)$ and $(1 - \rho)z$ is shown in Figure 1.

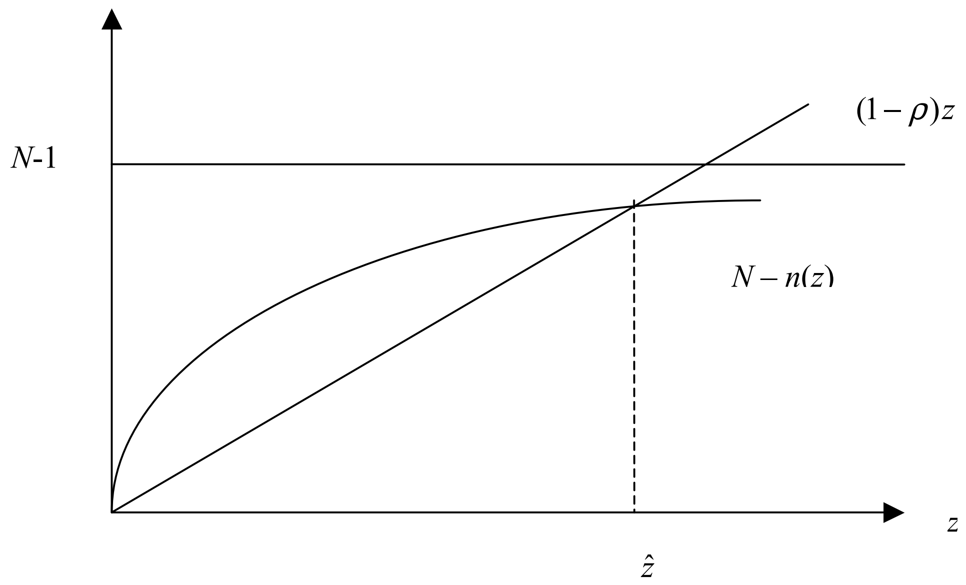

Figure 1

It is clear from Figure 1 that if $z_0 < (>)\, \hat{z}$ then the stock will steadily rise (fall) towards its steady state value and so membership will fall (rise) towards its steady-state value. Now the argument has been based on using *n(z)* rather than *m(z)* to denote membership, i.e. ignoring the fact that membership has to be an integer. This might lead to a situation where there is more than one steady-state membership and stock as shown in Figure 2.

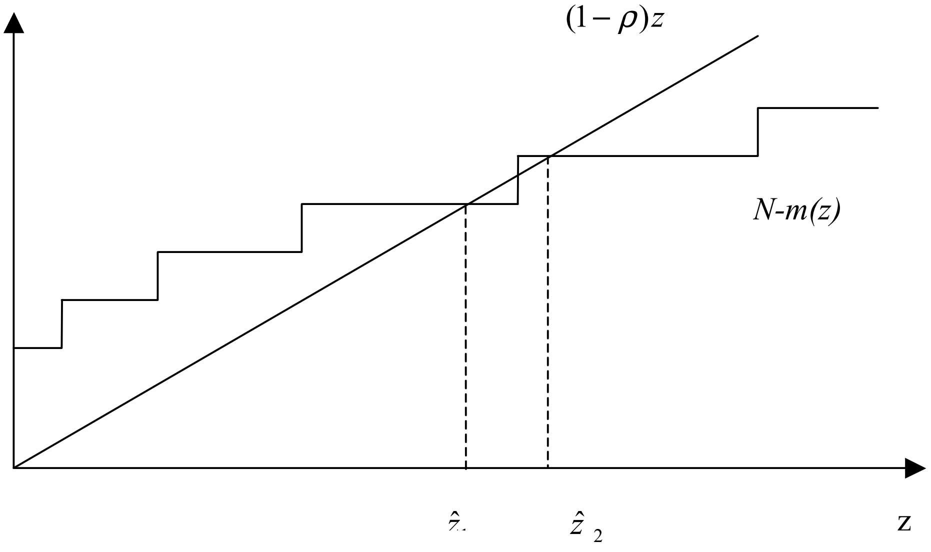

Figure 2

This suggests that if the initial stock is low (less than $\hat{z}_1$) the model will converge to a lower steady-state stock and higher steady-state membership than if the initial stock is high (greater than $\hat{z}_2$). However the steady-state membership levels are unlikely to differ by more than 1. We comment further on this possibility in the next section when we report on the numerical results.

4. Some Numerical Results.

We have conducted some numerical experiments for three reasons. First, we need to use numerical methods to solve for the value functions for the dynamic IEA membership model. Second, the numerical experiments allow us to compare the outcomes of the three different solutions, and in particular to say something about how many countries might join an IEA, how this might vary over time, and how much of the potential gains from cooperation might be delivered by dynamic IEAs. Finally, the numerical results allow us to say something more about the dynamics of the model and the possibility of there being more than one steady-state membership and stock level.

We start by specifying the parameters of the model: δ, ρ, *N,* $z_0, \phi, \theta_1, \theta_2$ (and hence λ, μ). We have chosen parameters *N* = 50, $\delta = 0.95$; $\rho = 0.9$. The discount factor and decay rate might seem quite low, but we have in mind that a period of time corresponds to a block of, say, 5 years. It turns out that the results are not very sensitive to these parameter values. We fix z_0 =50 and $\phi = 1.5$, which implies $\bar{z} = 750$. The key parameters are θ_1, θ_2 and hence λ, μ. It turns out that if $\theta_1 = \theta_2 > 0.25$, then in the dynamic IEA model, membership is always low (2,3) and does not vary much over time. Obviously the gains from cooperation are small (< 5%). To get more variation in outcomes it is necessary to choose smaller parameter values. So we have chosen values of $\theta_1 = 0.001, 0.102, 0.203$; θ_2 = 0.02, 0.14, 0.26. These give outcomes which span the range of sizes of IEA membership.

Having fixed parameter values, we now solve for the outcomes of the three models: non-cooperative, cooperative and dynamic IEA. For the non-cooperative model, we calculate directly the parameters of the value function, *a, b, c*, which allows us to calculate the present value of net benefits for a non-cooperative country $\tilde{U} \equiv U(z_0)$. We then simulate the model to determine $\tilde{T}$, the time it takes for the stock of emissions to move from z_0 to the steady-state value $\tilde{z}$. We do the same for cooperative model, computing the parameters β, χ for the value function and hence the present value benefits for a cooperative country $V^* = V(z_0)$. We simulate the model to determine *T**, the length of time it takes for the stock to decay to 0. Finally, for the dynamic IEA model, we first solve numerically the parameters of the value function *A, B, C* as described in the previous section. This allows us to compute the steady-state stock level $\hat{z}$, the corresponding steady-state membership and the present value of net benefits $\hat{W} \equiv W(z_0)$. Again we simulate the model, but using the actual function *n(z)* computed from the equilibrium parameters *A, B, C* rather than the quadratic approximation $\overline{n}(z)$; we also use the proper integer value for membership in each period, *m(z)*. The simulation allows us to determine $\hat{T}$, the time it takes the stock to reach steady-state. But it also allows us to compute directly the steady-state stock of pollution and steady-state membership, and to compare these with the values computed using the approximation. This provides some check on the reliability of the approximations. It also provides a check whether there might be multiple steady-states, as discussed in the last section.

In Table 1 for each of the nine pairs of values for θ_1, θ_2 we present, for the non-cooperative, cooperative and dynamic IEA models in turn, the parameters of the value functions, the steady state stock, the time to reach steady-state and the present value net benefits. For the dynamic IEA model we also present the initial and steady-state sizes of membership $m_0, \hat{m}$. Finally we compute $\Gamma_F = V^* - \widetilde{U}$, the absolute value of the *full* gains from cooperation (i.e. the difference in payoff per country between the cooperative and non-cooperative equilibria), and $\Gamma_P \equiv (\hat{W} - \widetilde{U})/(V^* - \widetilde{U})$, the *partial* gain from moving from the non-cooperative model to the dynamic IEA model, as a proportion of the full gains in cooperation. Γ_P is a measure of how successful IEAs might be.

The results show that, as cost parameters increase, the parameters of the value functions also increase (in absolute size) reflecting the fact that welfare falls as costs increase. The coefficients of *z* and z^2 are quite similar across the three cases: non-cooperative, cooperative and dynamic IEA. For the non-cooperative and cooperative models, steady-state stocks and the time to reach steady-state are independent of cost parameters, as we would expect, but for the dynamic IEA model, steady-state stock increases with cost parameters, although the time to reach steady-state is almost independent of costs.

Membership of IEAs falls as cost parameters rise, while the absolute (full) gains to cooperation rise. The (relative) partial gains to cooperation decline. These results are consistent with the well-known results of the static model (see Barrett (1994)) that IEA membership is greatest when gains to cooperation are lowest. In terms of dynamics of IEA membership, since we have chosen low values of initial stock, membership declines over time as the stock rises to steady-state, but as cost parameters rise membership becomes almost constant over time. So where cost parameters keep membership low, allowing for it to change over time does not add very much. So dynamics matter most when IEAs matter least.

Finally turning to dynamics, in all but the second case, the steady-state membership given by simulating the model was identical to the steady-state membership computed directly using the approximations, with corresponding steady-state stocks being within 2% of each other. But in the second case the simulations produced a steady-state membership of 9 while the steady-state membership computed directly was only 8, with the steady-state stock being 420 rather than 410. We interpret this as an example of the possibility of multiple steady-states because of the integer nature of membership. But the differences between the steady-states are not large.

5. Conclusions.

In this paper we have extended the familiar model of self-enforcing IEAs from the usual setting of a static pollution problem to a dynamic setting of a stock pollutant. Unlike previous models, which assumed that countries make a once-for-all decision at the outset whether or not to join an IEA, we have explored the implications of allowing IEA membership to vary over time as the stock of pollution changes. We have shown that there will exist a steady-state stock of pollution (usually unique) and corresponding steady-state IEA membership, and that if the initial stock of pollution is below (above) the steady-state then the stock will rise (fall) steadily towards steady-state, and IEA membership will fall (rise) towards steady-state. The intuition behind these results is that they are simply the dynamic generalisation of the pessimistic static results that the greater are the costs of damage, and hence the greater the potential gains from cooperation, the smaller is the size of a self-enforcing IEA.

The model is extremely simple and there are many possible lines for further research. One obvious question is how global welfare in this variable membership model would compare with global welfare in a fixed membership model. In our earlier two-period model we found that the variable membership model generally gave higher welfare than the fixed membership model. But in the two-period setting variable IEA membership rose over time and it would be interesting to know whether the welfare ranking of the two types of model would carry over to the infinite-horizon case. The model has also substantially simplified the emission choices for countries, and it would be interesting to know how our results would change if countries had a continuous choice of emissions.

Because of our assumption of symmetry, while we can determine how many countries might join an IEA at any date, we cannot determine which countries might join. So our modelling of the variable membership model of the IEA required us to specify what beliefs countries formed about whether or not they would be a signatory in future periods, and we have invoked a very simple assumption that each country believes that it has the same probability as any other country of being a signatory in a future period, independent of its past history of membership. This is clearly unsatisfactory and it would be desirable to either drop the symmetry assumption, and so try to determine which countries are likely to join in each period, or provide a proper justification for our assumption, and if that is not possible determine what might be a more appropriate assumption to make about beliefs.

Finally the model of self-enforcing agreements is itself a special model, and there are now a range of competing models of coalition stability (see Bloch (1997) and Finus (2001) for excellent surveys). Some of these models involve the concept of *sequential* formation of coalitions (so they allow for the possibility of more than one IEA group of countries). However the underlying game is static, and the notion of sequential formation is a conceptual rather temporal one. It would be interesting to consider how far such concepts could be integrated with the underlying dynamics of the pollution problem to assess how coalition structures might evolve in real time.

References

Barrett, S. (1994), "Self-enforcing international environmental agreements", *Oxford Economic Papers*, 46, 878-894.

Barrett, S. (1997), "Towards a Theory of international environmental cooperation" in C. Carraro and D. Siniscalco (eds.) *New Directions in the Economic Theory of the Environment*, CUP, 239-280.

Barrett, S. (2002), *Environment and Statecraft*, Oxford: OUP (forthcoming).

Bloch, F. (1997), "Non-cooperative models of coalition formation in games with spillovers" in C. Carraro and D. Siniscalco (eds.) *New Directions in the Economic Theory of the Environment*, CUP, 311-352.

Carraro, C. and D. Siniscalco (1993), "Strategies for the international protection of the environment", *Journal of Public Economics*, 52, 309-328.

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

Eyckmans, J. (2001), "On the farsighted stability of the Kyoto protocol", *CLIMNEG Working Paper* 40, CORE, Universite Catholique de Louvain

Finus, M. (2001), *Game Theory and International Environmental Cooperation*, Cheltenham, UK and Northampton, USA: Edward Elgar.

Germain, M., P. Toint,, H. Tulkens and A. de Zeeuw (2002) "Transfers to sustain core-theoretic cooperation in international stock-pollutant control", *CLIMNEG Working Paper 6* (Revised, May 2002) *,* CORE, Universite Catholique de Louvain; forthcoming in *Journal of Economic Dynamics and Control*

Karp, L. and S. Sacheti (1997), "Dynamics and limited cooperation in international environmental agreements", *mimeo*, University of California, Berkeley.

Rubio, S. and B. Casino (2001), “International cooperation in pollution control”, *mimeo*, University of Valencia.

Rubio, S. and A. Ulph (2002), “ A simple dynamic model of international environmental agreements with a stock pollutant”, *Discussion Paper in Economics and Econometrics 0209*, Department of Economics, University of Southampton.

TABLE 1: RESULTS OF NUMERICAL SIMULATIONS

θ_1	0.001			0.102			0.203		
θ_2	0.02	0.14	0.26	0.02	0.14	0.26	0.02	0.14	0.26
Non-Cooperative									
a (×-1)	7.124	43.99	80.85	105.5	138.6	171.8	203.8	233.2	262.7
b	0.028	0.064	0.100	0.132	0.164	0.196	0.235	0.264	0.292
c(×1000)	0.010	0.071	0.131	0.009	0.064	0.118	0.008	0.056	0.105
$\tilde{z}$	500.0	500.0	500.0	500.0	500.0	500.0	500.0	500.0	500.0
$\tilde{T}$	110	110	110	110	110	110	110	110	110
$\tilde{U}$ (×-1)	8.551	47.35	86.15	112.1	147.0	181.8	215.6	246.6	277.5
Cooperative									
β	0.022	0.022	0.022	0.126	0.126	0.126	0.230	0.230	0.230
χ(×1000)	0.010	0.071	0.131	0.009	0.064	0.118	0.008	0.056	0.105
z^*	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
T^*	169	169	169	169	169	169	169	169	169
V^* (×-1)	1.129	1.281	1.433	6.336	6.473	6.609	11.54	11.66	11.78
IEA									
A (×-1)	0.508	31.59	68.51	92.19	125.8	160.1	193.2	223.3	254.0
B	0.025	0.064	0.100	0.132	0.163	0.196	0.235	0.264	0.292
C(×1000)	0.015	0.082	0.144	0.009	0.065	0.120	0.008	0.057	0.106
$\hat{z}$	120.0	410.0	450.0	420.0	450.0	460.0	450.0	460.0	470.0
$\hat{T}$	121	110	110	109	110	109	109	109	109
$\hat{W}$ (×-1)	1.691	31.27	67.94	93.52	127.6	161.2	194.8	225.3	256.3
m_0	41	14	9	8	7	5	5	4	4
$\hat{m}$	38	9	5	8	5	4	5	4	3
Gains from Cooperation									
Γ_F	7.422	46.07	84.72	105.8	140.5	175.2	204.1	234.9	265.7
Γ_P	.924	.349	.215	.167	.138	.118	.102	.090	.080

Other Parameters:

$N = 50$; $\delta = 0.95$; $\rho = 0.9$; $z_0 = 50.0$; $\bar{z} = 750$