A comment on "International Cooperation for Sale"

Toshiyuki Fujita
Kyushu University

Abstract

We reexamine the analysis of Barrett (2001), that explores the size of a self-enforcing international environmental agreements. Barrett stresses that the key feature to realize the self-enforcing agreement is asymmetries among countries, but we get the following results; certain condition that usually does not hold is required for the Barrett's solution, so it is necessary to reconsider the model settings.

Citation: Fujita, Toshiyuki, (2006) "A comment on "International Cooperation for Sale"." *Economics Bulletin*, Vol. 8, No. 6 pp. 1-7

Submitted: December 2, 2005. **Accepted:** July 24, 2006.

URL: http://economicsbulletin.vanderbilt.edu/2006/volume8/EB-05H00095A.pdf

1 Introduction

The purpose of this article is to reexamine the analysis of Barrett (2001), which investigates the size of a self-enforcing international environmental agreement by considering *asymmetries* among countries. Earlier studies by Carraro and Siniscalco (1993), Barrett (1994) and others have shown that size of the self-enforcing agreement is typically very small. In these models, countries are assumed to make desisions simultaneously. Barrett (2001), however, shows when there is strong asymmetry among countries, the rule of the game changes so that countries that gain much from the agreement (developed countries) first make decisions to join an agreement and then offer side payments to countries that gain less (developing countries) to let them sign in. This change would expand the size of self-enforcing agreement substantially, thus Barrett (2001) insists that the key factor for self-enforcing agreements is asymmetries among countries. He also examines the cost-sharing rule belonging to the core proposed by Chander and Tulkens (1997) and points out that the Chander-Tulkens rule would not be effective without a rather unrealistic assumption that the agreement is terminated once a country withdraws from the agreement. Moreover, he stresses that his theory is consistent with what actually happened through the process of conclusion of the Montereal Protocol.

Although the analysis of Barrett (2001) is quite appealing and highly suggestive, it seems that the derivation of the equilibrium has not been sufficiently described in the paper. Therefore we attempt to reexamine his analysis rigorously. Unfortunately we get the following negative results; certain condition that usually does not hold is required for the most essential part of Barrett's solution, so it is necessary to reconsider the model settings. In Section 2, we briefly introduce the Barrett model and the solution of the model when side payments are not allowed. Section 3 is devoted to description of the solution after the change of the rules. Section 4 provides a summary.

2 Equilibria in the basic model

We describe a game of international cooperation on the pollution abatement formulated in Barrett (2001). Countries sharing the same environment are *players*. Let us assume that there are two *types* of countries, and let N_i denote the number of type i countries $(i = 1, 2)$. The game consists of two stages. In the first stage each country decides simultaneously whether to be a signatory to the agreement or not. In the second stage the actions of signatories are chosen so that they maximize the total payoff of the signatories; whereas each non-signatory behaves noncooperatively. It is assumed that there is at most one agreement at the same time.

There are only two actions to be chosen in the second stage: "Pollute" and "Abate." When type i countries play Pollute and Abate they get payoff of $\alpha_i(b_1z_1 + b_2z_2)$ and $-c+\alpha_i(b_1z_1+b_2z_2)$ respectively, where z_i is the number of type i countries playing Abate and c is the abatement cost which does not depend on the country type. In addition,

$\alpha_2 = 1 > \alpha_1 > 0$, and $b_2 > b_1 > 0$.[1] In the context of environmental issues such as ozone depletion, we can regard type 1 and type 2 players as developing and developed countries, respectively. In order to set the second stage as a "prisoner's dilemma" situation when all countries behave noncooperatively, assumptions $c > b_2$, and $\alpha_1 N_1 + N_2 > c/b_1$ are imposed.

First, equilibrium of the second stage is derived. It is clear that non-signatory countries play Pollute. When the number of type i signatory countries is denoted as k_i, equilibrium of the second stage is:

$$z_i^* = \begin{cases} 0 & \left(\alpha_1 k_1 + k_2 < \dfrac{c}{b_i}\right), \\ k_i & \left(\alpha_1 k_1 + k_2 \geq \dfrac{c}{b_i}\right). \end{cases} \tag{1}$$

Based on (1), equilibria of the first stage (k_1^*, k_2^*) become:

$$\frac{c}{\alpha_1 b_1} < k_1^* < \frac{c}{\alpha_1 b_1} + 1,\ k_2^* = 0 \left(\text{for } N_1 > \frac{c}{\alpha_1 b_1}\right), \tag{2a}$$

$$k_1^* = 0,\ \frac{c}{b_2} < k_2^* < \frac{c}{b_2} + 1 \left(\text{for } N_2 > \frac{c}{b_2}\right). \tag{2b}$$

Barrett (2001, p. 1840) derives the same equilibria. Here, a certain degree of difference between the values of b_1 and b_2 are assumed.[2] When there are strong asymmetries, it is difficult for (2a) type equilibrium to exist and thus only (2b) type is possible. In other words, a self-enforcing agreement consists of type 2 countries only.

3 Equilibria after changes in rules

As described, in the case where countries are strongly asymmetric so that only (2b) type equilibrium exists, type 1 countries do not have an incentive to become a signatory. The rule is then changed as follows: type 2 countries first decide whether to become a signatory and then the signatories offer side payment to type 1 countries to let them to accede. This game consists of three stages. In the first stage, type 2 countries decide whether to accede to the agreement. In the second stage, these countries discuss and choose actions and the amount of side payment m for type 1 countries which commit to accession and playing Abate. Then in the third stage, type 1 countries decide whether to be signatories by considering the offer from type 2 countries. Lastly both types of non-signatories decide their actions, but it is clear that all play Pollute.

[1] Barrett (2001) assumes $\alpha_2 = 1 \geq \alpha_1 > 0$, and $b_2 \geq b_1 > 0$, but we have changed the assumption slightly. The effect of this change is negligible.

[2] This solution is derived by using a similar method to one described in Section 3, but the details are not explained here. Barrett (2001) shows that there is another equilibrium which realizes the agreement with both types of countries when $b_1 \approx b_2$ and $\alpha_1 \approx 1$, that is, only a weak asymmetry is present. We do not show it because it is not directly related to this article.

We use suffix "$**$" to denote the equilibria of this game. When the number of type 2 signatories is k_2, solutions for the second and third stages are summarized as follows[3]:

$$z_2^{**} = 0, m^{**} = 0, z_1^{**} = k_1^{**} = 0 \ \left(k_2 < \frac{c}{b_2}\right), \tag{3a}$$

$$z_2^{**} = k_2, m^{**} = 0, z_1^{**} = k_1^{**} = 0 \ \left(\frac{c}{b_2} \leq k_2 < \frac{c}{b_1} - \alpha_1\right), \tag{3b}$$

$$z_2^{**} = k_2, m^{**} = c - \alpha_1 b_1, z_1^{**} = k_1^{**} = N_1 \ \left(k_2 \geq \frac{c}{b_1} - \alpha_1\right). \tag{3c}$$

In the first stage, k_2^{**} is determined to complete the solution. Barrett (2001) does not clearly describe how to derive the solution of the first stage, but it would be necessary to examine various cases based on the magnitude relation between k_2 and c/b_2, or between k_2 and $c/b_1 - \alpha_1$, and so on.

A set of one-dimensional positive real numbers is divided into three regions: $(0, c/b_2)$, $[c/b_2, c/b_1 - \alpha_1), [c/b_1 - \alpha_1, \infty)$, and each region is called region 1, 2 and 3 respectively. When the value of k_2 is within the region 1, type 2 countries play Pollute and also cooperation of type 1 countries is not available because side payment is not offered as shown in (3a). In region 2, (3b) holds, which means type 2 countries play Abate but side payment for type 1 countries is not offered. In region 3, (3c) holds, so type 2 countries play Abate and the additional cooperation of type 1 countries is available due to side payment $c - \alpha_1 b_1$.

Next let us define the self-enforcing agreement following Carraro and Siniscalco (1993). Factors which affect payoff of a type 2 country are number of signatories (the number of countries playing Abate depends on the number of signatories) and whether a country is a signatory or not. Let payoff of a type 2 signatory and non-signatory be denoted by $\pi_S(k_2)$ and $\pi_N(k_2)$ respectively, when the size of an agreement is k_2. When

$$\pi_S(k_2) > \pi_N(k_2 - 1) \tag{4}$$

holds, the agreement is defined to have *internal stability* or as internally stable. Also when

$$\pi_N(k_2) \geq \pi_S(k_2 + 1) \tag{5}$$

holds, the agreement is defined to have *external stability* or as externally stable. (4) means that the payoff of a signatory decreases when that country defects from agreement. (5) means that the payoff of a non-signatory does not increase by acceding to the agreement. When these conditions are met, the agreement can be judged as self-enforcing.[4] It is also clear that this self-enforcing agreement is consistent with the Nash equilibrium of a game dealing with the accession to an agreement.

[3]See (10a)-(10c) in Barrett (2001, p. 1843).

[4]It is assumed that each country does not accede when it is indifferent between being a signatory or a non-signatory.

Now we evaluate the situations where both internal stability and external stability hold. For internal stability, five cases shown in Table 1 can be considered according to the location of k_2 and $k_2 - 1$.

Table 1 Internal stability

Case	k_2	$k_2 - 1$	Benefit of defection	Cost of defection	Internal stability
1	region 1	region 1	0	0	no
2	region 2	region 1	c	$b_2 k_2$	yes (for $k_2 \neq c/b_2$)
3	region 2	region 2	c	b_2	no
4	region 3	region 2	$c + N_1(c - \alpha_1 b_1)/k_2$	$b_1 N_1 + b_2$	normally no
5	region 3	region 3	$c + N_1(c - \alpha_1 b_1)/k_2$	b_2	no

Let us explain the rationale for Table 1. Note that there is a sufficient distance between region 1 and region 3, so the case where k_2 belongs to region 3 while $k_2 - 1$ belongs to region 1 is not considered. In case 1, all signatories play Pollute, so benefit and cost of defection from the agreement are both zero, which means (4) does not hold[5] and thus the agreement is not internally stable. In case 2, signatories play Abate, therefore benefit of defecting from the agreement is c and cost becomes $b_2 k_2$, because all k_2 countries including the defecting country stop abatement. In this case, since $k_2 \geq c/b_2$, that is, $b_2 k_2 \geq c$, benefit is never greater than cost. Therefore, the agreement is internally stable except when $k_2 = c/b_2$. In case 3, should a signatory defect from the agreement, other signatory countries would continue to play Abate, so cost of defection is only b_2 and benefit is c. From the assumption, $c > b_2$. In case 4, benefit of defection is the sum of abatement cost and contribution for the side payment to type 1 countries, which is $c + N_1(c - \alpha b_1)/k_2$.[6] Meanwhile cost is calculated as $b_1 N_1 + b_2$ by adding $b_1 N_1$, which is due to loss of cooperation by type 1 countries, to b_2. In this case, which of benefit or cost is greater depends on the parameter values. In case 4, k_2 is located near boundary between the regions 2 and 3, so we get $k_2 \approx c/b_1 - \alpha_1$ and thus $c + N_1(c - \alpha_1 b_1)/k_2 - (b_1 N_1 + b_2) \approx c - b_2 > 0$ by simple calculation. Therefore, normally this agreement cannot be regarded as internally stable. In case 5, benefit of defection is the same as case 4 and cost is b_2, the same as case 3, so it is clear that the agreement is not internally stable.

For the external stability, cases can be divided into the following two cases depending on the location of k_2 and $k_2 + 1$. The results are described in Table 2.

Table 2 External stability

Case	k_2	$k_2 + 1$	Benefit of accession	Cost of accession	External stability
6	region 2	region 2	b_2	c	yes
7	region 3	region 3	b_2	$c + \dfrac{N_1(c - \alpha_1 b_1)}{k_2 + 1}$	yes

[5]In this case, $\pi_N(k_2 - 1) - \pi_S(k_2) = 0$.

[6]All type 2 signatories are assumed to bear equal side payment. If other payment rules are applied, the maximum value of benefit of defection would be greater.

According to previous analyses, it is not necessary to consider the case where k_2 is located in region 1. Moreover, when k_2 belongs to region 2, internal stability holds only when $k_2 - 1$ belongs to region 1, therefore the assumption previously made for region 1 and region 3 eliminates the case where $k_2 + 1$ is in the region 3. Derivations of cost and benefit are almost the same as in the case of internal stability, so we do not explain them here.

To summarize, as the equilibrium number of each type of signatories the following two cases are possible. One is:

$$k_1^{**} = 0,\ \frac{c}{b_2} < k_2^{**} < \frac{c}{b_2} + 1 \quad \left(\text{where } \ N_2 > \frac{c}{b_2}\right), \tag{6a}$$

which corresponds to cases 2 and 6, and the other is:

$$k_1^{**} = N_1,\ \frac{c}{b_1} - \alpha_1 < k_2^{**} < \frac{c}{b_1} - \alpha_1 + 1 \quad \left(\text{where } \ N_2 > \frac{c}{b_1} - \alpha_1\right), \tag{6b}$$

which corresponds to cases 4 and 7. Inequalities in the parentheses of (6a) and (6b) are conditions for $k_2 \leq N_2$ to exist in region 2 and 3, respectively. (6a) and (6b) are almost the same as those derived in Section 6 of Barrett (2001),[7] and Barrett insists that (6b) is the preferred type of equilibrium obtained through the change in the rule.

However, we should pay attention to the condition for the existence of (6b) type equilibrium:

$$c + N_1(c - \alpha_1 b_1)/k_2 - (b_1 N_1 + b_2) \leq 0. \tag{7}$$

Let us see whether (7) holds using a numerical example. As in an example of Barrett (2001), let $N_1 = N_2 = 50, c = 100, b_1 = 3, b_2 = 6$ and $\alpha_1 = 0.5$. In this case, Barrett shows the existence of an equilibrium of $k_1^{**} = 50$ and $k_2^{**} = 33$, but *it is not an equilibrium, because* (7) *does not hold.* Payoffs of type 2 countries before and after defecting from the agreement are actually calculated as $\pi_S(33) = 98.8$ and $\pi_N(32) = 192$ respectively, which means payoff after defection increases substantially. In order for (7) to hold, parameter b_2 must be greater than 99.24. We have $b_2 < c = 100$, so the value of b_2 is only within very narrow range. Considering the change of paremater N_1, (7) holds when N_1 is greater than 6204, but this is obviously unrealistic.

Additionally, if we fix b_1, b_2 and α_1, (7) tends to hold when c is smaller, as well as when N_1 is larger. A smaller c leads to a smaller benefit of free-riding, and a larger N_1 means higher effects of side payment. But in this case, $c/\alpha_1 b_1 < N_1$ is satisfied and thus the assumption of strong asymmetries does not hold any longer.

4 Summary

We get the result that the equilibrium practically does not change, even though the rule is changed so that the specific types of countries first decide whether to be signatories.

[7] Actually, the condition $N_2 < c/b_1 - \alpha_1$ is included to (6a) in Barrett (2001).

In other words, Barrett's (2001) argument that the asymmetries are the key to a larger self-enforcing agreement is questionable, therefore finding other factors is necessary.

As a matter of fact, success in the Montreal Protocol might be considered to be attributable to the presence of incentives of every single developed country, as explained by Barrett himself (in other words, $b_2 \geq c$ holds in the model).[8] As a political measure to expand the self-enforcing agreement, the side payment rule proposed by Chander and Tulkens (1997) criticized by Barrett seems to be worth reconsidering.

References

Barrett, S. (1994) "Self-Enforcing International Environmental Agreements" *Oxford Economic Papers* **46**, 878-894.

Barrett, S. (2001) "International Cooperation for Sale" *European Economic Review* **45**, 1835-1850.

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 (1997) "The Core of an Economy with Multilateral Environmental Externalities" *International Journal of Game Theory* **26**, 379-402.

[8] See Barrett (2001, p. 1846).