International Environmental Cooperation with Imperfect Monitoring

Marita Laukkanen*

October 23, 2009

Abstract

Many prominent environmental problems are plagued by uncertainty in the underlying biological and physical processes. Noise in the environmental process makes the link between current actions and future environmental conditions indirect. Where international policy coordination is called for, adherence to an agreement cannot be monitored unambiguously. This paper studies how to construct a self-enforcing international environmental agreement in the case of a stock pollutant with stochastic stock dynamics. The strategy profile proposed involves harsh punishments after a suspected deviation, followed by forgiveness. The model of environmental cooperation is illustrated with an application to a linear-quadratic problem with uniformly distributed additive shocks.

Keywords: International environmental agreements, cooperation, self-enforcing treaties, imperfect information, environmental uncertainty

*Agrifood Research Finland, Luutnantintie 13, 00410, Helsinki, Finland. E-mail marita.laukkanen@mtt.fi, telephone +358 40 743 8854.

1 Introduction

An extensive literature on game theoretic models analyzes strategic interactions among polluting countries and the design of international environmental agreements. With few exceptions, this literature neglects uncertainty in the underlying environmental process. Uncertainty in the development of a pollutant stock adds noise to the link between current actions and future environmental conditions, and deviations from an agreement cannot be unambiguously detected. The papers that do account for imperfect monitoring of past play consider a repeated game. In essence, the physical environment is the same in every period. This is a significant limitation in that environmental problems generally involve pollutants that accumulate. The past has a direct influence on current opportunities through the stock of pollution.

This paper studies how to construct a self-enforcing cooperative agreement on pollution control in the case of uncertainty and a stock pollutant when the physical environment evolves as a function of previous emissions and a stochastic environmental factor. Interaction is modeled as a dynamic game with imperfect information. Each country observes the current stock of pollution but does not observe other countries' emissions or the stochastic environmental shock. Thus, a country detecting a high stock level cannot be certain of whether the observation reflects relatively large emissions in the previous period, or simply a high pollution shock. A cooperative agreement in which the countries restrict their emission levels to reduce the accumulation of pollution then runs the risk that a country may be tempted to cheat on the agreement, increasing its emissions above the agreed upon level and hoping that the deviation will be masked by the randomness in the pollution process. Intertemporal incentives are required to support cooperation.

The approach that we consider here is similar in spirit to those in Tarui et al. (2008), Laukkanen (2003), Mason et al. (2008) and Haurie and Tolwinski (1990). Tarui et al. (2008) construct a self-enforcing agreement on harvesting a common property resource that is supported by optimal temporary punishments that constitute a worst perfect equilibrium. Monitoring of harvests is noisy, but there is no uncertainty in the stock dynamics. Thus, while players do not know who cheated, whether or not someone cheated is known by all players. Laukkanen (2003) studies a model with uncertainty in the stock dynamics, where adherence to the agreement cannot be unambiguously monitored. Cooperative harvesting is supported by the threat of reverting to Nash feedback harvest levels for a finite length of time if deviations are suspected. The stock dynamics in both Tarui et al. (2008) and Laukkanen (2003) are described by a stock transition function that has essentially one-step memory - the stock level in the next period depends only on current harvest level but not on the current stock. Mason et al. (2008) consider a model where part of the pollution stock carries over to the next period, with international cooperation on climate change mitigation as an application. Their results point to an important feature of international environmental cooperation - the incentives to cooperate may change as the pollutant stock level changes. Hence, where the pollutant accumulates over time, a dynamic game theoretic analysis is called for. Their approach also supports cooperation with temporary punishments, but assumes perfect information about each country's emissions. Finally, Haurie and Tolwinski (1990) account for both dynamic stock effects and imperfect monitoring of player's actions. Suspected violations of the agreement are punished with reversion to the Nash feedback emission levels. While their stochastic game format allows for the punishments to be temporary, cooperative equilibria are computed only for the grim trigger profile where punishments last forever. In games with imperfect monitoring, such punishments have two undesirable properties. First, with imperfect monitoring, punishments do occur along the equilibrium path. In the context of international environmental cooperation, it is questionable whether countries would actually carry out infinite punishments when all countries could clearly be made better off by renegotiating. Second, optimal punishments are in general more severe than the Nash feedback strategies and support equilibria that produce a higher equilibrium payoff (see e.g. Mailath and Samuelson 2006).

This paper contributes to the existing literature by developing a stochastic game theoretic model of international environmental agreements that incorporates (1) truly dynamic stock effects, (2) imperfect monitoring of adherence to the agreement, and (3) severe but temporary punishments to support cooperation. We apply the technique developed by Abreu, Pearce and Stacchetti (1986, 1990) for supporting efficient supergame equilibria under imperfect monitoring. We extend it for constructing cooperative equilibria in stochastic dynamic games. As suggested by Haurie and Tolwinski (1990), we define a monitoring statistic based on observed stock levels that serves as a signal in the game of imperfect information. "Bad" signals (i.e. monitoring statistics that are particularly likely to arise if a country breached the agreement) trigger a temporary punishment phase with low continuation payoffs. The trigger strategies are designed so that all players will find it in their self-interest to carry out the punishment, and the proposed agreement will be self-enforcing. A cooperative equilibrium is defined by solutions to two dynamic programming equations, one describing the cooperative phase and one the punishment phase. A nontrivial cooperative equilibrium exists if one can define a solution to the dynamic programming equations that produces cooperative phase expected payoffs dominating the Nash feedback payoffs. Whether or not such a solution exists is an empirical question. We apply the model to one of the parameterizations analyzed by Mason et al. (2008) and identify conditions under which a self-enforcing international agreement can be established.

2 The game

2.1 Stage game rewards

Consider $N \geq 2$ identical countries that share an environmental resource. The countries emit a pollutant that accumulates in the environment over time. In each period t, each country i, $i = 1, ..., N$, emits a pollution flow x_{it}. The domain of x_{it} is denoted by $\Delta = [0, x^U]$. The pollutant stock S_t changes from one period to the next as follows:

$$S_{t+1} = \beta S_t + \sum_{i=1}^{N} x_{it} + \theta_t, \tag{1}$$

where β is the fraction of the stock that persists until the next period, and $\{\theta_t\}$ is a sequence of independent identically distributed random variables. The random multipliers θ_t are distributed on some finite interval $[l, h]$, where $0 \leq l < h < \infty$, with a cumulative distribution F and a continuously differentiable density f.

Each country gets benefits from its own emissions and suffers damage from the accumulated pollution. Emission reductions entail abatement costs. Thus, the benefit of emissions is the amount of abatement costs that a country avoids. This benefit for country i is given by the quadratic function

$$B(x_{it}) = k + ax_{it} - \frac{b}{2}x_{it}^2, \tag{2}$$

where k, a, and b are positive parameters. The damage to each country is given by the quadratic function

$$D(S_t) = cS_t + \frac{d}{2}S_t^2, \tag{3}$$

where c and d are positive parameters. Each country's period-wise payoff equals benefits minus costs,

$$\pi(x_{it}, S_t) = k + ax_{it} - \frac{b}{2}x_{it}^2 - cS_t - \frac{d}{2}S_t^2. \tag{4}$$

The one-period discount factor for all countries is $\delta \in (0, 1)$.

2.2 Information structure and equilibrium concept

We assume that the countries are able to observe the state variable S_t directly, but cannot observe other countries' emissions or realizations of the random environmental shock θ_t. Each country's period-wise payoff depends only on its own emissions, not on those of the other countries - otherwise it would be possible for a country to infer information about other countries' emissions from the realized payoff. As a consequence, the only public information available in period t is the t-period history of observed stock levels,

$$H_t = \{S_0, S_1, ..., S_{t-1}, S_t\}, \qquad t = 0, 1, ... \tag{5}$$

We restrict our attention to public strategies, where a strategy σ_{it} for country i, in every period t, depends only on the public history H_t and not on i's private information on its own past emissions.

A strategy requiring a country to recall the entire sequence H_t for every t would be difficult to implement in practice (see e.g. Rubinstein 1986). Therefore, we further restrict our attention to stationary Markovian strategies, where each country chooses its emission level x_{it} based on the value of an extended state vector Z_t which in addition to S_t contains a summary of public information concerning past play. That is, σ_{it} does not explicitly depend on t. In what follows, we will use the symbol σ_i to denote both country i's cooperative decision rule in period t and its cooperative strategy, i.e. the whole infinite sequence of those decision rules. As for the summary of public information, we follow Haurie and Tolwinski (1990) and assume that the countries agree to condition their emissions in period t on the value of a monitoring statistic M_t, whose value is determined based on the observations on S_t and S_{t-1} and whose domain is given by Ω. We first define the extended state vector as $Z_t = \{S_t, M_t\}$ and let $\phi(Z_t) \equiv (\phi_1(Z_t), ..., \phi_N(Z_t))$ denote the N countries' cooperative emission profile given Z_t. The monitoring statistic M_t is then defined by

$$M_t = S_t - E(S_t \,|\phi(Z_{t-1})\,), \qquad \text{for } t = 0, 1, ... \tag{6}$$

where $E(S_t \,|\phi(Z_{t-1})\,)$ is the expected value of S_t when all countries employ their cooperative policies in period $t-1$. The monitoring statistic constitutes a public signal in the game of imperfect public monitoring. The signal provides only noisy information about past play, and deviations from cooperative emission levels cannot be unambiguously detected.

We assume that the countries' objectives are expressed in terms of discounted payoff functionals of the form

$$J_i\left[S_t, \sigma\right] = E_\sigma \left\{ \sum_{s=t}^{\infty} \delta^{s-t} \pi(\sigma_{it}(Z_t), S_t) \right\}. \tag{7}$$

The expectation in (7) is taken with respect to the probability measure induced by the strategy profile σ, where

$$\sigma = \left\{ \sigma_t = (\sigma_{1t}, ..., \sigma_{Nt}) : \; t = 0, 1, 2... \right\}.$$

The equilibrium concept we employ is that of perfect public equilibrium (PPE). A public strategy profile σ is a PPE if and only if for all Z_t,

$$\pi(\sigma_{it}, S_t) + \delta E_{\sigma_t} J_i\left[S_{t+1}, \sigma_{t+1}(Z_{t+1})\right] \geq \pi(\widehat{\sigma}_{it}, S_t) + \delta E_{\widehat{\sigma}_{it}, \sigma_{-it}} J_i\left[S_{t+1}, \sigma_{t+1}(Z_{t+1})\right],$$

that is, if and only if there are no profitable one-shot deviations (see e.g. Mailath and Samuelson 2006). A strategy profile σ is self-enforcing if it constitutes a PPE.

Equilibria in which countries reduce their emissions below the Nash feedback level require intertemporal incentives. A seminal feature of such incentives in the case of imperfect monitoring is that some realizations of the public signal must be followed by low continuation payoffs. As such, they are reminiscent of punishments. However, unlike in the case of perfect monitoring, low continuation payoffs now occur along the equilibrium path. They are needed to provide incentives for countries to adhere to the cooperative policies.

2.3 Intertemporal incentives to support cooperation

Abreu, Pearce and Stacchetti (1986, 1990) show that any supergame equilibrium can be constructed to have a bang-bang property that makes the intertemporal structure of the equilibrium relatively simple. We restrict our search of cooperative equilibria to ones with the bang-bang property. We propose a cooperative strategy profile that depends on the observed value of the monitoring statistic M_t and two subsets of Ω, $\overline{M}$ and $\widetilde{M}$. Let $\overline{V_i}(S)$ and $\widetilde{V_i}(S)$ be the maximum and minimum country i payoffs that can be supported as a symmetric PPE, given the current stock level S and the sets $\overline{M}$ and $\widetilde{M}$. By Theorem 3 in Abreu, Pearce and Stacchetti (1990), the maximum payoff $\overline{V}_i(S)$ can be supported by continuation payoffs drawn exclusively from the set $\left\{\overline{V}_i(S'), \widetilde{V_i}(S')\right\}$. The minimum payoff $\widetilde{V_i}(S)$ can also be supported by the same set of continuations. Play of the game can be viewed as an alteration between a reward phase, with the continuation payoff $\overline{V}_i(S)$ and emission profile $\overline{x}(S) = (\overline{x}_1, ..., \overline{x}_N)$, and a punishment phase, with the continuation payoff $\widetilde{V_i}(S)$ and emission profile $\widetilde{x}(S) = (\widetilde{x}_1, ..., \widetilde{x}_N)$. Play begins in the reward phase. The two sets of monitoring statistics, $\overline{M}$ and $\widetilde{M}$, govern transitions from one phase to another. In the reward phase, a monitoring statistic outside $\overline{M}$ prompts a switch to the punishment phase, with play otherwise continuing in the reward phase. In the punishment phase, a monitoring statistic in $\widetilde{M}$ prompts a switch back to the reward phase, with play otherwise continuing in the punishment phase.

Country i's reward phase continuation payoff then satisfies

$$\begin{aligned}\overline{V}_i(S) &= \pi(S,\overline{x}) + \delta\rho\left(\overline{M}\,|\overline{x}\right) E\left[\overline{V_i}\left(S'\right)|\overline{x}, M' \in \overline{M}\right] \\ &\quad + \delta\left(1-\rho\left(\overline{M}\,|\overline{x}\right)\right) E\left[\widetilde{V_i}\left(S'\right)|\overline{x}, M' \notin \overline{M}\right],\end{aligned} \tag{8}$$

where $\rho\left(\overline{M}\,|\overline{x}\right)$ is the probability that $M' \in \overline{M}$ given $\overline{x}$, and $S' = \beta S + \sum_i \overline{x}\,(S) + \theta$. The punishment phase continuation payoff satisfies

$$\begin{aligned} \widetilde{V}_i\,(S) \;\; &= \;\; \pi\,(S, \widetilde{x}) + \delta\rho\left(\widetilde{M}\,|\widetilde{x}\right) E\left[\overline{V}_i\left(S'\right)\Big|\widetilde{x}, M' \in \widetilde{M}\right] \\ &\quad + \delta\left(1 - \rho\left(\widetilde{M}\,|\widetilde{x}\right)\right) E\left[\widetilde{V}_i\left(S'\right)\Big|\widetilde{x}, M' \notin \widetilde{M}\right]. \end{aligned} \tag{9}$$

where $\rho\left(\widetilde{M}\,|\widetilde{x}\right)$ is the probability that $M' \in \widetilde{M}$ given $\widetilde{x}$, and $S' = \beta S + \sum_i \widetilde{x}\,(S) + \theta$.

Let $x^N(S)$ denote the symmetric Nash feedback solution. There is always an equilibrium in which $\overline{x}_i(S) = \widetilde{x}_i(S) = x^N(S)$ for all i. In this case, incentive constraints for equilibrium play are trivially satisfied. To achieve an equilibrium payoff $\overline{V}_i\,(S) > V_i^N(S)$, it must be that $\overline{x}_i(S) < x^N(S)$. An emission profile $\overline{x}(S)$ is said to be enforceable if, for all S and all feasible $\widehat{x}_i$, country i prefers to choose $\overline{x}_i$ rather than any alternative emission level $\widehat{x}_i$, in either case thereafter continuing with the equilibrium strategies:

$$\begin{aligned} &\pi\,(S, \widehat{x}_i, \overline{x}_{-i}) + \delta\rho\left(\overline{M}\,|\widehat{x}_i, \overline{x}_{-i}\right) E\left[\overline{V}_i\left(S'\right)|\widehat{x}_i, \overline{x}_{-i}, M' \in \overline{M}\right] \\ &+\delta\left(1 - \rho\left(\overline{M}\,|\widehat{x}_i, \overline{x}_{-i}\right)\right) E\left[\widetilde{V}_i\left(S'\right)|\widehat{x}_i, \overline{x}_{-i}, M' \notin \overline{M}\right] \\ \leq\;\; &\pi\,(S, \overline{x}) + \delta\rho\left(\overline{M}\,|\overline{x}\right) E\left[\overline{V}_i\left(S'\right)|\overline{x}, M' \in \overline{M}\right] \\ &+\delta\left(1 - \rho\left(\overline{M}\,|\overline{x}\right)\right) E\left[\widetilde{V}_i\left(S'\right)|\overline{x}, M' \notin \overline{M}\right], \end{aligned} \tag{10}$$

where $S'|\,\widehat{x}_i, \overline{x}_{-i} = \beta S + \widehat{x}_i(S) + \sum_{-i} \overline{x}\,(S) + \theta$. A basic equilibrium trade-off is apparent in this inequality. Deviating from the agreed upon emission strategy provides an immediate payoff gain and a loss in future payoffs. The loss in future payoffs works in two ways: overemitting increases both the future pollutant stock and the associated damage and the likelihood of triggering the punishment.

Green and Porter (1984) and Porter (1983) suggest punishments that involve reversion to the Nash feedback strategies for a finite number of periods; punishments are followed by a return to the reward phase. Optimal punishments of the kind suggested by Abreu, Pearce and Stacchetti (1986, 1990) instead minimize the continuation payoff associated with a bad signal. Such punishments are in general more severe than reversion to the Nash feedback strategies. As a consequence, optimal punishments allow for equilibria that are more cooperative, in the sense that they produce a higher equilibrium payoff than reversion to the Nash feedback solution. Punishments diverging from the Nash feedback emissions are enforceable if, for all S and all feasible $\widehat{x}_i$,

$$\begin{aligned} & \pi\left(S, \widehat{x}_i, \widetilde{x}_{-i}\right)+\delta \rho\left(\widetilde{M} \,|\widehat{x}_i, \widetilde{x}_{-i}\right) E\left[\overline{V}_i\left(S^{\prime}, M^{\prime}\right)\left|\widehat{x}_i, \widetilde{x}_{-i}, M^{\prime} \in \widetilde{M}\right.\right] \\ & +\delta\left(1-\rho\left(\widetilde{M} \,|\widehat{x}_i, \widetilde{x}_{-i}\right)\right) E\left[\widetilde{V}_i\left(S^{\prime}, M^{\prime}\right)\left|\widehat{x}_i, \widetilde{x}_{-i}, M^{\prime} \notin \widetilde{M}\right.\right] \\ \leq\; & \pi(S, \widetilde{x})+\delta \rho\left(\widetilde{M} \,|\widetilde{x}\right) E\left[\overline{V}_i\left(S^{\prime}, M^{\prime}\right)\left|\widetilde{x}, M^{\prime} \in \widetilde{M}\right.\right] \\ & +\delta\left(1-\rho\left(\widetilde{M} \,|\widetilde{x}\right)\right) E\left[\widetilde{V}_i\left(S^{\prime}, M^{\prime}\right)\left|\widetilde{x}, M^{\prime} \notin \widetilde{M}\right.\right], \end{aligned} \tag{11}$$

where $S'|\,\widehat{x}_i, \widetilde{x}_{-i} = \beta S + \widehat{x}_i(S) + \sum_{-i} \widetilde{x}\,(S) + \theta$. A strategy profile ϕ specifying that countries choose $\overline{x}(S)$ when play is in the reward phase and $\widetilde{x}(S)$ when play is in the punishment phase is a perfect public equilibrium if and only if $\overline{x}(S)$ and $\widetilde{x}(S)$ are enforceable, that is, conditions (10) and (11) hold for all countries $i = 1, ..., N$ and all possible stock levels given the initial stock S_0. In other words, a strategy profile ϕ is a perfect public equilibrium if and only if there are no profitable one-shot deviations for any player and any possible stock level.

2.4 Cooperative equilibria

In a symmetric equilibrium, the joint continuation payoffs corresponding to $\overline{V}_i(S)$ and $\widetilde{V}_i\,(S)$ are $\overline{V}(S) = N\overline{V}_i(S)$ and $\widetilde{V}\,(S) = N\widetilde{V}_i\,(S)$. For fixed sets $\overline{M}$ and $\widetilde{M}$, the joint continuation payoffs in reward and punishment phases must satisfy the following two equations:

$$\begin{aligned} \overline{V}\,(S) \;=\; & \max_{\overline{x}} \left\{ N\pi\,(S, \overline{x}) + \delta\rho\left(\overline{M}\,|\overline{x}\right) E\left[\overline{V}\left(S'\right)|\overline{x}, M' \in \overline{M}\right] \right. \\ & \left. +\delta\left(1-\rho\left(\overline{M}\,|\overline{x}\right)\right) E\left[\widetilde{V}\left(S'\right)|\overline{x}, M' \notin \overline{M}\right] \right\}, \end{aligned} \tag{12}$$

where $S' = \beta S + \sum_i \overline{x}\,(S) + \theta$, and

$$\begin{aligned} \widetilde{V}\,(S) \;=\; & \min_{\widetilde{x}} \left\{ N\pi\,(S, \widetilde{x}) + \delta\rho\left(\widetilde{M}\,|\widetilde{x}\right) E\left[\overline{V}\left(S'\right)\left|\widetilde{x}, M' \in \widetilde{M}\right.\right] \right. \\ & \left. +\delta\left(1-\rho\left(\widetilde{M}\,|\widetilde{x}\right)\right) E\left[\widetilde{V}\left(S'\right)\left|\widetilde{x}, M' \notin \widetilde{M}\right.\right] \right\}. \end{aligned} \tag{13}$$

where $S' = \beta S + \sum_i \widetilde{x}(S) + \theta$. The joint continuation payoffs $\overline{V}(S)$ and $\widetilde{V}(S)$ generated by the equilibrium strategies ϕ can be interpreted as Bellman value functions describing joint payoffs for the group of countries in reward and punishment phases, respectively. Equations (12) and (13) can be solved numerically.

The reward and punishment phase payoffs depend on the choice of the sets $\overline{M}$ and $\widetilde{M}$ that govern transitions from one phase to another. Defining these sets then becomes part of designing a cooperative equilibrium. The intuition is that $\overline{M}$ should be a set of monitoring statistics that makes the probability of remaining in the reward phase as high as possible while preserving incentives to cooperate. We consider sets of monitoring statistics of the form $\overline{M} = [r_l, r_u]$ and $\widetilde{M} = [p_l, p_u]$, where r_l, r_u, p_l, and p_u become design parameters. It is possible to define parameters r_l, r_u, p_l, and p_u that are optimal in the sense that the corresponding reward phase payoffs are maximized.

Given the sets $\overline{M} = [r_l, r_u]$ and $\widetilde{M} = [p_l, p_u]$, the probability of remaining in the reward phase is $P(M \in [r_l, r_u]) = P(\theta \in [r_l + E(\theta), r_u + E(\theta)])$ along the equilibrium path; the probability of returning to the reward phase once a punishment has been triggered is $P(M \in [p_l, p_u]) = P(\theta \in [p_l + E(\theta), p_u + E(\theta)])$. With the density f for the random variable θ, equations (12) and (13) then become

$$\begin{aligned} \overline{V}(S) &= \max_{\overline{x}} \Bigg\{ N\pi(S, \overline{x}) + \delta \int_{r_l+E(\theta)}^{r_u+E(\theta)} \overline{V}(S') f(\theta)\, d\theta \\ &\quad + \delta \int_{l}^{r_l+E(\theta)} \widetilde{V}(S') f(\theta)\, d\theta + \delta \int_{r_u+E(\theta)}^{h} \widetilde{V}(S') f(\theta)\, d\theta \Bigg\}, \end{aligned} \tag{14}$$

where $S' = \beta S + \sum_i \overline{x}(S) + \theta$, and

$$\begin{aligned} \widetilde{V}(S) &= \min_{\widetilde{x}} \Bigg\{ N\pi(S, \widetilde{x}) + \delta \int_{p_l+E(\theta)}^{p_u+E(\theta)} \overline{V}(S') f(\theta)\, d\theta \\ &\quad + \delta \int_{l}^{p_l+E(\theta)} \widetilde{V}(S') f(\theta)\, d\theta + \delta \int_{p_u+E(\theta)}^{h} \widetilde{V}(S') f(\theta)\, d\theta \Bigg\}. \end{aligned} \tag{15}$$

where $S' = \beta S + \sum_i \widetilde{x}(S) + \theta$.

The game has a linear-quadratic structure, which implies a quadratic form of the continuation payoffs. Therefore, we seek value functions in the form

$$\overline{V}(S) = \frac{\overline{P}}{2}S^2 + \overline{Q}S + \overline{R}, \tag{16}$$

$$\widetilde{V}(S) = \frac{\widetilde{P}}{2}S^2 + \widetilde{Q}S + \widetilde{R}, \tag{17}$$

where $\overline{P}, \overline{Q}, \overline{R}, \widetilde{P}, \widetilde{Q}$, and $\widetilde{R}$ are parameters to be determined. The dynamic programming equations (14) and (15) can be solved as follows. For the sets $[r_l, r_u]$ and $[p_l, p_u]$ given, we substitute the "trial solutions" (16) and (17) into (14) and (15), and find the optimal control rules for reward and punishment phases as functions of the stock S, the known model parameters, the design parameters r_l, r_u, p_l, p_u and the unknown parameters $\overline{P}, \overline{Q}, \overline{R}, \widetilde{P}, \widetilde{Q}, \widetilde{R}$. We then substitute these control rules into (14) and (15), which yields two quadratic equations in S. Equating coefficients of S^2, S and 1 gives equations determining the values of $\overline{P}, \overline{Q}, \overline{R}, \widetilde{P}, \widetilde{Q}, \widetilde{R}$. The equations for $\overline{P}$ and $\widetilde{P}$ can first be solved simultaneously, then those for $\overline{Q}$ and $\widetilde{Q}$, and at last those for $\overline{R}$ and $\widetilde{R}$. The expressions for $\overline{P}, \overline{Q}, \overline{R}, \widetilde{P}, \widetilde{Q}$ and $\widetilde{R}$ will depend on r_l, r_u, p_l and p_u. The choice of these design parameters may determine whether a solution of (14) and (15) exists, such that the incentive constraints (10) and (11) are satisfied and the reward phase payoff $\overline{V}(S)$ exceeds the Nash feedback payoff $V^N(S)$ for all S. The parameters r_l, r_u, p_l and p_u influence the value of the cooperative payoff $\overline{V}(S)$. The optimal r_l, r_u, p_l and p_u are determined numerically, by carrying out a grid search over these parameters where we first determine $\overline{V}(S)$ and $\widetilde{V}(S)$ for every point in the grid, and then choose the r_l, r_u, p_l, p_u combination that maximizes $\overline{V}(S_0)$ and satisfies the incentive constraints (10) and (11).

3 Numerical illustration

To characterize cooperative outcomes in a fluctuating environment and still get fairly detailed results, we give up on the elegance of analytic results and rely instead on numerical analysis. Also, whether or not a cooperative agreement can be sustained is in the end an empirical question. The precise parameter values used to obtain numerical results are similar to those in Mason et al. (2008) (a model that did not incorporate uncertainty). Here, we assume that the random multipliers θ are uniformly distributed:

$$f(\theta) = \begin{cases} \frac{1}{w_u - w_l} & \text{for } w_l \leq \theta \leq w_u, \\ 0 & \text{elsewhere,} \end{cases}$$

with $w_l = 0$ and $w_u = 0.03$. As the base case, we set the number of countries $N = 6$; the pollutant carryover $\beta = 0.99$; the parameters of the benefit function $k = 0$, $a = 10$, and $b = 2000$; the parameters of the damage function $c = 0$ and $d = 0.016$; and the discount factor $\delta = 0.99$.

The international environmental policy coordination game has cooperative equilibria that dominate the outcome corresponding to the Nash feedback solution. Table 1 summarizes the numerical results for the game under the optimal choice of monitoring sets, with $r_l = -0.015, r_u = 0.0147, p_l = 0.0099$ and $p_u = 0.015$. The punishment minimizing the punishment phase payoff is a corner solution, where the countries emit their maximum feasible emissions until play returns to the reward phase. For the purpose of comparison Table 1 also shows the Nash feedback solution and the first-best solution obtainable under complete information. The payoffs under complete information constitute an upper bound for payoffs that can be achieved with policy coordination under imperfect information. Figure 1 displays the payoffs upon implicit cooperation with noisy information, noncooperation (Nash feedback solution), and joint management for different stock levels. Under the assumed combination of parameter values the maximum feasible pollutant stock equals 9. While the payoffs upon cooperation exceed the payoffs upon noncooperation under all relevant stock levels, they fall short of those corresponding to the first-best solution. That is, efficiency cannot be achieved. Figure 2 shows the gains from cooperation for different stock levels, calculated as the percentage of the difference $V^*(S) - V^N(S)$ achieved by the cooperative equilibrium payoffs. The gains from cooperation range from 51% to 99%, taking on the highest values when the current stock level is low.

The optimal values of r_l, r_u, p_l, and p_u correspond to the 0th, 99th, 83th and 100th percentiles of the probability distribution $F(\theta)$, respectively.[1] That is, when in the

[1] Recell that the probability of remaining in the reward phase is $P(M \in [r_l, r_u]) = P(\theta \in [r_l + E(\theta), r_u + E(\theta)])$ along the equilibrium path; the probability of returning to the reward phase once a punishment has been triggered is $P(M \in [p_l, p_u]) = P(\theta \in [p_l + E(\theta), p_u + E(\theta)])$.

reward phase, the probability remaining in the reward phase in the next period is 99%. When in the punishment phase, the probability of returning to the reward phase is 17%.

Table 1. Policy and value functions for $\delta = 0.99$ and $N = 6$

	Policies	Value functions
Reward phase,		
noisy information	$\overline{x} = 10^{-3}(1.3 - 1.6S)$	$\overline{V}(S) = -2.6S^2 - 4.4S - 3.7$
Punishment phase,		
noisy information	$\widetilde{x} = x^u$	$\widetilde{V}(S) = -2.7S^2 - 5.0S - 5.5$
Feedback solution	$x^N = 10^{-3}(4.0 - 0.32S)$	$V^N(S) = -0.50S^2 - 0.97S - 1.6$
First-best solution	$x^* = 10^{-3}(1.5 - 1.6S)$	$V^*(S) = -2.6S^2 - 4.1S - 1.1$

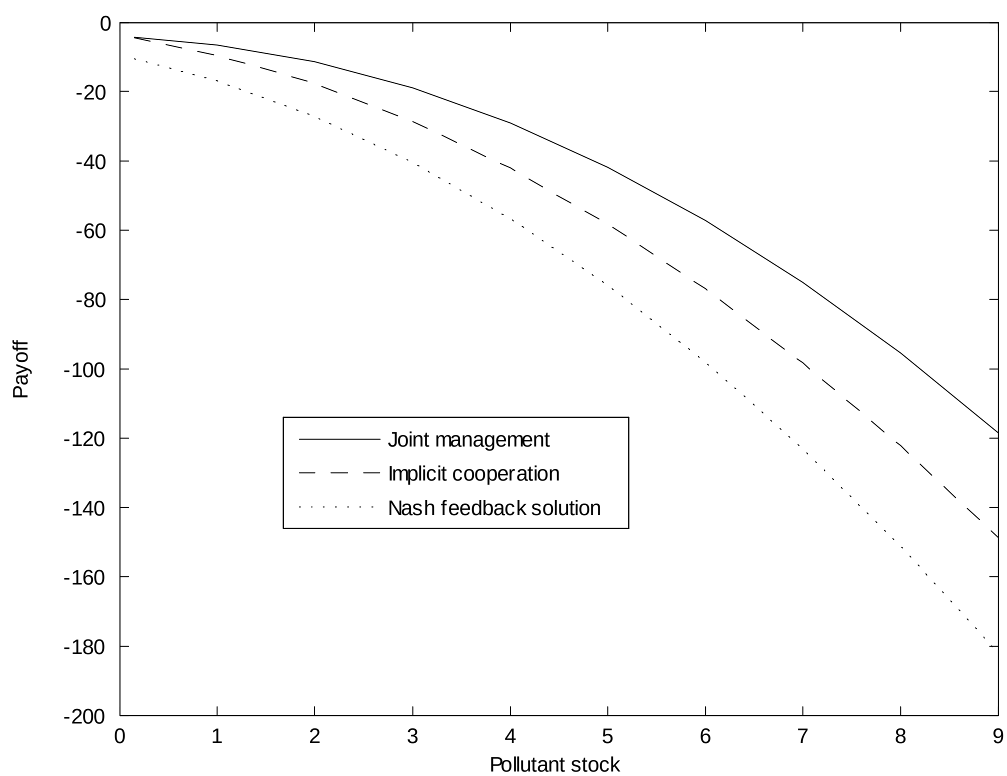

Figure 1. Payoff functions upon cooperation and noncooperation

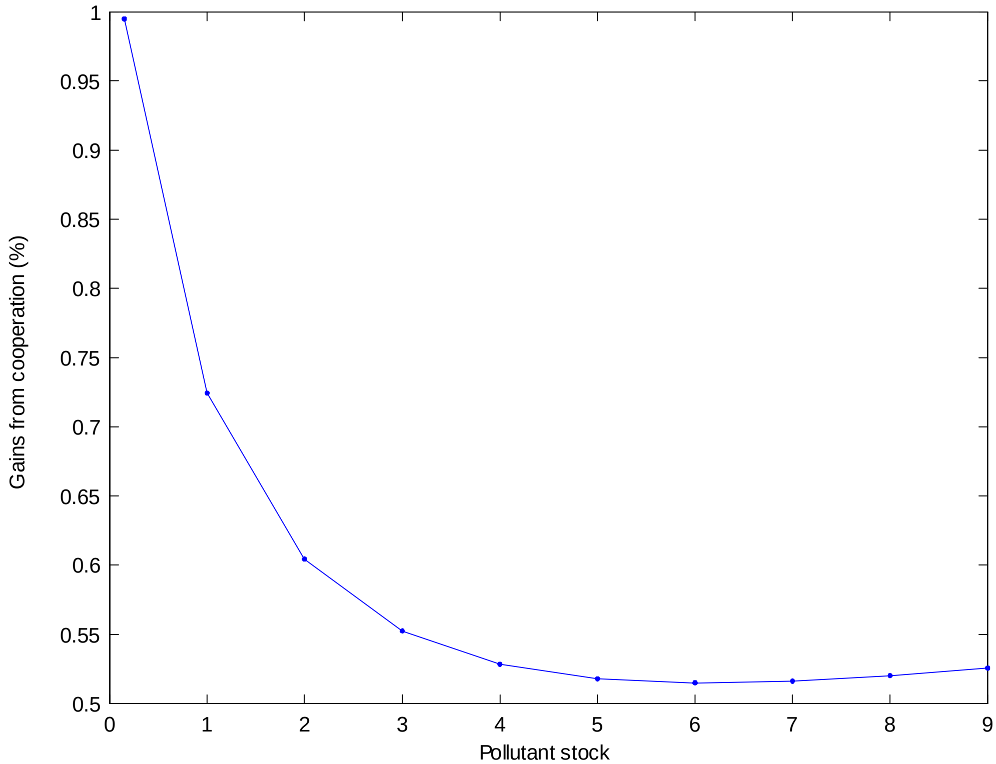

Figure 2. Gains from cooperation

We next examine how the discount factor and the number of players affect the possibility of supporting implicit cooperation under imperfect information. The results are shown in figures 3 and 4. It is not surprising that the benefits of cooperation increase when the discount factor approaches one. Players have to be sufficiently patient for implicit cooperation to be profitable; the lowest discount factor for which cooperation can be supported in our example is $\delta = 0.95$. For δ sufficiently large, implicit cooperation achieves up to 90% of the gain that cooperation would bring to the players under complete information. An interesting finding is that gains from cooperation are not necessarily monotonic in the number of players (e.g. Tarui et al. 2008 obtained a similar result). However, the benefits of cooperation do show a decreasing trend once the number of players exceeds 6. Somewhat surprisingly, cooperation can be supported with up to 35 players under the assumed combination of parameter values. Among the parameter defining the monitoring sets, only p_l changed in response to changes in the model parameters. The probability of returning to the reward phase increased with the gains from cooperation, reaching 25% when the discount factor approached one and 43% when the number of players was only two.

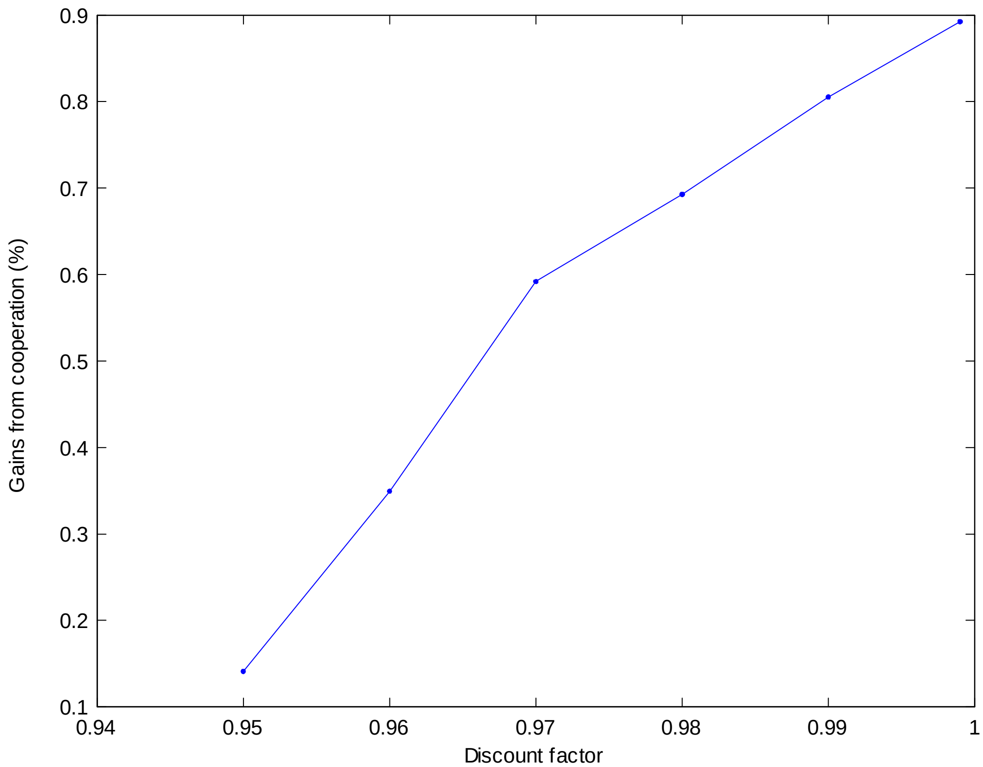

Figure 3. Gains from cooperation as a function of the discount factor

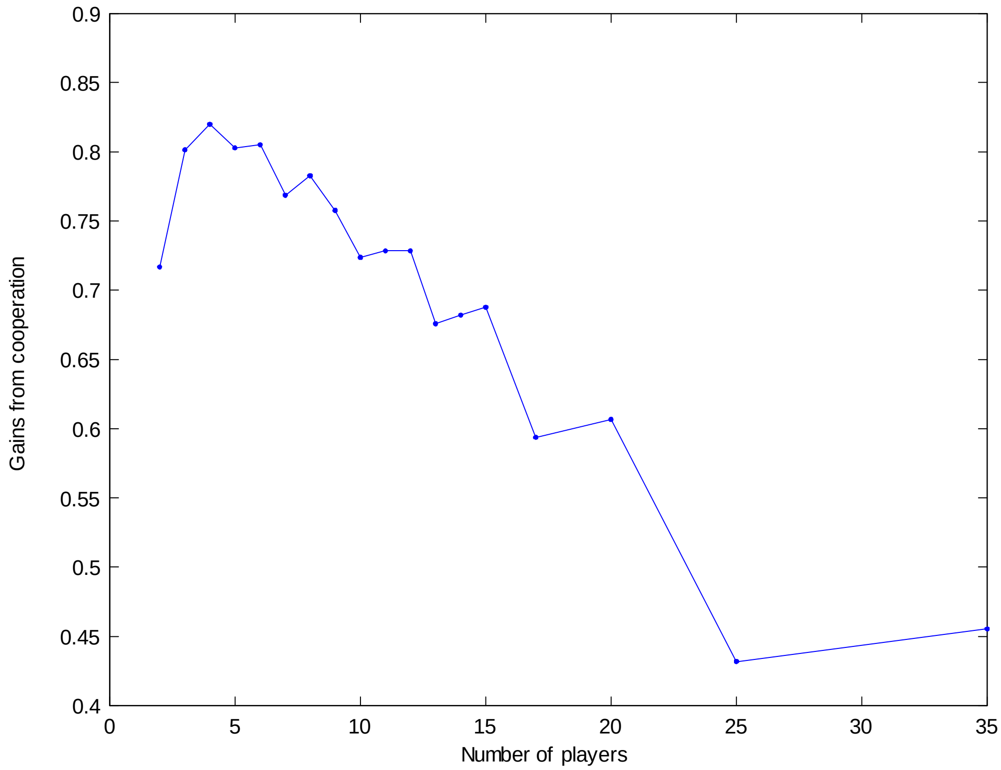

Figure 4. Gains from cooperation as a function of the number of players

4 Summary and discussion

Many international environmental problems involve pollutants that accumulate over time and cause damages to many countries. Reducing the emissions is costly. In order for sovereign countries to cooperate through an international agreement to cut back emissions, the agreement must be self-enforcing. Uncertainty in the environmental process makes it more difficult to achieve cooperation - adherence to the agreement cannot be monitored unambiguously. We proposed a model for an international agreement with harsh punishments when deviations of the agreement are suspected, followed by forgiveness. With a linear-quadratic model, we illustrated how to construct and compute cooperative equilibria. With a numerical example, we examined the conditions under which implicit cooperation can be supported as a perfect public equilibrium. In particular, we examined how the discount factor and the number of players influence the supportability of implicit cooperation. Further sensitivity analysis is still needed for our model, for example on the impact of the precision of the public signal and the nutrient carryover rate.

A natural extension of the model would be to allow for heterogeneity across countries and sanctions through means other than increased emissions, by linking several issues into one agreement. Uncertainty may also be present in terms of thresholds in the stock dynamics, and future research may study how the presence of such thresholds affects the potential for implicit cooperation.

Our results, despite their tentativeness, imply that the mechanism proposed by Abreu, Pearce and Stacchetti (1986, 1990) can be extended to the case of dynamic games and provides a useful approach to constructing international environmental agreements in the presence of stock uncertainty.

References

Abreu, D., Pearce, D. and Stacchetti, E. (1986). Optimal cartel equilibria with imperfect monitoring. *Journal of Economic Theory, 39, 251-269.*

Abreu, D., Pearce, D. and Stacchetti, E. (1990). Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica, 58(5), 1041-1063.*

Green, E. and Porter, R. (1984). Noncooperative collusion under imperfect price information. *Econometrica, 52, 87-100.*

Haurie, A. and Tolwinski, B. (1990). Cooperative equilibria in discounted stochastic sequential games. *Journal of Optimization Theory and Applications, 64(3), 511-535.*

Laukkanen, M. (2003). Cooperative and non-cooperative harvesting in a stochastic sequential fishery. *Journal of Environmental Economics and Management, 45, 454-473.*

Mailath, G, and Samuelson, L. (2006). *Repeated Games and Reputations.* Oxford University Press, New York.

Mason, C., Polasky, S. and Tarui, N. (2008). Cooperation on climate change mitigation. Unpublished manuscript.

Porter, R. (1983). Optimal cartel trigger-price strategies. *Journal of Economic Theory, 29, 313-338.*

Rubinstein, A. (1986), Finite automata play the prisoner's dilemma. *Journal of Economic Theory, 39, 83-96*.

Tarui, N., Mason, C., Polasky, S. and Ellis, G. (2008). Cooperation in the commons with unobservable actions. *Journal of Environmental Economics and Management, 55, 37-51.*