Designing International Environmental Agreements under Participation Uncertainty

Mao, Liang

College of Economics, Shenzhen University

15 May 2017

Designing International Environmental Agreements under Participation Uncertainty

Liang Mao*

May 2017

Abstract

We analyze the design of optimal international environmental agreement (IEA) by a three-stage coalition formation game. A certain degree of participation uncertainty exists in that each country choosing to sign the IEA for its best interest has a probability to make a mistake and end up a non-signatory. The IEA rule, which specifies the action of each signatory for each coalition formed, is endogenously determined by a designer, whose goal is to maximize the expected payoff of each signatory. We provide an algorithm to determine an optimal rule and compare this rule to some popular rules used in the literature.

Keywords: International environmental agreement, coalition formation, participation uncertainty, stable coalition
JEL codes: Q54, C72, H41

*College of Economics, Shenzhen University, Shenzhen, Guangdong, 518060, China. Email: maoliang@szu.edu.cn.

1 Introduction

The human society is facing a serious threat of climate change, mainly due to the emission of greenhouse gases (GHG). For a country, reducing the emission of GHG can be regarded as providing a public good benefiting the whole world. However, voluntary abatement of GHG is typically not sufficient, because every country has an incentive to free ride on the abatement effort of other countries. One method to overcome this free-riding problem is to form a coalition wherein the members sign a self-enforcing international environmental agreement (IEA) and follow certain abatement rules. The Kyoto Protocol and the Paris Agreement are examples of such IEAs.

The formation of these coalitions is sometimes modeled as a two-stage game, or its variant, played by some self-interested countries.[1] In stage one (participation stage), each country decides whether to join the coalition and sign the IEA. In stage two (abatement stage), those signing the IEA have to follow the IEA rule, while each non-signatory can decide its own abatement level.

Nevertheless, it is reported that IEAs do not work very well. For instance, Kellenberg and Levinson (2014) suggest that "IEAs appear to do little more than ratify what countries would have done absent the agreements." There could be many reasons for the failure of IEAs, but in this paper, we focus only on the following two of them.

First, the IEA rule is typically exogenously given and may not provide for much incentive to overcome the free-riding problem. For instance, a large body of studies assume that in stage two of the game, all coalition members should coordinate their actions and maximize the total payoffs of the coalition formed in stage one. We call this IEA rule the maximal total payoff (MTP) rule. In section 5, we will show that the MTP rule is generally not optimal.

In order to overcome the problem raised by exogenous IEA rules, several studies analyzed the endogenous determination of the IEA rules. For example, Carraro et al. (2009) discuss the MTP rule with an additional restriction

[1] For example, see Carraro and Siniscalco (1993), Barrett (1994), Thoron (1998), Finus (2001), Masoudi and Zaccour (2017).

of minimal participation; here, the threshold of forming the coalition is endogenously determined. Köke and Lange (2017) considers an endogenous rule that simultaneously determine the threshold of cooperation and the signatories' abatement level. However, these studies analyze only certain special cases of endogenous rules and hence cannot be considered fully general. In particular, a signatory's abatement level as specified by these rules need not vary endogenously according to the coalition formed in stage one, except when the change involves the minimal participation condition.

The second reason for the IEAs' failure is participation uncertainty: a country initially intending to sign the IEA for its own interest has a chance to make a mistake and end up a non-signatory. This uncertainty, which makes it more difficult to form a large coalition, may be due to various reasons under different cases. For example, ratification of the IEA may be prevented by some interest groups[2], or a newly elected leader may overturn the decision made by his predecessor. In contrast, we assume that the probability that a country not intending to sign the IEA becomes a signatory, which rarely happens in reality, is zero. Also note that participation uncertainty is unlike several other types of risk and uncertainty discussed in the IEA literature.[3]

The main purpose of this study is to extend the traditional coalition formation models of IEA to allow for participation uncertainty and fully general rules that are endogenously determined. We hope these extensions will help us design a better IEA rule than the ones used in reality and those in the literature. To this end, we employ a three-stage coalition formation game. In stage one (designing stage), a designer[4] launches an coalition and announces an IEA rule, which is a function specifies the abatement level of a signatory (coalition member) for each possible coalition formed in stage two of the game. As the initiator of the coalition, the designer's goal is to maximize the expected payoff of each signatory. Stage two extends the usual participation stage by assuming a given probability $\varepsilon \geq 0$ that each country choosing to sign the IEA would finally end up not being a coalition member.

[2]See Köke and Lange (2017).

[3]See, among others, Kolstad (2007), Dellink et al. (2008), Kuiper and Olaizola (2008), Hong and Karp (2014), Cazals and Sauquet (2015), and Masoudi et al. (2016).

[4]For example, the United Nations.

Stage three is the usual abatement stage that determines each country's abatement level and payoff.

This three-stage game can be solved by backward induction. Thus, we determine the optimal rule that the designer would announce in stage one and the coalition of countries that intend to sign the IEA in stage two. Note that participation uncertainty would make it more difficult to determine the coalition that will form in stage two. Given the IEA rule announced in stage one, a coalition will be formed if and only if it is stable; that is, no country would change its participation decision both before and after observing the mistakes made by some other countries. We prove that given any IEA rule, the cardinality of a stable coalition can be uniquely determined (Proposition 1). Furthermore, we provide an algorithm to determine an optimal rule for the designer (Theorem 1).

Some IEA rules, for example, the MTP rule, the minimal participation rule, and the coalition unanimity rule, are commonly used in the literature. We show through an example that these rules are generally not optimal for the designer. Additionally, we briefly discuss the conditions under which these rules are optimal (Proposition 2, 3).

The remainder of this paper is organized as follows. Section 2 presents the setup of the model and the three-stage coalition formation game. We solve this game and derive an optimal IEA rule in section 3 and 4. Some traditional IEA rules are discussed and compared with the optimal rule in section 5. Finally, section 6 concludes the paper.

2 The model

Let $N = \{1, 2, \ldots, n\}$ be a set of homogeneous countries, where $n \geq 2$. There is a perfectly divisible good with negative externalities, for example, greenhouse gas. Furthermore, let x_i denote country i's abatement level of the good and $x = (x_1, \ldots, x_n)$ be an abatement combination.

Given x, country i's payoff is

$$u_i(x) = \alpha \sum_{j \in N} x_j - \frac{1}{2} x_i^2, \tag{1}$$

where $\alpha > 0$ is the constant marginal benefit from total abatement $\sum_{j \in N} x_j$ due to negative externalities of the good, and $x_i^2/2$ is country i's abatement cost. Assume that payoffs are transferable, and therefore social welfare is the total payoffs of all countries:

$$U(x) = \sum_{i \in N} u_i(x) = n\alpha \sum_{i \in N} x_i - \sum_{i \in N} \frac{x_i^2}{2}.$$

An abatement combination $(x_1^*, \ldots, x_n^*)$ is said to be socially optimal if it maximizes social welfare. The first-order conditions $\partial U(x)/\partial x_i = 0$ yield

$$x_i^* = n\alpha, \quad \forall i \in N. \tag{2}$$

On the other hand, if each country i chooses x_i to maximize its own payoff u_i given the other countries' abatement levels, the first-order conditions $\partial u_i(x)/\partial x_i = 0$ lead to

$$\bar{x}_i = \alpha, \quad \forall i \in N. \tag{3}$$

Note that $\bar{x}_i$ is a dominant abatement level of i, regardless of other countries' actions. From this, it follows that $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_n)$ is the unique Nash equilibrium of this non-cooperative abatement game.

Since $x_i^* > \bar{x}_i$, the world suffers from too much emission of the good. This is a commonly known social dilemma due to externalities. One possible method to partially overcome this problem is to form a coalition that regulates the countries' actions by a self-enforcing IEA. The formation of the coalition follows a three-stage game.

- Stage one. A designer announces an IEA rule e, which is a function assigning a real value $e(\overline{m}) \geq 0$ to each integer $\overline{m} \in [1, n]$, where $\overline{m}$ is the cardinality of the coalition $\overline{M}$ that will be formed in stage two. A rule can be denoted by the vector $e = \big(e(1), \ldots, e(n)\big) \in \mathbb{R}_+^n$.

- Stage two. All countries in N simultaneously decide whether or not to sign the IEA. Let M denote the set of countries that choose to sign and $m = |M|$ denote its cardinality. However, there is a one-way uncertainty with regard to each country's final participation decision. Specifically, each country $i \in M$ has a probability ε of making a mistake and failing to sign the IEA, where $0 \leq \varepsilon < 1$ is exogenously given. However, each $j \notin M$ never makes a mistake and would certainly not sign the IEA. Let $\overline{M}$ denote the set of signatory countries that choose to sign the IEA and make no mistake, and $\overline{m} = |\overline{M}|$ denote its cardinality. Given m and $\varepsilon \in (0,1)$, $\overline{m}$ follows a binomial distribution so that the probability that $\overline{m} = k$ is

$$b(k; m, 1-\varepsilon) = \frac{m!}{k!(m-k)!}\varepsilon^{m-k}(1-\varepsilon)^k, \quad \forall k = 0, 1, \ldots, m.$$

  Additionally, if $\varepsilon = 0$, then $b(m; m, 1) = 1$, $b(k; m, 1) = 0$, $\forall k < m$.

- Stage three. Given rule e and the coalition $\overline{M}$, each signatory $i \in \overline{M}$ carries out its abatement $x_i = e(\overline{m})$ according to e, while each non-signatory $j \notin \overline{M}$ chooses its dominant abatement level $x_j = \alpha$. All countries receive their respective payoffs according to (1).

Now, let $G(n, \alpha, \varepsilon)$ denote this three-stage game. Assume that each country is risk neutral and chooses its action to maximize its expected payoff. We also assume that a country will choose to sign the IEA if it is indifferent between signing and not signing. The designer will choose $e \in \mathbb{R}^n_+$ to maximize the (identical) expected value of the payoff of each signatory $i \in \overline{M}$.

3 Stable coalition and equilibrium scale

We solve game $G(n, \alpha, \varepsilon)$ by backward induction. Consider stage three first. Given e and $\overline{m}$, let

$$X(\overline{m}, e) = \overline{m}e(\overline{m}) + (n - \overline{m})\alpha$$

denote the total abatement of all countries. Now, a signatory's payoff is

$$\overline{u}^C(\overline{m}, e) = \alpha X(\overline{m}, e) - \frac{e(\overline{m})^2}{2}, \quad \text{if } \overline{m} \geq 1. \tag{4}$$

Additionally, let

$$\overline{u}^C(0, e) = u_i(\bar{x}) = (n - 1/2)\alpha^2. \tag{5}$$

A non-signatory's payoff is

$$\overline{u}^I(\overline{m}, e) = \alpha X(\overline{m}, e) - \frac{\alpha^2}{2}, \quad \text{if } \overline{m} < n.$$

In stage two of $G(n, \alpha, \varepsilon)$, given e and $m_{-i} = |M \backslash \{i\}|$, country i's expected payoff is

$$u^C(m_{-i} + 1, e) = \sum_{k=0}^{m_{-i}} b(k; m_{-i}, 1 - \varepsilon) \left[\varepsilon \overline{u}^I(k, e) + (1 - \varepsilon)\overline{u}^C(k + 1, e)\right]$$

if i chooses to sign the IEA ($i \in M$), and it is

$$u^I(m_{-i}, e) = \sum_{k=0}^{m_{-i}} b(k; m_{-i}, 1 - \varepsilon)\overline{u}^I(k, e)$$

if i chooses not to sign ($i \notin M$). In other words, $u^C(m, e)$ and $u^I(m, e)$ are the expected payoffs of a country that chooses to sign and not to sign the IEA, respectively, when exactly m countries choose to sign the IEA.

We use the concept of stable coalition to predict which countries choose to sign the IEA in stage two. Roughly speaking, a coalition M is stable if the countries in the coalition are the only ones choosing to sign the IEA before any mistake occurs (*ex ante* stable), and countries that do not make mistakes will not change their decisions after some other countries have made mistakes (*ex post* stable).

Formally, following d'Aspremont et al. (1983) and many others, coalition M is said to be *ex ante* stable relative to e if no country $i \in M$ is willing to unilaterally leave M and no country $j \notin M$ is willing to unilaterally join M before uncertainty is realized. Hence, a coalition $M \notin \{\emptyset, N\}$ is *ex ante*

stable relative to e if

$$u^C(m,e) \geq u^I(m-1,e),\ u^C(m+1,e) < u^I(m,e).$$

In addition, $M = \emptyset$ is *ex ante* stable relative to e if $u^C(1,e) < u^I(0,e)$, while $M = N$ is *ex ante* stable relative to e if $u^C(n,e) \geq u^I(n-1,e)$.

A coalition $M \neq \emptyset$ is said to be *ex post* stable relative to e if no signatory will regret its decision to sign the IEA and withdraw after uncertainty is realized, no matter how many countries in M have made mistakes; that is,

$$u^C(k+1,e) \geq u^I(k,e),\ \forall k \in [0, m-1].$$

In addition, $M = \emptyset$ is trivially *ex post* stable relative to any $e \in \mathbb{R}^n_+$.

Finally, M is said to be stable relative to e if it is both *ex ante* stable relative to e and *ex post* stable relative to e. Ultimately, a stable coalition will not provide any incentive for any country to change its decision regarding participation under any circumstance. Consequently,

(a) $M = \emptyset$ is stable relative to e if

$$u^C(1,e) < u^I(0,e); \tag{6}$$

(b) $M \notin \{\emptyset, N\}$ is stable relative to e if

$$u^C(m+1,e) < u^I(m,e),\ u^C(k+1,e) \geq u^I(k,e),\ \forall k \in [0, m-1]; \tag{7}$$

(c) $M = N$ is stable relative to e if

$$u^C(k+1,e) \geq u^I(k,e),\ \forall k \in [0, n-1]. \tag{8}$$

Because of the symmetry of countries, whether a coalition M is stable relative to rule e depends only on $m = |M|$. If a coalition M is stable relative to e, then we say that m is an equilibrium scale relative to e.

The following proposition establishes the existence and uniqueness of an equilibrium scale relative to any given rule. Let $m(e)$ denote this unique

equilibrium scale relative to e. In other words, if the designer announces rule e in stage one, then in stage two there will be $m(e)$ countries choosing to sign the IEA. In the proof of this proposition, we provide an algorithm to derive $m(e)$ for each $e \in \mathbb{R}^n_+$.

Proposition 1. *There is a unique equilibrium scale relative to each $e \in \mathbb{R}^n_+$.*

Proof. First, we prove that there exists at most one equilibrium scale relative to any e. Assume for a contradiction that both m_1 and m_2 are equilibrium scales relative to some e, where $m_1 < m_2$. Since m_1 is an equilibrium scale, we have $u^C(m_1+1, e) < u^I(m_1, e)$, which contradicts the assumption that m_2 is also an equilibrium scale.

If $u^C(1, e) < u^I(0, e)$, then $m(e) = 0$; otherwise, we have $u^C(1, e) \geq u^I(0, e)$. Furthermore, if $u^C(2, e) < u^I(1, e)$, then $m(e) = 1$; otherwise, we have $u^C(1, e) \geq u^I(0, e)$, $u^C(2, e) \geq u^I(1, e)$. Proceeding in this manner, we shall either find an equilibrium scale $m(e) < n$, or eventually have $u^C(k, e) \geq u^I(k-1, e)$, $k = 1, 2, \ldots, n$, which implies that $m(e) = n$. □

Given $e \in \mathbb{R}^n_+$, let $Eu^C(e)$ denote the expected payoff of a signatory. Since there are $m(e)$ countries intending to sign the IEA, the probability that there are exactly k signatories is $b\big(k; m(e), 1-\varepsilon\big)$ for all $k \leq m(e)$. Thus, we have

$$Eu^C(e) = \sum_{k=0}^{m(e)} b\big(k; m(e), 1-\varepsilon\big)\overline{u}^C(k, e). \tag{9}$$

It follows from (5) that that when no country signs the IEA, the designer will take a non-signatory's payoff as a substitute for a signatory's payoff; we make this trivial assumption only to ensure that the objective of the designer is always well-defined.

4 An optimal rule

Now, consider stage one of $G(n, \alpha, \varepsilon)$. The objective of the designer in this stage is to maximize $Eu^C(e)$ by choosing an appropriate rule e. If a rule e exists such that $Eu^C(e) \geq Eu^C(e')$ for all $e' \in \mathbb{R}^n_+$, then we say that e is

optimal. The following theorem shows that there always exists an optimal rule. The proof explicitly demonstrates how to construct an optimal rule.

Theorem 1. *For any game $G(n, \alpha, \varepsilon)$, there exists an optimal rule e^*.*

Proof. For each integer $s \in [0, n]$, let $E(s) = \{e \in \mathbb{R}^n_+ \,|\, m(e) = s\}$ be the set of rules whose equilibrium scale is s. Additionally, for each $s \in [1, n]$, define

$$\overline{E}(s) = \big\{\big(e(1), \ldots, e(s)\big) \in \mathbb{R}^s_+ \,|\, u^C(k{+}1, e) \geq u^I(k, e),\ \forall k \in [0, s{-}1]\big\}. \quad (10)$$

Now, it is obvious that, if $e = \big(e(1), \ldots, e(n)\big) \in E(s)$, then $\big(e(1), \ldots, e(s)\big) \in \overline{E}(s)$. The proof of the following lemma is in the appendix.

Lemma 1. *For each $s \in [1, n]$, $\overline{E}(s)$ is a non-empty bounded closed set.*

If $m(e) = 0$, then from (6), we have $e(1) \neq \alpha$. Therefore, $Eu^C(e) = (n - \frac{1}{2})\alpha^2$ for all $e \in E(0) = \{e \in \mathbb{R}^n_+ \,|\, e(1) \neq \alpha\}$.

If $m(e) = 1$, then from (7), we can easily obtain $E(1) = \{e \in \mathbb{R}^n_+ \,|\, e(1) = \alpha,\ e(2) \in (-\infty, \alpha) \cup (3\alpha, \infty)\}$, $Eu^C(e) = \varepsilon\overline{u}^C(0, e) + (1 - \varepsilon)\overline{u}^C(1, e) = (n - \frac{1}{2})\alpha^2$ for all $e \in E(1)$.

If $m(e) = 2$, then again from (7), we have $E(2) = \{e \in \mathbb{R}^n_+ \,|\, e(1) = \alpha,\ e(2) \in [\alpha, 3\alpha],\ e(3) \in (-\infty, f_1(e(2))) \cup (f_2(e(2)), \infty)\}$, where f_1 and f_2 are two functions that can be easily determined, but are irrelevant to our analysis. When $e \in E(2)$, $Eu^C(e) = \varepsilon^2\overline{u}^C(0, e) + 2\varepsilon(1 - \varepsilon)\overline{u}^C(1, e) + (1 - \varepsilon)^2\overline{u}^C(2, e)$ depends only on $e(1)$ and $e(2)$. Therefore, we have $\max_{e \in E(2)} Eu^C(e) = \max_{(e(1), e(2)) \in \overline{E}(2)} Eu^C(e)$. Similarly, for $s \in [2, n]$, we have

$$\max_{e \in E(s)} Eu^C(e) = \max_{(e(1), \ldots, e(s)) \in \overline{E}(s)} Eu^C(e).$$

Now, for any $s \in [1, n]$, there exists $e^*_s = \big(e^*_s(1), \ldots, e^*_s(n)\big) \in E(s)$ such that $Eu^C(e^*_s) \geq Eu^C(e')$ for all $e' \in E(s)$. That is, $Eu^C(e^*_s) = \max_{e \in E(s)} Eu^C(e)$. In fact, from Lemma 1, we can derive $\big(e^*_s(1), \ldots, e^*_s(s)\big)$ by solving the constrained optimization problem $\max_{(e(1), \ldots, e(s)) \in \overline{E}(s)} Eu^C(e)$,[5] and $\big(e^*_s(s{+}1), \ldots, e^*_s(n)\big)$ can be any vector as long as $u^C(s{+}1, e^*_s) < u^I(s, e^*_s)$.

[5] Apply the Kuhn–Tucker theorem.

Finally, the maximal value of $Eu^C(e)$ equals $\max_{1\le s\le n}\max_{e\in E(s)} Eu^C(e)$. An optimal rule e^* can thus be found, where

$$Eu^C(e^*)=\max_{1\le s\le n}\{Eu^C(e_1^*),Eu^C(e_2^*),\ldots,Eu^C(e_n^*)\}.$$

This ends the proof of the proposition. □

As an example, consider game $G(5,2,0.1)$. In Table 1, we list for each $s\in[1,5]$ the value of $Eu^C(e_s^*)$, which is the maximal value of $Eu^C(e)$ under the condition $m(e)=s$, as well as the corresponding rules $\big(e_s^*(1),\ldots,e_s^*(s)\big)$. Since $Eu^C(e_5^*)>Eu^C(e_4^*)>Eu^C(e_3^*)>Eu^C(e_2^*)>Eu^C(e_1^*)$, we have $e^*=e_5^*=(2,2,3.61,6.44,10)$, and $Eu^C(e^*)=42.78$.

Table 1: Calculating optimal rule e^* for $G(5,2,0.1)$

s	$\big(e_s^*(1),\ldots,e_s^*(s)\big)$	$Eu^C(e_s^*)$
1	(2)	18
2	(2, 4)	19.62
3	(2, 4, 6)	24.32
4	(2, 2.95, 5.23, 8)	32.13
5	(2, 2, 3.61, 6.44, 10)	**42.78**

5 Discussion

Now, we discuss some special rules commonly used in the literature and compare them to the optimal rule e^*.

(a) A rule e^a is called the MTP rule if it always aims to maximize the total payoffs of all signatories. Because of the symmetry of players, we have $\overline{m}\cdot\overline{u}^C(\overline{m},e^a)\ge\overline{m}\cdot\overline{u}^C(\overline{m},e')$, or $\overline{u}^C(\overline{m},e^a)\ge\overline{u}^C(\overline{m},e')$, for all $\overline{m}\in[1,n]$ and $e'\in\mathbb{R}_+^n$. That is, for all $\overline{m}\in[1,n]$, e^a maximizes $\overline{u}^C(\overline{m},e)$, and thus $e^a(\overline{m})=\alpha\overline{m}$.

(b) A rule e^b is called a minimal participation rule[6] if there exists $m^*\in[2,n]$ such that $e^b(\overline{m})=\alpha$ when $1\le\overline{m}<m^*$, and $e^b(\overline{m})=q>\alpha$ when

[6]See Köke and Lange (2017).

$\overline{m} \geq m^*$. In other words, this rule requires an abatement level q for signatories when at least m^* countries sign the IEA. In particular, if $m^* = n$, e^b is called the coalition unanimity rule[7].

(c) A rule e^c is called an MTP rule with minimal participation[8] if there exists $m^* \in [2, n]$ such that $e^c(\overline{m}) = \alpha$ for all $1 \leq \overline{m} < m^*$, and $e^c(\overline{m}) = \alpha\overline{m}$ for all $\overline{m} \geq m^*$. Hence, e^c is a combination of e^a and e^b.

To compare these rules, we reconsider the example in the previous section where $n = 5$, $\alpha = 2$. First, the MTP rule e^a satisfies $e^a(\overline{m}) = 2\overline{m}$ for all $\overline{m} \in [1, n]$. Next, consider a coalition unanimity rule e^b where $e^b(\overline{m}) = 2$ when $\overline{m} < 5$ and $e^b(5) = 10$. Finally, consider an MTP rule with minimal participation e^c where $e^c(4) = 8$, $e^c(5) = 10$, and $e^c(\overline{m}) = 2$ when $\overline{m} < 4$. For these rules and the optimal rule e^*, we list the corresponding $m(e)$ and $Eu^C(e)$ for some particular value of ε in Table 2.[9] We shall explain and discuss the data in this table.

Table 2: Simulation for $G(5, 2, \varepsilon)$

	e^a		e^b		e^c		e^*	
ε	$m(e)$	$Eu^C(e)$	$m(e)$	$Eu^C(e)$	$m(e)$	$Eu^C(e)$	$m(e)$	$Eu^C(e)$
0	3	26	5	**50**	4	36	5	50
0.1	3	24.32	5	**36.90**	4	29.81	5	42.78
0.2	3	22.86	5	28.49	5	**35.86**	5	36.88
0.3	3	21.63	5	23.38	5	**29.86**	5	32.02
0.4	3	20.59	5	20.49	5	**25.15**	5	28.03
0.5	4	**21.88**	5	19	5	21.81	5	24.79
0.6	4	**20.38**	5	18.33	5	19.71	5	22.22
0.7	5	**20.26**	5	18.08	5	18.59	5	20.26
0.8	5	**18.94**	5	18.01	5	18.13	5	18.94
0.9	5	**18.22**	5	18.00	5	18.01	5	18.22

For the designer, a rule e has two important aspects that may affect the value of his objective $Eu^C(e)$. On the one hand, the designer may wish

[7]See Chander and Tulkens (1997).

[8]See Carraro et al. (2009).

[9]For each ε in the table, the maximal value of $Eu^C(e)$, where $e \in \{e^a, e^b, e^c\}$, is highlighted in bold to show which of the three rules is best for the designer.

more countries to sign the IEA and that the signatories engage in a high abatement level. Thus, the rule should provide for a strong incentive for cooperation or strong punishment for free riding by creating a large payoff gap between signing and not signing. On the other hand, the designer may also wish to reduce the harm that uncertainty brings on the expected payoffs of signatories. This can be accomplished only by designing a rule by which even when some countries do not sign the IEA due to mistakes, other signatories can still maintain a relatively high level of abatement, leading to a small payoff gap between signing and not signing.

We call these two aspects of the rules as incentive effect and uncertainty effect respectively. A rule has a strong/weak incentive effect if it provides strong/weak incentives for countries to sign the IEA; a rule has a strong/weak uncertainty effect if a certain ε has a small/large impact on $Eu^C(e)$.

The incentive effect and uncertainty effect are typically contradictory. For example, a rule with a strong uncertainty effect usually has a weak incentive effect. This is because any factor of the rule protecting the signatories from harm caused by uncertainty will require those signatories to maintain a high level of abatement regardless of the other countries' mistakes. However, this requirement would also reduce the incentive for cooperation. An appropriate rule should have a good balance between the two conflicting effects.

From Table 2, of the three special rules we discussed above, the coalition unanimity rule e^b is optimal when $\varepsilon = 0$. This is because e^b obviously has a strong incentive effect and weak uncertainty effect, but the latter is irrelevant when $\varepsilon = 0$. Moreover, the next proposition shows that the coalition unanimity rule is almost optimal when uncertainty is sufficiently small.

Proposition 2. *Suppose $e^b(n) = \alpha n$, and $e^b(m) = \alpha$ for all $m < n$. For any $\mu > 0$, there exists $\gamma > 0$ such that if $\varepsilon < \gamma$, $Eu^C(e^b) > Eu^C(e') - \mu$, for all $e' \in \mathbb{R}^n_+$.*

Proof. It is obvious that $m(e^b) = n$. Given any $\mu > 0$, when ε is sufficiently small, $Eu^C(e^b) = \sum_{k=0}^{n} b\big(k; n, 1-\varepsilon\big)\overline{u}^C(k, e^b)$ can be arbitrarily close to $\overline{u}^C(n, e^b)$, and thus $Eu^C(e^b) > \overline{u}^C(n, e^b) - \mu$. From (4), it is easy to verify that $\overline{u}^C(n, e^b) \geq \overline{u}^C(m, e')$ for all $m \leq n$ and all $e' \in \mathbb{R}^n_+$. Hence, $Eu^C(e^b) >$

$\overline{u}^C(n, e^b) - \mu \geq \sum_{k=0}^{m(e')} b\big(k; m(e'), 1-\varepsilon\big)\overline{u}^C(k, e') - \mu = Eu^C(e') - \mu$, for all $e' \in \mathbb{R}^n_+$. □

In contrast, from Table 2, the MTP rule e^a is an optimal rule only when ε is large enough. This turns out to be a general outcome according to the next proposition, which suggests that the MTP rule has a relatively strong uncertainty effect.

Proposition 3. *There exists $\theta > 0$ such that if $\varepsilon > \theta$, e^a is optimal.*

Proof. See the appendix. □

Consider game $G(5, 2, \varepsilon)$ again and suppose that one country deviates from the grand coalition $\overline{M} = N$ because of a mistake. This deviation will cause each remaining signatory to reduce its abatement level by $e(5) - e(4)$, which is $e^a(5) - e^a(4) = 2$ under the MTP rule, and is $e^b(5) - e^b(4) = 8$ under the coalition unanimity rule. This example illustrates why the MTP rule has a stronger uncertainty effect than the coalition unanimity rule.

The fact that the MTP rule may not be optimal under a small uncertainty seems to be counterintuitive at first glance. Once a coalition is formed, it is quite natural to require all signatories to act as one player and maximize their total payoffs. This explains why the MTP rule is so popular in the coalition formation literature. However, a shortcoming of the MTP rule is that it has a weak incentive effect and hence cannot effectively overcome the free-riding problem. Indeed, when ε is sufficiently small, the designer should require the maximization of total payoffs of coalition members for only a stable coalition, rather than for all coalitions. These redundant requirements lead to a weak incentive effect and undermine the MTP rule.

Finally, from Table 2, the MTP rule with minimal participation e^c can be regarded as a mixture of e^a and e^b. Hence, for the designer, e^c is better than e^a and e^b when ε is neither too large nor too small.

6 Concluding remarks

In this study, we introduce a three-stage coalition formation game to analyze the endogenous determination of the IEA rule under participation uncertainty. We provide an algorithm to derive an optimal rule, which reaches an appropriate balance between providing sufficient incentive for cooperation and reducing the losses caused by participation uncertainty.

We find that some commonly used rules are generally not optimal. In particular, the MTP rule has a weak incentive effect and is not optimal unless the uncertainty is very large; while the coalition unanimity rule has a weak uncertainty effect, it is optimal only when there is no participation uncertainty. Some of the failures of the IEAs in reality or in theory can be attributed to the inappropriate rules used under certain situations.

Some further works and extensions may be worth studying in future research. First, an open question is whether optimal rules are always (*ex ante*) efficient in the sense that they result in full participation and induce enough abatement level before uncertainty is realized; that is, $m(e^*) = n$ and $e^*(n) = n\alpha$. This question is important, because if the answer is positive, then we can be fairly optimistic about what IEAs may accomplish as long as their rules are properly designed. However, by now the author can neither prove the statement nor find an counterexample.

Second, we can study models with more general settings, for example, models with heterogeneous countries, or models with more general payoff function. Third, we may consider more complex IEA rules. For example, a rule may contain an emission function $e_i(\overline{M})$ specifying the abatement level of $i \in \overline{M}$ and a transfer function $t_i(\overline{M})$ characterizing the amount of money transferred to country i when coalition $\overline{M}$ is formed. Last but not least, some other goals of the designer can be studied. For instance, sometimes it makes more sense to assume that the designer will maximize expected social welfare rather than the signatories' welfare.

Finally, note that in addition to the IEA issue, the MTP rule is widely applied in some other areas involving the voluntary provision of goods with externalities, such as cartel formation in oligopoly markets, cooperation in

R&D, and sharing natural resource.[10] In a typical application, players first decide whether to join a coalition, and then all coalition members act according to the MTP rule. However, in most of these works, participation uncertainty is implicitly assumed to be zero, which implies that the MTP rule may not be an optimal rule for coalition members and the designer. Therefore, it is reasonable and necessary to re-examine the outcome of these works by endogenizing the choice of the coalition rules.

Appendix

Proof of Lemma 1.

(a) From (10), $\overline{E}(s)$ is obviously a closed set in $\mathbb{R}_+^s$ for each $s \in [1, n]$.

(b) Now, we prove that $\overline{E}(s)$ is a bounded set in $\mathbb{R}_+^s$ by induction on s. We can easily see that $\overline{E}(1) = \{\alpha\}$ is bounded in $\mathbb{R}_+^1$. Assume inductively that $\overline{E}(k)$ is bounded in $\mathbb{R}_+^k$, $1 \leq k \leq n-1$. That is, there exist $T_1, T_2, \ldots, T_k > 0$, such that for each $\big(e(1), \ldots, e(k)\big) \in \overline{E}(k)$: $e(q) < T_q$, $1 \leq q \leq k$.

Now, consider $\overline{E}(k+1)$. According to (10), for each $\big(e(1), \ldots, e(k+1)\big) \in \overline{E}(k+1)$, we have $e(q) < T_q$, $1 \leq q \leq k$. Additionally, $e(k+1)$ satisfies $u^C(k+1, e) \geq u^I(k, e)$; that is,

$$-\frac{1}{2}e(k+1)^2 + a(k+1)e(k+1) + A(k) \geqslant 0,$$

where $A(k)$ depends on $\big(e(1), \ldots, e(k)\big)$. Thus, $e(k+1)$ is also bounded, implying that $\overline{E}(k+1)$ is bounded in $\mathbb{R}_+^{k+1}$. Consequently, $\overline{E}(s)$ is bounded in $\mathbb{R}_+^s$ for each $s \in [1, n]$.

(c) It remains to be proved that $\overline{E}(s)$ is not empty. Given $s \in [1, n]$, we can construct $\big(\hat{e}(1), \ldots, \hat{e}(s)\big)$ as follows:

(n1) $\hat{e}(s) = \alpha s$.
(n2) $\hat{e}(k) = \alpha$, $1 \leq k \leq s-1$.

[10] See, for example, d'Aspremont et al. (1983), Katz (1986), Poyago-Theotoky (1995), Ray and Vohra (2001), Miller and Nkuiya (2016).

For any $m < n$ and any rule e, we have

$$\begin{aligned} & u^C(m+1,e) - u^I(m,e) \\ =&(1-\varepsilon)\sum_{k=0}^{m} b(k;m,1-\varepsilon)\left[\overline{u}^C(k+1,e) - \overline{u}^I(k,e)\right]. \end{aligned} \tag{11}$$

Note that from (n2), $\overline{u}^C(k+1,\hat{e}) - \overline{u}^I(k,\hat{e}) = 0$, $k \in [1, s-2]$; from (n1) and (n2), $\overline{u}^C(s,\hat{e}) - \overline{u}^I(s-1,\hat{e}) = \frac{1}{2}\alpha^2(s-1)^2 \geq 0$. Hence, from (11), $u^C(m+1,\hat{e}) \geq u^I(m,\hat{e})$, $m \in [0, s-1]$. Therefore, $\big(\hat{e}(1),\ldots,\hat{e}(s)\big) \in \overline{E}(s)$, implying $\overline{E}(s) \neq \emptyset$. □

Proof of Proposition 3.

From the definition of e^a, we can easily verify that

$$\overline{u}^C(\overline{m},e^a) - \overline{u}^I(\overline{m}-1,e^a) = -\frac{1}{2}\alpha^2(\overline{m}-1)(\overline{m}-3) = \begin{cases} = 0, & \text{if } \overline{m} = 1,3 \\ > 0, & \text{if } \overline{m} = 2 \\ < 0, & \text{if } 3 < \overline{m} \leq n \end{cases}.$$

Further, from (11), when ε is sufficiently large, $u^C(m+1,e^a) - u^I(m,e^a) \geq 0$ for all $m \in [0, n-1]$, implying that $m(e^a) = n$.

When ε is very large, we have $b\big(0;n,1-\varepsilon\big) \gg b\big(1;n,1-\varepsilon\big) \gg \cdots \gg b\big(n;n,1-\varepsilon\big)$. According to (9), a necessary condition for rule e^0 to be optimal is that $e^0(1)$ maximizes $\overline{u}^C(1,e)$; that is, $e^0(1) = \alpha$, since otherwise we can find e' such that $\overline{u}^C(1,e') > \overline{u}^C(1,e^0)$, and hence $u^C(1,e') > u^C(1,e^0)$, which implies that $Eu^C(e') > Eu^C(e^0)$ for a sufficiently large ε.

Now, assume that $e^0(k)$ maximizes $\overline{u}^C(k,e)$ for all $k \in [1,m]$, where $m < n$. If e^0 is optimal for a sufficiently large ε, $e^0(k+1)$ also maximizes $\overline{u}^C(k+1,e)$, since otherwise let e' be such that $e'(s) = e^0(s)$, $s \leq k$, and $\overline{u}^C(k+1,e') > \overline{u}^C(k+1,e^0)$, implying that $u^C(k+1,e') > u^C(k+1,e^0)$ and $Eu^C(e') > Eu^C(e^0)$, which contradicts the assumption that e^0 is optimal.

Thus, we have proved that if e^0 is optimal when ε is large enough, then $e^0(k)$ maximizes $\overline{u}^C(k,e)$ for all $k \in [1,n]$, which implies that $e^0 = e^a$. That is, e^a is optimal when ε is sufficiently large. □

References

Barrett, S., 1994. Self-enforcing international environmental agreements. Oxford Economic Papers 46, 878–894.

Carraro, C., Marchiori, C., Oreffice, S., 2009. Endogenous minimum participation in international environmental treaties. Environmental and Resource Economics 42, 411–425.

Carraro, C., Siniscalco, D., 1993. Strategies for the international protection of the environment. Journal of Public Economics 52, 309–328.

Cazals, A., Sauquet, A., 2015. How do elections affect international cooperation? Evidence from environmental treaty participation. Public Choice 162, 263–285.

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

d'Aspremont, C., Jacquemin, A., Gabszewicz, J.J., Weymark, J., 1983. On the stability of collusive price leadership. Canadian Journal of Economics 16, 17–25.

Dellink, R., Finus, M., Olieman, N., 2008. The stability likelihood of an international climate agreement. Environmental and Resource Economics 39, 357–377.

Finus, M., 2001. Game theory and international environmental cooperation. Edward Elgar.

Hong, F., Karp, L., 2014. International environmental agreements with endogenous or exogenous risk. Journal of the Association of Environmental and Resource Economists 1, 365–394.

Katz, M.L., 1986. An analysis of cooperative research and development. The RAND Journal of Economics 17, 527–543.

Kellenberg, D., Levinson, A., 2014. Waste of effort? International environmental agreements. Journal of the Association of Environmental and Resource Economists 1, 135–169.

Köke, S., Lange, A., 2017. Negotiating environmental agreements under ratification constraints. Journal of Environmental Economics and Management 83, 90–106.

Kolstad, C., 2007. Systematic uncertainty in self-enforcing international environmental agreements. Journal of Environmental Economics and Management 53, 68–79.

Kuiper, J., Olaizola, N., 2008. A dynamic approach to cartel formation. International Journal of Game Theory 37, 397–408.

Masoudi, N., Santugini, M., Zaccour, G., 2016. A dynamic game of emissions pollution with uncertainty and learning. Environmental and Resource Economics 64, 349–372.

Masoudi, N., Zaccour, G., 2017. Adapting to climate change: Is cooperation good for the environment? Economics Letters 153, 1–5.

Miller, S., Nkuiya, B., 2016. Coalition formation in fisheries with potential regime shift. Journal of Environmental Economics and Management 79, 189–207.

Poyago-Theotoky, J., 1995. Equilibrium and optimal size of a research joint venture in an oligopoly with spillovers. Journal of Industrial Economics 43, 209–226.

Ray, D., Vohra, R., 2001. Coalitional power and public goods. Journal of Political Economy 109, 1355–1384.

Thoron, S., 1998. Formation of a coalition-proof stable cartel. Canadian Journal of Economics 31, 63–76.