Environmental Issues and Game Theory

Toshiyuki FUJITA
Faculty of Economics, Kyushu University
6-19-1, Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan

Abstract
The model of transboundary pollution is formulated as a game played by countries. As the voluntary and sufficient abatement by an individual country cannot be expected, it is necessary to conclude international environmental agreements (IEAs) to solve the problems. Simple IEAs only on abatements, however, do not function well to extend the size of self-enforcing agreements. Matching schemes are proposed as an additional rule, and it has been shown that the full agreement which consists of all countries becomes self-enforcing.

1. Introduction

Various transboundary environmental issues have attracted much public attentions nowadays. For domestic environmental issues, the government can implement policies such as direct regulation, pollution emission tax, and emissions trading. However, there is no organization with supranational power to control anthropogenic pollutants. Therefore, *international environmental agreements* (*IEAs*) are essential to developing measures to protect the global environment. Without IEAs, each country typically emits more pollution than an efficient level for all countries, because it does not take into account the negative effects on other countries caused by its pollution.

IEA refers to an agreement among countries that specifies what each country should do to improve the environment, which is decided upon negotiation, and, if necessary, supplemental rules. If the IEA in which all related countries participate is reached and each country complies with it, an efficient level of abatement will be attained. A certain amount of financial burden, however, is incurred by abatement, and abatements by one country benefit all countries. This causes an incentive for some countries to enjoy "*free riding*" without implementing their own abatement efforts. As a result, it is often the case that some countries do not sign the agreement and do not commit to the sufficient abatement.

Agreements that generate no incentives for any countries to enjoy free riding are called *self-enforcing*. In other words, in a self-enforcing IEA, participation and national interest do not contradict for all countries, and each country voluntarily signs it without coercion from other countries. The design of the agreement should prevent any free riding and thus realize the large efficient self-enforcing IEA.

Below, game theory and its equilibrium are introduced briefly in Section 2. In Section 3, a simple game model of transboundary pollution is formulated and it is shown that the equilibrium leads to an inefficient outcome. Therefore, we need some agreements to achieve an efficient outcome. In Section 4, IEAs are introduced in the model and it turns out that IEAs which refer to only abatements do not function well in general settings. In Section 5, matching schemes are introduced in the model of IEA, and the analysis of a matching agreement game suggests the existence of a self-enforcing agreement leading to an efficient and equitable outcome, and thus shows that matching schemes are effective.

2. Game Theory

IEAs are typically analyzed in the game theoretic framework. *Game theory* is a study field for analyzing decision-making processes in the so-called game situations, where many rational agents' decision-makings interact with each other. A decision-making agent is called a "*player*," whereas decision-making is called a "*strategy*." The outcome of a game is decided upon the strategies of all the players, whereby players' evaluation value for outcome is called "*payoff*." Players try to maximize their own payoffs.

If every player maximizes his/her payoff against the set of strategies by other players, thus adopting an optimal response strategy, this set of strategies of all the players is said to constitute the *Nash equilibrium*. In Nash equilibrium, on the assumption that the strategies of other players would not change, there are no incentives for every player to change his/her strategies. In this sense, Nash equilibrium is a stable situation. It is an important solution concept of a game.

One well-known game is called "*prisoner's dilemma*," in which two prisoners who have committed a crime together and are subject to police inspection have strategies: "to confess the crime" and "to deny the crime." Under some assumptions, both prisoners choose to confess the crime in Nash equilibrium, although both of them would be better-off if they both denied it. This game shows an example of the inconsistency between individual rationality and collective (social) rationality.

3. Model of transboundary pollution

In this section, it is shown that the model of transboundary pollution without IEA can be the prisoner's dilemma. Starting with a simple example, let there be two countries, Country 1 and Country 2, sharing the same environment. They have a binary choice between "Abate (A)" and "Pollute (P)." If one country, say Country 1, takes A, two countries both gain the benefit of 2 because of better environment, but it is Country 1 who bears the cost burden of 3. The benefit and the cost of taking P are assumed to be 0.

Figure 1 shows the payoff matrix of this game. Both countries take P in Nash equilibrium and payoff to each country is 0 although if they take A the payoff would be 1. Thus, the equilibrium leads to an inefficient outcome and this situation turns out to be prisoner's dilemma.

Let us consider more general models by setting a number of countries as n, and the benefit and the cost of taking A to all countries as bn and c, respectively. It is assumed that $b < c < bn$. As $c < bn$ holds, all countries should take A in an efficient solution, but taking P is always a rational choice for individual country, since $b < c$. Again in this generalized model, it is easy to see that the situation is the prisoner's dilemma.

4. Model of IEA

IEAs are introduced in the game of Section 3. Let us consider the model of IEAs as two-stage games played by countries following Barrett (2003). In the first stage, each country decides independently whether it will accede to an agreement or not. It is assumed that only one agreement would be formed, and cases whereby two or more agreements exist simultaneously are not considered. In the second stage, signatories, through negotiation, make binary choice (A or P) to maximize the total gain of the signatories. Non-signatories, meanwhile, make the same choice independently. The payoff to a country is determined by all countries' decision makings. In the first stage, each country decides whether to accede to an agreement by taking into account its final payoff. This decision is made rationally without any external coercion. Agreements formed in equilibrium are considered to be self-enforcing.

		2: A	2: P
1	A	(1, 1)	(-1, 2)
	P	(2, -1)	(0, 0)

Figure 1. Payoff matrix of transboundary pollution game

Let us derive the equilibrium of the two-country model in Section 3 with IEA. The equilibria of multi-stage games are typically solved by backward induction. Let us first focus on what would happen in the second stage. There are three possible outcomes of the first stage.

(i) The case where both countries accede

Signatories collectively determine that both countries should take A so that the total payoff to signatories is maximized. This gives payoff of 1 to each country.

(ii) The case where one country accedes

A game between one signatory and one non-signatory is played, and the equilibrium is that both countries take P as is shown in Section 3. This gives payoff of 0 to each country.

(iii) The case where no country accedes

The situation is exactly the same as (ii) leading to the equilibrium of (P, P) and each country's payoff is 0.

Now consider the decision-makings in the first stage. Each country decides whether to accede to the agreement taking into account its final payoff. From the payoff matrix shown in Fig. 2, it is easy to see that the only equilibrium of the first stage is (Accede, Accede) and we have an efficient final outcome. In this example, introduction of IEA has changed the equilibrium to one leading to an efficient outcome. This result, however, does not hold in the n-country model. Consider the solution of n-country model. If it is assumed that k countries have acceded in the first stage $(0 \leq k \leq n)$, non-signatories always take P, and signatories take A if $k \geq c/b$ and take P if $k < c/b$ in the equilibrium of the second stage.

Based on this, consider the solution of the first stage. The payoff of a country depends on the number of signatories and on whether it accedes to the agreement or not. If the number of signatories is k, and if the payoffs of signatory and non-signatory are denoted by $\pi^S(k)$ and $\pi^N(k)$ respectively, the agreement of size k is self-enforcing as long as the following two inequalities are established:

		2: Accede	2: Not accede
1	Accede	(1, 1)	(0, 0)
1	Not accede	(0, 0)	(0, 0)

Figure 2. Payoff matrix of IEA game

$$\pi^S(k) \geq \pi^N(k-1),\ \pi^N(k) > \pi^S(k+1).$$

The former inequality means that any signatory cannot gain from withdrawal, which is called the *internal stability*. The latter one means that, if non-signatory accedes to the agreement, it will suffer loss, which is called the *external stability*. Agreements with these two characteristics are the outcome lead by the equilibrium of the first stage.

From the equilibrium of the second stage, it follows that

$$\pi^S(k) = \begin{cases} bk - c & \left(k \geq \dfrac{c}{b}\right) \\ 0 & \left(k < \dfrac{c}{b}\right) \end{cases},$$

$$\pi^N(k) = \begin{cases} bk & \left(k \geq \dfrac{c}{b}\right) \\ 0 & \left(k < \dfrac{c}{b}\right) \end{cases}.$$

By conducting straightforward calculations, we obtain k^* which satisfies $\frac{c}{b} \leq k^* < \frac{c}{b} + 1$ as a size of self-enforcing IEA. In the two-country model, parameter values are $b = 2$, $c = 3$, and $n = 2$, so we get the result of $k^* = 2 = n$, which means the agreement by all countries are formed. However, it is usually not the case in the n-country model because, for example, when $b = 2$ and $c = 3$, an agreement which consists of three countries or more is not self-enforcing. The difference between one country's payoff in an efficient outcome and that in a noncooperative situation is $bn - c$. Therefore, the potential gain of the agreement is greater if b is large and c is small, but in such cases, the size of the self-enforcing IEA is quite small.

In more general settings, it is known that the size of self-enforcing IEAs is extremely small if agreements refer to only abatements and they have no supplemental rules (Carraro and Siniscalco, 1993 and Barrett, 1994). For example, if each country's abatement cost is a convex function of the amount of abatement, and if damage from pollution is proportional to the emission of pollutants, it is proved that only agreements by up to three countries can be self-enforcing (Fujita, 2011). In an agreement signed by four countries or more, individual countries, provided that other countries remain in the agreement, decide not to join the agreement because they can get higher payoffs by doing so.

Let us look at a numerical example. Let there be 10 identical countries. Assume that when the abatement of country i (=1,⋯,10) is x_i, the final payoff of country i is

$$\pi_i = \sum_{j \in S} x_j - \frac{{x_i}^2}{2},$$

where S is a set of signatories. Abatements selected in the second stage by signatories and non-signatories are $k(\equiv |S|)$ and 1, respectively.

Table 1. Numerical result of international agreement game

Number of signatories	Payoff of signatories	Payoff of non-signatories	Internal stability	External stability
0	–	9.5	N.A.	Not stable
1	9.5	9.5	Stable	Not stable
2	10.0	11.5	Stable	Not stable
3	11.5	15.5	Stable	Stable
4	14.0	21.5	Not stable	Stable
5	17.5	29.5	Not stable	Stable
6	22.0	39.5	Not stable	Stable
7	27.5	51.5	Not stable	Stable
8	34.0	65.5	Not stable	Stable
9	41.5	81.5	Not stable	Stable
10	50.0	–	Not stable	N.A.

In Table 1, the final payoffs of signatory and non-signatory after the second stage are given as a function of the number of signatories. Also given are the internal stability and the external stability, which are examined based on these payoffs. For example, if the number of signatories is 6, the payoff of signatory is 22.0. If one signatory withdraws from the agreement and acts as a non-signatory, international stability is not satisfied, because payoff increases to 29.5 (the payoff of non-signatory when the number of signatories is 5). On the contrary, if one non-signatory joins the agreement, payoff will decrease from 39.5 to 27.5, so the external stability is satisfied. As can be seen from Table 1, both of two characteristics are satisfied only if the number of signatories is 3. If the number of signatories is 4 or more, the internal stability is not satisfied.

The arguments in this section reveal that some additional rules other than just asking each country a voluntary participation to the agreement are necessary to make IEAs effective.

5. Matching agreements

The author has recently explored the effectiveness of "*matching agreements*" in several papers (Fujita, 2011 and Fujita, 2013). Matching in this context means that an agent coordinates its own action with the actions of others. Under matching agreements, countries negotiate on so-called *matching rate*, instead of abatements. Afterwards, all countries independently determine their *flat abatements*. Eventually, each country is imposed not only its flat abatement but also an additional abatement which depends on the total of all the other countries' flat abatements and its matching rate. The concept of matching is first proposed by Guttman (1978). The first application of matching scheme in the context of IEAs is done by Rübbelke (2006).

Matching agreement games comprise of three stages. In the first stage, after the announcement of rules on matching, each country decides whether it will accede to the agreement or not. In the second stage, through negotiation among signatories, the common matching rate that maximizes the total payoff of the signatories is determined. In the third stage, both signatories and non-signatories decide their flat abatements based on the matching rate determined in the second stage. The values of matching rates and flat abatements are assumed to be non-negative.

The signatories of matching agreements commit themselves to their own flat abatement, as well as additional abatement, the amount equal to the multiplication by the matching rates determined in the second stage of the total of flat abatement of all the countries including non-signatories. It is assumed that a matching agreement once concluded has a certain binding authority, whereby signatories comply with additional abatements determined by the rule of agreements after the third stage. Non-signatories, meanwhile, do not have to follow the matching rule, and their actual abatements are just their own flat abatements determined in the third stage. The final payoff of each country will be decided after the actual abatements.

Solving for equilibria of matching agreement games reveals that a socially efficient agreement comprising all countries (hereafter referred to as "full agreement") is self-enforcing, regardless of the number of countries, both for symmetric and asymmetric cases. A country that withdraws from the full agreement is exempted from additional abatement decided upon the matching rule. In this case, however, remaining signatories will set their flat abatements as zero. The deviator, after all, is forced to increase flat abatement, and departure from the full agreement does not lead to an increase in gains. That is, matching rates and the arrangement of flat abatement function as penalties for countries withdrawing from the full agreement. Also, in the full agreements of symmetric case, the matching rate determined in the second stage is unity, and all countries obtain the same payoffs. Therefore, it is shown that the equilibria of matching agreements lead to the outcomes desirable also from the perspective of equity.

Table 2. Numerical result of matching agreement game

Number of signatories	Matching rate	Payoff of signatories	Payoff of non-signatories	Internal stability	External stability
0	–	–	9.5	N.A.	Not stable
1	0.17	10.7	11.6	Stable	Not stable
2	0.25	13.5	16.9	Stable	Not stable
3	0.30	17.3	23.7	Stable	Stable
4	0.35	21.9	31.7	Not stable	Stable
5	0.40	27.0	40.5	Not stable	Stable
6	0.46	32.6	49.5	Not stable	Stable
7	0.54	38.6	57.6	Not stable	Stable
8	0.68	44.6	61.7	Not stable	Stable
9	0.97	50.1	47.6	Not stable	Not stable
10	1.00	50.0	–	Stable	N.A.

Let us look at the same numerical example used in Section 4 to examine the self-enforcement of the matching agreements. Table 2 shows matching rates and the final payoffs of signatories and non-signatories as a function of the number of signatories. The internal stability and external stability are examined in the same way as in Section 4. The full agreement allows every country to gain the payoff of 50. If one country withdraws from the full agreement, its payoff decreases to 47.6, whereas those of signatories slightly increase to 50.1. Therefore, the full agreement is self-enforcing. At the same time, however, note that agreements comprising of three countries are also self-enforcing. This means that once an agreement of this scale is formed, it is difficult to expand the agreement.

6. Summary

It is necessary to conclude IEAs to solve transboundary problems. It turns out, however, that simple IEAs regarding abatements do not function well to extend the size of self-enforcing agreements. Matching schemes are proposed as an additional rule, and it has been shown that the full agreement which consists of all countries becomes self-enforcing.

References

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

Barrett, S. (2003), *Environment and Statecraft: The*

Strategy of Environmental Treaty-Making, Oxford University Press.

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

Fujita, T. (2011), "The effectiveness of environmental matching agreements among asymmetric countries," *International Journal of Environmental, Cultural, Economic and Social Sustainability*, **7**, 17-24.

Fujita, T. (2013), "A self-enforcing international environmental agreement on matching rates: Can it bring about an efficient and equitable outcome?," *Strategic Behavior and the Environment*, **3**, 329-345.

Guttman, J.M. (1978), "Understanding collective action: Matching behavior," *American Economic Review*, **68**, 251-255.

Rübbelke, D.T.G. (2006), "Analysis of an international environmental matching agreement," *Environmental Economics and Policy Studies*, **8**, 1-31.