Constructing the Most Stable Climate Coalitions Based on the Lindahl Principle *

Zili Yang

Department of Economics, Binghamton University

January 2025

Abstract: Coalition stability is a crucial concern in the studies of international environmental agreements (IEA) and climate negotiations where externalities are present. In this paper, we demonstrate that the climate coalitions based on the Lindahl principle possess the best incentive properties and the fairest cost-sharing structure. Such properties lead to the most stable coalitions among all alternatives. Through exhaustive simulations in the 6-region and 12-region RICE2020 models, we show that the coalitions built with the Lindahl principle are far superior to those with the utilitarian principle on which the literature has focused. The sophisticated algorithms to implement the Lindahl principle in the RICE model pave the way for the broad application of the Lindahl equilibrium concept in environmental problems. Finally, this paper also connects coalition stability based on the Lindahl principle with climate negotiation in the real world.

Keywords: The Lindahl equilibrium, international environmental agreement (IEA), integrated assessment model (IAM), stable coalitions

JEL code: C71, D62, Q54, Q58

Author information:

Department of Economics

Binghamton University, SUNY

Phone: (607) 777-4726

Email: zlyang@binghamton.edu

* The author thanks the financial support of China's National Key R&D Program (No. 2016YFA0602603) and the Whitney and Betty MacMillan Center for International and Area Studies at Yale.

Constructing the Most Stable Climate Coalitions Based on the Lindahl Principle

1. Introduction

Climate change is a global environmental externality phenomenon. International cooperation is the most desirable approach to internalizing climate externality. If an all-inclusive global coalition is politically infeasible, various regional coalitions of greenhouse gas (GHG) mitigation are preferable to the non-cooperative Cournot-Nash equilibrium because they partially internalize climate externality and represent the Pareto improvement upon the inefficient Cournot-Nash equilibrium. Over the past three decades, international environmental agreements (IEA), i.e., the coalition theory related to environmental externalities, have received increasing attention. Climate negotiations under the auspice of the UNFCCC also prompt the IEA theory research. The investigation of IEA also bridges the gap between cooperative game theories and pressing environmental issues in the real world.

Since the early 1990s, scholars have paid increasing attention to IEA issues. The pioneer and path-breaking works include the following: Barrett (1994) examined the "self-enforcing" IEAs under the symmetric-agent assumption. Carraro and Siniscalco (1998) studied the stability

of IEAs under the symmetric-agent assumption. Chander and Tulkens (1997, 2008) examined the core properties of environmental coalitions. Barrett (2003) is a comprehensive approach to the institutional aspects of IEAs in the real world. Recently, research on climate coalitions and other IEAs has been active, and they aim at various analytical and empirical issues. For example, McGinty (2007) extends Barrett (1994) into asymmetric situations. Kolstad (2007), De Zeeuw (2008), Finus and Rundshagen (2009), and Bosetti et al. (2012) focus on some specific aspects of IEAs, respectively. Osmani and Tol (2009) examine the farsightedness of IEA. Nordhaus (2015) suggests using trade tools to establish "climate clubs" and overcome the free-riding in greenhouse gas (GHG) mitigation. Lessmann et al. (2015) is a comparative study incorporating the outcomes of coalition stabilities of several integrated assessment models (IAMs) for climate change. None of the six IAMs examined and compared have good coalition stability. Volt (2016) studies the stability of climate coalitions under certain assumptions, tests the results in a 12-region version of the RICE model, and finds that only two 4-member coalitions are internally stable. Kersting et al. (2017) examined the impacts of macro-linkage on the stability of the global climate coalition. They concluded that there is no stable climate coalition in their model.

The literature on IEAs is extensive, and while we cannot provide a comprehensive survey here, it is important to note that research on the stability of IEA based on full-scale IAMs is limited. No consensus related to empirical modeling results has been reached. However, the overall results on climate coalitions' stabilities, the most significant IEA in the real world, are often pessimistic. They suggest that holding a climate coalition together requires side payments among the members (sometimes it needs reaching out to outsiders); the coalition stability is sensitive to the parametric assumptions of the model, and the identified "stable coalitions" are small and varied in different IAMs. In this research, we challenge the above conclusions in the

literature, i.e., the necessity of side payments, parametric variations, and the small size of stable coalitions. We demonstrate that stable climate coalitions do not need side payments, are invariant to the model's parametric assumptions, and can be much larger if the coalitions are formed under the Lindahl principle to be elaborated in this research.

A coalition's "internal stability" is an appealing concept in IEA analysis. Originating from d'Aspremont et al. (1983), a coalition is internally stable if no coalition members can gain by defecting from the coalition, thus incentivizing them to stay inside. An IEA could only rely on the self-interest of sovereign countries. A country's incentive to stay or defect from it is based on internal welfare calculation. Undoubtedly, such a calculation affects the stability of an IEA. The possible incentive to hold a coalition together, particularly the large ones, is the substantive efficiency gains from internalizing environmental externalities (even partially). More importantly, all coalition members must share these gains fairly and equitably, given heterogeneous mitigation costs and environmental damage across members. What should the members do to make the coalition stable? We argue that they should share the mitigation costs according to the Lindahl principle within the coalition and abandon the universally adopted utilitarian principle in the literature.

This paper introduces an analytical approach and empirical methods to identify the most stable coalition structure in environmental externality provision in the above argument's spirit. Such stable environmental coalitions are established on the Lindahl principle and implemented numerically in the 6-region and 12-region RICE2020 models developed by Nordhaus and Yang (2021). The invariant "cost-share" ratio within a coalition is based on the Lindahl equilibrium without transfers for the global (grand) coalition. The exhaustive simulation results from the RICE2020-R12 model show that ***ALL*** coalitions up to 8 regions are internally stable (the world

consists of 12 regions here), and ***NO*** side payments are needed. The incentive to "free ride" reduces to the minimum under the Lindahl principle, and the stable global (grand) coalition is attainable with minor enforceable rules. In contrast, a set of simulations in the same model based on the utilitarian principle reiterates the pessimistic findings in the literature.

The remainder of the paper is organized as follows: Section 2 explains the Lindahl principle's analytical rationales for guiding IEAs; Section 3 introduces the RICE2020 model and the algorithmic methods for implementing the Lindahl principle in its coalition solutions; Section 4 presents the simulation results in RICE2020; Section 5 discusses the policy implications of our results in IEA analysis; and Section 6 contains concluding remarks.

2. The analytical foundation of IEA formation based on the Lindahl principle

Legislation and executive power can mandate domestic environmental policies. Thus, the "coalitions" and "agreements" for pollution control among private entities are often not applicable domestically. However, "agreements" between sovereign countries are essential to deal with global or transboundary environmental issues, and the only enforceable factor of an IEA is the "selfish" incentive of sovereign countries. The studies mentioned in the previous Section explore the limitations of such incentives for forming a climate coalition under the widely accepted assumption of the *utilitarian* principle. In problems of externality provision, the utilitarian principle is represented by the *equal-weight* social optimum of all agents. In this study, we point out that strengthening the "incentives" for forming a voluntary IEA cannot be achieved through the utilitarian principle. Instead, the most robust possible incentives are from the *Lindahl* principle based on the Lindahl equilibrium without transfers in the presence of externalities.

In essence, the Lindahl principle replaces the equal-weight social optimum (the utilitarian principle) with a specific *weighted* social optimum consisting of all agents (such weight is labeled the "Lindahl weights" in this paper) that represents the Lindahl equilibrium without transfers. Yang (2021) establishes the analytical foundation and constructive approach to find the Lindahl weights. They are the foundation for the Lindahl principle for forming stable coalitions we relied on in this study. In this section, we outline the analytical components of the Lindahl principle and its connections to the empirical part of this study.

Externality provision problems have two significant benchmarks: the inefficient Cournot-Nash equilibrium and the set of all Pareto efficient outcomes that fully internalize the externality. In the set of all efficient outcomes, the literature has focused on a single outcome based on the utilitarian principle, i.e., the equal-weight social optimum. While efficient, this outcome neither has attractive incentive properties nor is an equilibrium. Under the utilitarian principle, economic agents (regions in the IEA context) could be worse off in this social optimum than their inefficient Cournot-Nash equilibrium positions. We observe such phenomena in the original RICE model (Nordhaus and Yang, 1996) and the RICE2020 model. The simulation results from other models also concur with this observation (Lessmann et al., 2015). All empirical works surveyed in the previous section are based on the utilitarian principle. Their results on the incentives for climate coalition formation are poor, as shown in the RICE model.

"Free riding" behavior is a ubiquitous disincentive to the efficient provision of externality, and it weakens the internal stability of an IEA. The utilitarian principle cannot resolve these incentive deficiencies when agents are heterogeneous. Instead, it acerbates "free riding" problems. If a region's welfare in the equal-weight social optimum is worse off than its Cournot-Nash equilibrium payoff, defecting the grand coalition is the dominant strategy. In contrast, the

Lindahl principle in externality provision minimizes the free-riding behavior by negotiating a set of the Lindahl weights. Such Lindahl weights allow regions' mitigation cost-sharing and efficiency gains to be allocated fairly in the social optimum. The coalition structure explored in this paper is constructed on the Lindahl principle.

The "Lindahl equilibrium" concept in a public good/externality economy has been morphed through generations of economists since Lindahl (1919) and formally labeled as "Lindahl equilibrium" by Foley (1970). A clear explanation of the chronicle evolution of the "Lindahl equilibrium" concepts is given by van der Nouweland (2014). More than a dozen "Lindahl equilibrium" definitions are presented and compared in that paper. A common feature of these definitions is the "personalized price" for the public good. However, these definitions are not equivalent, and nuances can be detected. The author suggested that the "ratio equilibrium" by Kaneko (1977) is closer to Lindahl's original ideas. In sum, the literature on the Lindahl equilibrium focused on various analytical features of the Lindahl equilibrium and provided some ad hoc examples. The algorithmic identification of the Lindahl equilibrium in empirical environmental models was not explored.

In Yang (2021), the Lindahl equilibrium without transfer is defined in a generic framework of environmental externality provision. The Lindahl equilibrium is analytically identified as the solution of a weighted social optimum. Just like the Negishi weighted social optimum is the Walrasian equilibrium without transfers for an Arrow-Debreu economy, the "Lindahl weighted" social optimum is the Lindahl equilibrium with the "personalized price" and "cost-sharing" ratios succinctly identified. The results in Yang (2021) are compatible with the Lindahl equilibrium in Foley (1970), the ratio equilibrium in Kaneko (1977), and the cost-sharing equilibrium in Mas-Colell and Silvestre (1989).

The Lindahl weights reflect such a cost-sharing scheme in global cooperation on GHG mitigation in an IAM. Although the Lindahl equilibrium is an outcome of a global or grand coalition that fully internalizes externality, it can also directly guide the smaller coalitions involved with fewer agents. Therefore, one of the IEA analysis's most crucial problems is constructively identifying the Lindahl equilibrium. We outline a stylized modeling framework and its main results related to the constructive identification of the Lindahl equilibrium without transfers and the Lindahl weights below.

The social optimums of an economy with an externality (pollution) consisting of N heterogeneous agents can be expressed as follows:

$$\underset{\{x_i\}}{Max} \quad W = \sum_{i=1}^{N} \varphi_i U^i(x_i, y_i, B), \quad \varphi_i > 0, \quad \sum_{i=1}^{N} \varphi_i = N. \qquad (1)$$

$$s.t. \quad F^i(x_i, b_i) = 0, \quad i = 1, 2, \dots, N. \qquad (2)$$

$$B = \sum_{i=1}^{N} b_i \qquad (3)$$

$$G^i(x_i, y_i, \omega_i, \boldsymbol{\omega}) = 0, \quad i = 1,2, \dots, N. \quad \boldsymbol{\omega} = \sum_{i=1}^{N} \omega_i \qquad (4)$$

$$\frac{\partial U^i}{\partial x_i} > 0, \frac{\partial U^i}{\partial y_i} > 0, \quad \frac{\partial U^i}{\partial B} < 0, \quad \frac{db_i}{dx_i} > 0, \quad \frac{\partial^2 U^i}{\partial x_i^2} < 0, \frac{\partial^2 U^i}{\partial y_i^2} < 0, \quad \forall\, i. \qquad (5)$$

In System (1) through (5), W is the weighted social welfare function. U^i is an individual utility function with three arguments: x_i (a compound dirty private good), y_i (a compound clean private good), and B (externality or pollution). Equ. (2) is an agent-specific transformation function between the dirty private good (x_i) and pollution emission (b_i). Equ. (3) defines the

collective contribution to the externality (pollution) level. Equ. (4) is an implicit budget constraint between private goods and private endowment (ω_i) and global endowment (**ω**). This constrained social optimum is a model of efficient provision of externality. The solution of this model under a specific social welfare weight $\{\varphi_i\}$, maps to a point on the efficiency frontier of the model. In particular, the utilitarian solution is the model's solution when $\varphi_i \equiv 1$. Most empirical models in the literature are formulated by assuming $\varphi_i \equiv 1$.

The set of social welfare weights is a simplex defined by

$$S^N = \left\{ \{\varphi_i\} \mid \sum_{i=1}^{N} \varphi_i = N, \qquad \varphi_i \geq 0 \right\}. \tag{6}$$

The optimal solution of (1) under any point on S^N is efficient — or S^N maps to the efficient frontier of the model. The Lindahl equilibrium without transfers is an efficient solution and thus corresponds to a specific point on S^N. The task is to identify this Lindahl weight $\{\varphi_i^L\}$ on S^N.

To calculate the Lindahl weight $\{\varphi_i^L\}$, we need to define the inefficient Cournot-Nash equilibrium that corresponds to the above social optimums as the simultaneous solutions of the following system:

$$\underset{\{x_i\}}{Max} \qquad U^i(x_i, y_i, B), \qquad i = 1, 2, \ldots, N. \tag{7}$$

s.t. Equ. (2), (4), and (5)

$$B = b_i + \sum_{j \neq i}^{N} \overline{b_j} \tag{8}$$

It is easy to see that System (7) is a non-cooperative Cournot-Nash game of externality provisions or emission mitigations. Each agent decides the emissions by the level of dirty good x_i

through transformation function (2) while taking other agents' emissions as given. When N optimal solutions of (7) are compatible with one another, the outcome is the Cournot-Nash equilibrium. This Cournot-Nash equilibrium is inefficient and thus does not correspond to any weighted social optimum.

The environmental externalities are dynamic in most Global environmental problems. The RICE model described in the Appendix is a concrete example of a dynamic model in which climate change is treated as a stock externality. The analytical results for the Lindahl equilibrium related to the above stylized static model (Equ. (1) to (8)) are applicable to complicated dynamic IAMs, such as the RICE model. Here, we introduce the basic conclusions regarding the Lindahl equilibrium in the above stylized framework and implement the conclusions in the RICE2020 model detailed in the next section.

Proposition 1: The Lindahl equilibrium without transfers is the model's Nash bargaining solution, with the payoffs of the non-cooperative Cournot-Nash equilibrium as the disagreement point. Namely, the model's Lindahl equilibrium without transfers is the solution of the following problem:

$$\underset{\{x_i\}}{Max} \quad V = \sum_{i=1}^{N} \log\left(U^i - \underline{U}^i\right) \qquad (9)$$

s.t. Equ. (2) to (5).

Here, $\underline{U}^i$ is agent i's payoff in the solution of System (7) and is a constant in (9).

Proof: see Yang (2021). [1]

[1] The derivation of Proposition 1 requires extensive preparatory work that exceeds the length of a typical section of a paper. They are presented clearly in Yang (2021).

Proposition 2: The model's Lindahl equilibrium without transfers, or the solution of the Nash bargaining problem defined by (9), is also the solution of a social optimum under a specific Lindahl social welfare weight given by:

$$\varphi_i^L = \left(\frac{N}{\sum_{j=1}^{N} \left(\frac{1}{U^{j*} - \underline{U}^j} \right)} \right) \left(\frac{1}{U^{i*} - \underline{U}^i} \right), \qquad \sum_{i=1}^{N} \varphi_i^L = N. \qquad (10)$$

Here, U^{i*} is the payoff of the optimal solution in (9), or the Nash bargaining solution of the model.

Proof: see Appendix 1.

Propositions 1 and 2 establish a triangular equivalence relationship related to the Lindahl equilibrium without transfers through a constructive approach. This triangular relationship can be reiterated as follows. In a model of externality provision, the Lindahl equilibrium without transfers is the model's Nash bargaining solution with the payoffs of the non-cooperative Cournot-Nash equilibrium as the disagreement point; this Lindahl equilibrium is also the social optimum under a specific Lindahl social welfare weight (10). This Lindahl weight can be calculated from the above Nash bargaining solution.

The triangular equivalence relationship regarding the Lindahl equilibrium without transfers directly impacts the IEA analysis. Notably, the Lindahl weights in the social optimum are crucially valuable for the coalitions formed under the Lindahl principle. The Lindahl equilibrium without transfers, i.e., the social optimum under the Lindahl weight, is the grand (global) coalition with the strongest incentives. Because the Lindahl equilibrium without transfers is in the economy's core, all members of this grand coalition are better off than their

respective payoffs in the Cournot-Nash equilibrium, among other desirable properties (individual rationality, or IC property). In addition, such strong incentives are preserved in the smaller coalitions if they keep the same Lindahl weights within the coalition.

Forming a coalition of externality provisions or IEA has a unique feature that may be absent in other coalition problems. The feature is the efficiency gains when the coalition increases in size—the members of an expanding coalition of externality provision gain through added partial internalization of externality. Similarly, when a coalition's size reduces due to defection, the surviving coalition experiences efficiency losses. To maintain the internal stability of such an expanding coalition, each member must receive a fair portion of such gain and share the mitigation cost for further internalizing externality. The arrangement should make every member better off through the expansion. We should emphasize that the payoff betterment is through efficiency gains, not side payments. The incentive-compatible cost-sharing scheme that allows all members to enjoy the efficiency gains fairly must be based on the Lindahl principle within the coalition.

When studying externality coalitions, the strategic interaction between a coalition and outsiders is unavoidable. Our framework assumes that a coalition plays a non-cooperative Cournot-Nash game of externality provision with the outsiders as a single player while partially internalizing the externality within the coalition under the Lindahl principle. We call this "internal cooperation and external competition" outcome a "hybrid" Nash equilibrium.[2] As we practiced here, a comprehensive coalition analysis should exhaustively identify all "hybrid" Nash equilibriums.

[2] The "hybrid" Nash equilibrium here bears some resemblance with the "PANE" equilibriums in Chander and Tulkens (1997) but differs in some important respects.

The literature related to coalitions of environmental externality provisions, such as climate coalitions, is sketchy on what members do inside to preserve the coalition. So far, the empirical modeling always assumes that members maximize a social welfare function with an equal social welfare weight inside the coalition. This maximization is the "cooperation" vaguely referred to in the literature. Such an arrangement is precisely the *utilitarian principle*. In climate change modeling, where regions are heterogeneous in GHG mitigation costs and climate damage, the coalitions' incentive dilemma begins with this equal-weight assumption: not all regions are happy with this arrangement, and some do not want to cooperate under this condition. From 2-member minimum sub-coalitions to N-member grand coalition, equal weight cost-sharing creates unhappy campers. A necessary condition for an internally stable coalition is that all coalition members are better off than their non-cooperative Cournot-Nash equilibrium payoffs; otherwise, they would prefer the non-cooperative outcome. The utilitarian principle fails in this necessary condition almost universally.

On the other hand, the social welfare weights play vital roles in mitigation cost-sharing in a weighted social optimum of environmental externality provision. We take "equal weights" as a benchmark (all weights equal to 1 and add up to N) in an N-agent economy for the argument. Because the social welfare weights sum up to N, if some agents' weights are larger than 1, some agents' weights must be less than 1. An agent with a relatively higher weight (> 1) bears lower mitigation responsibility and benefits more from others' mitigation efforts in this social optimum than in the equal-weight benchmark. The effects of lower weight (< 1) are the opposite.

The Lindahl weights enable all members of a sub-coalition to be better off than the Cournot-Nash equilibrium and worse off than the Lindahl equilibrium (the grand coalition). The *Lindahl principle* recognizes the social welfare weights' proxy role in mitigation cost-sharing

and allocates the Lindahl weights through bargaining within a coalition of any size. The Lindahl principle offers the strongest incentive to hold the coalition together by allocating the efficiency gain fairly. More specifically, the formation of an externality provision coalition should be based on the following steps: the grand (global) coalition with N agents, i.e., the social optimum under the Lindahl weights, is the guidance for any M-member sub-coalitions ($1 < M < N$). In such sub-coalitions, members solve an M-agent social optimum under the social welfare weights proportional to their respective Lindahl weights in the grand coalition. This coalition, as a "single" player, plays a non-cooperative Cournot-Nash game with (N-M) outsiders (called "singletons"). The internal cooperation under the Lindahl weights and external Cournot-Nash game are "played" out simultaneously. The solution is the "hybrid" Nash equilibrium. There are 2^N "hybrid" Nash equilibriums. The exhaustive simulation results from the RICE2020 model presented in the next Section will demonstrate such maximal internal stability of the Lindahl principle for all possible "hybrid" Nash equilibriums.

Lastly, this study uses the term "the most stable coalition structure." We want to give a precise interpretation of the notion here. In an N-agent economy, the most stable coalition structure is the mechanism or structural design that makes the largest number of sub-coalitions out of 2^N internally stable. Because of finite combinatory possibilities, such a structure is identifiable. The Lindahl principle provides such a mechanism in theory; the simulations from empirical models, such as RICE, demonstrate it. Thus, the coalition structure under the Lindahl principle is the most stable one.

3. The RICE2020-R12 model and its coalition solutions

The RICE model is an influential integrated assessment model (IAM) for climate change. In the RICE model, regions emit GHGs in their economic activities and suffer damage from

climate change. The regions consider the GHG mitigation costs and the benefits of mitigated climate impacts to decide their optimal mitigation efforts. Meanwhile, the effects of other regions' emissions and mitigation efforts (the external effects) also play a crucial role in a region's optimal decision. Thus, the RICE model captures the externality characteristics of climate change.

Since its original appearance (Nordhaus and Yang, 1996), the RICE model has experienced several updating rounds. The version used in this study is the RICE2020 under the development of Nordhaus and Yang (2021). The mathematical description of the base model is in the Appendix. The model has two basic settings (corresponding to two benchmarks of externality provision mentioned in the previous section): one is the weighted social optimum that fully internalizes the climate externality. It is a dynamic optimal control problem (Appendix 2. Part I). Another is the non-cooperative Cournot-Nash game of GHG emissions (Appendix 2. part II). It is a differential game, and its solution is the Cournot-Nash equilibrium. In the model description, variable $\Sigma E(n,t)$ in (A-12) and (A-12'), or aggregate GHG emissions by the regions, captures the external effect, and $\mu(n,t)$ is the regional GHG control rate --- the critical control variable of the model. Finally, the heterogenous damages to the regions caused by temperature increase $T_{AT}(t)$, and the temperature increases are the impacts of climate externality.

The RICE2020 model has three versions with identical model structures but different regional disaggregation with N = 6, 12, and 16. To thoroughly explore the ramifications of climate coalitions and recognize the "curse" of dimensionality, the study here adopts N = 12 and a time horizon T = 50 (in a 5-year step). The model is labeled RICE2020-R12, as in the section head. The algorithm designs are from Yang (2008, 2021).

In RICE2020-R12, we solved the whole gamut of $2^{12} = 4096$ "hybrid" Nash equilibrium solutions. The calculations are carried out through several computationally intensive procedures in the GAMS language. The calculation steps and their explanations are outlined as follows:[3]

(i) Solve the inefficient non-cooperative Cournot-Nash equilibrium of the RICE model (as in Appendix Part II) and preserve the payoff results.

(ii) Solve the Nash bargaining solution of the RICE model by using the regional payoffs of the Cournot-Nash equilibrium retrieved from (i) as the disagreement point.

(iii) Using the formula given by (10), calculate the Lindahl social welfare weights based on the results from (ii) and solve the social optimum of the RICE model (as in Appendix Part I) under the Lindahl weights (the solutions of (ii) and (iii) are identical as predicted by Propositions 1 and 2).

(iv) Controlled by a looping structure, solve all "hybrid" Nash equilibriums of the RICE model, $2^{12} = 4096$ in total, one by one. The same Lindahl weights identified in (iii) are applied within all proper sub-coalitions (as in Appendix Part III) to identify all coalition solutions under the Lindahl principle. For comparison, the whole spectrum of the "hybrid" Nash equilibriums is also solved under the utilitarian principle, i.e., equal-weights sub-coalitions.

The Nash bargaining solution of the RICE2020 in step (ii) is the solution of the RICE model by replacing objective function (A-1) in Appendix Part I with a new objective function analog to (9). The disagreement point is the regional payoff of the non-cooperative Cournot-Nash equilibrium obtained by solving Appendix Part II as in step (i).

[3] The algorithms coded in the GAMS language are available upon requests.

The Lindahl social welfare weights used in step (iii) are calculated according to proposition 2, with $\{\varphi_i^L\} = \{Nr_i\}$. The value of $\{\varphi_i^L\}$ is an *ex-post* off-line calculation before implementation in step (iv).

Step (iv) solves all combinations of the hybrid Nash equilibriums in which an M-member sub-coalition plays a non-cooperative Cournot-Nash game with (N-M) outsiders. M runs from 0 (the non-cooperative Cournot-Nash game) to N (full cooperation under the Lindahl weights and N = 12 in this paper). The total outcomes are $2^{12} = 4096$. When M = 1 (12 outcomes), the hybrid Nash equilibriums are identical to the Cournot-Nash equilibrium (M = 0). Therefore, there are 4096 - 1-1 -12 = 4082 proper hybrid Nash equilibriums. A coalition M (1 < M < 12) plays the Cournot-Nash game with other (12 – M) outsiders. Simultaneously, the members inside the coalition solve a social optimum under the Lindahl weights given by:

$$\varphi_i^M = \frac{M\varphi_i^L}{\sum_{m\subset M}\varphi_m^L}, \quad \forall i \subset M \qquad (11)$$

The calculations of 4096 outcomes are computationally demanding. We use a 2^{12} x 12 design matrix (1 = in and 0 = out) to control the sequential loops of the calculations. By decomposing the task into four parallel programs, each containing 1024 coalition combinations, we obtain all solutions in step (iv) in around 10 hours on a multi-core PC workstation.

Finally, to compare our approach with the results in the literature, i.e., the utilitarian outcome, we also execute the computational-demanding step (iv) with equal social welfare weights inside the coalitions. After that, we obtained the complete coalitional solutions of RICE2020-R12 based on Lindahl and utilitarian principles.

4. Main results on the stability of coalitions in RICE2020-R12

The critical factor affecting the stability of a coalition is the regional payoff in or out of it. From a game-theoretic perspective, the RICE2020 model has a non-linear objective function or a non-transferable utility (NTU). Therefore, cross-regional comparison of payoffs is not meaningful. Instead, we examine the individual regional payoff changes from the benchmarks in different coalition outcomes. Furthermore, the magnitudes of these payoff changes are less critical because of the NTU setting, and the changes in directions (or signs) matter. If the payoff change is negative, the region is worse off, and vice versa. Therefore, the reporting results are 2^{12} = 4096 (coalition possibilities) x 12 (regions) matrices of regional payoffs and their changes from the benchmarks.

Presenting multiple 4096 x 12 tables in paper-size research is impossible, yet the information in these tables is crucial to our analysis. To offer an effective alternative for communication, we conducted the same set of calculations as steps (i) to (iv) in RICE2020-R06 (the RICE2020 model with six regions: USA, JPN, EEC, CHN, FSU, and ROW) and presented the same tables in RICE2020-R06 in this paper (Tables 2 to 7) for illustration purposes. These tables are 2^6 x 6 = 64 x 6 by size and fit the paper size presentation. The rich simulation results from RICE2020-R12 are in a "data.zip" file available upon request. Each of Tables 2 to 7 in this paper has a 4096 x 12 size counterpart in the data file. Our economic analyses are based on the information from 4096 x 12 tables generated in RICE2020-R12; our explanation of the table structure and analyses of coalition stability are based on Tables 2 to 7 in the size of 64 x 6 from RICE2020-R06 illustratively.

As a part of reporting, Table 1 contains the Lindahl weights of the RICE2020-R12 model obtained in step (iii) and is utilized in step (iv). The accuracy of the Lindahl weights is preserved

to 10^{-12} decimal points. Table 2 is a 64 x 6 design matrix. Entry "1" indicates that the region is in the coalition; entry "0" means that the region is outside the coalition. Row #1 is "all-in" or the grand coalition; row #64 is "all-out" or the non-cooperative Cournot-Nash equilibrium; each row of the design matrix represents a specific hybrid Nash equilibrium. The counterpart of Table 2 in RICE2020-R12 (a 4096 x 12 design matrix) controls the execution sequence in step (iv). All results related to the regional payoffs presented in other tables are in the same order as the design matrix.

Table 3 is a 64 x 6 matrix containing 6 regions' payoffs in 64 hybrid Nash equilibriums under the Lindahl principle from step (iv). The original results in RICE2020-R12 are preserved at 10^{-10} accuracy (all regional payoff levels are scaled around 10^{0} range in the calculations). Here in Table 3, we present 6 decimal digits for clarity. Table 4 is derived from Table 3. It is a 64 x 6 matrix that shows the regional payoff differences between the grand coalition in row #1 and the hybrid Nash equilibrium in that row of Table 3. Table 4's entries in row # are 10^{3} x ($W^{\#}(n) - W^{1}(n)$). Here, superscript "1" indicates the payoff of row #1. The values of entries are multiplied by 10^{3} to magnify the significant digits (preserved at 10^{-4} for easy reading). If the difference is negative, the cell entry is in black; if the difference is positive, it is colored either in red or orange. Table 5, also derived from Table 3, shows the regional payoff differences between the non-cooperative Cournot-Nash equilibrium in row #64 and the hybrid Nash equilibrium in that row of Table 3. Table 5's entries in row # are 10^{3} x ($W^{\#}(n) - W^{64}(n)$). Here, superscript "64" indicates the payoff of row #64. We also present two tables (Tables 6 and 7), corresponding to Tables 4 and 5, for the hybrid Nash equilibrium outcomes under the utilitarian principle. Tables 6 and 7 are obtained by executing steps (i) (irrelevant to the weight), step (iii) under the equal weight, and step (iv) under the equal weight. Because our focus is the Lindahl outcomes, the

results from the utilitarian principle are simplified to the "signs." Tables 6 and 7 show the signs of the payoff differences between a hybrid Nash equilibrium and the utilitarian social optimum (Table 6) or the Cournot-Nash equilibrium (Table 7).

Tables 4 and 5 show superior stability properties based on the Lindahl principle. Table 4 shows only a few instances where a region is better off than the Lindahl grand coalition. In six "5-in and 1-out" cases, the sole outsider gains (colored in red). This is the typical free riding. If the other five regions stay in the coalition based on the Lindahl principle, the outsider plays a two-player Cournot-Nash game with the coalition and gains through the coalition's internalization of climate externality. However, in fifteen "4-in and 2-out" cases, the two outsiders cannot gain simultaneously (colored in orange and only one orange color in the same row). In two cases, both outsiders lose (rows #6 and #19). *All* "2-in and 4-out" and "3-in and 3-out" cases do not have any outsiders gaining from the grand coalition. In addition, all coalition members gain monotonically as the coalition size increases (from M = 2 to M =6); equivalently, all coalition members lose monotonically as the coalition size decreases (from M = 6 to M =2).

In all twenty "3-in and 3-out" cases, all 3 outsiders lose from the Lindahl grand coalition, and the 3-member coalition can expand voluntarily. For example, {USA, JPN, EEC} is a 3-member coalition with three outsiders, CHN, FSU, and ROW, in Row #8. All three outsiders' payoffs are lower than in the grand coalition; three coalition members' payoffs are also lower due to efficiency loss. However, CHN and ROW would find their payoff improving by (re)joining a "5-in and 1-out" sub-coalition in Row #3. Their payoff losses reduce from -0.0537 to -0.0151 and -0.2334 to -0.0597, respectively. For all "3-in and 3-out" hybrid Nash equilibriums, at least one outsider wants to return to a coalition with more members (we checked

them individually). Such patterns in "3-in and 3-out" hybrid equilibriums guarantee the internal stability of all "4-in and 2-out" coalitions.

In Table 5, *all* the differences are positive.[4] Particularly, all regions in the grand coalition based on the Lindahl principle (row #1) are strictly better off than the Cournot-Nash equilibrium. The pattern indicates that regions are better off in all hybrid Nash equilibriums (in or out of the coalition) than non-cooperative Cournot-Nash equilibrium. Regions inside a coalition gain monotonically as the coalition's size increases (checked individually). The uniformly positive signs and monotonic payoff improvements inside coalitions are also necessary conditions for an internally stable coalition. They are satisfied under the Lindahl principle.

In contrast, the stability property of the utilitarian solutions is extremely weak. No identifiable sign pattern can be found in Tables 6 and 7, i.e., the outcome from the utilitarian principle. Tables 6 and 7 echo the amorphous and pessimistic results on the stabilities of climate coalitions in the literature. Some members of many coalitions, even in the small size ones, want to defect. For example, the 2-member coalition's internal stability is violated in {CHN, ROW}(Row # 59) and {FSU, ROW}(Row #61) because CHN and FSU are worse off in the coalition than in the Cournot-Nash equilibrium. They prefer the non-cooperative Cournot-Nash equilibrium to a 2-member coalition with ROW. Stable coalitions are hard to find in RICE2020-R06 under the utilitarian principle.

With the visual help of miniature Tables 2 to 7 from RICE2020-R06, we highlight some significant findings in the RICE2020-R12 model related to the coalition stabilities under the

[4] Rows #32, 48, 56, 60, 62, and 63 are six "1-in and 5-out" cases. They are identical to the Cournot-Nash equilibrium conceptually (0-in and 6-out). Numerically, the difference entries are zero. They are kept in the execution loops for programming simplicity and for testing the algorithm's robustness.

Lindahl and utilitarian principles. The results are based on the data in 4096 x 12 size counterparts of Tables 3 to 7. They could be summarized in the following six points:

(1) Under the Lindahl principle, *all* regions are better off, inside or outside a coalition, than their respective Cournot-Nash equilibrium outcome. In particular, *all* regions are better off in the grand coalition or the Lindahl equilibrium (the first row of the 4096 x 12 table corresponding to Table 3) than in the non-cooperative Cournot-Nash equilibrium (the last row of the 4096 x 12 table). The results are in a 4096 x 12 counterpart of Table 5. In that table, *all* entries are positive except twelve rows (# 2048, 3072, 3584, 3840, 3968, 4032, 4064, 4080, 4088,4092, 4094, and 4095). These rows are twelve "1-in and 11-out" scenarios. These 12 scenarios are equivalent to the non-cooperative Cournot-Nash equilibrium, and the entries in these 12 rows are numerically zeros. This 4096 x 12 table shows that all sub-coalitions meet the necessary conditions for internal stability.

(2) Under the Lindahl principle, all regions are worse off in a proper sub-coalition ($1<M<12$) than their respective payoffs in the grand coalition due to efficiency loss. *All* entries inside sub-coalitions are negative in the 4096 x 12 counterpart of Table 4. In addition, all members' payoffs monotonically increase as the coalition's size expands (from $M = 2$ to $M = 12 = N$) due to the efficiency gain. Equivalently, all members' payoffs monotonically decrease as the coalition's size shrinks (from $M = 12 = N$ to $M = 2$) due to the efficiency loss. The results indicate that *no* proper sub-coalition can block the grand coalition based on the Lindahl principle. The result shows that the Lindahl equilibrium without transfers (the grand coalition based on the Lindahl principle) is in the core of the RICE2020-R12 model.[5]

[5] In the analytical framework here, a singleton is not a "proper" sub-coalition.

(3) In the 4096 x 12 counterpart of Table 4, i.e., the payoff differences between hybrid equilibrium and the grand coalition based on the Lindahl principle, no row has more than three positive entries. Only six rows have three positive entries or gains from the Lindahl grand coalition (corresponding to the colored entries in Table 4). Most rows are uniformly negative. The occurrences of positive entries are sparse in the table. Besides, all positive entries are outsiders. The pattern implies that no more than three outsiders can gain simultaneously by leaving the grand coalition under the Lindahl principle.

(4) We find that when there is only one outsider ("11-in and 1-out", 12 occurrences), these sole outsiders gain (corresponding to the red entries in Table 4). For all "10-in and 2-out" cases (66 occurrences), both outsiders gain in 42 cases, and one gains and one loses in the remaining 24 cases. As for "9-in and 3-out" cases (220 occurrences), all three gain in only 6 cases, and at least one or more lose in all remaining 214 cases. When there are four outsiders ("8-in and 4-out"), at most one gains, and all others lose. In the dominant majority of "7-in and 5-out" cases, all outsiders lose with a few exceptions of one gaining. All outsiders lose in all "6-in and 6-out" and smaller coalition cases. We enlist all sweeping gainers of one, two, and three outsiders in Table 8.

(5) Under the Lindahl principle, the large GHG emitters, USA, CHN, and ROW, have higher "positive" frequencies than others in the 4096 x 12 counterpart of Table 4. They are more likely to gain as outsiders when there are not too many other "peer" outsiders. Thus, they are more motivated to stay outside a coalition as "singletons." The "mass" may influence the incentives to remain in a coalition. This observation seems consistent with what we saw in climate negotiation in the real world.

(6) Similar to RICE2020-R06, the coalition stability in RICE2020-R12 is weak and unpredictable under the utilitarian principle. No identifiable sign patterns can be found in the payoff differences in the 4096 x 12 counterparts of Tables 6 and 7. Substantial portions of regional payoffs are better off than the grand coalition, either in or outside a specific sub-coalition. In contrast, such outcomes do not happen under the Lindahl principle. Unlike uniformly positive signs in the 4096 x 12 counterpart of Table 5, many negative entries are in the 4096 x 12 counterpart of Table 7. Substantial portions of regional payoffs are worse off than the Cournot-Nash equilibrium payoff. Notably, six regions (USA, EUS, JPN, OHI, RUS, and CHN) are worse off in the grand coalition than their Cournot-Nash equilibrium payoffs. These six regions are major stakeholders of climate change. Such patterns indicate that the necessary condition for an internally stable coalition is violated in most proper sub-coalitions. In fact, small-sized stable coalitions are difficult to find in RICE2020-R12 under the utilitarian principle.

Lastly, the observations summarized in items (1) to (6) are consistent with RICE2020-R06, as shown in Tables 3 to 7. Using the logic of the "3-in and 3-out" example discussed above to "6-in and 6-out" or "7-in and 5-out" cases in RICE2020-R12, we can establish that *all* 7-member and most 8-member coalitions are internally stable under the Lindahl principle. To sketch the coalition stability in the RICE2020-R12 model, we can identify up to 3575 ($=C_{12}^{8}$ -13) internally stable coalitions out of 4096 hybrid Nash equilibriums under the Lindahl principle. In contrast, we cannot characterize any predictable patterns of coalition stability under the utilitarian principle.

Under the Lindahl principle, we expect all 2- to 9-member (maybe many 10-member) coalitions to be internally stable in RICE2020-R16 (not examined in this study). The outcome from the utilitarian principle would show more amorphous coalition patterns in RICE2020-R16.

The total number of agents is a factor of incentives in externality problems. The more agents there are, the stronger the incentive to free ride. The full range of hybrid Nash equilibriums in the RICE2020-R16 model consists of $2^{16} = 65536$ scenarios. The exhaustive simulations in RICE2020-R16 are demanding computationally but doable.

5. The policy and methodology implications

The exhaustive simulations of RICE2020-R12's coalition solutions reveal several vital implications for applying the IEA theory to climate change policies. These results based on the Lindahl principle fundamentally expand our understanding of the stability of IEA, which has been primarily confined to the utilitarian principle in the literature. Our study here establishes the superiority and feasibility of the Lindahl principle in forming stable climate coalitions.

The results from the RICE2020-R12 model in step (iv) demonstrate the superior internal stability of coalitions under the Lindahl principle. *All* 2- to 7-member and *most* (if not all) 8-member sub-coalitions are internally stable. Such superiority is entirely built upon the fair and incentive-compatible cost-sharing inside the coalition designated by the Lindahl weights (21). The grand coalition is the Lindahl equilibrium without transfers. The Lindahl principle's excellent incentive properties are based on the consensus of all members. This consensus is that everyone's willingness to contribute is compatible with and conditional upon others' willingness to contribute. Individual and collective GHG mitigation targets must be moderate to achieve such consensus. Such moderation is "the greatest common divisor" to preserve the internal stability of a coalition and the essence of the Lindahl equilibrium.

The Lindahl principle does not eliminate the possibility of free-riding as it is ubiquitous in the presence of externality. However, the chance of free-riding gains is few and far between

when coalitions are built under the Lindahl principle. If agents (countries in the IEA setting) do not conceive themselves as the "only smart guy in the room," they would stay in the grand coalition under the Lindahl principle. Multiple "smart" agents defect simultaneously (more than 3 out of 12 in our simulations) would guarantee "lose-lose" outcomes for everyone because the efficiency loss from higher emissions to the whole economy outweighs the individual gain from free-riding. Thinking about others' reactions to one's action is the "farsightedness" assumption in the coalition theory (see Ray, 2007; Osmani and Tol, 2009). Our simulations in RICE2020-R12 show that the agents' farsightedness plus the Lindahl principle in cost-sharing can ensure the internal stability of large coalitions up to the grand coalition. The dissolution of the grand coalition built on the Lindahl principle into the non-cooperative Cournot-Nash equilibrium makes everyone worse off. Under the Lindahl principle, the grand coalition's superiority and stability do not need transfers across agents (regions). The incentive-compatible mitigation cost-sharing under the Lindahl principle guarantees that this efficient outcome is acceptable to everyone.

In contrast, the coalitions built on the utilitarian principle portray a pessimistic view of coalitions' stability. Based on the utilitarian principle, 6 out of 12 regions are worse off in the grand coalition than in the Cournot-Nash equilibrium. Their payoffs do not improve from the Cournot-Nash equilibrium in most sub-coalitions for several regions. The signs of the entries are mixed in the 4096 x 12 counterpart of Table 7. If the grand coalition's maximum efficiency gains could not make a region better off than its Cournot-Nash equilibrium, it likely does not gain in a smaller coalition under the same cost-sharing arrangement. Furthermore, the stable coalitions are small and sparse. Identifying them does not have any policy relevance. We would also like to point out that the results under the utilitarian principle here echo the empirical results on the

stability of IEAs in the literature surveyed in Section 1. The pattern of stable IEAs across models under the utilitarian principle is ambiguous and amorphous.

The utilitarian approach is a modeling convenience and seems the only choice in complicated IAMs for climate change. However, the utilitarian approach is misleading. A weighted social optimum solution reflects a specific mitigation cost-sharing outcome in empirical studies of externality provisions. Individual mitigation burden varies with the different weights. The heterogeneity across agents (regions) regarding mitigation costs and climate damage creates various incentives. The incentives to cooperation under the utilitarian principle or equal weights are generally weak and mixed ---such weak incentives of the utilitarian principle call for remedies. Despite its political difficulties in the international setting, unilateral transfers or side payments are among the prescriptions to hold IEAs under the utilitarian principle. The simulation results in RICE2020-R12 show that we do not need unilateral transfers to form a stable coalition under the Lindahl principle.

Furthermore, this paper outlines that the methodological approach to finding the most stable coalitions is invariant across models of externality provisions under the Lindahl principle. The Lindahl equilibrium without transfers is a unique weighted social optimum in well-behaved externality provision models. IAMs belong to this category. Once the Lindahl equilibrium is identified, the coalitions based on the Lindahl principle can always be found with the algorithms explained in the previous section (steps (i) to (iv)). In contrast, the incentive outcomes under the utilitarian principle are unreconcilably different from model to model. The puzzles, such as “why a bloc of countries can form a stable coalition in model A but becomes unstable in model B,” cannot be resolved under the widely accepted utilitarian principle, and comparative studies across models are often futile.

In modeling climate change and other environmental problems, any weighting scheme other than equal weights requires justification and a constructive way to identify it. A set of arbitrary social welfare weights is not necessarily better than equal weights. The good incentive property is only held within the close neighborhood of the Lindahl weights (Table 1) due to the smallness of the core (Starrett, 1973). Moving away from the Lindahl weights, the incentives of some agents begin to deteriorate, and the coalition stability under those weights may no longer hold. Therefore, the Lindahl weights are the only meaningful alternative to the equal weights.

As implemented in RICE2020-R12, steps (i) to (iv) are closely connected, and any missing link would make identifying the Lindahl weights impossible. Many environmental externalities or IAM models do not solve the non-cooperative Cournot-Nash equilibrium (step (i)). The Lindahl equilibrium without transfers, as in steps (ii) and (iii), is impossible without knowing the Cournot-Nash equilibrium. Although the Lindahl equilibrium is a well-established public and environmental economics concept, we would attribute its absence in empirical research, such as IEA analysis, to the lack of verifiable and robust algorithms to identify it. Yang (2021) and this paper show that the algorithm works in the popular RICE model despite its complexity. Consequently, it should be implementable in other IAMs and models of environmental externality provisions.

Finally, the coalition analysis in the RICE2020-R12 model has implications for our understanding of international negotiations of climate treaties. The "nationally determined contributions (NDCs)" policy in the Paris Agreement (UNFCCC, 2016) is close (but not necessarily equivalent) to the non-cooperative Cournot-Nash equilibrium in the framework here. Acting independently, the outcome does not internalize much climate externalities, and global GHG emissions are excessive compared to all efficient results. Any stringent mitigation target is

probably unachievable under NDCs. The reason is simple: NDCs miss an essential element of the efficient provision of climate externality---the global cooperation and the mechanism to enforce it. Implementing the Lindahl principle requires the explicit commitment of bargained (or negotiated) mitigation cost-sharing and individual/aggregate mitigation targets through cooperation. The lengthy international negotiations of GHG mitigation under the auspice of the UNFCCC have proven that full cooperation would be challenging when such a principle is not followed.

Despite the difficulties in the real world, the Lindahl principle we explored through RICE2020-R12 offers a promising direction for climate negotiation. As discussed in detail in Yang (2021), the Lindahl equilibrium is a more "conservative" or "moderate" choice of mitigation efforts on the efficiency frontier, both in aggregate externality level and the institutional requirements (no transfers or side payments needed). The reason is that such full cooperation must be agreed upon and welcomed by all regions. Still, full cooperation, even partial coalitions under the Lindahl principle, are more "proactive" than NDCs. The cost-sharing specified by the Lindahl principle should be on the negotiation table for regional cooperation related to GHG mitigation. At least no region is worse off in a proper sub-coalition than its Cournot-Nash equilibrium position.

6. Concluding remarks

In this paper, we argue for the superiority of the Lindahl principle in forming stable climate coalitions through elaborative calculations in the RICE2020-R12 model. The simulation results show that if a coalition is built with the Lindahl principle, i.e., sharing the mitigation costs under the Lindahl social welfare weights by each agent (region), the size of an internally stable coalition can be large. In RICE2020-R12, *all* coalitions with up to 8 members are internally

stable. The coalition's stability is hopeful for the grand coalition (the model's Lindahl equilibrium). While the payoff gains from free-riding may occur in the 19- to 11-member coalitions, the incentive to hold the grand coalition together is the strongest, and the enticement to defect is the weakest under the Lindahl principle. To support a climate coalition's internal stability, the superiority of the Lindahl principle over the utilitarian principle is evident.

Due to lacking calculation algorithms, the Lindahl equilibrium concept has rarely been applied in empirically modeling environmental issues. In this paper, we implement it in the RICE model, a well-known climate change model, through a sequence of meticulously designed algorithms. The approach here draws a clear distinction from the conventional methods that take the utilitarian principle for granted. Comparisons here show that the utilitarian results have very weak incentives from the IEA perspective, albeit its calculation convenience. We would promote using the Lindahl equilibrium without transfers as the benchmark solution for externality problems, such as the stability analysis of IEA.

The results in this paper also have significant implications for the IEAs in the real world, like climate negotiations. The incentive to cooperate or form a GHG mitigation coalition is crucial in (partially) internalizing the climate externality. The fair cost-sharing scheme embodied in the Lindahl principle is the most effective way to boost the incentive, enhance the stability of coalition, and achieve full cooperation in climate negotiation. Finally, we would argue that the methodologies explored here are helpful in other coalitional issues. The social optimum under the Lindahl weights is the most crucial solution concept when an externality exists.

Table 1 The Lindahl social welfare weights of regions

Region	Lindahl Weight
USA	1.469395011608
EUS	1.214492951320
JPN	2.242938536240
OHI	1.568705161742
RUS	2.158136927960
EEC	1.401363859359
CHN	0.614877933229
IND	0.158812137738
MDE	0.578223694828
SSA	0.140973680291
LAA	0.285980764836
ROW	0.166099340850
Total	12

Definition of regional acronyms:

USA: — the United States of America; EUS: — the European Union countries; JPN: — Japan; OHI: — other high-income countries; RUS: — Russia; EEC: — Eastern European countries; CHN: — China; IND: — India; MDE: — Middle East countries; SSA: — Sub-Sahara African countries; LAA: — Latin American countries; ROW: — the rest of the world.

Table 2 Design matrix of the RICE2020-R06 model

	USA	JPN	EEC	CHN	FSU	ROW
1	1	1	1	1	1	1
2	1	1	1	1	1	0
3	1	1	1	1	0	1
4	1	1	1	1	0	0
5	1	1	1	0	1	1
6	1	1	1	0	1	0
7	1	1	1	0	0	1
8	1	1	1	0	0	0
9	1	1	0	1	1	1
10	1	1	0	1	1	0
11	1	1	0	1	0	1
12	1	1	0	1	0	0
13	1	1	0	0	1	1
14	1	1	0	0	1	0
15	1	1	0	0	0	1
16	1	1	0	0	0	0
17	1	0	1	1	1	1
18	1	0	1	1	1	0
19	1	0	1	1	0	1
20	1	0	1	1	0	0
21	1	0	1	0	1	1
22	1	0	1	0	1	0
23	1	0	1	0	0	1
24	1	0	1	0	0	0
25	1	0	0	1	1	1
26	1	0	0	1	1	0
27	1	0	0	1	0	1
28	1	0	0	1	0	0
29	1	0	0	0	1	1
30	1	0	0	0	1	0
31	1	0	0	0	0	1
32	1	0	0	0	0	0
33	0	1	1	1	1	1
34	0	1	1	1	1	0
35	0	1	1	1	0	1
36	0	1	1	1	0	0
37	0	1	1	0	1	1
38	0	1	1	0	1	0
39	0	1	1	0	0	1
40	0	1	1	0	0	0

41	0	1	0	1	1	1
42	0	1	0	1	1	0
43	0	1	0	1	0	1
44	0	1	0	1	0	0
45	0	1	0	0	1	1
46	0	1	0	0	1	0
47	0	1	0	0	0	1
48	0	1	0	0	0	0
49	0	0	1	1	1	1
50	0	0	1	1	1	0
51	0	0	1	1	0	1
52	0	0	1	1	0	0
53	0	0	1	0	1	1
54	0	0	1	0	1	0
55	0	0	1	0	0	1
56	0	0	1	0	0	0
57	0	0	0	1	1	1
58	0	0	0	1	1	0
59	0	0	0	1	0	1
60	0	0	0	1	0	0
61	0	0	0	0	1	1
62	0	0	0	0	1	0
63	0	0	0	0	0	1
64	0	0	0	0	0	0

Note: 1 = in the coalition; 0 = outside the coalition. The design matrix of RICE2020-R12 is a 2^{12} x 12 = 4096 x 12 matrix that represents the “in” and “out” cases.

Table 3 The payoff level under the Lindahl principle (RICE2020-R06)

	USA	JPN	EEC	CHN	FSU	ROW
1	1.112712	1.114568	0.925852	2.991788	0.771536	11.443000
2	1.112690	1.114545	0.925831	2.991732	0.771511	11.443341
3	1.112705	1.114561	0.925845	2.991772	0.771546	11.442940
4	1.112686	1.114541	0.925826	2.991726	0.771516	11.443124
5	1.112694	1.114549	0.925835	2.991845	0.771516	11.442760
6	1.112682	1.114536	0.925821	2.991753	0.771501	11.442899
7	1.112689	1.114544	0.925829	2.991808	0.771521	11.442725
8	1.112680	1.114533	0.925819	2.991734	0.771502	11.442766
9	1.112707	1.114563	0.925857	2.991780	0.771530	11.442984
10	1.112688	1.114542	0.925832	2.991730	0.771508	11.443181
11	1.112701	1.114556	0.925848	2.991764	0.771538	11.442918
12	1.112684	1.114538	0.925827	2.991724	0.771510	11.442977
13	1.112691	1.114546	0.925837	2.991817	0.771512	11.442756
14	1.112681	1.114534	0.925822	2.991738	0.771499	11.442798
15	1.112686	1.114540	0.925830	2.991781	0.771514	11.442715
16	1.112680	1.114532	0.925818	2.991722	0.771499	11.442685
17	1.112705	1.114578	0.925844	2.991771	0.771527	11.442934
18	1.112686	1.114549	0.925826	2.991725	0.771506	11.443116
19	1.112698	1.114568	0.925838	2.991756	0.771534	11.442872
20	1.112683	1.114543	0.925822	2.991719	0.771508	11.442922
21	1.112689	1.114555	0.925829	2.991806	0.771510	11.442720
22	1.112680	1.114537	0.925819	2.991733	0.771499	11.442761
23	1.112684	1.114546	0.925824	2.991772	0.771512	11.442684
24	1.112679	1.114533	0.925817	2.991718	0.771499	11.442661
25	1.112701	1.114570	0.925847	2.991763	0.771522	11.442913
26	1.112684	1.114544	0.925826	2.991723	0.771503	11.442970
27	1.112694	1.114560	0.925839	2.991747	0.771527	11.442845
28	1.112681	1.114538	0.925821	2.991717	0.771503	11.442791
29	1.112686	1.114548	0.925830	2.991780	0.771506	11.442711
30	1.112680	1.114534	0.925818	2.991721	0.771497	11.442681
31	1.112681	1.114540	0.925823	2.991747	0.771506	11.442668
32	1.112679	1.114531	0.925816	2.991710	0.771496	11.442604
33	1.112726	1.114558	0.925842	2.991763	0.771525	11.442890
34	1.112696	1.114539	0.925824	2.991721	0.771505	11.443058
35	1.112715	1.114551	0.925836	2.991749	0.771531	11.442833
36	1.112689	1.114535	0.925821	2.991716	0.771506	11.442874
37	1.112701	1.114542	0.925827	2.991796	0.771508	11.442692
38	1.112684	1.114533	0.925818	2.991729	0.771498	11.442733
39	1.112693	1.114537	0.925822	2.991763	0.771510	11.442661
40	1.112681	1.114532	0.925817	2.991716	0.771498	11.442648

41	1.112717	1.114553	0.925845	2.991755	0.771520	11.442872
42	1.112691	1.114536	0.925825	2.991719	0.771502	11.442920
43	1.112707	1.114546	0.925837	2.991741	0.771524	11.442808
44	1.112685	1.114533	0.925820	2.991715	0.771502	11.442751
45	1.112695	1.114538	0.925828	2.991771	0.771504	11.442686
46	1.112681	1.114532	0.925818	2.991719	0.771497	11.442664
47	1.112687	1.114533	0.925822	2.991739	0.771504	11.442649
48	1.112679	1.114531	0.925816	2.991710	0.771496	11.442604
49	1.112714	1.114564	0.925835	2.991748	0.771517	11.442828
50	1.112689	1.114541	0.925821	2.991716	0.771501	11.442868
51	1.112704	1.114555	0.925829	2.991734	0.771521	11.442769
52	1.112683	1.114535	0.925817	2.991712	0.771500	11.442711
53	1.112692	1.114544	0.925822	2.991761	0.771502	11.442657
54	1.112681	1.114533	0.925817	2.991716	0.771497	11.442646
55	1.112685	1.114537	0.925818	2.991731	0.771502	11.442625
56	1.112679	1.114531	0.925816	2.991710	0.771496	11.442604
57	1.112706	1.114557	0.925837	2.991740	0.771512	11.442804
58	1.112684	1.114536	0.925820	2.991714	0.771498	11.442746
59	1.112696	1.114548	0.925829	2.991725	0.771513	11.442737
60	1.112679	1.114531	0.925816	2.991710	0.771496	11.442604
61	1.112686	1.114538	0.925822	2.991738	0.771498	11.442646
62	1.112679	1.114531	0.925816	2.991710	0.771496	11.442604
63	1.112679	1.114531	0.925816	2.991710	0.771496	11.442604
64	1.112679	1.114531	0.925816	2.991710	0.771496	11.442604

Note: The “hybrid” Nash equilibriums are calculated based on the Lindahl principle. There are 6 shaded rows: #32, #38, #56, #60, #62, and #63. They are “1 in” and “5 out” cases. Conceptually and algorithmically, they are equivalent to the Cournot-Nash equilibrium (#64). The simulation results confirm the identities. The same table for RICE2020-R12 is a 4096 x 12 matrix and is included in the data zip file of this paper.

Table 4 The payoff differences from the Lindahl equilibrium

	USA	JPN	EEC	CHN	FSU	ROW
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	-0.0216	-0.0229	-0.0213	-0.0553	-0.0243	**0.3416**
3	-0.0067	-0.0074	-0.0071	-0.0151	**0.0105**	-0.0597
4	-0.0253	-0.0273	-0.0257	-0.0616	-0.0200	*0.1247*
5	-0.0174	-0.0185	-0.0171	**0.0578**	-0.0197	-0.2401
6	-0.0298	-0.0324	-0.0306	-0.0343	-0.0345	-0.1008
7	-0.0225	-0.0242	-0.0226	*0.0202*	-0.0145	-0.2749
8	-0.0314	-0.0345	-0.0331	-0.0537	-0.0331	-0.2334
9	-0.0043	-0.0050	**0.0047**	-0.0079	-0.0054	-0.0160
10	-0.0239	-0.0259	-0.0194	-0.0575	-0.0275	*0.1810*
11	-0.0109	-0.0122	-0.0041	-0.0234	*0.0022*	-0.0813
12	-0.0275	-0.0301	-0.0252	-0.0640	-0.0254	-0.0224
13	-0.0207	-0.0224	-0.0150	*0.0298*	-0.0239	-0.2435
14	-0.0307	-0.0338	-0.0302	-0.0492	-0.0361	-0.2018
15	-0.0257	-0.0279	-0.0219	-0.0062	-0.0211	-0.2843
16	-0.0320	-0.0357	-0.0334	-0.0657	-0.0361	-0.3150
17	-0.0070	**0.0105**	-0.0074	-0.0164	-0.0082	-0.0661
18	-0.0255	-0.0185	-0.0260	-0.0625	-0.0293	*0.1164*
19	-0.0134	-0.0001	-0.0141	-0.0313	-0.0011	-0.1274
20	-0.0289	-0.0253	-0.0300	-0.0680	-0.0274	-0.0773
21	-0.0227	-0.0133	-0.0229	*0.0187*	-0.0261	-0.2797
22	-0.0315	-0.0310	-0.0333	-0.0546	-0.0370	-0.2385
23	-0.0275	-0.0217	-0.0280	-0.0160	-0.0236	-0.3156
24	-0.0324	-0.0345	-0.0350	-0.0692	-0.0370	-0.3384
25	-0.0112	*0.0025*	-0.0044	-0.0247	-0.0134	-0.0872
26	-0.0277	-0.0237	-0.0254	-0.0648	-0.0323	-0.0297
27	-0.0176	-0.0078	-0.0127	-0.0400	-0.0091	-0.1552
28	-0.0309	-0.0300	-0.0305	-0.0706	-0.0323	-0.2091
29	-0.0259	-0.0196	-0.0222	-0.0075	-0.0301	-0.2883
30	-0.0321	-0.0338	-0.0335	-0.0664	-0.0382	-0.3187
31	-0.0306	-0.0276	-0.0285	-0.0405	-0.0298	-0.3314
32	-0.0327	-0.0366	-0.0357	-0.0776	-0.0391	-0.3955
33	**0.0138**	-0.0104	-0.0097	-0.0243	-0.0109	-0.1093
34	-0.0158	-0.0291	-0.0274	-0.0666	-0.0309	*0.0581*
35	*0.0027*	-0.0172	-0.0162	-0.0384	-0.0043	-0.1666
36	-0.0226	-0.0328	-0.0311	-0.0712	-0.0292	-0.1254
37	-0.0107	-0.0264	-0.0246	*0.0083*	-0.0280	-0.3081
38	-0.0279	-0.0353	-0.0339	-0.0588	-0.0376	-0.2666
39	-0.0192	-0.0314	-0.0295	-0.0250	-0.0258	-0.3391
40	-0.0311	-0.0363	-0.0353	-0.0713	-0.0375	-0.3521

41	*0.0056*	-0.0151	-0.0070	-0.0320	-0.0159	-0.1276
42	-0.0210	-0.0316	-0.0269	-0.0682	-0.0336	-0.0802
43	-0.0051	-0.0217	-0.0150	-0.0466	-0.0120	-0.1914
44	-0.0272	-0.0351	-0.0316	-0.0729	-0.0337	-0.2488
45	-0.0171	-0.0299	-0.0240	-0.0169	-0.0318	-0.3133
46	-0.0305	-0.0360	-0.0341	-0.0689	-0.0385	-0.3359
47	-0.0252	-0.0348	-0.0301	-0.0483	-0.0317	-0.3512
48	-0.0327	-0.0366	-0.0357	-0.0776	-0.0391	-0.3955
49	*0.0023*	-0.0037	-0.0165	-0.0395	-0.0185	-0.1718
50	-0.0229	-0.0273	-0.0313	-0.0718	-0.0350	-0.1315
51	-0.0081	-0.0134	-0.0226	-0.0534	-0.0149	-0.2312
52	-0.0287	-0.0328	-0.0345	-0.0755	-0.0352	-0.2890
53	-0.0195	-0.0242	-0.0297	-0.0262	-0.0336	-0.3423
54	-0.0312	-0.0351	-0.0353	-0.0716	-0.0388	-0.3542
55	-0.0272	-0.0314	-0.0341	-0.0562	-0.0337	-0.3749
56	-0.0327	-0.0366	-0.0357	-0.0776	-0.0391	-0.3955
57	-0.0055	-0.0110	-0.0153	-0.0476	-0.0233	-0.1960
58	-0.0274	-0.0316	-0.0318	-0.0734	-0.0375	-0.2535
59	-0.0155	-0.0204	-0.0228	-0.0621	-0.0221	-0.2632
60	-0.0327	-0.0366	-0.0357	-0.0776	-0.0391	-0.3955
61	-0.0254	-0.0298	-0.0303	-0.0493	-0.0372	-0.3536
62	-0.0327	-0.0366	-0.0357	-0.0776	-0.0391	-0.3955
63	-0.0327	-0.0366	-0.0357	-0.0776	-0.0391	-0.3955
64	-0.0327	-0.0366	-0.0357	-0.0776	-0.0391	-0.3955

Note: the entries are the difference between the cells in each row and the first row in Table 3. To provide a better visual effect, the results are multiplied by 1000. All except colored entries are negative. The 6 red and bold entries are “5 in” and “1 out” cases where the sole outsider gains. The 13 orange and italic entries are “4 in” and “2 out” cases in which one of the outsiders gains and the other loses.

Table 5 The payoff differences from the Cournot-Nash equilibrium

	USA	JPN	EEC	CHN	FSU	ROW
1	0.0327	0.0366	0.0357	0.0776	0.0391	0.3955
2	0.0111	0.0137	0.0145	0.0223	0.0148	0.7371
3	0.0260	0.0292	0.0287	0.0625	0.0496	0.3358
4	0.0074	0.0093	0.0100	0.0161	0.0191	0.5202
5	0.0153	0.0181	0.0187	0.1355	0.0194	0.1554
6	0.0029	0.0042	0.0051	0.0433	0.0046	0.2947
7	0.0102	0.0124	0.0131	0.0979	0.0246	0.1206
8	0.0013	0.0020	0.0026	0.0239	0.0060	0.1621
9	0.0284	0.0316	0.0405	0.0697	0.0337	0.3795
10	0.0088	0.0107	0.0163	0.0201	0.0116	0.5765
11	0.0218	0.0243	0.0317	0.0542	0.0413	0.3141
12	0.0052	0.0065	0.0106	0.0137	0.0137	0.3731
13	0.0120	0.0142	0.0208	0.1074	0.0152	0.1519
14	0.0020	0.0028	0.0055	0.0284	0.0030	0.1937
15	0.0070	0.0086	0.0138	0.0715	0.0180	0.1112
16	0.0007	0.0009	0.0023	0.0119	0.0030	0.0805
17	0.0257	0.0471	0.0283	0.0612	0.0309	0.3294
18	0.0072	0.0181	0.0097	0.0151	0.0098	0.5119
19	0.0193	0.0364	0.0216	0.0464	0.0380	0.2680
20	0.0038	0.0113	0.0058	0.0096	0.0117	0.3181
21	0.0100	0.0233	0.0129	0.0963	0.0130	0.1158
22	0.0012	0.0056	0.0025	0.0231	0.0021	0.1569
23	0.0052	0.0149	0.0077	0.0617	0.0155	0.0798
24	0.0003	0.0021	0.0007	0.0084	0.0021	0.0571
25	0.0215	0.0391	0.0313	0.0530	0.0257	0.3082
26	0.0050	0.0129	0.0103	0.0129	0.0068	0.3658
27	0.0151	0.0288	0.0230	0.0376	0.0300	0.2403
28	0.0018	0.0066	0.0053	0.0070	0.0068	0.1863
29	0.0068	0.0169	0.0135	0.0701	0.0090	0.1071
30	0.0006	0.0027	0.0022	0.0113	0.0009	0.0768
31	0.0021	0.0090	0.0072	0.0372	0.0093	0.0641
32	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
33	0.0465	0.0262	0.0260	0.0533	0.0282	0.2862
34	0.0169	0.0074	0.0083	0.0110	0.0082	0.4536
35	0.0354	0.0194	0.0195	0.0392	0.0348	0.2289
36	0.0101	0.0038	0.0046	0.0065	0.0099	0.2701
37	0.0220	0.0102	0.0111	0.0859	0.0111	0.0873
38	0.0048	0.0013	0.0019	0.0188	0.0015	0.1288
39	0.0135	0.0051	0.0062	0.0527	0.0133	0.0564
40	0.0016	0.0003	0.0005	0.0064	0.0016	0.0434

41	0.0383	0.0215	0.0288	0.0456	0.0232	0.2679
42	0.0117	0.0049	0.0088	0.0095	0.0055	0.3153
43	0.0276	0.0148	0.0208	0.0310	0.0271	0.2041
44	0.0055	0.0015	0.0041	0.0047	0.0054	0.1467
45	0.0156	0.0066	0.0117	0.0608	0.0073	0.0822
46	0.0022	0.0005	0.0017	0.0087	0.0006	0.0596
47	0.0075	0.0018	0.0057	0.0293	0.0074	0.0443
48	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
49	0.0350	0.0329	0.0192	0.0381	0.0206	0.2237
50	0.0098	0.0092	0.0044	0.0059	0.0041	0.2640
51	0.0246	0.0232	0.0131	0.0242	0.0242	0.1643
52	0.0040	0.0037	0.0012	0.0021	0.0039	0.1064
53	0.0132	0.0124	0.0060	0.0515	0.0055	0.0532
54	0.0015	0.0015	0.0004	0.0060	0.0003	0.0413
55	0.0055	0.0052	0.0016	0.0214	0.0054	0.0206
56	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
57	0.0272	0.0255	0.0204	0.0300	0.0158	0.1995
58	0.0053	0.0049	0.0040	0.0042	0.0016	0.1420
59	0.0172	0.0162	0.0129	0.0155	0.0170	0.1323
60	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
61	0.0073	0.0068	0.0055	0.0283	0.0019	0.0419
62	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
63	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
64	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Note: the entries are the difference between the cells in each row and the last row in Table 3. To provide a better visual effect, the results are multiplied by 1000. The entries in this table are uniformly positive: all regions are better off than their Cournot-Nash equilibrium, in or out of a coalition.

Table 6 The payoff differences from social optimum (utilitarian principle)

	USA	JPN	EEC	CHN	FSU	ROW
1						
2	+	+	+	+	+	-
3	-	-	-	-	+	-
4	+	+	+	+	+	-
5	+	+	-	+	-	-
6	+	+	+	+	+	-
7	+	-	-	+	+	-
8	+	+	+	+	+	-
9	+	-	+	-	-	-
10	+	+	+	+	+	-
11	+	-	+	-	+	-
12	+	+	+	+	+	-
13	+	+	+	+	-	-
14	+	+	+	+	+	-
15	+	-	+	+	+	-
16	+	+	+	+	+	-
17	-	+	-	-	-	-
18	+	+	+	+	+	-
19	-	+	-	-	+	-
20	+	+	+	+	+	-
21	+	+	-	+	-	-
22	+	+	+	+	+	-
23	+	+	-	+	+	-
24	+	+	+	+	+	-
25	-	+	+	-	-	-
26	+	+	+	+	+	-
27	-	+	+	-	+	-
28	+	+	+	+	+	-
29	+	+	+	+	-	-
30	+	+	+	+	+	-
31	+	+	+	+	+	-
32	+	+	+	+	+	-
33	+	-	-	-	-	-
34	+	+	+	+	+	-
35	+	-	-	-	+	-
36	+	+	+	+	+	-
37	+	-	-	+	-	-
38	+	+	+	+	+	-
39	+	-	-	+	+	-
40	+	+	+	+	+	-

41	+	-	+	-	-	-
42	+	+	+	+	+	-
43	+	-	+	-	+	-
44	+	+	+	+	+	-
45	+	-	+	+	-	-
46	+	+	+	+	+	-
47	+	-	+	+	+	-
48	+	+	+	+	+	-
49	+	+	-	-	-	-
50	+	+	+	+	+	-
51	+	+	-	-	+	-
52	+	+	+	+	+	-
53	+	+	-	+	-	-
54	+	+	+	+	+	-
55	+	+	-	+	+	-
56	+	+	+	+	+	-
57	+	+	+	-	-	-
58	+	+	+	+	+	-
59	+	+	+	-	+	-
60	+	+	+	+	+	-
61	+	+	+	+	-	-
62	+	+	+	+	+	-
63	+	+	+	+	+	-
64	+	+	+	+	+	-

Note: "+" indicates the positive payoff difference between the hybrid Nash equilibrium and the equal-weight social optimum; "-" denotes the negative payoff difference between the hybrid Nash equilibrium and the equal-weight social optimum.

Table 7 The payoff differences from Cournot-Nash equilibrium (utilitarian principle)

	USA	JPN	EEC	CHN	FSU	ROW
1	-	-	-	-	-	+
2	-	+	+	+	+	+
3	-	-	-	-	+	+
4	-	+	+	+	+	+
5	-	-	-	+	-	+
6	+	+	+	+	+	+
7	-	-	-	+	+	+
8	+	+	+	+	+	+
9	-	-	+	-	-	+
10	-	+	+	+	+	+
11	-	-	+	-	+	+
12	-	+	+	+	+	+
13	-	-	+	+	-	+
14	+	+	+	+	+	+
15	-	-	+	+	+	+
16	+	+	+	+	+	+
17	-	+	-	-	-	+
18	-	+	+	+	+	+
19	-	+	-	-	+	+
20	-	+	+	+	+	+
21	-	+	-	+	-	+
22	+	+	+	+	+	+
23	-	+	-	+	+	+
24	+	+	+	+	+	+
25	-	+	+	-	-	+
26	-	+	+	+	+	+
27	-	+	+	-	+	+
28	-	+	+	+	+	+
29	-	+	+	+	-	+
30	+	+	+	+	+	+
31	-	+	+	+	+	+
32						
33	+	-	-	-	-	+
34	+	-	+	+	+	+
35	+	-	-	-	+	+
36	+	-	+	+	+	+
37	+	-	-	+	-	+
38	+	+	+	+	+	+
39	+	-	-	+	+	+
40	+	+	+	+	+	+

41	+	-	+	-	-	+
42	+	-	+	+	-	+
43	+	-	+	-	+	+
44	+	-	+	+	+	+
45	+	-	+	+	-	+
46	+	+	+	+	+	+
47	+	-	+	+	+	+
48						
49	+	+	-	-	-	+
50	+	+	+	+	-	+
51	+	+	-	-	+	+
52	+	+	-	+	+	+
53	+	+	-	+	-	+
54	+	+	+	+	+	+
55	+	+	-	+	+	+
56						
57	+	+	+	-	-	+
58	+	+	+	+	-	+
59	+	+	+	-	+	+
60						
61	+	+	+	+	-	+
62						
63						
64						

Note: “+” indicates the positive payoff difference between the hybrid Nash equilibrium and the Cournot-Nash equilibrium; “-” denotes the negative payoff difference between the hybrid Nash equilibrium and the Cournot-Nash equilibrium.

Table 8 Gaining Singleton Regions

	Row #	Regions	Row #	Regions	Row #	Regions	Row #	Regions
11-member coalition vs. one outsider	2	ROW	3	LAA	5	SSA	9	MDE
	17	IND	33	CHN	65	EEC	129	RUS
	257	OHI	513	JPN	1025	EUS	2049	USA
10-member coalition vs. two outsiders	4	LAA, ROW	7	SSA, LAA	10	MDE, ROW	11	MDE, LAA
	18	IND, ROW	19	IND, LAA	25	IND, MDE	34	CHN, ROW
	41	CHN, MDE	69	CHN, SSA	130	EEC, ROW	131	EEC, LAA
	133	EEC, SSA	137	EEC, MDE	145	EEC, IND	258	RUS, ROW
	259	RUS, LAA	261	RUS, SSA	265	RUS, MDE	273	RUS, IND
	385	OHI, RUS	515	JPN, LAA	517	JPN, SSA	577	JPN, EEC
	641	JPN, RUS	769	JPN, OHI	1026	EUS, ROW	1027	EUS, LAA
	1029	EUS, SSA	1033	EUS, MDE	1041	EUS, IND	1153	EUS, RUS
	1281	EUS, OHI	1537	EUS, JPN	2050	USA, ROW	2051	USA, LAA
	2057	USA, MDE	2065	USA, IND	2081	USA, CHN	2177	USA, RUS
	2305	USA, OHI	3073	USA, EUS				
9-member coalition vs. three outsiders	26	IND, MDE, ROW	275	OHI, CHN, LAA	387	OHI, RUS, LAA	1155	EUS, RUS, LAA
	1283	EUS, OHI, LAA	1409	EUS, OHI, RUS				

Appendix:

1. <u>The Proof of Proposition 2</u>

Proof: Rewriting (2) as a global transformation relationship as:

$$F(x, B) = 0, \quad x = \sum_{i=1}^{N} x_i\,, \quad B = \sum_{i=1}^{N} b_i \qquad (A-1)$$

The Lagrangian of problem (9) is:

$$\mathcal{L} = \sum_{i=1}^{N} \log\left(U^i - \underline{U}^i\right) - \mu F(x, B), \qquad (A-2)$$

The first order conditions are (after eliminating y_i through (4)):

$$\frac{\partial \mathcal{L}}{\partial x_i} = \frac{1}{U^i(x_i, B) - \underline{U}^i}\frac{\partial U^i}{\partial x_i} - \mu\frac{\partial F}{\partial x} = 0, \qquad i = 1, 2, \dots, N. \quad (A-3)$$

$$\frac{\partial \mathcal{L}}{\partial B} = \sum_{i=1}^{N}\left(\frac{1}{U^i(x_i, B) - \underline{U}^i}\ \frac{\partial\, U^i}{\partial B}\right) - \mu\frac{\partial F}{\partial B} = 0 \quad (A-4)$$

Re-organizing (A-4), we have,

$$\left(\, U^i(x_i, B) - \underline{U}^i\right) = \frac{\dfrac{\partial\, U^i}{\partial\, x_i}}{\mu\dfrac{\partial F}{\partial x}}, \qquad i = 1, 2, \dots, N. \quad (A-5)$$

Substituting $\left(U^i(x_i, B) - \underline{U}^i\right)$ in (A-4) with (A-5), we obtain the following Samuelson rule:

$$\sum_{i=1}^{N}\left(\frac{\partial\, U^i}{\partial B} \Big/ \frac{\partial U^i}{\partial x_i}\right) = \frac{\partial F}{\partial B} \Big/ \frac{\partial F}{\partial x}, \qquad or \ \sum_{i=1}^{N} MRS_{x_i,B} = MRT_{x,B} \qquad (A-6)$$

In addition, if private good x is the numeraire and denotes the shadow price of externality as p_B, we have,

$$\frac{\frac{\partial F}{\partial B}}{\frac{\partial F}{\partial x}} = MRT_{x,B} = \frac{p_B}{p_x} = p_B \qquad (A-7)$$

Denote $\{r_i\}$ as the cost-share arrangement such that $\sum r_i = 1$. Using the share in (A-3):

$$r_i(x_i) = \left(\frac{1}{\sum_{j=1}^{N}\left(\frac{1}{U^j(x_i,B) - \underline{U}^j}\right)}\right)\left(\frac{1}{U^i(x_i,B) - \underline{U}^i}\right), \quad \sum_{i=1}^{N} r_i = 1 \qquad (A-8)$$

Alternatively, r_i is proportional to the inverse of $(U^i(x_i,B) - \underline{U}^i)$. Furthermore, let

$$\frac{\frac{\partial U^i}{\partial B}}{\frac{\partial U^i}{\partial x_i}} = r_i(x_i)p_B, \quad i = 1, 2, \dots, N. \qquad (A-9)$$

Assessing (1) under social welfare weights $\{r_i\}$ given by (A-8) (after eliminating y_i from (4)):

$$Max \qquad W = \sum_{i=1}^{N} r_i U^i(x_i, B), \qquad (A-10)$$

$$s.t. \quad F(x,B) = 0, \quad x = \sum_{i=1}^{N} x_i\,, \quad B = \sum_{i=1}^{N} b_i \qquad (A-1)$$

The first-order conditions of objective function (A-10) are:

$$\frac{\partial \mathcal{L}}{\partial\, x_i} = r_i \frac{\partial\, U^i}{\partial x_i} - \mu \frac{\partial F}{\partial x} = 0, \qquad i = 1, 2, \dots, N. \quad (A-3')$$

$$\frac{\partial \mathcal{L}}{\partial B} = \sum_{i=1}^{N} \left(r_i \frac{\partial U^i}{\partial B} \right) - \mu \frac{\partial F}{\partial B} = 0 \quad (A-4')$$

Comparing r_i in (A-3') and (A-4') with $1/(U^i(x_i, B) - \underline{U}^i)$ in (A-3) and (A-4), we have that, if $r_i \propto 1/(U^i(x_i,B) - \underline{U}^i)$, Nash bargaining problem (9) and social optimum (A-10) have the identical first-order condition. Therefore, the solution of social optimum (A-10) is identical to the solution of the Nash bargaining problem (9) if $r_i \propto 1/(U^i(x_i,B) - \underline{U}^i)$. #

2. The Mathematical Description of the RICE Model in Discrete Form

I. The RICE model as a social planner's problem (efficient solutions fully internalize climate externality):

1 The objective function of the social planner that is subject to the constraints (B-2) to (B-17):

$$\underset{\{I(n,t),\mu(n,t)\}}{Max} \quad W = \sum_{n=1}^{N} \sum_{t=0}^{T} \varphi_n L(n,t) U(c(n,t)) (1+\rho)^{-t}, \quad \sum_{n=1}^{N} \varphi_n = N \qquad (B-1)$$

2 The economic module (regional Solow growth model with embedded economy-climate interaction):

Definition of utility function:

$$U(c(n,t)) = \begin{cases} \dfrac{c(n,t)^{1-\alpha}}{1-\alpha}, & 0 \le \alpha < 1 \\ \log(c(n,t)), & \alpha = 1 \end{cases} \qquad (B-2)$$

Definition of per capita consumption:

$$c(n,t) = \frac{C(n,t)}{L(n,t)} \qquad (B-3)$$

Aggregate production function without climate damage and mitigation cost:

$$Q(n,t) = A(n,t) L(n,t)^{\gamma} K(n,t)^{1-\gamma} \qquad (B-4)$$

Aggregate production function including climate damage and mitigation costs:

$$Y(n,t) = \Omega(n,t)Q(n,t) \qquad (B-5)$$

Net GDP identity:

$$Y(n,t) = C(n,t) + I(n,t) \qquad (B-6)$$

The motion equation of Capital formation:

$$K(n,t+1) = I(n,t) + (1-\delta_K)K(n,t), \quad K(n,0) = K_{n,0} > 0, \quad 0 < \delta_K < 1 \qquad (B-7)$$

3 Carbon emission and economy-climate linkage module:

Industrial carbon emission equation:

$$E_{Ind}(n,t) = \sigma(n,t)(1-\mu(n,t))Q(n,t), \quad \mu(n,t) \geq 0. \qquad (B-8)$$

Total regional carbon emission:

$$E(n,t) = E_{Ind}(n,t) + E_{Land}(n,t) \qquad (B-9)$$

Global carbon stock constraint:

$$\sum_{t=0}^{T}\sum_{n=1}^{N} E_{Ind}(n,t) \leq CCum \qquad (B-10)$$

Climate to economy adjustment function:

$$\Omega(n,t) = \frac{1-MC(\mu(n,t))}{1+CD(n,T_{AT}(t))}, \quad 0 \ll \Omega(n,t) < 1 \qquad (B-11)$$

4 Carbon cycle module:

$$M_{AT}(t+1) = \sum_{n=1}^{N} E(n,t) + \phi_{11}M_{AT}(t) + \phi_{21}M_{UP}(t) \qquad (B-12)$$

$$M_{UP}(t+1) = \phi_{12}M_{AT}(t) + \phi_{22}M_{UP}(t) + \phi_{32}M_{LO}(t) \qquad (B-13)$$

$$M_{LO}(t+1) = \phi_{23}M_{UP}(t) + \phi_{33}M_{LO}(t) \qquad (B-14)$$

$$F(t) = \eta\{log_2[M_{AT}(t) / M_{AT}(1750)]\} + F_{EX}(t) \qquad (B-15)$$

$$T_{AT}(t+1) = T_{AT}(t) + \xi_1\{F(t) - \xi_2 T_{AT}(t) - \xi_3[T_{AT}(t) - T_{LO}(t)]\} \qquad (B-16)$$

$$T_{LO}(t+1) = T_{LO}(t) + \xi_4\{T_{AT}(t) - T_{LO}(t)]\} \qquad (B-17)$$

II. The RICE model as a non-cooperative Cournot-Nash game (inefficient and unique Cournot-Nash equilibrium):

$$\underset{\{I(n,t),\mu(n,t)\}}{Max} \quad V^n = \sum_{t=0}^{T} U(c(n,t))(1+\rho)^{-t}, \quad n = 1,2,...,N. \qquad (B-18)$$

s.t. constraints (B-2) to (B-17) except for (B-12) that is revised as:

$$M_{AT}(t+1) = E(n,t) + \sum_{j \neq n}^{N} \overline{E}(j,t) + \phi_{11} M_{AT}(t) + \phi_{21} M_{UP}(t) \qquad (B-12')$$

III. The RICE model as a coalitional game (the "hybrid" Nash equilibrium):

Suppose a coalition S contains s regions such that S $\subset$ N and N\S regions do not belong to S. Then, S solves a social planner's problem like (1) internally and simultaneously plays the non-cooperative game as a single player with other N\S regions as in (B-18).

The Definition of notations and symbols in I and II:

Definitions of sets:

n --- regions; t --- time (in discrete steps, 5-year period).

Definition of the objective functions:

W, V^n --- in A and B, respectively.

Definitions of control variables and state variables:

(a) Control variables (underneath "Max" in (B-1) and (B-18)):

I(n,t) --- investment of regions; (trillions of 2015 U.S. dollars)

μ(n,t) --- carbon control rates of regions. (in percentage, maybe greater than 1 if carbon sequestration technology is allowed)

State variables:

U(c(n,t)) --- instantaneous payoff functions of regions; (in scalable and nontransferable "utils")

C(n,t) --- consumptions of goods and services of regions; (trillions of 2015 U.S. dollars)

c(n,t) --- per capita consumptions; (thousands of U.S. dollars per person)

K(n,t) --- capital stocks of regions; (trillions of 2015 U.S. dollars)

Q(n,t) --- output of goods and services of regions, gross of abatement and damages; (trillions of 2015 U.S. dollars)

Y(n,t) --- output of goods and services of regions, net of abatement and damages; (trillions of 2015 U.S. dollars)

$E_{Ind}(n,t)$ --- industrial carbon emissions by regions; (billions of metric tons C per period)

E(n,t) --- total carbon emissions by regions; (billions of metric tons C per period)

MC(μ(n,t)) --- mitigation cost of regions; (trillions of 2015 U.S. dollars)

CD(n, $T_{AT}(t)$) --- climate damage of regions; (trillions of 2015 U.S. dollars)

Ω(n,t) --- gross-net output converting factor

$M_{AT}(t)$, $M_{UP}(t)$, $M_{LO}(t)$ --- carbon concentration in the atmosphere, upper oceans, and deep oceans, respectively; (billions of metric tons C at the beginning of each period)

$T_{AT}(t)$, $T_{LO}(t)$ --- global mean surface temperature and temperature of upper oceans (°C from 1900)

F(t) --- total radiative forcing; (watts per m^2 from 1900)

Definitions of exogenous trends:

A(n,t) --- total factor productivity of regions; (productivity units)

L(n,t) --- population and proportional to labor inputs of regions; (billions)

$E_{Land}(n,t)$ --- carbon emissions from land use; (billions of metric tons C per period)

$\sigma(t)$ --- ratio of uncontrolled industrial emissions to output (a proxy for AEEI)

$F_{EX}(t)$ --- exogenous radiative forcing; (watts per m^2 from 1900)

Definitions of parameters:

φ_n --- social welfare weights;

α --- elasticity of marginal utility of consumption; (pure number)

ρ --- pure rate of social time preference (per year)

$CCum$ --- maximum consumption of fossil fuels (billions of metric tons of carbon)

γ --- elasticity of output with respect to capital (pure number)

δ_K --- rate of depreciation of capital (per period)

η --- temperature-forcing parameter (°C per watts per meter squared)

$\phi_{11}, \phi_{21}, \phi_{22}, \phi_{32}, \phi_{12}, \phi_{33}, \phi_{23}$ --- parameters of the carbon cycle (flows per period)

$\xi_1, \xi_2, \xi_3, \xi_4$ --- parameters of climate equations (flows per period)

ψ_1, ψ_2 = parameters of the damage function

References:

Barrett, S. (1994), "Self-enforcing international environmental agreements," *Oxford Economic Papers*, **46**, 111-122.

Barrett, S. (2003), *Environment and Statecraft--- the Strategy of Environmental Treaty-making*, Oxford University Press, Oxford, UK.

Bosetti, V., C. Carraro, E. DeCian, E. Massetti, and M. Tavoni, (2012) "Incentives and Stability of International Climate Coalitions: An Integrated Assessment." *Energy Policy*, **55**: 44–56.

Carraro, C. and D. Siniscalco, (1998) "International environmental agreements: incentives and political economy," *European Economic Review*, **42**, 561-572

Chander P. and H. Tulkens, (1997) "A core of an economy with multilateral environmental externalities," *International Journal of Game Theory*, **26**, 379-401

Chander P. and H. Tulkens, (2008) "Cooperation, stability, and self-enforcement in international environmental agreements: conceptual discussion," in *The Design of Climate Policy*, eds. By R. Guesnerie and H. Tulkens, The MIT Press.

D'Aspremont, C., A. Jacquemin, J. J. Gabszwwicz, and J. A. Weymark (1983) "On the stability of collusive price leadership," *The Canadian Journal of Economics*, **16**, 17-25.

De Zeeuw, A. (2008) "Dynamic effects on the stability of international environmental agreements," *Journal of Environmental Economics and Management*, **55**(2), 163-174.

Eyckmans J. and H. Tulkens, (2003) "Simulating coalitionally stable burden sharing agreements for the climate change problem," *Resource and Energy Economics*, **25**, 299–327.

Finus, M. and B. Rundshagen, (2009) "Membership Rules and Stability of Coalition Structures in Positive Externality Games." *Social Choice and Welfare*, **32** (3): 389–406.

Foley, D. (1970) “Lindahl's solution and the core of an economy with public goods." *Econometrica*, **38**(1), 66-72.

Kaneko, M. (1977) “The ratio equilibrium and a voting game in a public goods economy.” *Journal of Economic Theory*, **16**, 123–136.

Kersting, J., V. Duscha, and M. Weitzel, (2017) “Cooperation on climate change under economic linkages: how the inclusion of macroeconomic effects affects stability of a global climate coalition,” *The Energy Journal*, **38**(4), 19-42.

Kolstad, C. (2007) “Systematic uncertainty in self-enforcing international environmental agreements,” *Journal of Environmental Economics and Management*, **53**(1), 68-79.

Lessmann, K., U. Kornek; V. Bosetti, R. Dellink, J. Emmerling, J. Eyckmans, M. Nagashima, H. Weikard, Z. Yang (2015) “The Stability and effectiveness of climate coalitions: a comparative analysis of multiple integrated assessment models,” *Environmental and Resource Economics*, **62**, 811-834.

Lindahl, E. (1919) “Just taxation - a positive solution." In *Classics in the Theory of Public Finance* (Eds. R. Musgrave and A. Peacock) (1958), Macmillan, London, UK.

Mas-Colell, A. and J. Silvestre, (1989), “Cost share equilibria: a Lindahl approach,” *Journal of Economic Theory*, **47**, 239-56.

McGinty, M. (2007) “International environmental agreements among asymmetric nations,” *Oxford Economic Papers*. **59**(1), 45-62.

Nordhaus, W. D. and Z. Yang, (1996) “A Regional Dynamic General-Equilibrium Model of Alternative Climate-Change Strategies,” *American Economic Review*, **86**(4), 741-765.

Nordhaus, W. D., (2015), “Climate club: overcoming free-riding in climate policy,” *American Economic Review*, **105**(4), 1339-1370.

Nordhaus, W. D. and Z. Yang, (2021) “The new DICE2020/RICE2020 modeling platform,” mimeo.

Osmani, D. and R. Tol (2009) “Toward farsightedly stable international environmental agreement,” *Journal of Public Economic Theory*, **11**(3), 455-492.

Ray, D., (2007), *A Game-Theoretic Perspective on Coalition Formation*. Oxford University Press, Oxford, UK.

Starrett, D. (1973), “A note on Externalities and the core,” *Econometrica*, **41**, 179-83.

UNFCCC (2016), https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement

van den Nouweland, A. (2015). “Lindahl and Equilibrium”, In *Individual and Collective Choice and Social Welfare*, Binder, C., Codognato, G., Teschl, M., Xu, Y. (eds). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-46439-7_18

Volt, C. (2016) “Climate coalition formation when players are heterogeneous and inequality averse” *Environmental and Resource Economics*, **65**, 33-59.

Yang, Z. (2008). *Strategic Bargaining and Cooperation in Greenhouse Gas Mitigation --- A, n Integrated Assessment Modeling Approach*, The MIT Press, Cambridge, MA.

Yang, Z. (2021). *The Environment and Externality --- The Theory, Algorithms, and Applications*, Cambridge University Press, Cambridge, UK.