Essays on International Environmental Agreements

Fadi Harb

A Thesis

In the Department

of

Economics

Presented in Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy (Economics) at

Concordia University

Montréal, Québec, Canada

December 2022

CONCORDIA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Fadi Harb

Entitled: Essays on International Environmental Agreements

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Economics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

______________________________ Chair

Dr. Peter Pawelek

______________________________ Supervisor

Dr. Effrosyni Diamantoudi

______________________________ External Examiner

Dr. George Deltas

______________________________ Examiner

Dr. Ming Li

______________________________ Examiner

Dr. Majumdar Dipjyoti

______________________________ Examiner

Dr. Huan Xie

Approved by ______________________________

Dr. Christian Sigouin, Graduate Program Director

____

Dr. Pascale Sicotte, Dean, Faculty of Arts and Science

ABSTRACT

Essays on International Environmental Agreements

Fadi Harb, PhD
Concordia University, 2023

The following dissertation consists of three essays that tackle topics in Environmental Economics. It focuses on the formation of International Environmental Agreements (IEAs) in three respective essays: (i) the first essay "IEAs - Contingency plans for all coalition sizes" uses the Stackelberg assumption to build a contingency plan for a remaining coalition when faced with a potential exit by one or more of its members. Homogeneous agents and quadratic benefit and environmental damage functional forms are used to compute contingency plans for all coalition sizes in order to immunise a coalition against single or multiple unilateral deviations, (ii) the second essay "IEAs - Contingency Plans under foresight" studies the series of strategies taken by a coalition of any size when members are foresighted, in that a member considers the possibility that once it acts, another member might react. It presents the strategies taken by different coalition sizes for subsequent unilateral deviations, and (iii) the third essay "IEAs - Choice of net emissions" adds abatement choice as a separate variable. Given that countries have two choice variables, the framework examines agreements on net emissions, in which countries commit to either emissions or abatement and choose the other variable independently in a subsequent stage of the game. Using numerical simulations, results suggest that cooperation on net emissions is possible even with a high degree of heterogeneity among two countries. Comparing results to the pure Nash non-cooperative benchmark case, cooperation on net emissions show that the model achieves lower aggregate net emissions and allow gains from cooperation to both countries.

ACKNOWLEDGEMENT

I would like to start by thanking Dr. Effrosyni Diamantoudi, in her roles as my supervisor, Professor and former Graduate Program Director. I am grateful for her guidance and unwavering support throughout my graduate studies at Concordia University. I would also like to thank Dr. Eftichios Sartzetakis (Department of Economics, University of Macedonia) for his generous assistance and kind feedback. This dissertation would not have been possible without their support.

I would like to thank all the Professors at the Department of Economics for sharing their knowledge and expertise. I would like to express my utmost gratitude to Dr. Tatyana Koreshkova, Dr. Szilvia Papai, Dr. Dipjyoti Majumdar, Dr. Ming Li, Dr. Damba Lkhagvasuren, and Dr. Prosper Dovonon. I am thankful for all their teachings and knowledge. I would like to extend my gratitude to Dr. Maria Da Palma, Dr. Ivan Tchinkov, and Dr. Moshe Lander.

I would like to thank members of the examining committee, external examiner, Dr. George Deltas, examiners, Dr. Huan Xie, Dr. Ming Li and Dr. Dipjyoti Majumdar, and Chair, Dr. Peter Pawelek. Their intuitive comments have greatly contributed to the improvement of this dissertation.

I would like to extend my appreciation and thank the former and current staff members of the Department of Economics; Elise Melancon, Lucy Gilson, Lise Gosselin, Bonnie Janicki, Mary-Anne Jirjis, Melissa Faisal, Sandra Topisirovic, Kelly Routly, and Emilie Martel for their support and assistance. I would also like to thank my friend Olha Hnatsyhyn for her support and cheer through the final stages of my research. I finally like to thank my parents for their encouragement and support.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

INTRODUCTION

International environmental agreements (IEAs) are treaties that establish a set of rules between countries to achieve an environmental goal. IEAs fall under the branch of public good provision theory, since the environment is a public good.[1] Thus, pollution is a "public bad" and global pollutants like greenhouse-gas emissions, mainly caused by human activities, we have "global public bad" problems, the solutions of which require IEAs. William Nordhaus, winner of the 2018 Nobel Prize in economics for his work on climate change, states that: "It is only by designing, implementing, and enforcing cooperative multinational policies that nations can ensure effective climate-change policies" (Nordhaus, 2019). One of the most successful IEAs is the Montreal Protocol, adopted in 1987, which aimed to regulate and reduce the global production, consumption, and emissions of ozone-depleting substances (Velders et al., 2007). More recently, the Paris Agreement, adopted in 2015, addresses climate change with a goal to limit global warming to well below 2 degrees Celsius (UNFCCC, 2015). As damages from pollutants are incurred by all countries, global mitigation efforts are required. Climate action requires global cooperation and coordination and, in the absence of a global supranational authority, IEAs are designed to be self-enforcing, that is, no country is forced to sign an lEA, and countries that sign, can always withdraw from the coalition (Barrett, 1994). The concept of self-enforcement was first employed in IEAs by Barrett in 1994.

An IEA is considered stable if none of the countries that joined the agreement,

[1] The theory of public goods was developed by Samuelson (1954).

called signatories, has an incentive to pull out of the agreement (internal stability) and none of the countries that decided not join the agreement, called non-signatories, has an incentive to participate in the agreement (external stability). The concept of self-enforcing IEAs is embedded in the above definition of stability, where essentially no signatory has an incentive to deviate from the collective agreement. Such a coalition stability notion was originally introduced by D'Aspremont, Jaquemin, Gabszewicz and Weymark (1983) in a price leadership model to study cartel stability, and has been extended to IEAs by Hoel (1992), and Carraro and Siniscalco (1993). Barrett (1994) and Diamantoudi and Sartzetakis (2006) study the problem of deriving the size of a stable IEA. The main difference is the choice variable: countries in Diamantoudi and Sartzetakis (2006) choose emission levels whereas in Barrett's (1994) paper, the choice variable is abatement efforts. Emissions are a byproduct of production, whereas abatement is the elimination of emissions.[2] The concept considers only unilateral deviations, i.e., it assumes that potential deviators expect no other country to follow. This assumption creates a strong incentive for an individual country to free ride on the cooperation of the rest. Free-riding incentives generally lead countries to stay out of an agreement, and thus hinder the efforts of global cooperation. This comes from the circumstance that costs avoided by not abating outweigh marginal environmental damages caused to the country when all other countries in the agreement choose to control their emissions (Diamantoudi and Sartzetakis, 2015). Both works depart from the standard assumption that all countries choose their emissions si-

[2]The main debate between the two papers is that abatement should not be a stand alone model. The reality is that countries do both, emit and abate.

multaneously,[3] and follow the Stackelberg assumption; i.e., signatories of a coalition, assumed to act as a leader, collectively maximise the coalition's aggregate welfare taking the non-signatories' reaction into account, who act as followers. Both works assume homogeneous agents, i.e., countries have identical benefits from emissions and identical damages from global pollution. Diamantoudi and Sartzetakis (2006), assuming homogenous agents and quadratic functional forms, find that if the total number of countries is greater than four, for the IEA to be self-enforcing, a stable coalition will consist of either two, three, or four members. In addition, they show that when the IEA is stable, the welfare level of signatories is very close to its lowest value (Diamantoudi and Sartzetakis, 2006). The main driving force of these results is the assumption that when a country contemplates leaving the coalition, it makes two assumptions: 1.) that no other country will follow, and 2.) the countries that remain in the coalition will adjust their emissions to maximise their joint welfare.

When the choice variable is pollution abatement, Barrett (1994) shows that a self-enforcing IEA can have any number of signatories between two and the grand coalition. Barrett obtains this result using numerical simulations in a pollution abatement model, where he did not constrain emissions as non-negative (Rubio and Ulph, 2006). Rubio and Ulph (2006) use Kuhn-Tucker conditions to derive the equilibrium of the emissions game and show that the key results from Barrett's paper are maintained by providing an analytical proof of the main results of the model introduced

[3] The assumption of countries choosing their emission levels simultaneously is known as the Nash-Cournot assumption. Under this assumption, most of the literature, with specific benefit and damage functional forms, find pessimistic results of small stable coalitions. For more discussions on Nash equilibria in simultaneous games, see, for instance, Hoel (1991, 1992), and Carraro and Siniscalco (1993).

by Barrett (1994). If corner solutions are not considered, coalitions are very small, as Diamantoudi and Sartzetakis (2006) show.

Comparing the assumption of homogeneous agents to the heterogeneous, Barrett (1997) finds that there is no substantial difference in the size of stable coalition in the heterogeneous case relative to the homogeneous (Barrett, 1997). Diamantoudi and Sartzetakis (2015) point out that with the introduction of heterogeneous agents, the results of the literature are mixed, with most papers finding that the size of IEAs may be higher. However, this might not always lead to increases in aggregate welfare (Diamantoudi and Sartzetakis, 2015). Diamantoudi, Sartzetakis and Strantza (2018) assume quadratic functional forms to examine the stability of self-enforcing IEAs among heterogenous agents. They conclude that the assumption of homogeneity is not the determining factor driving pessimistic results of small agreements, and introducing heterogeneity does not enhance the size of a stable agreement compared to the homogeneous case (Diamantoudi, Sartzetakis and Strantza, 2018).

Results in the theoretical literature on IEAs vary depending on the deviator's expectations over the reaction of the remaining coalition's members.[4] Chander and Tulkens (1995, 1997) assume that if one country deviates, the coalition collapses to the Nash equilibrium. Supposing a potential deviator expects that if it leaves, the coalition will collapse, then a much larger coalition, including the grand coalition, are stable (see Chander and Tulkens, 1995, 1997). Furthermore, Diamantoudi and Sartzetakis (2015) find that when endogenising the reaction of the IEA's members

[4]For more literature on coalition formation and the stability of IEAs, see Diamantoudi and Sartzetakis (2006), Chander (2007), Finus and Rundshagen (1998) and Finus and McGinty (2019).

to a deviation by a group of members, new larger farsighted stable coalitions can be formed without any assumption regarding the behaviour of remaining members of the coalition (Diamantoudi and Sartzetakis, 2015).

The first essay in this thesis utilizes the model of Diamantoudi and Sartzetakis (2006) as a benchmark, and follows the Stackelberg assumption. As mentioned earlier, Diamantoudi and Sartzetakis (2006), using the non-cooperative solution concept, assume that when a member contemplates exiting an agreement, the contemplating member makes two assumptions; that no other member will follow and that the remaining coalition's members will adjust their emission so as to maximise their joint welfare (Diamantoudi and Sartzetakis, 2006). The first essay loosens the latter assumption that the remaining signatories will respond to a deviation by adjusting their emissions so as to maximise the joint welfare. Instead, they will increase their emissions so as to increase the deviator's damages to a level such that it nullifies the envisioned benefits from free-riding.

It assumes that when a number of countries agree to cooperate in reducing their emissions, they draft an agreement that contains a contingency plan prescribing the response of coalition members when faced with a potential deviation. In order to discourage deviations, countries commit to the contingency plan to respond to single or multiple countries potentially exiting the agreement and renders any potential exit non-profitable.

The second essay extends the work of contingency plans from the first, and allows an agreement to be signed by countries with a more foresighted approach. The assumption that no other country will follow when a signatory contemplates exiting an agreement is loosened. The framework studies the formation of IEAs taking into account possible subsequent deviations by members. It captures the sequence of strategies that can keep all coalitions stable and proposes a contingency plan at every coalition size to allow an agreement to be signed at the proposed emission targets in order to mitigate any unilateral exit by one of its signatories. The comprehensive cooperation outcome encompasses an optimal level of output and full participation; given any number of signatories in the agreement, the sequence of contingency plans for coalitions is unfolded from full cooperation.

When all countries approach the negotiation table and build a contingency plan accounting for any unilateral member's deviation from the grand coalition, they take into account that if a member exits, another member might possibly follow. The contingency plans built by the coalition allow members to adjust their level of emissions in case of such subsequent unilateral deviations. An example with ten countries is provided that illustrates the contingency plan levels relative to the optimal solutions of Diamantoudi and Sarztetakis (2006). The essay contributes to the understanding that coalition members realizing the potential gains of cooperation, give more substance to their leadership role, and instead of choosing emissions through joint welfare maximisation, they commit to emission levels that can sustain the coalition and their welfare level.

The third essay considers abatement effort as a separate choice variable from emissions to examine the case of heterogeneity in abatement and emission technologies across two countries. The production of emissions generates a benefit for a country, whereas abatement incurs a cost. That is, each country can produce, generate emissions as a byproduct, and incur the costs of abatement. Given that countries have two choice variables, we examine agreements on net emissions, in which countries commit to either emissions or abatement and choose the other variable independently at a subsequent stage of the game. Net emissions are the difference between the level of pollutants emitted and the level of abatement by a country. In the IEA literature on abatement technologies, Carraro and Siniscalco (1997) show that the size of a stable coalition will grow if the coalition members link an environmental agreement with an R&D cooperation with large technology spillovers by securing extra positive externalities among coalition members (Carraro and Siniscalco, 1997). Hoel and de Zeeuw (2010) show that it can be beneficial for IEAs to consider breakthrough technologies and R&D; they show that a large stable coalition can be achieved, resulting in welfare improvement (Hoel and de Zeeuw, 2010). Barrett (2006) examines whether a climate treaty system that relies on targeted R&D and the adoption of breakthrough technologies can improve the performance of IEAs, and finds that except for breakthrough technologies that exhibit increasing returns to scale, a focus on breakthrough technologies cannot enhance the performance of IEAs. Barrett argues that the same forces that undermine Kyoto also challenge the R&D and technology approach (Barrett, 2006).

The theoretical literature on the choice of net emissions is not extensive. Diamantoudi, Sartzetakis, and Strantza (2022) examine the formation and size of stable IEAs on net emissions taking into account countries' choice of emissions, abatement and adaptation strategies. They consider a leadership three-stage game with abatement and adaptation choices, in which countries simultaneously choose their level of adaptation activities independently in the third stage after observing the global net emissions in the second stage. In the second stage, signatories maximise their joint welfare to choose emission and abatement levels taking non-signatories' reactions into account. In the first stage, countries choose whether or not to be part of the agreement (Diamantoudi, Sartzetakis, and Strantza, 2022). Quadratic functional forms are utilised where damages from pollution are a function of net emissions. Given the assumption that adaptation is not effective, introducing abatement as a separate choice variable from emissions will allow a larger stable coalition than in the case where countries choose emission levels only. This is achieved without violating the net emission positivity constraint. The lower the cost of abatement relative to environmental damages, the larger the size of the stable coalition. As the effectiveness of adaptation efforts increases, countries have less incentive to join the coalition and the size of the stable coalition returns to the levels reported in the literature without using abatement efforts (Diamantoudi, Sartzetakis, and Strantza, 2022).

The framework in the third essay utilises the model of Diamantoudi, Sartzetakis, and Strantza (2022) in the absence of adaptation. Four cases are presented, the pure Nash non-cooperative case, the case of cooperation with a commitment to both abatement and emissions, and two cases where countries cooperate on net emissions but commit only to one of the two choice variables. The last two cases introduce a two-stage game with two countries cooperating on net emissions in the first stage and committing to either of the two choice variables. In the second stage, countries simultaneously choose their other choice variable independently, given the level of net emissions agreed upon in the first stage.

This is the first paper in the literature that highlights analysis on net emissions. The main significance of the results is that cooperation on net emissions is possible even with a high degree of heterogeneity among countries. As each country free rides on the other country's net emission reduction efforts, it increases its emissions as the other country reduces its net emissions. It also increases its abatement efforts given the other country's reduction in net emissions. Comparing results to the pure Nash non-cooperative benchmark case, we show that the model achieves lower aggregate net emissions and allows gains from cooperation for both countries.

ESSAY ONE

INTERNATIONAL ENVIRONMENTAL AGREEMENTS: CONTINGENCY PLANS FOR ALL COALITION SIZES

1.1 Introduction

The present essay considers the development of environmental agreements that contain a contingency plan prescribing the emission levels that coalition members will adopt if one or more of its members simultaneously exit the coalition. Under this one-step framework, countries that form any size of coalition agree not only on the emission level that maximises their joint welfare but also on how to respond to a deviation by adjusting their emissions, not to the level that maximises the sum of the remaining members' welfare but instead, increase their emissions to a level that renders a potential deviation non-profitable. The notion of Strong-Nash equilibrium introduced by Aumann (1959) requires stability against deviations by every conceivable coalition. An equilibrium is strong if no coalition, taking the actions of its complement as given, can cooperatively deviate in a way that benefits all of its members (Bernheim et al, 1987). In other words, a Nash Equilibrium is strong if no coalition of players can jointly deviate for all players to still get better payoffs. The Nash concept defines equilibrium in terms of unilateral deviations, while Strong Nash equilibrium allows for deviations by any coalition size (Bernheim et al, 1987).

In the present analysis, homogenous agents and quadratic benefit and environmental damage functional forms similar to Diamantoudi and Sartzetakis (2006) are utilised to compute the contingency plan for the coalition. Diamantoudi and Sartzetakis (2006) assume that when a member contemplates exiting an agreement, it makes two assumptions; that no other member will follow and that the remaining coalition's members will adjust their emission to maximize the joint welfare (Diamantoudi and Sartzetakis, 2006). The analysis under this framework loosens the latter assumption.

The framework develops as follows. First, it assumes that a number of countries agree to cooperate in reducing their emissions and that only one coalition can be formed. One-step deviations are considered, i.e., when one member considers exiting the agreement, it assumes no other member will follow. The framework is similar but not the same as that of Chander and Tulkens (1995, 1997). As mentioned in the general Introduction, the framework of Chander and Tulkens has the embedded assumption that if one country deviates, the coalition collapses to the Nash equilibrium outcome. Instead of an exogenous assumption, in the present paper, we assume that the agreement contains a contingency plan prescribing the response of coalition members when faced with a potential deviator. This response will be an increase in the remaining members' emissions to a level that makes the deviator's damages large enough to nullify the envisioned benefits from free-riding. In the classical Nash equilibrium, the deviator leaves because it is profitable given that no other member is leaving the agreement. When one member deviates, there is no contingency plan, the contemplating member also assumes that other members will behave in a way that they maximize their joint welfare. The reality is that the remaining coalition may not behave in this manner, there are several assumptions one can make upon a deviation. With a country deviating from the agreement, another country might leave, or multiple countries might exit the agreement. The framework on contingencies takes the analysis a step further, if there's a stipulation over what should happen, it specifies the assumption that members can commit to a contingency plan that discourages deviations. The framework begins with the case of a unilateral deviation, and is extended to examine the emission adjustments required to address simultaneous deviations by multiple countries. The remainder of the essay is structured as follows. Section 1.2 describes the model for n identical countries and solves for countries' choice of emissions under the contingency plan. Section 1.3 presents an example of the contingency plan in the case of a unilateral deviation. Section 1.4 studies agreements in the case of multiple simultaneous deviations. Section 1.5 concludes.

1.2 The Model

The model in Diamantoudi and Sartzetakis (2006) is utilised as a benchmark in order to build a contingency plan for members of the coalition. The assumption of homogeneity is maintained, i.e., there exist n identical countries, $N = \{1, \ldots, n\}$, where all countries incur the same costs and benefits. As a result of production and consumption activities, each country i generates positive emissions levels $e_i \geq 0$ of global pollutants. The social welfare of country i, W_i, is defined as total benefits from country i's emissions, $B_i(e_i)$ minus damages from total global emissions, $D_i\,(E)$, where $E = \sum_{i \in N} e_i$.

Given n homogeneous countries, the subscript can be dropped for the benefit and damage parameters. The framework uses the following particular specific functional forms. The benefit function of country i is given by $B(e_i) = b[ae_i - \frac{1}{2}e_i^2]$, where a

and b are positive parameters. The damage function of country i from total emissions are $D(E) = \frac{1}{2}cE^2$. Given the above benefit and damage functions, each country's original welfare function, W_i, can be defined as,

$$W_i(e_i) = b\left[ae_i - \frac{1}{2}e_i^2\right] - \frac{c}{2}(\sum_{i \epsilon N} e_i)^2. \tag{1.0}$$

In the pure non-cooperative case, country i behaves in a Cournot fashion maximising eq. (1.0). The simultaneous solution of the N first order conditions, delivers the non-cooperative Nash equilibrium. Given the assumption of identical countries; all countries generate the same level of emissions. Each country emits a level of emission, e_{nc}, given by

$$e_{nc} = \frac{a}{\gamma n + 1}, \tag{1.1}$$

where $\gamma = \frac{c}{b}$.

In the case of full cooperation, the grand coalition maximises its joint welfare, which yields a per country emissions, e_c. Eq. (1.2) presents the full cooperation level of emissions given by

$$e_c = \frac{a}{\gamma n^2 + 1}. \tag{1.2}$$

Following the joint welfare maximisation and choice of emissions e_c, the grand coalition is not stable as any single country has incentives to exit and attain a higher welfare level when it expects that it will be the only deviator and that the remaining members will adjust their emissions so as to maximise their joint welfare. Under the non-cooperative assumption, the grand coalition generally collapses to a very low participation level, and the agreement is stable only for a small number of countries (Diamantoudi and Sartzetakis, 2006). However, if the signatories to an agreement incorporate a contingency plan that forces them to a level nullifying any benefit from exiting, this agreement could sustain the initial number of signatories. In this analysis, one-step deviations are considered, i.e., when one (or a number of) member(s) consider exiting the coalition, it (they) simultaneously assumes no other member(s) will follow. Within this framework, a contingency plan is built for the remaining coalition to raise its level of emissions when members contemplate exiting. This contingency can be built for potential simultaneous deviations by a single country or multiple countries.

The game starts with an arbitrary set $S \subset N$ of countries that sign an agreement and $N \backslash S$ that do not. Let the size of coalition S be denoted by s. Every signatory of the coalition emits e_s and each non-signatory emits e_{ns} yielding a total emission level $E = E_s + E_{ns} = se_s + (n-s)e_{ns}$. A coalition is said to be internally stable if no signatory has an incentive to exit the agreement. Formally, the internal stability condition is given by,

$$W_{ns}(s-1) \leq W_s(s). \tag{1.3}$$

The coalition is said to be externally stable if no country outside the coalition want to join. Formally, the external stability condition is given by,

$$W_s(s+1) \leq W_{ns}(s). \tag{1.4}$$

In the leadership model, non-signatories act non-cooperatively after having observed the choice of signatories. The level of emissions by non-signatories e_{ns} can be determined as a function of signatories' emission level e_s (Diamantoudi and Sartzetakis, 2006). Each non-signatory's best response function for its level of emissions e_{ns} is given by,

$$e_{ns} = \left[\frac{a - \gamma s e_s}{1 + \gamma(n - s)} \right] \tag{1.5}$$

In this model, signatories adjust their emission level to account for a potential exit, where the action taken by the remaining coalition avoids any payoff to the deviator. Given that in this approach, the size of the coalition is taken as arbitrarily given, the game reduces to a single stage of choosing emission levels. First, we examine the case of a single potential deviation. Let ω_i denote the indirect welfare function of country i. The coalition not only chooses the level of emissions that maximise its collective welfare at s, $e_s^*(s)$, which is the level of emissions derived by Diamantoudi and Sartzetakis (2006), but also chooses the contingency level of emissions at $s - 1$, $\widehat{e}_s(s-1)$, derived from the constraint $\omega_{ns}(\widehat{e}_{ns}(\widehat{e}_s(s-1))) \leq \omega_s(e_s^*(s))$. At each arbitrarily given size of coalition, the optimal level of emissions e_s^* is derived as well as the contingency values $\widehat{e_s}$ associated with signatories' contingency plan. That is, given s, the coalition identifies $e_s^*(s)$ and $\widehat{e}_s(s-1)$ and agrees upon both of them. In this manner, the maximisation problem is formulated as follows

$$\max_{\widehat{e}_s(s-1)} sW_s(s) \tag{1.6}$$

$$\textit{subject to} \qquad \omega_{ns}(\widehat{e}_{ns}(\widehat{e}_s(s-1))) \leq \omega_s(e_s^*(s)).$$

Signatories maximise their joint welfare at s in addition to setting up a contingency plan for their emissions at $s-1$. That is, signatories agree on e_s^* at s, and the contingency plan at $s-1$ determines the level of emissions that signatories include in the agreement for a potential unilateral exit when the size of the coalition is at s. For example, if $n = 10$ and all ten countries sign an agreement, the contingency plan at $s = 9$, $\widehat{e}_s(s = 9)$, presents the level of emissions the coalition of the nine remaining members emit if a single country deviates from the grand coalition. Hence, the remaining signatories increase their level of emissions so as to raise the deviator's damages to a level such that it nullifies the envisioned benefits from free-riding if the coalition drops to $s = 9$. In other words, the welfare of the potential deviator will be no different from its welfare as a member of the grand coalition. Recall each county's welfare function is,

$$W(e_i) = b[ae_i - \tfrac{1}{2}e_i^2] - \tfrac{c}{2}(\textstyle\sum_{i\epsilon N} e_i)^2$$

From the maximisation of the above welfare function, assuming s countries join the coalition, Diamantoudi and Sartzetakis (2006) derive the optimal level of emissions given by,

$$e_s^*(s) = a\left(1 - \frac{\gamma s n}{\psi}\right)$$

where $\psi = X^2 + \gamma s^2$ and $X = 1 + \gamma(n-s)$.

The optimal solutions require the condition $\gamma < \frac{4}{n(n-4)}$ for positive level of emissions

(Diamantoudi and Sartzetakis, 2006). The indirect welfare of each signatory is given by,

$$\omega_s(e_s^*(s)) = ba^2 \left(\frac{1}{2} - \frac{n^2\gamma}{2\psi} \right). \tag{1.7}$$

The adjusted level of emissions $\widehat{e_s}(s-1)$ is determined from the condition $\omega_{ns}(\widehat{e}_{ns}(\widehat{e_s}(s-1))) \leq \omega_s(e_s^*(s))$. The reaction function of a non-signatory at $s-1$ is given by

$$\widehat{e}_{ns}\left(\widehat{e_s}(s-1)\right) = \frac{a - \gamma(s-1)\widehat{e_s}(s-1)}{1+\gamma\left[n-(s-1)\right]}. \tag{1.8}$$

Given the above reaction function, the welfare of non-signatories at $s-1$ can be expressed as a function of $\widehat{e_s}(s-1)$. With the assumption that non-signatories will make their choice after observing the level of emissions generated by the coalition, the coalition determines the level of its emissions at $s-1$ necessary to nullify a member's deviation and deter exit. The indirect welfare of non-signatories at $s-1$ is given by

$$\omega_{ns}(s-1) = \\ ba\widehat{e}_{ns}\left(\widehat{e_s}(s-1)\right) - \tfrac{b}{2}\widehat{e}_{ns}\left(\widehat{e_s}(s-1)\right)^2 - \tfrac{c}{2}\left[(s-1)\widehat{e_s}(s-1) + t\widehat{e}_{ns}\left(\widehat{e_s}(s-1)\right)\right]^2, \tag{1.9}$$

where $t = n-(s-1)$. Substituting (1.7) and (1.9) in the condition $\omega_{ns}(\widehat{e}_{ns}(\widehat{e_s}(s-1))) \leq \omega_s(e_s^*(s))$, generates the contingency plan for the coalition. The following inequality is solved to determine the minimum level of emissions for $\widehat{e_s}\,(s-1)$ adopted by the coalition at $s-1$,

$$a\tfrac{[a-(s-1)\gamma\widehat{e}_s(s-1)]}{1+\gamma t} - \tfrac{[a-(s-1)\gamma\widehat{e}_s(s-1)]^2}{2(1+\gamma t)^2} - \tfrac{\gamma}{2}\left[(s-1)\widehat{e}_s(s-1) + t\tfrac{a-\gamma(s-1)\widehat{e}_s(s-1)}{1+\gamma t}\right]^2 \leq \\ a^2[\tfrac{1}{2} - \tfrac{n^2\gamma}{2\psi}].$$

The solution to the maximisation uses the binding constraint, $\omega_{ns}(e_{ns}(\widehat{e}_s(s-1))) = \omega_s(e_s^*(s))$.[5] The level of emissions resulting from the solution of the equality is presented in the following Proposition.

Proposition 1.1 *A group of s countries, $s \leq n$, can form a stable coalition, immune to unilateral deviations, if they can credibly commit to respond to the deviation by one of its members by emitting $\widehat{e}_{s-1} > e^*_{s-1}$, where,*

$$\widehat{e}_s(s-1) = \frac{a}{n-t}\left[\frac{n(1+\gamma t)}{\sqrt{\psi(1+\gamma)}} - t\right], \tag{1.10}$$

where $\psi = X^2 + \gamma s^2$, $X = 1 + \gamma(n-s)$, and $t = n - (s-1)$.

Proposition 1.1 presents the contingency plan for the case of a single deviator. Equation (1.10) displays the level of emissions that a coalition of s countries should commit to, if one single signatory exits the agreement. The level of emissions $\widehat{e}_{s-1}$ is the contingency plan taken by signatories at $s-1$. The contingency plan is drafted into the agreement and made enforceable to ensure no unilateral deviation is favourable to any member. The optimal solutions require $\gamma < \frac{4}{n(n-4)}$ for positive level of emissions as per Diamantoudi and Sartzetakis (2006), and the contingency plan requires the condition $s \neq 1$. Comparing the contingency plan level $\widehat{e}_s(s-1)$ to the optimal solution at $s-1$, $e_s^*(s-1) = a(1 - \frac{\gamma(s-1)n}{\psi_{s-1}})$, where $\psi_{s-1} = X^2_{s-1} + \gamma(s-1)^2$, $X_{s-1} = 1 + \gamma(n-s+1)$, the remaining coalition raises emissions further in the contingency plan to emit a higher level of emissions, that is, $\widehat{e}_s(s-1) > e_s^*(s-1)$.

[5] Diamantoudi and Sartzetakis (2006) show that the welfare levels of both signatories and non-signatories do not monotonically increase in the size of the coalition. The proofs and derivations of the model are found in Appendix 1.6 at the end of the essay.

To illustrate the results, a numerical example is presented. We assume that the total number of countries is ten ($n = 10$), and all countries approach the agreement and agree to cooperate. We consider the contingency plan of coalition members embedded into the agreement in order to discourage deviation from the grand coalition.

1.3 An Example

To facilitate direct comparison, we employ the same parameter values used in Diamantoudi and Sartzetakis (2006). We consider the following numerical example with $n = 10$, $a = 10$, $b = 6$ and $c = 0.39999$, which results in $\gamma = \frac{c}{b} = 0.066665$. The results of emissions and welfare levels are illustrated graphically. Examining the emission levels, the solid and dashed curves in Figure 1.1 illustrate the optimal emission levels of signatory e_s^* and non-signatory e_{ns}^* countries, respectively. Assuming that we start from the grand coalition, Figure 1.1 also depicts the level of emission $\widehat{e}_s(s = 9)$ to which all signatories commit to adjusting if any one of them decides to free ride. It also depicts the emission response of the single deviator $\widehat{e}_{ns}(s = 9)$ when the remaining coalition members adjust their emissions to $\widehat{e}_s(s = 9)$.

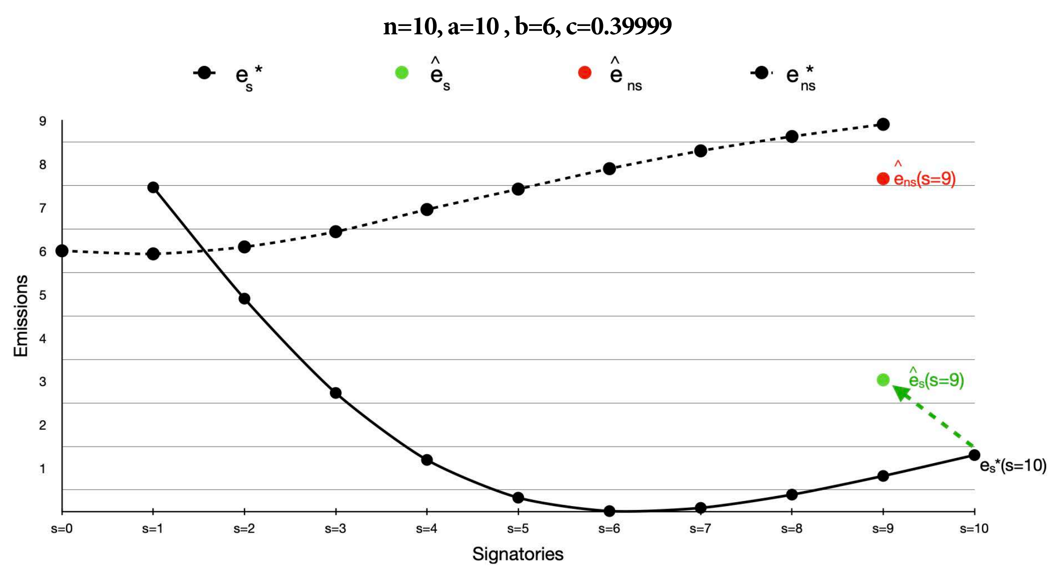

Figure 1.1 Optimal emission levels and contingency plan at s=9

It is evident that the increase of the remaining members' emissions to $\widehat{e}_s(9) > e_s^*(9)$ induces a decrease of the deviator's emissions $\widehat{e}_{ns}(9) < e_{ns}^*(9)$, which reduces its benefits while its damages are increased due to the fact that $\widehat{E}(9) > E^*(9)$.[6] The combination of these two effects renders deviation by a single member non-profitable. Notice that since $\widehat{E}(9) > E^*(9)$, damages to the remaining coalition members are also higher, that is why $\widehat{e}_s(9)$ is not welfare maximising and needs to be enforced.

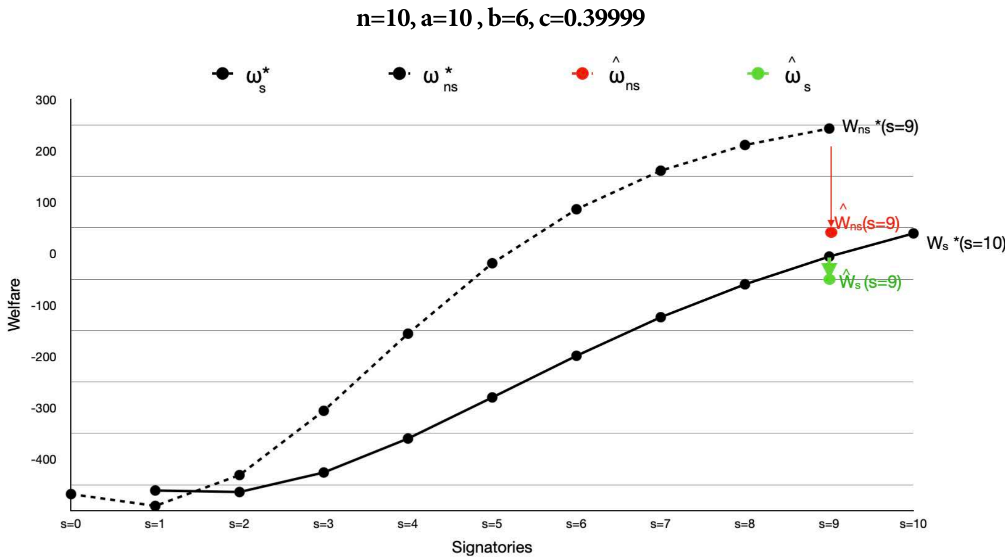

Figure 1.2 Welfare levels given optimal level of emissions and contingency plan at s=9

Figure 1.2 illustrates the indirect welfare of signatories and non-signatories facilitating welfare comparisons. The solid (dashed) curve in Figure 1.2 illustrates the welfare level of signatories ω_s^*(non-signatories ω_{ns}^*). The figure also presents the levels of welfare that correspond to $(\widehat{e}_s(9), \widehat{e}_{ns}(9))$ for signatories $\widehat{\omega}_s(9)$ and non-signatories $\widehat{\omega}_{ns}(9)$, respectively. At the grand coalition, $s = 10$, the optimal level of emissions is $e_{s=10}^* = 1.3044$ from the joint profit maximisation. In case a signatory contemplates

[6] $\widehat{E}(9) = 9\widehat{e}_s(9) + \widehat{e}_{ns}(9) = 9(3.03) + (7.66) = 34.93$ whereas $E^*(9) = 9e_s^*(9) + e_{ns}^*(9) = 9(0.82) + (8.91) = 16.29$.

defecting, the adjustment of emissions by signatories required to eliminate any incentive to deviate is $\widehat{e}_s(9)$ leading to $\widehat{e}_{ns}(9)$ which yield $\widehat{\omega}_s(9)$ and $\widehat{\omega}_{ns}(9)$. The coalition at $s-1=9$ would not emit at the level resulting from the joint welfare maximisation $e_9^*=0.8226$, but it will choose a much higher level of emissions $\widehat{e}_s(9)=3.034$ sufficient to offset any benefits from exiting. This is evident since, by construction, $\omega_s^*(10)=\widehat{\omega}_{ns}(9)$.

It is clear from Figure 1.2 that $\widehat{e}_s(9)$ is not the level of emissions that maximise the nine members' aggregate welfare, that is, $\omega_s^*(9)>\widehat{\omega}_s(9)$. The coalition members commit to an action that will result in lower welfare in order to immunise their agreement from deviation. For the contingency plan to be credible, it requires an enforcement mechanism. One may argue that credibility can come from repeated behaviour; countries make threats all the time but are not necessarily enforced. One method in which countries can encourage compliance and cooperation is through the tightening of an environmental decision to a trade agreement. Members that will not execute the terms could be penalised through trade restrictions.

The next section allows a coalition to build a contingency plan to prevent multiple countries from potentially exiting the agreement simultaneously. An example with two countries or three contemplating simultaneous deviations is provided to illustrate the contingency values calculated by the coalition.

1.4 Agreements with Multiple Deviations

This section studies the case of multiple simultaneous deviations from any coalition of size s. In order to increase the robustness of the agreement, the coalition builds a contingency plan to eliminate incentives for multiple countries exiting the coalition. Let j denote the number of coalition members contemplating exit. With $j < s$, the signatories' welfare maximisation problem is formulated as follows,

$$\max_{\widehat{e}_s(s-j)} sW_s(s) \tag{1.11}$$

$$\textit{subject to} \qquad \omega_{ns}(e_{ns}(\widehat{e}_s(s-j))) \leq \omega_s(e_s^*(s)).$$

The contingency plan provides the emission level of coalition members when a coalition is faced with j potential signatories simultaneously contemplating exit. Recall, assuming s countries join the coalition, Diamantoudi and Sartzetakis (2006) derive $e_s^*(s) = a(1 - \frac{\gamma sn}{\psi})$, and the indirect welfare given in eq (1.7). The adjusted level of emissions $\widehat{e}_s(s-j)$ is determined from the condition $\omega_{ns}(\widehat{e}_{ns}(\widehat{e}_s(s-j))) \leq \omega_s(e_s^*(s))$. The reaction function of non-signatories at $s-j$ is given by

$$\widehat{e}_{ns}\left(\widehat{e}_s(s-j)\right) = \frac{a - \gamma(s-j)\widehat{e}_s(s-j)}{1 + \gamma\left[n - (s-j)\right]}, \tag{1.12}$$

which is a generalisation of the reaction in eq. (1.8).

The welfare of non-signatories at $s-j$ can be expressed as a function of $\widehat{e}_s(s-j)$. Given that non-signatories make their choice after observing the choice of the coalition, the coalition determines its level of emissions at $s-j$ necessary to nullify j members' deviation and deter exit. The welfare of non-signatories at $s-j$ is given by

$$\omega_{ns}(s-j) =$$

$$ba\widehat{e}_{ns}\left(\widehat{e}_s(s-j)\right) - \tfrac{b}{2}\widehat{e}_{ns}\left(\widehat{e}_s(s-j)\right)^2 - \tfrac{c}{2}\left[(s-j)\widehat{e}_s(s-j) + t_j\widehat{e}_{ns}\left(\widehat{e}_s(s-j)\right)\right]^2, \quad (1.13)$$

where $t_j = n - (s-j)$. Substituting equations (1.7) and (1.13) in the condition $\omega_{ns}(\widehat{e}_{ns}(\widehat{e}_s(s-j))) \leq \omega_s(e^*_s(s))$, yields the contingency plan for the coalition. Substitution yields the following inequality which is solved to determine the level of emissions adopted by the coalition at $s-j$,

$$a\tfrac{[a-(s-j)\gamma\widehat{e}_s(s-j)]}{1+\gamma t_j} - \tfrac{[a-(s-j)\gamma\widehat{e}_s(s-j)]^2}{2(1+\gamma t_j)^2} - \tfrac{\gamma}{2}\left[(s-j)\widehat{e_s} + t_j\tfrac{a-\gamma(s-j)\widehat{e}_s(s-j)}{1+\gamma t_j}\right]^2 \leq a^2[\tfrac{1}{2} - \tfrac{n^2\gamma}{2\psi}].$$

The next Proposition summarises the contingency plan for the remaining members of the coalition when faced with a potential simultaneous deviation by multiple countries.

Proposition 1.2 *A group of $s \leq n$ countries can form a stable coalition immune to deviations, if they can credibly commit to responding to simultaneous deviations by a number $j < s$ of its members by emitting $\widehat{e}_{s-j} > e^*_{s-j}$, where,*

$$\widehat{e}_s\left(s-j\right) = \frac{a}{n-t_j}\left[\frac{n(1+\gamma t_j)}{\sqrt{\psi(1+\gamma)}} - t_j\right], \quad (1.14)$$

where $\psi = X^2 + \gamma s^2$, $X = 1 + \gamma(n-s)$, and $t_j = n - (s-j)$.

Equation (1.14) specifies the contingency plan taken by the $s-j$ remaining members in case j members of the initial coalition of size s contemplate deviating. The contingency plan computed by coalition members drafts the emission adjustment necessary to eliminate incentives of the coalition members contemplating exiting. The contingency plan is drafted into the agreement and made enforceable to ensure that no simultaneous deviation by multiple countries is favourable. The next section presents three examples of developing a contingency plan in case two, three or four countries simultaneously deviate from the grand coalition.

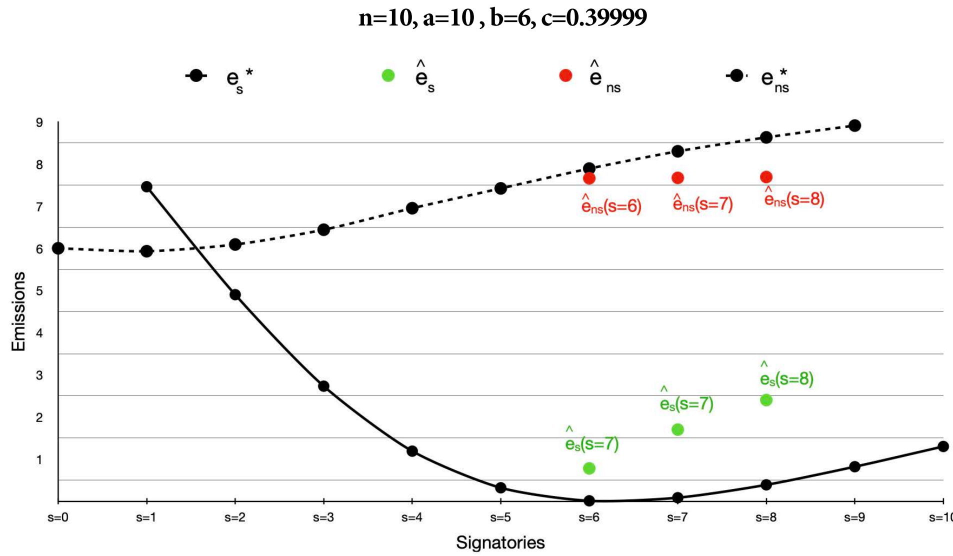

Figure 1.3 Optimal level of emissions and contingency plans at s=8, s=7, and s=6

Using the same parameter values as in the previous section, the solid and dashed lines in Figure 1.3 depict the optimal emission levels of signatory e_s^* and non-signatory e_{ns}^* countries, respectively, as in Figure 1.1. If ten countries sign an agreement, the optimal level of emissions is given by $e_s^*(10) = 1.3044$ as before. The coalition of ten builds a contingency plan for the case of multiple countries deviating. Examining the cases where two, three, or four of the grand coalition's members contemplate exiting simultaneously, Figure 1.3 also depicts the levels of emissions $\widehat{e}_s(8)$, $\widehat{e}_s(7)$, and $\widehat{e}_s(6)$ to which all signatories commit to adjusting to if two, three or four countries potentially contemplate exit respectively. The figure also depicts the emission response of non-signatories at $\widehat{e}_{ns}(8)$, $\widehat{e}_{ns}(7)$, and $\widehat{e}_{ns}(6)$ when the remaining coalition members commit to adjusting to $\widehat{e}_s(8)$, $\widehat{e}_s(7)$, and $\widehat{e}_s(6)$ respectively.

Given the adjustment in emissions by the remaining coalition when two countries contemplate leaving the agreement, the coalition of eight would not emit $e_s^*(8) = 0.392$ but would instead increase their emissions to $\widehat{e}_s(8) = 2.454 > e_s^*(8) = 0.392$. Similarly, in the case that three countries contemplate exit, the remaining coalition's members, a coalition of seven, raise their emission level to $\widehat{e}_s(7) = 1.709 > e_s^*(7) = 0.085$. With four countries contemplating exit, the remaining coalition's members, a coalition of six, emit $\widehat{e}_s(6) = 0.7157 > e_s^*(6) = 0.011$.

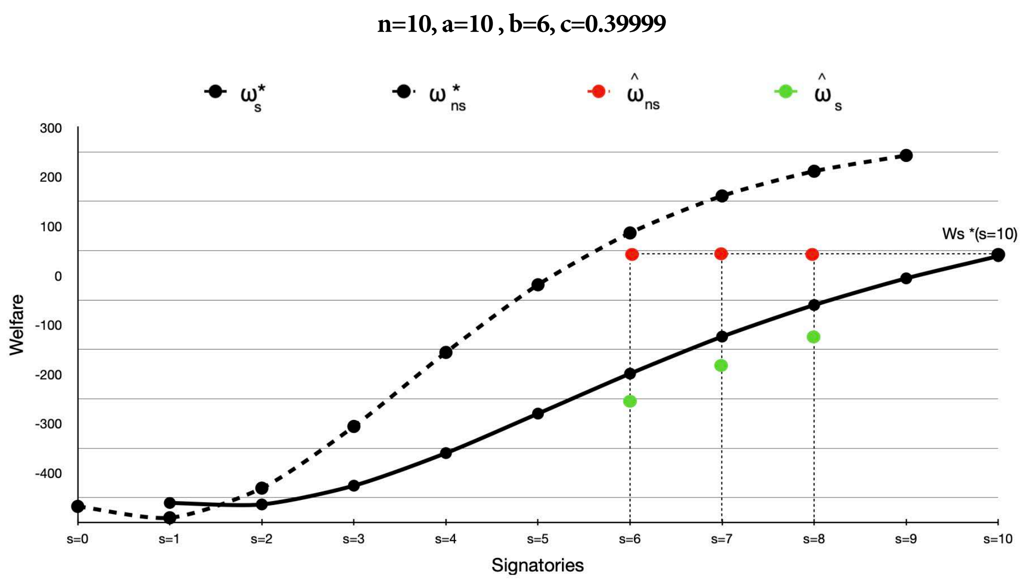

Figure 1.4 Welfare levels given optimal level of emissions and contingency plans for s=8, s=7 and s=6

The two curves in Figure 1.4 illustrate the level of indirect welfare of signatories ω_s^* and non-signatories ω_{ns}^*, same as in Figure 1.2 above. It presents the indirect levels of welfare that correspond to $(\widehat{e}_s(8), \widehat{e}_{ns}(8))$, for signatories $\widehat{\omega}_s(8)$ and non-signatories $\widehat{\omega}_{ns}(8)$, when two countries contemplate exiting the agreement simultaneously. It also presents the levels of welfare that correspond to $(\widehat{e}_s(7), \widehat{e}_{ns}(7))$, for signatories $\widehat{\omega}_s(7)$ and non-signatories $\widehat{\omega}_{ns}(7)$, when three countries contemplate exiting the agreement simultaneously, and the levels of welfare that correspond to $(\widehat{e}_s(6), \widehat{e}_{ns}(6))$, when four countries simultaneously contemplate exit. The welfare of a non-signatory at $s = 8$ by construction will be the same as that obtained when it is a member of the grand coalition, $\omega_s^*(10) = \widehat{\omega}_{ns}(8) = 39.1$. The remaining members of the coalition commit to raise their emissions to $\widehat{e}_s(8)$ when faced with a potential simultaneous deviation by two countries. Figure 1.4 shows that it is only necessary to develop contingency plans for multiple deviations until non-signatory welfare levels drop lower than full cooperation.

When three countries contemplate exit, the remaining coalition's members raise their emission level to $\widehat{e}_s(7) = 1.709$. The commitment to this level of emissions is sufficient to nullify any additional welfare gains that the three deviators would expect from exiting the grand coalition, i.e., they are receiving $\widehat{\omega}_{ns}(7) = 39.1$ instead of $\omega_{ns}^*(7) = 161$. Given the adjustment taken by the remaining coalition, this disincentivizes contemplating signatories from exiting the agreement. With a similar intuition, the coalition builds a contingency plan for four members simultaneously exiting the agreement.

1.5 Conclusion

The framework developed in this section assumes that a number of countries agree to cooperate in reducing their emissions and draft an agreement that contains a contingency plan for all coalition sizes. In the benchmark model by Diamantoudi and Sartzetakis (2006), a deviator leaves a coalition because it is profitable given that no other country leaves the agreement. The reality is that the remaining coalition might not act in this way, there are a number of inferences that can be made upon a deviation. The framework for contingencies takes the analysis a step further by specifying the assumption that members can commit to a contingency plan that discourages deviations. The assumption that coalition members will maximise their joint welfare following a potential deviation is loosened. Assuming one-step deviations, i.e., simultaneous deviations (either unilateral or multiple), in this analysis, we allow signatories to build a contingency plan prescribing the emission levels that its members should adopt when one or more of its members contemplate exiting the agreement. The framework examines that when countries form any size of coalition, they also agree to respond to any deviations by adjusting their level of emissions to a level that will render deviations non-profitable, i.e., increasing rather than decreasing their emissions so as to increase the deviator's damages to a level such that it nullifies the envisioned benefits from free-riding. As mentioned earlier, the contingency plan's credibility requires some enforcement mechanism.

The contingency plans in this section do not take into account possible subsequent deviation by signatories. In the next essay, the assumption that no other country will follow when a country contemplates exiting the agreement is loosened. It extends the methodology of contingencies to build a sequence of emission targets to study the formation of IEAs in a more foresighted approach that takes into account possible subsequent deviation by members.

1.6 Appendices

The derivations of the model and proofs of propositions are presented in the following section.

Appendix A. Calculations

Calculation 1. This presents the contingency plan maximisation in the case of a unilateral deviation by a signatory.

Recall each county's welfare function is,

$$W(e_i) = b[ae_i - \tfrac{1}{2}e_i^2] - \tfrac{c}{2}(\textstyle\sum_{i\epsilon N} e_i)^2$$

From the maximisation of the above welfare function, assuming s countries join the coalition, Diamantoudi and Sartzetakis (2006) derive $e_s^*(s) = a(1 - \frac{\gamma sn}{\psi})$, and then the indirect welfare is given by,

$$\omega_s\left(e_s^*(s)\right) = ba^2\left(\frac{1}{2} - \frac{n^2\gamma}{2\psi}\right) \tag{1.7}$$

where $\psi = X^2 + \gamma s^2$ and $X = 1 + \gamma(n - s)$. The adjusted level of emissions $\widehat{e}_s\left(s-1\right)$ is determined from the condition $\omega_{ns}(\widehat{e}_{ns}(\widehat{e}_s(s-1))) \leq \omega_s(e_s^*(s))$. The reaction function of non-signatories at $s-1$ is given by

$$\widehat{e}_{ns}(s-1) = \frac{a - \gamma(s-1)\widehat{e}_s(s-1)}{1 + \gamma\left[n - (s-1)\right]} \tag{1.8}$$

Given the reaction function of non-signatories in eq (1.8), the maximisation problem can be written as

$$\max_{\widehat{e}_s(s-1)} sW_s$$

$$\text{st } \omega_{ns}(\widehat{e}_{ns}(\widehat{e}_s(s-1))) \leq \omega_s(e_s^*(s))$$

We can write the Lagrangian for this problem as

$$\mathcal{L} = sW_s + \lambda \begin{bmatrix} ba^2(\frac{1}{2} - \frac{n^2\gamma}{2\psi}) - ba\frac{[a-(s-1)\gamma\widehat{e}_s(s-1)]}{1+\gamma t} - \frac{[a-(s-1)\gamma\widehat{e}_s(s-1)]^2}{2(1+\gamma t)^2} + \\ \frac{c}{2}[(s-1)\widehat{e}_s(s-1) + (n-s+1)\frac{a-\gamma(s-1)\widehat{e}_s(s-1)}{1+\gamma t}]^2] \end{bmatrix}$$

Here λ is the Lagrange multiplier on the contingency constraint in case of a unilateral deviation. Substituting eq. (1.7), the Lagrangian can be written as

$$\mathcal{L} = s\left[(bae_s - \frac{b}{2}e_s^2) - \frac{c}{2}(se_s + (n-s)e_{ns})^2\right] +$$

$$\lambda\left[ba^2(\frac{1}{2} - \frac{n^2\gamma}{2\psi}) - ba\frac{[a-(s-1)\gamma\widehat{e}_s(s-1)]}{1+\gamma t} - \frac{[a-(s-1)\gamma\widehat{e}_s(s-1)]^2}{2(1+\gamma t)^2} + \frac{c}{2}[(s-1)\widehat{e}_s(s-1) + (t)\frac{a-\gamma(s-1)\widehat{e}_s(s-1)}{1+\gamma t}]^2\right]$$

When we differentiate with respect to $\widehat{e}_s(s-1)$ and λ, we have the following first-order conditions:

$$\frac{\partial\mathcal{L}}{\partial\widehat{e}_{s-1}} = \frac{\lambda(\gamma^2+\gamma)(s-1)(a(n-s+1)+\widehat{e}_s(s-1))}{(\gamma+n\gamma-s\gamma+1)^2} = 0, \text{ rejected.}$$

$$\frac{\partial\mathcal{L}}{\partial\lambda} = a^2[\tfrac{1}{2}-\tfrac{n^2\gamma}{2\psi}]-a\tfrac{[a-(s-1)\gamma\widehat{e}_s(s-1)]}{1+\gamma t}-\tfrac{[a-(s-1)\gamma\widehat{e}_s(s-1)]^2}{2(1+\gamma t)^2}-\tfrac{\gamma}{2}\left[(s-1)\widehat{e}_s(s-1)+t\tfrac{a-\gamma(s-1)\widehat{e}_s(s-1)}{1+\gamma t}\right]^2 = 0.$$

The first-order condition with respect to λ yields the contingency plan constraint in the case of a unilateral deviation.

Calculation 2. This presents the contingency plan maximisation in the case of simultaneous deviation by multiple signatories. The adjusted level of emissions $\widehat{e_s}(s-j)$ is determined from the condition $\omega_{ns}(\widehat{e}_{ns}(\widehat{e}_s(s-j))) \leq \omega_s(e_s^*(s))$. The reaction function of non-signatories at $s-j$ is given by

$$\widehat{e}_{ns}(s-j) = \frac{a-\gamma(s-j)\widehat{e}_s(s-j)}{1+\gamma\left[n-(s-j)\right]} \tag{1.8}$$

Given the reaction function of non-signatories in eq (1.8), the maximisation problem can be written as

$$\max_{\widehat{e}_s(s-j)} sW_s$$

$$\text{st } \omega_{ns}(\widehat{e}_{ns}(\widehat{e}_s(s-j))) \leq \omega_s(e_s^*(s))$$

We can write the Lagrangian for this problem as

$$\mathcal{L} = sW_s + \lambda \left[\begin{array}{c} ba^2(\frac{1}{2} - \frac{n^2\gamma}{2\psi}) - ba\frac{[a-(s-j)\gamma\widehat{e}_s(s-j)]}{1+\gamma t_j} - \frac{[a-(s-j)\gamma\widehat{e}_s(s-j)]^2}{2(1+\gamma t_j)^2} + \\ \frac{c}{2}[(s-j)\widehat{e}_s(s-j) + (t_j)\frac{a-\gamma(s-j)\widehat{e}_s(s-j)}{1+\gamma t_j}]^2] \end{array} \right]$$

Here λ is the Lagrange multiplier on the contingency constraint in case of a simultaneous multiple deviation by j countries. Substituting eq. (1.7), the Lagrangian can be written as

$$\mathcal{L} = s\left[(bae_s - \frac{b}{2}e_s^2) - \frac{c}{2}(se_s + (n-s)e_{ns})^2\right] +$$
$$\lambda \left[\begin{array}{c} ba(\frac{1}{2} - \frac{n^2\gamma}{2\psi}) - ba\frac{[a-(s-j)\gamma\widehat{e}_s(s-j)]}{1+\gamma t_j} - \frac{[a-(s-j)\gamma\widehat{e}_s(s-j)]^2}{2(1+\gamma t_j)^2} + \\ \frac{c}{2}[(s-j)\widehat{e}_s(s-j) + (t_j)\frac{a-(s-j)\gamma\widehat{e}_s(s-j)}{1+\gamma t_j}]^2] \end{array} \right]$$

When we differentiate with respect to $e_s(s-j)$ and λ, we have the following first-order conditions:

$$\frac{\partial \mathcal{L}}{\partial \widehat{e}_{s-j}} = \frac{\lambda(\gamma^2+\gamma)(s-j)(a(n-s+j)+\widehat{e}_s(s-j))}{(j\gamma+n\gamma-s\gamma+1)^2} = 0, \text{ rejected.}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = a^2[\frac{1}{2} - \frac{n^2\gamma}{2\psi}] - a\frac{[a-(s-j)\gamma\widehat{e}_s(s-j)]}{1+\gamma t_j} - \frac{[a-(s-j)\gamma\widehat{e}_s(s-j)]^2}{2(1+\gamma t_j)^2} - \frac{\gamma}{2}\left[(s-j)\widehat{e}_s\,(s-j) + t_j\frac{a-\gamma(s-j)\widehat{e}_s(s-j)}{1+\gamma t_j}\right]^2 = 0.$$

The first-order condition with respect to λ yields the contingency plan constraint in the case of a simultaneous deviation by multiple signatories.

Appendix B. Proofs

Proof. Proposition 1.1 Case of a single deviation

The signatories' indirect welfare when choosing $e_s^*(s)$ is,

$$\omega_s^*(s) = ba^2[\frac{1}{2} - \frac{n^2\gamma}{2\psi}]$$

where $\psi = X^2 + \gamma s^2$, $X = 1 + \gamma(n - s)$.

Reaction function of non-signatories at $s - 1$:

$$e_{ns}(s-1) = \frac{a-\gamma(s-1)(e_s(s-1))}{1+\gamma(n-(s-1))}$$

Welfare of non-signatories at $s - 1$:

$$\widehat{\omega}_{ns}(s-1) =$$
$$ba\frac{[a-(s-1)\gamma e_s(s-1)]}{1+\gamma(n-(s-1))} - b\frac{[a-(s-1)\gamma e_s(s-1)]^2}{2[1+\gamma(n-(s-1))]^2} - \frac{c}{2}[(s-1)e_s(s-1) + (n-(s-1))\frac{[a-\gamma(s-1)e_s(s-1)]}{1+\gamma(n-(s-1))}]^2$$

Plugging the above defined welfare levels in the constraint: $\omega_{ns}(s-1) \leq \omega_s(s)$ yields inequality (A.1):

$$a^2[\frac{1}{2} - \frac{n^2\gamma}{2\psi}] \geq$$
$$a\frac{[a-(s-1)\gamma\widehat{e}_s(s-1)]}{1+\gamma(n-s+1)} - \frac{[a-(s-1)\gamma\widehat{e}_s(s-1)]^2}{2[1+\gamma(n-s+1)]^2} - \frac{\gamma}{2}[(s-1)\widehat{e}_s(s-1) + (n-s+1)\frac{[a-\gamma(s-1)\widehat{e}_s(s-1)]}{1+\gamma(n-s+1)}]^2$$

Examining the right hand side, define $F = (s-1)\widehat{e}_s(s-1)$ and $t = n - (s-1)$

$$\frac{a(a-\gamma F)}{1+\gamma t} - \frac{(a-\gamma F)^2}{2(1+\gamma t)^2} - \frac{\gamma}{2}[F + \frac{t(a-\gamma F)}{1+\gamma t}]^2$$

$$= \frac{a^2}{1+\gamma t} - \frac{a\gamma F}{1+\gamma t} - \frac{a^2}{2(1+\gamma t)^2} + \frac{a\gamma F}{(1+\gamma t)^2} - \frac{\gamma^2 F^2}{2(1+\gamma t)^2} - \frac{\gamma}{2}F^2 - \frac{\gamma F t(a-\gamma F)}{1+\gamma t} - \frac{\gamma t^2(a-\gamma F)^2}{2(1+\gamma t)^2}$$

$$= \frac{a^2}{1+\gamma t} - \frac{a\gamma F}{1+\gamma t} - \frac{a^2}{2(1+\gamma t)^2} + \frac{a\gamma F}{(1+\gamma t)^2} - \frac{\gamma^2 F^2}{2(1+\gamma t)^2} - \frac{\gamma}{2}F^2 - \frac{\gamma F t a}{1+\gamma t} + \frac{\gamma^2 F^2 t}{1+\gamma t} - \frac{\gamma t^2 a^2}{2(1+\gamma t)^2} + \frac{a t^2\gamma^2 F}{(1+\gamma t)^2} - \frac{\gamma^2 t^2 F^2}{2(1+\gamma t)^2}$$

$= F^2[-\frac{\gamma^2}{2(1+\gamma t)^2} - \frac{\gamma}{2} + \frac{\gamma^2 t}{1+\gamma t} - \frac{\gamma^3 t^2}{2(1+\gamma t)^2}] + F[-\frac{a\gamma}{1+\gamma t} + \frac{a\gamma}{(1+\gamma t)^2} - \frac{\gamma ta}{1+\gamma t} + \frac{at^2\gamma 2}{(1+\gamma t)^2}] + \frac{a^2}{1+\gamma t} - \frac{\gamma t^2 a^2}{2(1+\gamma t)^2} - \frac{a^2}{2(1+\gamma t)^2}$

Therefore, the inequality (A.1) can be reduced to:

$$F^2[-\frac{\gamma^2}{2(1+\gamma t)^2} - \frac{\gamma}{2} + \frac{\gamma^2 t}{1+\gamma t} - \frac{\gamma^3 t^2}{2(1+\gamma t)^2}] - F[\frac{a\gamma t(1+\gamma)}{(1+\gamma t)^2}] - \frac{a^2\gamma t^2(1+\gamma)}{2(1+\gamma t)^2} + \frac{a^2 n^2 \gamma}{2\psi} \leq 0,$$

Solving for the quadratic roots of F:

Solutions: $-\frac{\left(at+an\sqrt{\frac{(\gamma+1)}{\psi}}+at\gamma+ant\gamma\sqrt{\frac{(\gamma+1)}{\psi}}\right)}{\gamma+1}, -\frac{\left(at-an\sqrt{\frac{(\gamma+1)}{\psi}}+at\gamma-ant\gamma\sqrt{\frac{(\gamma+1)}{\psi}}\right)}{\gamma+1}$

Root 1. $F \leq -\frac{\left(at+an\sqrt{\frac{(\gamma+1)}{\psi}}+at\gamma+ant\gamma\sqrt{\frac{(\gamma+1)}{\psi}}\right)}{\gamma+1} = -\frac{at}{\gamma+1} - \frac{an\sqrt{\frac{\gamma}{\psi}+\frac{1}{\psi}}}{\gamma+1} - \frac{at\gamma}{\gamma+1} - \frac{ant\gamma\sqrt{\frac{\gamma}{\psi}+\frac{1}{\psi}}}{\gamma+1}$ rejected

Root 2. $F \geq -\frac{\left(at-an\sqrt{\frac{(\gamma+1)}{\psi}}+at\gamma-ant\gamma\sqrt{\frac{(\gamma+1)}{\psi}}\right)}{\gamma+1} = \frac{an\sqrt{\frac{\gamma}{\psi}+\frac{1}{\psi}}}{\gamma+1} - \frac{at}{\gamma+1} - \frac{at\gamma}{\gamma+1} + \frac{ant\gamma\sqrt{\frac{\gamma}{\psi}+\frac{1}{\psi}}}{\gamma+1} = \frac{a\left(n\sqrt{\frac{\gamma}{\psi}+\frac{1}{\psi}}-t-t\gamma+nt\gamma\sqrt{\frac{\gamma}{\psi}+\frac{1}{\psi}}\right)}{\gamma+1}$

Root 2 can be rewritten as:

$F \geq \frac{-t\psi(1+\gamma)+n(1+\gamma t)\sqrt{\psi(1+\gamma)}}{\frac{\psi}{a}(1+\gamma)}$

$F = (s-1)\widehat{e}_s(s-1)$

Thus, to make a coalition of s countries immune to unilateral deviations, the contingency plan requires the adjustment in signatories emission levels at $s-1$ to be:

$\widehat{e}_s(s-1) = \frac{-t\psi(1+\gamma)+n(1+\gamma t)(\sqrt{\psi(1+\gamma)})}{\frac{\psi}{a}(1+\gamma)(s-1)}$

which can be rewritten as,

$\widehat{e}_s(s-1) = \frac{a}{n-t}\left[\frac{n(1+\gamma t)}{\sqrt{\psi(1+\gamma)}} - t\right]$. ■

Proof. Proposition 1.2 Case of multiple deviation

The signatories' indirect welfare when choosing $e_s^*(s)$ is

$$\omega_s^*(s) = ba^2[\tfrac{1}{2} - \tfrac{n^2\gamma}{2\psi}]$$

where $\psi = X^2 + \gamma s^2$, $X = 1 + \gamma(n-s)$.

Reaction function of non-signatories at *s*-*j*:

$$e_{ns}(s-j) = \tfrac{a-\gamma(s-j)(e_s(s-j))}{1+\gamma(n-(s-j))}$$

Welfare of non-signatories at *s*-*j*:

$$\widehat{\omega}_{ns}(s-j) = ba\tfrac{[a-(s-j)\gamma\widehat{e}_s(s-j)]}{1+\gamma(n-(s-j))} - b\tfrac{[a-(s-j)\gamma\widehat{e}_s(s-j)]^2}{2[1+\gamma(n-(s-j))]^2} - \tfrac{c}{2}[(s-j)\widehat{e}_s(s-j) + (n-s+j)\tfrac{[a-\gamma(s-j)\widehat{e}_s(s-j)]}{1+\gamma(n-(s-j))}]^2$$

Plugging the above defined welfare levels in the constraint: $\omega_{ns}(s-j) \leq \omega_s(s)$ yields inequality (A.2):

$$a^2[\tfrac{1}{2} - \tfrac{n^2\gamma}{2\psi}] \geq a\tfrac{[a-(s-j)\gamma\widehat{e}_s(s-j)]}{1+\gamma(n-s+j)} - \tfrac{[a-(s-j)\gamma\widehat{e}_s(s-j)]^2}{2[1+\gamma(n-s+j)]^2} - \tfrac{\gamma}{2}[(s-j)\widehat{e}_s(s-j) + (n-s+j)\tfrac{[a-\gamma(s-j)\widehat{e}_s(s-j)]}{1+\gamma(n-s+j)}]^2$$

Examining the right hand side, define $F = (s-j)\widehat{e}_s(s-j)$ and $t_j = n-(s-j)$

Thus, the inequality (A.2) can be reduced to:

$$F^2[-\frac{\gamma^2}{2(1+\gamma t_j)^2} - \frac{\gamma}{2} + \frac{\gamma^2 t_j}{1+\gamma t_j} - \frac{\gamma^3 t_j^2}{2(1+\gamma t_j)^2}] - F[[\frac{a\gamma t_j(1+\gamma)}{(1+\gamma t_j)^2}]] - \frac{a^2\gamma t_j^2(1+\gamma)}{2(1+\gamma t_j)^2} + \frac{a^2 n^2 \gamma}{2\psi} \leq 0,$$

Solving for the quadratic roots of F:

Solutions: $\frac{\left(at_j + an\sqrt{\frac{(\gamma+1)}{\psi}} + at_j\gamma + ant_j\gamma\sqrt{\frac{(\gamma+1)}{\psi}}\right)}{\gamma+1}, -\frac{\left(at_j - an\sqrt{\frac{(\gamma+1)}{\psi}} + at_j\gamma - ant_j\gamma\sqrt{\frac{(\gamma+1)}{\psi}}\right)}{\gamma+1}$

Root 1. $F \leq -\frac{\left(at_j + an\sqrt{\frac{(\gamma+1)}{\psi}} + at_j\gamma + ant_j\gamma\sqrt{\frac{(\gamma+1)}{\psi}}\right)}{\gamma+1} = -\frac{at_j}{\gamma+1} - \frac{an\sqrt{\frac{\gamma}{\psi}+\frac{1}{\psi}}}{\gamma+1} - \frac{at_j\gamma}{\gamma+1} - \frac{ant_j\gamma\sqrt{\frac{\gamma}{\psi}+\frac{1}{\psi}}}{\gamma+1}$ rejected

Root 2. $F \geq -\frac{\left(at_j - an\sqrt{\frac{(\gamma+1)}{\psi}} + at_j\gamma - ant_j\gamma\sqrt{\frac{(\gamma+1)}{\psi}}\right)}{\gamma+1} = \frac{an\sqrt{\frac{\gamma}{\psi}+\frac{1}{\psi}}}{\gamma+1} - \frac{at_j}{\gamma+1} - \frac{at_j\gamma}{\gamma+1} + \frac{ant_j\gamma\sqrt{\frac{\gamma}{\psi}+\frac{1}{\psi}}}{\gamma+1}$

Root 2 can be rewritten as:

$$F \geq \frac{-t_j\psi(1+\gamma) + n(1+\gamma t_j)\sqrt{\psi(1+\gamma)}}{\frac{\psi}{a}(1+\gamma)}$$

$$F = (s-j)\widehat{e}_s(s-j)$$

Thus, to make a coalition of s countries immune to simultaneous deviation by j countries, the contingency plan requires the adjustment in signatories emission levels at $s-j$ to be

$$\widehat{e}_s(s-j) = \frac{a}{n-t_j}\left(\frac{n(1+\gamma t_j)}{\sqrt{\psi(1+\gamma)}} - t_j\right). \blacksquare$$

ESSAY TWO

INTERNATIONAL ENVIRONMENTAL AGREEMENTS: CONTINGENCY PLANS UNDER FORESIGHT

2.1 Introduction

The present essay extends the work of the first, and allows an agreement to be signed by members to study contingency plans of coalition members in a more foresighted approach. Chwe (1994) takes foresight into account to define the largest consistent set. His definition applies to situations in which coalitions freely form and are fully farsighted, in that a coalition considers the possibility that once it acts, another coalition might react and so on (Chwe, 1994). Eyckmans (2001) explores foresight when studying the coalitional stability of the Kyoto Protocol on reducing greenhouse gas emissions. His analysis assumes that potential deviators are farsighted and take into account possible subsequent deviations by the remaining players. He argues that conventional myopic stability analysis suggests that several signatories of the Kyoto Protocol would have a profitable free-riding strategy and introducing foresight strongly restricts the number of credible free-riding strategies (Eyckmans, 2001).

Another concept in the game theory literature is the notion of the Coalition-Proof Nash equilibrium introduced by Bernheim, Peleg and Whinston (1987). It is designed to capture the notion of an efficient self-enforcing agreement for environments with non-binding pre-play communication (Bernheim et al., 1987). An agreement is coalition-proof if and only if it is Pareto efficient within the class of self-enforcing agreements (Bernheim et al., 1987). In turn, an agreement is self-enforcing if and only if no subset coalition of countries, taking the actions of other countries as fixed, can agree to deviate in a way that makes all of its countries better off. In contrast to the strong equilibrium concept, they do not entertain all possible deviations by such coalitions (Bernheim et al., 1987). Diamantoudi and Sartzetakis (2018) endow countries with foresight, i.e., they endogenize the reaction of the coalition's members to a deviation by one member. They assume that when a country contemplates withdrawing or joining the agreement, it takes into account the reactions of other countries given its own action (Diamantoudi and Sartzetakis, 2018). They identify conditions under which there always exists a unique set of farsighted stable IEAs, and find that new farsighted IEAs can be much larger but are not necessarily always Pareto efficient (Diamantoudi and Sartzetakis, 2018).

In non-cooperative coalition formation models, most of the literature follows the embedded assumption that the coalition members determine their cooperative level of emissions by maximising their joint welfare. Few studies in the literature have argued that the coalition's assumption of joint profit maximisation can eliminate other noteworthy equilibriums. In a study on stable cartels, Mao (2018) highlights through an example that this assumption is problematic because it imposes some unnecessary restrictions on cartel members' actions. A simple open membership cartel formation model is presented where the agreement of maximising joint profit will lead to a stable cartel in which all members are willing to adopt a different agreement (Mao, 2018). With this observation, the same intuition can be applied to coalition formation models in international environmental agreements to study other equilibriums.

The present analysis captures strategies for all coalition sizes in the case of subsequent unilateral deviations. The assumption that no other country will follow when a country contemplates exiting the agreement is loosened. It takes into account all possible and subsequent unilateral deviations by members of a coalition, and finds a sequence of contingency plans for all coalition sizes under homogenous agents and quadratic functional forms. Given any number of signatories in the agreement, the sequence of contingency plans for coalition members is unfolded from full cooperation, where the contingency constraint internally stabilises each coalition size by construction.

The analysis starts with full cooperation in which all countries approach the negotiation table and build a contingency plan accounting for a unilateral deviation from the grand coalition, which nullifies any gains to the deviator. In other words, the contemplating deviator is indifferent between being in or out of the agreement. This work has been developed independently and runs parallel to Masoudi (2022). Masoudi (2022) starts with full cooperation and argues that in the face of a potential unilateral deviation, the remaining coalition increases their emissions so that the welfare of the defector is no different from being in the grand coalition. Such a treaty can remove the free-riding problem and the grand coalition becomes self-enforcing. However, it makes all partial coalitions weakly stable (Masoudi, 2022). Under the Stackelberg assumption, with homogenous agents and quadratic functional forms, Masoudi (2022) finds the farsighted emission profile of a coalition with one country outside the agreement and runs numerical simulations to capture stability for the full cooperative solution (Masoudi, 2022).

In this analysis, the sequence of contingency plans is constructed to take into account all possible subsequent unilateral deviations. This work differs from Masoudi (2022) in that it captures the loss in welfare for a coalition compared to the joint welfare maximisation setup and identifies the contingency plans for any coalition size. With one country outside the agreement, an alternative welfare (later defined as shadow welfare) of the contingency plan captures the loss in welfare for the coalition compared to the optimal welfare levels obtained in Diamantoudi and Sartzetakis (2006). If a signatory unilaterally exits, a new contingency plan is agreed upon that nullifies any benefit to a subsequent deviation by another signatory that wants to follow the action of the first deviator and exit the agreement. Our work suggests that instead of choosing emissions through joint welfare maximization, they commit to emission levels that can sustain the coalition and, thus, their welfare level. Instead of choosing emissions defensively, which leads to very low emission levels when the coalition is large, they choose higher emissions that will pose the necessary threat to any member looking forward to large free-riding welfare gains. The remainder of the essay is structured as follows. Section 2.2 presents the model. Section 2.3 introduces a comparative equilibrium analysis in the general case of a unilateral deviation from the grand coalition to the optimal solutions derived by Diamantoudi and Sartzetakis (2006). Section 2.4 shows the sequence of coalition contingency plans that signatories take into account in the agreement for cases of subsequent unilateral deviations. Section 2.5 presents an example with ten countries and shows that no signatory benefits from exit at any participation outcome. Section 2.6 concludes.

2.2 The Model

Similar to the first essay, the assumption of homogeneity is maintained with n identical countries, $N = \{1, \dots, n\}$. Each country i generates positive emission levels $e_i \geq 0$ of global pollutants as a result of the production and consumption processes. The social welfare of country i, W_i, is defined as the difference between the total benefits from country i's emissions, $B_i(e_i)$, and the damages from total global emissions, including country i's emissions, $D_i(E)$, where $E = \sum_{i \in N} e_i$. We consider the same quadratic benefit and damage functions in Diamantoudi and Sartzetakis (2006), and given that countries are identical, the subscripts of both functions can be dropped. For each country i, $i \in N$, given positive parameters a and b, the benefit function is assumed to be $B(e) = b[ae_i - \frac{1}{2}e_i^2]$. Country i's damages from pollution are the damages generated by aggregate pollution, E. The damage function for each country i is assumed to be quadratic, $D(E) = \frac{1}{2}cE^2$. Given the following benefit and damage functions, each county i's welfare function W_i is given by

$$W_i(e_i) = b[ae_i - \frac{1}{2}e_i^2] - \frac{c}{2}(\sum_{i \epsilon N} e_i)^2 \tag{2.0}$$

In the case of full cooperation, countries maximise the joint welfare of the grand coalition, and this yields a per country emission level, e_c, and indirect welfare levels, ω_c, given by

$$e_c = \frac{a}{\gamma n^2 + 1}, \qquad \omega_c = \frac{a^2 b}{2(1 + \gamma n^2)}, \tag{2.1}$$

where $\gamma = \frac{c}{b}$.

In the non-cooperative case, each country maximises its own welfare, defined in eq. (2.2). The value of emissions in this case is given by,

$$e_{nc} = \frac{a}{1 + \gamma n}. \tag{2.2}$$

We consider an agreement that contain contingency plans prescribing the response of the coalition members when faced with a potential deviator at each coalition size, starting from the grand coalition. The response will be an increase in the remaining members' emissions to a level that makes the deviator's damages large enough to nullify the envisioned benefits from free-riding. This top-down approach allows us to study subsequent deviation by members when faced with a potential unilateral exit.

In the previous essay, we extended the standard model by allowing the coalition's members to adjust their emissions to prevent unilateral or simultaneous group deviations. In this section, we consider signatories' emission adjustments that will take into account not only one-step deviation, be it unilateral or in a group, but also possible subsequent unilateral deviations. Given that in this approach the size of the coalition is taken as arbitrarily given, the contingency plans examine the case of a unilateral deviation, starting from the grand coalition. Let ω_s denote the indirect welfare level of an individual signatory and ω_{ns} that of an individual non-signatory. The grand coalition not only chooses the level of emissions that maximise its collective welfare at n, e_c, but also chooses the contingency level of emissions at $n-1$, $\tilde{e}_s(n-1)$, derived from the constraint $\omega_{ns}(e_{ns}(e_s(n-1))) \leq \omega_c(e_c)$. Suppose $n = 10$, the contingency level of emissions at the grand coalition takes into account a deviation that will lead to $s = 9$. If for any reason the coalition reduces to nine members, a new contingency plan exists to deter a possible subsequent deviation by another signatory from exiting and the coalition size becomes $s = 8$.

The contingency plans are computed recursively from the previous contingency plan realized. That is, given n, the coalition identifies the full cooperation output e_c and the series of contingency plans starting with a single deviation from the grand coalition, $\tilde{e}_s(n-1)$, and agrees upon all of them. If the coalition drops to $n-1$, a new contingency plan is agreed upon at $s = n-2$ that nullifies any benefit to a subsequent deviation by another signatory that wants to follow the action of the first deviator and exit. In this manner, the sequence of contingency plans is constructed to take into account all possible subsequent deviations.

Starting with the case of full cooperation, $s = n$, the coalition builds a contingency plan for $s = n-1$ in case of a potential unilateral exit. The contingency plan computes the response of the remaining coalition at $n-1$ by making the unilateral deviator indifferent between being in the agreement or opting out.

Recall the indirect welfare level of each signatory at full cooperation is given by,

$$\omega_c = \frac{a^2 b}{2(1 + \gamma n^2)}. \tag{2.3}$$

The minimum level of emissions at $n-1$ need to ensure the deviator does not benefit from exiting the agreement. $\widetilde{\omega}_{ns}(n-1)$ represents the indirect welfare of a single non-signatory outside the agreement. After plugging the reaction function,

$e_{ns}\left(n-1\right)=\frac{a-\gamma(n-1)e_s(n-1)}{1+\gamma}$ in $W_{ns}(n-1)$, given in eq. (2.1), the indirect welfare of a single non-signatory, $\widetilde{\omega}_{ns}(n-1)$, is formally given by

$$\widetilde{\omega}_{ns}\left(n-1\right)=\tfrac{ba(a-\gamma(n-1)\widetilde{e}_s(n-1))}{1+\gamma}-\tfrac{b(a-\gamma(n-1)\widetilde{e}_s(n-1))^2}{2(1+\gamma)^2}$$

$$-\tfrac{c}{2}((n-1)\widetilde{e}_s\left(n-1\right)+\tfrac{(a-\gamma(n-1)\widetilde{e}_s(n-1))}{1+\gamma})^2\ . \tag{2.4}$$

From the constraint, $\omega_c=\widetilde{\omega}_{ns}(n-1)$, the contingency plan at $n-1$ is computed to be

$$\widetilde{e}_s(n-1)=\frac{a}{n-1}\left[\sqrt{\frac{\gamma n^2+n^2}{\gamma n^2+1}}-1\right], \tag{2.5}$$

which is positive for any $n>1$. Differentiating signatories' level of emissions with respect with γ at $n-1$ yields

$$\tfrac{\partial}{\partial\gamma}\left[\widetilde{e}_s\left(n-1\right)\right]=-\tfrac{n^2(n-1)(n+1)}{2(\gamma n^2+1)^2\sqrt{\frac{n^2(\gamma+1)}{\gamma n^2+1}}}<0.$$

As γ increases, i.e., as the ratio of the marginal cost of pollution to the marginal benefit from emissions rises, the remaining coalition drafts a lower level of emissions at $n-1$. This is intuitively clear, since when damages becomes more important than benefits, the free-riding incentives reduce and thus, the required threat of increasing the coalition's emissions. The level of emissions by a non-signatory at $n-1$, $\widetilde{e}_{ns}\left(n-1\right)$, is computed from the reaction function in eq. (2.3) and is given by

$$\widetilde{e}_{ns}(n-1)=\frac{a}{\gamma+1}\left[\gamma+1-\gamma n\sqrt{\frac{\gamma+1}{\gamma n^2+1}}\right]. \tag{2.6}$$

The level of emissions of non-signatories at $n-1$ is positive for all values of γ given positive paramaters a, b, c. Given $\widetilde{e}_s(n-1)$ and $\widetilde{e}_{ns}(n-1)$, the indirect welfare of signatories at $n-1$ under the contingency plan, $\widetilde{\omega}_s(n-1)$, can be expressed as

$$\widetilde{\omega}_s(n-1) = -ba^2\left[\frac{(2n-1)(\gamma+1)+n^2(2\gamma+1)+n^3\gamma(n+2\gamma)}{2(\gamma n^2+1)(n-1)^2(\gamma+1)} - \frac{n^2\sqrt{(\gamma n^2+1)(\gamma+1)}}{(\gamma n^2+1)(n-1)^2}\right] \tag{2.7}$$

The next section covers the equilibrium comparisons at $n-1$ to that of Diamantoudi and Sartzetakis (2006).

2.3 Equilibrium Comparisons

The optimal solution derived by Diamantoudi and Sartzetakis (2006) from the joint profit maximisation of the coalition at s is given by $e_s^*(s) = a\left[1-\frac{\gamma sn}{\psi}\right]$ where $\psi = X^2+\gamma s^2$ and $X = 1+\gamma(n-s)$. At $n-1$, the optimal level of emissions is given by

$$e_s^*(n-1) = a\left[1-\frac{\gamma(n-1)n}{(1+\gamma)^2+\gamma(n-1)^2}\right]. \tag{2.8}$$

The solution generates an indirect welfare to signatories of size $n-1$, $\omega_s^*(n-1)$, equal to

$$\omega_s^*(n-1) = -ba^2\left[\frac{\gamma(2n-1)-(\gamma+1)^2}{2\gamma(\gamma+1)^2(n-1)^2}\right]. \tag{2.9}$$

The difference in the level of emissions from the contingency plan and the optimal solution represents the increase in emissions under the contingency plan. This is given by

$$\widetilde{e}_s(n-1) - e_s^*(n-1) =$$

$$\frac{a}{n-1}\left[\sqrt{\frac{n^2\gamma+n^2}{1+n^2\gamma}}-1\right] - a\left[1-\frac{\gamma(n-1)n}{(1+\gamma)^2+\gamma(n-1)^2}\right].$$

The contingency plan drafts a higher level of emissions relative to e_s^* as signatories increase the level of their emissions to offset any gains to a unilateral deviation. The indirect welfare difference between the optimal solution and the contingency plan represents the sacrifice in welfare that signatories make under the contingency plan. This sacrifice is due to the increase in emissions under the contingency plan. The sacrifice in signatory welfare at $n-1$ is the difference between the optimal welfare level, $\omega_s^*(n-1)$, and the shadow welfare, $\widetilde{\omega}_s(n-1)$, given by

$$\omega_s^*(n-1) - \widetilde{\omega}_s(n-1) =$$

$$-ba^2\left[\frac{\gamma(2n-1)-(\gamma+1)^2}{2\gamma(\gamma+1)^2(n-1)^2} - \frac{(2n-1)(\gamma+1)+n^2(2\gamma+1)+n^3\gamma(n+2\gamma)}{2(\gamma n^2+1)(n-1)^2(\gamma+1)} + \frac{n^2\sqrt{(\gamma n^2+1)(\gamma+1)}}{(\gamma n^2+1)(n-1)^2}\right].$$

As mentioned earlier, eq. (2.5) computes the contingency plan for a remaining coalition of $n-1$ if one signatory contemplates exiting the agreement. This drafts an increase in the level of signatories' emissions in case of a unilateral deviation. This leads to a loss or sacrifice in welfare for $n-1$ under the contingency plan compared to the optimal welfare level, and leads to a larger decrease in non-signatories' welfare when we start with a large coalition. The sacrifice in welfare, $\omega_s^*(n-1) - \widetilde{\omega}_s(n-1)$, is positive given $n > 1$ and positive parameters a, b, and c.[7] If the coalition drop to $n-1$, the remaining coalition increases its emissions so as to raise the deviator's

[7]Given the long terms in the equation, the sacrifice in welfare can be shown to be positive by numerical simulations taking different values of n and γ.

damages to a level such that it nullifies the envisioned benefits from free-riding. In other words, the welfare of the potential deviator will be no different from its welfare as a member of the grand coalition.

However, when one signatory exits the agreement, this action might lead to a subsequent deviation by another signatory, so there exists a different contingency plan for that coalition size that takes into account the possibility of a further unilateral deviation. The next section computes the sequence of contingency plans for all coalition sizes.

2.4 Sequence of Coalition Equilibriums

To find the contingency plan for a coalition of size s, the sequence of strategies for coalitions of sizes $s+1$ to n need to be computed. In this study, we focus on the sequence of contingency plans unfolding from full cooperation. It is noteworthy to mention that the sequence of contingency plans can be unfolded from a different starting point, that is, a different coalition size other than the grand coalition. Let k denote the number of countries that do not participate in the agreement, i.e., $k = n - s$. The analysis starts with the case of full cooperation, $s = n$, where the welfare of each country is denoted by $\omega_s\,(k = 0)$. The coalition builds a strategy for $k = 1$, i.e., given the welfare of each country at the grand coalition, $\omega_s\,(k = 0)$, the strategy is to find the level of emissions at $k = 1$, $e_s\,(k = 1)$ that the remaining coalition needs to emit in order to make the unilateral deviator indifferent between remaining in the grand coalition and opting out. Let $\omega_{ns}(k = 1)$ represent the indirect welfare of a single non-signatory when one member defects from the agreement and the size of non-signatories becomes $k = 1$. The level of emissions at $k = 1$ is determined from the constraint $w_s\,(k = 0) = \omega_{ns}\,(k = 1)$. As mentioned earlier, the remaining coalition members raise their emissions in response to a unilateral deviation by a signatory from the agreement.

From the constraint, $\omega_s(k = 0) = \omega_{ns}(k = 1)$, the level of signatories' emissions required to eliminate free rider incentives at $k = 1$ is defined by

$$\widetilde{e}_s(k = 1) = \frac{\alpha_{k=1}\,(1+\gamma) - a\gamma\,(1+\gamma)}{\gamma(\gamma+1)\,(n-1)}, \tag{2.10}$$

where $\alpha_{k=1} = \sqrt{\gamma(\gamma+1)(a^2 - \frac{2\omega_s(k=0)}{b})} = a\gamma n\sqrt{(\frac{\gamma+1}{\gamma n^2+1})}$.

The value of $\widetilde{e}_s(k = 1)$ is the level of emissions that signatories adjust their emissions to at $k = 1$ making the deviator indifferent between being in the grand coalition ($k = 0$) or out ($k = 1$). The level of signatories' emissions at $k = 1$ generates the indirect welfare $w_s\,(k = 1)$. A new strategy is implemented for $k = 2$. Given the level of welfare of non-signatories at $k = 2$, the constraint $\omega_s\,(k = 1) = \omega_{ns}(k = 2)$ gives the signatories' contingency plan at $k = 2$, given by

$$\widetilde{e}_s(k = 2) = \frac{\alpha_{k=2}\,(1+2\gamma) - 2a\gamma\,(1+\gamma)}{\gamma(\gamma+1)\,(n-2)}, \tag{2.11}$$

where $\alpha_{k=2} = \sqrt{\gamma(\gamma+1)(a^2 - \frac{2\widetilde{\omega}_s(k=1)}{b})}$.

The value of $\widetilde{e}_s(k = 2)$ is the level of emissions that signatories adjust to at $k = 2$ in order to make the 2nd deviator indifferent between staying in the agreement and maintaining the same welfare whether in the coalition ($k = 1$) or out ($k = 2$). If at $k = 2$ another deviator wants to opt out, the strategy of signatories at $k = 3$ is computed by $\omega_s(k = 2) = \omega_{ns}(k = 3)$, and the signatories' contingency plan at $k = 3$ is

$$\widetilde{e}_s(k = 3) = \frac{\alpha_{k=3}\,(1 + 3\gamma) - 3a\gamma\,(1 + \gamma)}{\gamma(\gamma + 1)\,(n - 3)}, \tag{2.12}$$

where $\alpha_{k=3} = \sqrt{\gamma(\gamma + 1)(a^2 - \frac{2\widetilde{\omega}_s(k=2)}{b})}$.

The same procedure can be followed to compute the contingency plan of the coalition for any size. It is important to note that if we want to compute the strategy for the coalition, given k countries outside the agreement, the sequence of strategies of the coalition need to be computed starting from $k = 0$ to $k + 1$. This generates every contingency plan starting from full cooperation, which computes the contingency plan at $k = 1$, all the way to the strategy taken by the coalition at $k + 1$ to keep all coalitions stable. The boundary is hit at $k = n - 2$, i.e., $s = 2$, when a possible coalition of three signatories find the level of emissions for $s = 2$. In the case of a unilateral deviation at $s = 3$, the contingency plan finds the equilibrium level at $s = 2$ where $n - 2$ behaves as non-signatories.

The reaction function of non-signatories at $k + 1$ is given by,

$$e_{ns}\,(s|k + 1) = \frac{a - \gamma\,(n - k - 1)\,(e_{s|k+1})}{1 + \gamma(k + 1)}. \tag{2.13}$$

After plugging eq. (2.13) in the welfare function, the indirect welfare of non-signatories at $k + 1$ is formally given by

$$\omega_{ns}(k+1) = \frac{ba(a-\gamma(n-k-1)e_{s|k+1})}{1+\gamma(k+1)} - \frac{b(a-\gamma(n-k-1)e_{s|k+1})^2}{2(1+\gamma(k+1))^2}$$

$$-\frac{c}{2}((n-k-1)e_{s|k+1} + \frac{(k+1)(a-\gamma(n-k-1)e_{s|k+1})}{1+\gamma(k+1)})^2 \ . \tag{2.14}$$

The above equation, part of the constraint, determines the level of emissions of the coalition at $k+1$. The results can be summarised in the following Proposition.

Proposition 2.1 *There exists a stable coalition at s, given the assumption that coalition members commit to the sequence of contingency plans drafted from the grand coalition subject to the constraint, $\omega_s(s) = \omega_{ns}(s-1)$. The contingency plan for signatories at s is computed from the coalition $s+1$ dropping to s, that is, the sequence of equilibriums $\{\widetilde{e}_{s|n}, \widetilde{e}_{s|n-1}, ..., \widetilde{e}_{s|s+1}\}$ give rise to $\widetilde{e}_s(s)$ given by*

$$\widetilde{e}_s\left(s\right) = \frac{X\alpha_{n-s} - a\gamma\left(1+\gamma\right)\left(n-s\right)}{s\gamma\left(\gamma+1\right)},$$

where $X = 1+\gamma(n-s)$, and $\alpha_{n-s} = \sqrt{\gamma(\gamma+1)(a^2 - \frac{2\widetilde{\omega}_s(s+1)}{b})}$.

For instance, the contingency plan at $n-2$ allows signatories to adjust their emissions to $\widetilde{e}_s(n-2)$ to offset a subsequent deviation by another signatory from $n-1$. The value of $\widehat{e}_s(k=2)$ is the level of emissions that signatories adjust to at $k=2$ in order to make the 2nd deviator indifferent between staying in the agreement and maintaining the same welfare whether in the coalition ($k=1$) or out ($k=2$). The next section covers an example with ten homogenous countries and discusses the sequence of equilibriums for each coalition size.

2.5 An Example

We utilise the same parameter values as in Diamantoudi and Sartzetakis (2006), $n = 10$, $a = 10$, $b = 6$, $c = 0.39999$. We denote welfare and emissions levels derived in Diamantoudi and Sartzetakis (2006) with a star (*), and we compare them to the results of this model, which are denoted by a tilde (˜). Table 2.1 reports all the results of emissions and welfare for different coalition sizes. The first column indicates the size of the coalition. The four subsequent columns report welfare and emissions levels of signatories and non-signatories when the coalition commits to the contingency plan. The sequence of contingency plans are computed working backwards from the grand coalition. The last four columns report the welfare and emission levels of signatories and non-signatories when coalition members maximise their joint welfare.

Table 2.1 Emission and welfare levels
n=10, a=10 , b=6, c=0.39999

s	$\tilde{\omega}_s$	$\tilde{\omega}_{ns}$	$\tilde{e}_s$	$\tilde{e}_{ns}$	ω^*_s	ω^*_{ns}	e^*_s	e^*_{ns}
s=0		-468		6		-468		6
s=1					-461	-491	7.46	5.93
s=2					-464	-431	4.9	6.09
s=3	-444	-421	5.1	6.12	-426	-306	2.73	6.44
s=4	-421	-383	4.81	6.22	-360	-156	1.18	6.95
s=5	-383	-333	4.53	6.36	-280	-19	0.32	7.42
s=6	-333	-270	4.24	6.55	-199	86	0.011	7.89
s=7	-270	-190	3.93	6.8	-124	161	0.085	8.3
s=8	-190	-90	3.55	7.14	-60	211	0.392	8.63
s=9	-90	39	3.03	7.67	-6	243	0.82	8.91
s=10	39.1		1.3044		39.1		1.3044	

Figure 2.1 and Figure 2.2 illustrate emission and welfare levels presented in Table 2.1, respectively. Figure 2.1 illustrates emission levels of signatories (solid black line) and non-signatories (dashed black line) under the joint welfare maximisation to that of signatories (solid green line) and non-signatories (dashed red line) under the contingency plan for different coalition sizes. Figure 2.2 illustrates the welfare levels associated with their respective emission levels at each coalition size.

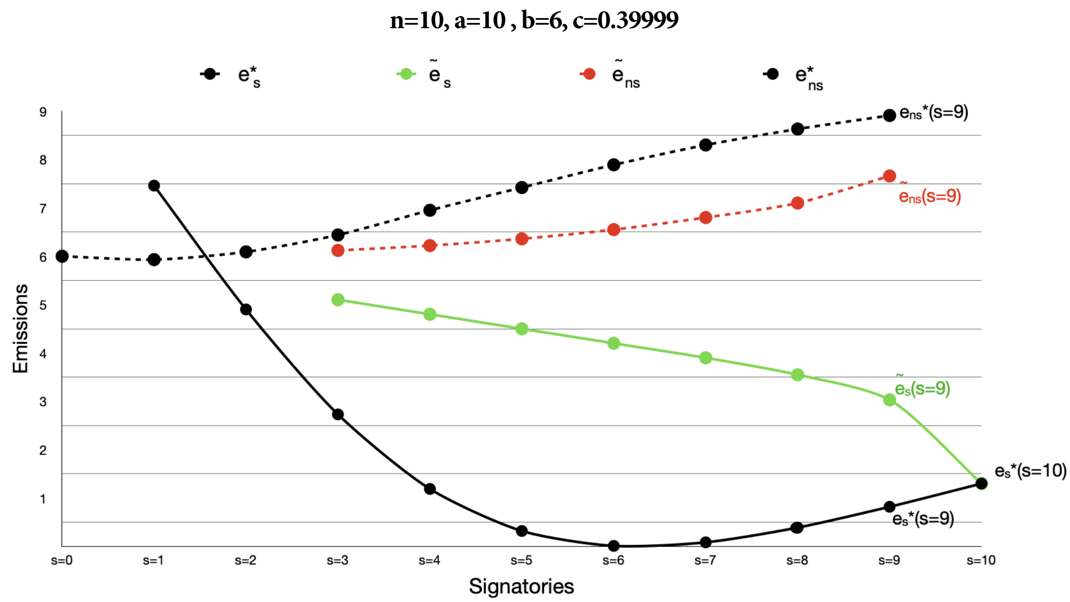

Figure 2.1 Comparative statics: emission levels

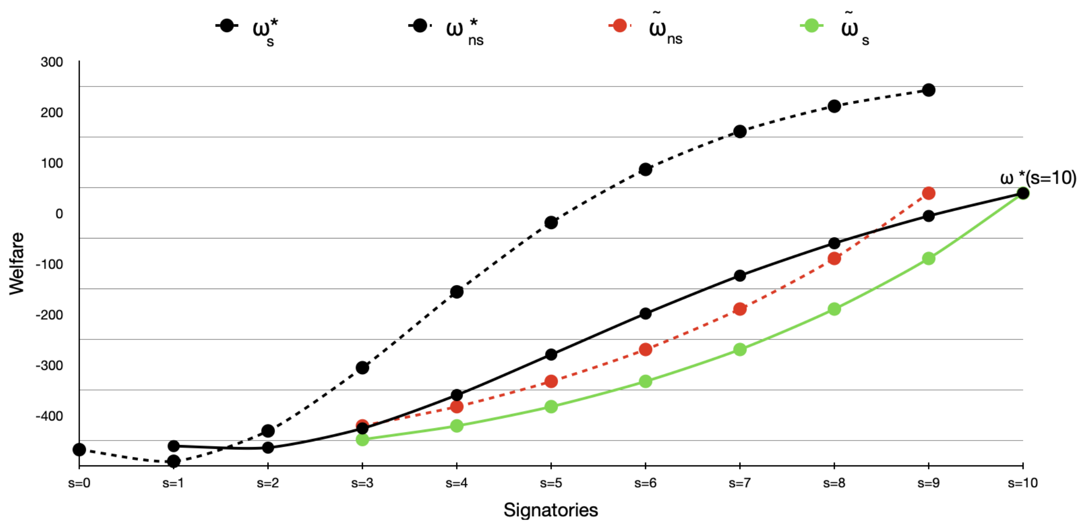

Figure 2.2 Comparative statics: welfare levels

We start from the largest stable coalition under the joint welfare maximisation rule reported in Diamantoudi and Sartzetakis' (2006) simulations, i.e., $s^* = 3$. Then, we examine if larger coalitions can become stable when coalition's member agree on the contingency plans proposed in this section. Under the joint welfare maximisation, a coalition of three signatories will agree to each emitting $e_s^*(3) = 2.71$, and non-signatories to the agreement, $n - s = 7$, will emit $e_{ns}^*(3) = 6.44$. This generates an indirect welfare for each signatory $\omega_s^*(3) = -426$, and that of non-signatories', $\omega_{ns}^*(3) = -306$, as reported in Table 2.1. Since $\omega_{ns}^*(2) = -431 < \omega_s^*(3)$ and $\omega_{ns}^*(4) = -156 > \omega_s^*(3)$, the coalition $s^* = 3$ is both internally and externally stable. From Table 2.1, we conclude that under the joint welfare maximising behaviour, a coalition of four countries is not stable.

However, if the coalition follows the contingency plan examined in this section, a coalition of four countries could become stable. That is, if a coalition of four countries commits that in the case one of its members exits, the remaining three will each emit $\widetilde{e}_s(3) = 5.1$, and non-signatories' welfare becomes $\widetilde{\omega}_{ns}(3) = -421$, which is the same as the welfare it attains as a member of the coalition of four under the contingency plan. Therefore, there are no incentives under the contingency plan to exit the coalition of four.

Examining the case of full cooperation, each member's level of emissions is $e_s^*(10) = 1.3044$. This generates a welfare of $\omega_s^*(s = 10) = 39.1$. If one member exits, i.e., $s = 9$, the level of emissions generated under the joint profit maximisation is $e_s^*(9) = 0.82$. This generates a welfare for signatories, $\omega_s^*(9) = -6$. The level of the remaining members' emissions necessary to prevent the unilateral exit of one member is $\widetilde{e}_s(9) = 3.03$.

This adjusted level of the remaining members' emissions leads to a sharp decrease in their welfare to $\widetilde{\omega}_s(9) = -90$. However, by committing to this welfare sacrifice, they are able to eliminate the potential deviator's incentives to free-ride, since the welfare it attains, $\widetilde{\omega}_{ns}(9) = 39$ is no larger than that it enjoys as a member of the grand coalition. If members of the grand coalition can commit to increasing their emissions to $\widetilde{e}_s(9) = 3.03$, instead of decreasing them to $e_s^*(9) = 0.82$, they can successfully threaten any potential deviator and retain the grand coalition.

If one member is no longer part of the agreement, or the game starts with one individual country outside the agreement, a new contingency plan has to be derived by the remaining coalition members to ensure no other member unilaterally exits. In our example, the coalition of nine countries will commit that, in case one member exits, the remaining eight countries will increase their emissions to $\widetilde{e}_s\,(8) = 3.55$ instead of decreasing them to $e_s^*\,(8) = 0.89$ as the joint welfare maximisation dictates. By committing to react in such a way to a potential deviation, the eight remaining countries eliminate any gains from exiting, and the coalition of nine is immune to unilateral deviations. Table 2.1 and Figures 2.1 and 2.2 present the contingency plans to which members of a coalition of any size have to commit in order to prevent deviations.

Under the assumption that coalition members maximise joint welfare, starting from the grand coalition, their response to a deviation is to choose a lower level of emissions, which provides sufficient incentives for further shrinking the size of the coalition. This is evident from the data reported in Table 2.1's second to the last column, with $e_s^*\,(10) = 1.3044$ going to $e_s^*\,(6) = 0.011$. For smaller coalition sizes, coalition members start increasing their emissions because the leadership effect prevails over the environmental damage effect: the coalition internalises the externality over fewer members while it takes advantage of its leadership over more followers. However, it takes much for a very small coalition to increase their emissions enough to prevent free-riding. In the particular example, this happens at $s^* = 3$, yielding the highest stable coalition under joint welfare maximisation.

These findings suggest that coalition members realizing the potential gains of cooperation, give more substance to their leadership role and instead of choosing emissions through joint welfare maximisation, they commit to emission levels that can sustain the coalition and thus their welfare level. As shown above, there is an emission level to which all members of a coalition of any size can commit to in case of a deviation, such that no member has an incentive to free-ride.

2.6 Conclusion

We propose that coalition members use their leadership more proactively in order to preserve higher welfare to all their members. Instead of choosing emissions defensively, which leads to very low emission levels when the coalition is large, they act proactively, choosing much higher emissions that will pose the necessary threat to any member looking forward to large free-riding welfare gains.

Coalition members, instead of just exploiting their leadership role over non-members, which leads only to small stable coalitions and thus a low level of welfare, can protect their commonly achieved higher welfare level by devising any organisational framework that allows them to credibly commit to punishing potential deviators by choosing higher emission levels. It is understood that when we start with a large coalition, the required increase in emissions will inflict welfare losses on coalition members. However, if coalition members accept these welfare sacrifices and credibly commit to adjust their emissions in the face of deviations, they can prevent deviations and continue enjoying high welfare levels.

An example with ten countries is provided to show no signatory benefits from exiting at any participation outcome. The contemplating deviator is indifferent between being in the agreement or out. If a signatory unilaterally exits, a new contingency plan is adopted that nullifies any benefit to a subsequent deviation by another signatory that wants to follow the action of the first deviator and exit the agreement. In this manner, the sequence of contingency plans for the coalition are constructed to take into account all possible subsequent deviations from the grand coalition. This work can be extended to look at subsequent unilateral deviations from a different stage of a game, that is, the sequence of contingency plans can be unfolded starting at a different coalition size other than the grand coalition. Given a different starting point in the game, the emission and welfare levels can be compared to the sequence of contingency plans unfolded from full cooperation.

2.7 Appendix

The calculations and proofs of the contingency plan for a unilateral deviation from grand coalition is presented in this section.

The indirect welfare of signatories at n is given by

$$\omega_c = \frac{ba^2}{2(1+\gamma n^2)}.$$

The reaction function of non-signatories at $n-1$ is given by

$$e_{ns}(n-1) = \frac{a-\gamma(n-1)e_s(n-1)}{1+\gamma} .$$

Plugging the reaction function in the indirect welfare of a non-signatory at $n-1$, $\widetilde{\omega}_{ns}(n-1)$, gives

$$\widetilde{\omega}_{ns}(n-1) = \frac{ba(a-\gamma(n-1)e_s)}{1+\gamma} - \frac{b(a-\gamma(n-1)e_s)^2}{2(1+\gamma)^2} - \frac{c}{2}((n-1)e_s + \frac{(a-\gamma(n-1)e_s)}{1+\gamma})^2.$$

From the constraint, $\omega_c = \widetilde{\omega}_{ns}(n-1)$, the indirect welfares can be rewritten as

$$\frac{a^2}{2(1+\gamma n^2)} = \frac{a(a-\gamma(n-1)e_s)}{1+\gamma} - \frac{(a-\gamma(n-1)e_s)^2}{2(1+\gamma)^2} - \frac{\gamma}{2}((n-1)e_s + \frac{(a-\gamma(n-1)e_s)}{1+\gamma})^2.$$

From the constraint, $\omega_c = \widetilde{\omega}_{ns}(n-1)$, the contingency plan of signatories at $n-1$ is computed to be

Root 1. $\widetilde{e}_s(n-1) = \frac{\left(-a+an\sqrt{\frac{1}{n^2\gamma^2+n^2\gamma+\gamma+1}}+an\gamma\sqrt{\frac{1}{n^2\gamma^2+n^2\gamma+\gamma+1}}\right)}{n-1}.$

Root 2. $\widetilde{e}_s(n-1) = -\frac{\left(a+an\sqrt{\frac{1}{n^2\gamma^2+n^2\gamma+\gamma+1}}+an\gamma\sqrt{\frac{1}{n^2\gamma^2+n^2\gamma+\gamma+1}}\right)}{n-1}$ rejected.

$\widetilde{e}_s(n-1) = \frac{a\left(n\sqrt{\frac{1}{n^2\gamma(1+\gamma)+\gamma+1}}+n\gamma\sqrt{\frac{1}{n^2\gamma(1+\gamma)+\gamma+1}}-1\right)}{n-1}$

$= \frac{a}{n-1}\left[n\sqrt{\frac{\gamma+1}{1+\gamma n^2}} - 1\right]$

$$= \frac{a}{n-1}\left[\sqrt{\frac{n^2\gamma+n^2}{1+n^2\gamma}}-1\right].$$

With $n > 1$, the above solution is always positive for all values of γ given positive parameters a, b, c. The level of emissions by non-signatory at $n-1$ is computed to be

$$\widetilde{e}_{ns}(n-1) = \frac{a-\gamma(n-1)e_s}{1+\gamma}$$

$$= \frac{a\left(\gamma-\gamma\sqrt{\frac{(n^2\gamma+n^2)}{\gamma n^2+1}}+1\right)}{\gamma+1}$$

$$= \frac{a}{\gamma+1}\left(\gamma-\gamma\sqrt{\frac{(n^2\gamma+n^2)}{\gamma n^2+1}}+1\right).$$

The level of emissions of a non-signatory is positive for all values of γ given $n > 1$ and positive parameters a, b, c.

Proof.

$$\gamma+1 > \gamma\sqrt{\frac{(n^2\gamma+n^2)}{\gamma n^2+1}}$$

$$\gamma+1 > \gamma\left(\frac{\gamma n^2}{\gamma n^2+1}\right)$$

Always true given $\frac{\gamma n^2}{\gamma n^2+1} < 1$, $\gamma > 0$ and $n > 1$. ■

ESSAY THREE

INTERNATIONAL ENVIRONMENTAL AGREEMENTS: CHOICE OF NET EMISSIONS

3.1 Introduction

The present essay highlights and examines an agreement with the choice of net emissions as a coalition's policy instrument. Net emissions of a country are defined as the difference between its emission and abatement levels. Global net emissions is the sum of each country's net emissions. In the 2015 Paris Agreement, countries pledge to reduce their greenhouse gas emissions, where pledges are known as Nationally Determined Contributions (NDCs). Article 6.4 of the Paris Agreement (UNFCCC, 2015) places the following requirement on parties:

"A mechanism to contribute to the mitigation of greenhouse gas emissions and support sustainable development... shall aim: (a) To promote the mitigation of greenhouse gas emissions while fostering sustainable development; (b) To incentivize and facilitate participation in the mitigation of greenhouse gas emissions by public and private entities authorized by a Party; (c) To contribute to the reduction of emission levels in the host Party, which will benefit from mitigation activities resulting in emission reductions that can also be used by another Party to fulfil its nationally determined contribution; and (d) To deliver an overall mitigation in global emissions."

One can argue that the agreement is, in fact, setting parties to act cooperatively not only in emission reductions but also in joint mitigation efforts, i.e., parties are encouraged to act cooperatively strengthening action on sharing abatement technology development with policy coordination on net emissions. With a growing majority of countries pushing toward carbon neutrality, achieving net zero emissions, it highlights the importance of introducing abatement efforts in modelling countries' choice variables using game theory as the tool of analysis. Abatement measures adopted domestically or through joint efforts with other countries reduce global emissions and damages from aggregate pollution.

The present study utilises the model of Diamantoudi, Sartzetakis, and Strantza (2022) in the absence of adaptation to examine the impact of heterogeneity in emission benefits and abatement costs on the choices of emissions, abatement, and net emissions. Compared to the previous two essays, it adds abatement efforts as a choice variable. This is the first paper in the literature that highlights analysis on net emissions. If we try to emulate negotiations on countries' commitments in the Paris Agreement, countries do not collectively choose what each country emits and what each country abates, each country commits to one net target collectively in the first stage, they choose net emissions and then go on to individually decide on emission and abatement levels. For countries to sit together and collectively choose certain variables by maximizing total welfare, the welfare functions have to be shared with each other so they become common information. However, which part of the welfare function countries share depends on the choice variables chosen.

The framework examines four cases, beginning with two benchmark cases: the non-cooperative pure Nash case and the case of cooperation on emissions and abatement. Next, it examines two cases when countries cooperate on net emissions. Both cases utilise the Cournot setup and introduce a two-stage game to study two countries cooperatively choosing their level of net emissions by maximising their joint welfare in the first stage.

The first case examines two countries agreeing to cooperate on net emissions and capture emission benefits in the first stage. That is, countries choose net emissions with revealed information on preferences. In the second stage, countries simultaneously choose their levels of abatement independently, given the level of net emissions agreed upon in the first stage. Given each country's net emissions and abatement levels, each country deduces its emission levels.

The second case examines two countries agreeing to cooperate on net emissions and capture abatement technologies in the first stage. That is, countries choose net emissions with revealed information on abatement. In the second stage, countries simultaneously choose their levels of emissions independently, given the level of net emissions agreed upon in the first stage. Given each country's net emissions and emission levels, each country deduces its abatement levels.

Comparing results to the pure non-cooperative Nash benchmark case, cooperation on net emissions allows countries to abate more and agree on lower global net emissions. Results show that cooperation on net emissions, even with a high degree of heterogeneity in emission benefits and abatement technology, can bring gains to cooperation in both countries. Numerical simulations are presented to show the impact of heterogeneity on net emission decisions. The remainder of the essay is structured as follows. Section 3.2 presents the model, with subsections 3.2.1 and 3.2.2 presenting the generalisation for the benchmarks and net emission cases. Section 3.3 introduces an illustrative example with specific functional forms to examine the differences between cases. Section 3.4 examines numerical simulations, and section 3.5 concludes.

3.2 The Model

We assume there exist two heterogenous of countries, $i \in \{A, B\}$. Each country i generates emission levels, $e_i \geq 0$, of global pollutant as a result of its production and consumption activities, and engages in abatement efforts, $x_i \geq 0$, to reduce damages from aggregate global pollutants. Net emissions of a country, NE_i, is defined by $NE_i = e_i - x_i$. The benefit function is strictly concave, $B_i' \geq 0$, and $B_i'' < 0$, that is, increasing at a decreasing rate. Each country can engage in abatement, $x_i \geq 0$, which is costly. The abatement cost function is strictly convex, that is, $C_i' \geq 0$, and $C_i'' \geq 0$. Each country suffers from global emissions $E = \sum_{i \in A,B} e_i$ and enjoys benefits from global abatement, $X = \sum_{i \in A,B} x_i$. Benefits from abatement are spread globally, while country i bears complete the cost of its abatement. That is, while each country's emissions create a negative externality, its abatement generates a positive externality. Each country i suffers from damages that depend on aggregate net emissions, $NE = E - X$. That is, aggregate net emissions are defined as the difference between global emissions and global abatement. The damage function is strictly convex in net emissions, $D'(NE) \geq 0$, and $D''(NE) \geq 0$ such that damages from global net pollutants increase at an increasing rate.

The social welfare of country i, $W_i\left(e_i, x_i\right)$, is expressed as the difference between total benefits from country i's emissions, $B_i\left(e_i\right)$ and damages from aggregate net emissions, $D(NE)$, where $NE = \sum_i NE_i$, minus country i's cost of abatement, $C_i(x_i)$, given by

$$W_i = B_i(e_i) - D(NE) - C_i(x_i), \tag{3.0}$$

where $i \in \{A, B\}$.

To simplify the analysis and focus on differences in emission benefits and abatement costs, we assume that both countries suffer the same damages D from aggregate net emissions. We begin by examining the two benchmark cases, the non-cooperative pure Nash case and the case of cooperation on abatement and emission levels.

3.2.1 Benchmark Cases

M0. The non-cooperative pure Nash case

In the non-cooperative pure Nash case, denoted as model (M0), each country i simultaneously decides on its optimal emissions and abatement levels, taking the other country's emission and abatement levels as given. That is, countries maximise their individual welfare function and simultaneously choose their optimal levels of emissions and abatement by solving the maximization problem given by,

$$\max_{e_i, x_i} \left(B_i(e_i) - D(NE) - C_i(x_i) \right) \ , \tag{3.1}$$

where $i \in \{A, B\}$. The first-order conditions of eq. (3.1) with respect to e_i and x_i yields country i's equilibrium levels of emissions, e^{nc}, and abatement, x^{nc}, given by,

$$B_i'(e_i^{nc}) = D'(NE) \tag{3.2}$$

and

$$C_i'\left(x_i^{nc}\right) = D'\left(NE\right) \tag{3.3}$$

where $NE = \sum_i \left(e_i - x_i\right)$.

Each country determines its level of emissions by equating its marginal benefit from its emissions to the marginal damage inflicted on it when global emissions increases as a result of its emissions increase. Each country also determines its levels of abatement by equating its marginal abatement cost to the decreased marginal damage as net emissions decrease due to its increase in abatement.

M1. Cooperation on emissions and abatement

In this case, denoted as model (M1), emissions and abatement decisions are made together by both countries. That is, countries decide collectively on the optimal levels of both emissions and abatement to maximise the aggregate welfare given their benefit from emissions and the cost of abatement. This model has been previously introduced by Sartzetakis and Strantza (2013), where it examines coalition formation between countries on emissions and abatement in an n-country model. Here, both countries collectively choose optimal levels of emissions and abatement by solving the following maximisation problem,

$$\max_{e_i, x_i} \sum_{i \in A,B} \left[B_i\left(e_i\right) - D\left(NE\right) - C_i(x_i)\right]. \tag{3.4}$$

The first-order conditions yield country i's equilibrium levels of emissions and abatement,

$$B_i^{'}\left(e_i^c\right) = 2D^{'}\left(NE\right), \tag{3.5}$$

and

$$C_i^{'}(x_i^c) = 2D^{'}\left(NE\right), \tag{3.6}$$

where $i \in \{A, B\}$ and $NE = \sum_i \left(e_i - x_i\right)$.

When countries cooperate, they take into account the increase (decrease) that their own emission (abatement) generates on the other country's damages. Given the assumption of homogenous damage function, marginal damage is multiplied by two in both eq. (3.5) and (3.6). Notice that since we assume countries have the same damage function, both the above models M0 and M1 set marginal benefits and marginal abatement costs equal between the two countries. The difference is that the cooperative game results in lower emissions and higher abatement and, thus, higher welfare. That is, only M1 achieves efficiency by internalizing the externality.

3.2.2 Agreements on net emissions

When countries agree to cooperate, they choose one net target collectively, net emissions, without committing to both abatement and emissions in the first stage but to either of them, and we examine each case separately. The first case, denoted as model (M2), assumes both countries reveal information on preferences[8] and collectively decide on optimal net emission levels in order to maximise their joint welfare. In the second stage, countries simultaneously and independently choose their optimal

[8]We assume that the rate of emission per unit of output is constant. Therefore, $B(e_i)$ reflects preferences of enjoying goods and services that generate emissions.

levels of abatement constrained by their first-stage commitments on net emissions. The second case, denoted as model (M3), assumes both countries reveal information on abatement technologies and collectively decide on optimal net emission levels in the first stage. In the second stage, countries simultaneously and independently choose their optimal levels of emissions constrained by their first-stage commitments on net emissions.

M2. Information on preferences

In the first stage, countries collectively choose their optimal level of net emissions, NE_i, by solving the maximisation problem,

$$\max_{NE_A, NE_B} \sum_{i \in A,B} [B_i\left(e_i\right) - D\left(NE\right) - C_i(x_i)], \tag{3.7}$$

where $e_i = NE_i + x_i$ and thus $\frac{\partial e_i}{\partial NE_i} = 1$. The first order conditions are given by,

$$B_i'(e_i\left(NE_i^x\left(x_A, x_B\right), x_i\right)) = 2D'\left(NE_A^x\left(x_A, x_B\right) + NE_B^x\left(x_A, x_B\right)\right). \tag{3.8}$$

The solution of the above yields $NE_A^x\left(x_A, x_B\right)$ and $NE_B^x\left(x_A, x_B\right)$, the agreed upon levels of net emissions which are functions of both countries' abatement.

In the second stage, each country i independently chooses its optimal level of abatement by taking the optimal level of net emissions, NE_i^x, as given, by solving the following maximisation problem,

$$\max_{x_i} B_i\left(e_i\left(NE_i^x\left(x_A, x_B\right), x_i\right)\right) - D\left(NE^x\right) - C_i(x_i), \tag{3.9}$$

where $NE^x = \sum_i NE_i^x$ and $e_i = NE_i^x + x_i$. The first order condition for country A is given by,

$$\frac{\partial B_A}{\partial e_A}\left(\frac{\partial NE_A}{\partial x_A} + \frac{\partial x_A}{\partial x_A}\right) - \frac{\partial D}{\partial NE}\left(\frac{\partial NE_A^x}{\partial x_A} + \frac{\partial NE_B^x}{\partial x_A}\right) - \frac{\partial C_A}{\partial x_A} = 0$$

$$\Rightarrow B_A' + (B_A' - D')\,\frac{\partial NE_A^x}{\partial x_A} - D'\frac{\partial NE_B^x}{\partial x_A} - C_A' = 0. \tag{3.10}$$

Similarly, country B's first order condition is given by,

$$B_B' + (B_B' - D')\,\frac{\partial NE_B^x}{\partial x_B} - D'\frac{\partial NE_A^x}{\partial x_B} - C_B' = 0. \tag{3.11}$$

Solving (3.10) and (3.11) yields the optimal x_A^* and x_B^*. Taking eq. (3.10) and differentiating both sides by x_A,

$$B_A''\left(\frac{\partial e_A}{\partial x_A}\right) = B_A''\left(\frac{\partial NE_A^x}{\partial x_A} + 1\right) = 2D''\left(\frac{\partial NE_A^x}{\partial x_A} + \frac{\partial NE_B^x}{\partial x_A}\right)$$

$$\Rightarrow \frac{\partial NE_A^x}{\partial x_A}\,(B_A'' - 2D'') = -B_A'' + 2D''\frac{\partial NE_B^x}{\partial x_A}. \tag{3.12}$$

Similarly, from country B's first order conditions,

$$\frac{\partial NE_B^x}{\partial x_B}\,(B_B'' - 2D'') = -B_B'' + 2D''\frac{\partial NE_A^x}{\partial x_B} \tag{3.13}$$

We then differentiate eq. (3.10) by x_B,

$$B_A''\left(\frac{\partial e_A}{\partial x_B}\right) = B_A''\left(\frac{\partial NE_A^x}{\partial x_B}\right) = 2D''\left(\frac{\partial NE_A^x}{\partial x_A} + \frac{\partial NE_B^x}{\partial x_A}\right)$$

$$\Rightarrow \frac{\partial NE_A^x}{\partial x_B}\left(B_A'' - 2D''\right) = 2D''\frac{\partial NE_B^x}{\partial x_B}. \tag{3.14}$$

Similarly, from country B's first order conditions,

$$\frac{\partial NE_B^x}{\partial x_A}\left(B_B'' - 2D''\right) = 2D''\frac{\partial NE_A^x}{\partial x_A}. \tag{3.15}$$

Substituting eq. (3.15) into eq. (3.12), we can determine the sign of $\frac{\partial NE_A^x}{\partial x_A}$,

$$\frac{\partial NE_A^x}{\partial x_A} = \frac{-B_A''\left(B_B'' - 2D''\right)}{\left[B_A''B_B'' - 2D''\left(B_A'' + B_B''\right)\right]} \tag{3.16}$$

We note that the denominator is positive while the numerator is negative. Therefore, we can deduce from the above expression that $\frac{\partial NE_A}{\partial x_A} < 0$. Similarly, from eqs. (3.13) and (3.14), we can deduce in a similar manner that $\frac{\partial NE_B}{\partial x_B} < 0$. The intuition is trivial: each country's net emissions' commitment decreases as a result of an increase in its own abatement.

Given the sign of $\frac{\partial NE_B}{\partial x_B}$ is negative, we can examine eq. (3.15),

$$\frac{\partial NE_A^x}{\partial x_B} = \frac{2D''}{B_A'' - 2D''}\frac{\partial NE_B^x}{\partial x_B}. \tag{3.17}$$

Given that the first term on the right hand side of eq. (3.17) is negative, we can deduce that $\frac{\partial NE_A}{\partial x_B} > 0$ and in a similar way $\frac{\partial NE_B}{\partial x_A} > 0$. That is, each country increases its own net emissions' commitment given an increase in abatement by the other country. Substituting $\frac{\partial NE_B}{\partial x_B}$ into (3.17) we derive,

$$\frac{\partial NE_A}{\partial x_B} = \frac{2D''}{B_A'' - 2D''} . \frac{-B_B''\left(B_A'' - 2D''\right)}{B_A''B_B'' - 2D''\left(B_A'' + B_B''\right)}. \tag{3.18}$$

We can deduce from the above expression that $\frac{\partial NE_A}{\partial x_B} > 0$. From (3.17) and (3.18) we derive,

$$\frac{\partial NE_A^x}{\partial x_A} - \frac{\partial NE_A^x}{\partial x_B} = -1 \tag{3.19}$$

and in a similar way we can derive,

$$\frac{\partial NE_B^x}{\partial x_B} - \frac{\partial NE_B^x}{\partial x_A} = -1. \tag{3.20}$$

In this case, marginal benefits are equalized, $B_A' = B_B'$, capturing exchange efficiency but not technological efficiency. That is, the ratio of marginal benefits of the two countries is equal to 1 from eq. (3.8), given our assumption that damages are the same across both countries. The effect of a change in a country's abatement on its own net emissions is negative. The cross effect, the effect of a change in a country's abatement on the other country's net emissions, is positive, i.e., when one country increases its abatement, the other country's net emissions commitment increases. From eqs. (3.19) and (3.20), the total effect of the adjustment of x_A and x_B on NE_i^e is -1. That is, the total effect is a weighted sum of the individual effects. When the two countries cooperate under model (M2), the countries collectively choose their net emission levels in the first stage and their levels of abatement individually in the second stage given the agreed upon level of net emissions in the first stage, where model (M2) captures equality in marginal benefits from emissions of both countries.

M3. Information on abatement technologies

In the first stage, similar to the previous section, the two countries agree to cooperate and maximise their joint welfare with respect to net emissions. The maximisation problem is given by,

$$\max_{NE_A, NE_B} \sum_{i \in A,B} B_i\left(e_i\right) - D\left(NE\right) - C_i(x_i). \tag{3.21}$$

where $x_i = e_i - NE_i$, and thus $\frac{\partial x_i}{\partial NE_i} = -1$. The first order conditions are given by,

$$C_i'(x_i\left(NE_i^e\left(e_A, e_B\right), e_i\right)) = 2D'\left(NE_A^e\left(e_A, e_B\right) + NE_B^e\left(e_A, e_B\right)\right). \tag{3.22}$$

The solution of the above yields $NE_A^e\left(e_A, e_B\right)$ and $NE_B^e\left(e_A, e_B\right)$, the agreed upon levels of net emissions which are functions of both countries' emissions.

In the second stage, each country i independently chooses its optimal level of emissions taking the optimal level of net emissions, NE_i, as given, by solving the following maximisation problem,

$$\max_{e_i} B_i\left(e_i\right) - D\left(NE\right) - C_i(x_i\left(NE_i^e\left(e_A, e_B\right), e_i\right)), \tag{3.23}$$

where $NE = \sum_i NE_i^e$ and $x_i = e_i - NE_i^e$. The first order condition for country A is given by,

$$\frac{\partial B_A}{\partial e_A} - \frac{\partial D}{\partial NE}\left(\frac{\partial NE_A^e}{\partial e_A} + \frac{\partial NE_B^e}{\partial e_A}\right) - \frac{\partial C_A}{\partial e_A}\left(\frac{\partial e_A}{\partial e_A} - \frac{\partial NE_A}{\partial e_A}\right) = 0$$

$$\Rightarrow B_A' + (C_A' - D')\frac{\partial NE_A}{\partial e_A} - D'\frac{\partial NE_B}{\partial e_A} - C_A' = 0 \tag{3.24}$$

Similarly, country B's first order condition is given by,

$$B_B' + (C_B' - D')\frac{\partial NE_B}{\partial e_B} - D'\frac{\partial NE_A}{\partial e_B} - C_B' = 0 \tag{3.25}$$

from which we derive the optimal e_A^* and e_B^*.

Taking eq. (3.24) and differentiating both sides by e_A,

$$C_A''\left(\frac{\partial x_A}{\partial e_A}\right) = C_A''\left(1 - \frac{\partial NE_A^e}{\partial e_A}\right) = 2D''\left(\frac{\partial NE_A^e}{\partial e_A} + \frac{\partial NE_B^e}{\partial e_A}\right)$$

$$\Rightarrow C_A'' - \frac{\partial NE_A^e}{\partial e_A}(C_A'' + 2D'') = 2D''\frac{\partial NE_B^e}{\partial e_A}. \tag{3.26}$$

Similarly, from country B's first order condition,

$$C_B'' - \frac{\partial NE_B^e}{\partial e_B}(C_B'' + 2D'') = 2D''\frac{\partial NE_A^e}{\partial e_B}. \tag{3.27}$$

We can differentiate eq. (3.24) by e_B, to obtain,

$$C_A''\left(\frac{\partial x_A}{\partial e_B}\right) = C_A''\left(-\frac{\partial NE_A^e}{\partial e_B}\right) = 2D''\left(\frac{\partial NE_A^e}{\partial e_B} + \frac{\partial NE_B^e}{\partial e_B}\right)$$

$$\Rightarrow \frac{\partial NE_A^e}{\partial e_B} = \frac{-2D''}{C_A'' + 2D''} \frac{\partial NE_B^e}{\partial e_B}. \tag{3.28}$$

Similarly,

$$\frac{\partial NE_B^e}{\partial e_A} = \frac{-2D''}{C_B'' + 2D''} \frac{\partial NE_A^e}{\partial e_A}. \tag{3.29}$$

Substituting eq. (3.29) into eq. (3.26),

$$C_A'' - \frac{\partial NE_A^e}{\partial e_A} \left(C_A'' + 2D''\right) = 2D'' \frac{-2D''}{C_B'' + 2D''} \frac{\partial NE_A^e}{\partial e_A} \tag{3.30}$$

$$\Rightarrow \left[\left(C_A'' + 2D''\right)\left(C_B'' + 2D''\right) - \left(2D''\right)^2\right] \frac{\partial NE_A^e}{\partial e_A} = C_A''\left(C_B'' + 2D''\right) \tag{3.31}$$

$$\Rightarrow \frac{\partial NE_A^e}{\partial e_A} = \frac{C_A''\left(C_B'' + 2D''\right)}{C_A''C_B'' + 2D''\left(C_A'' + C_B''\right)}. \tag{3.32}$$

We can deduce from the above expression that $\frac{\partial NE_A^e}{\partial e_A} > 0$. Similarly we get $\frac{\partial NE_B^e}{\partial e_B} > 0$. Substituting (3.32) into eq. (3.29) yields,

$$\frac{\partial NE_A}{\partial e_B} = \frac{-2D''}{C_A'' + 2D''} . \frac{C_B''\left(C_A'' + 2D''\right)}{C_A''C_B'' + 2D''\left(C_A'' + C_B''\right)}. \tag{3.33}$$

From eq. (3.33), we can deduce that $\frac{\partial NE_A}{\partial e_B} < 0$ and similarly $\frac{\partial NE_B}{\partial e_A} < 0$. Each country reduces its net emissions as the other country increases its emissions. From (3.32) and (3.33) we can conclude that,

$$\frac{\partial NE_A^e}{\partial e_A} - \frac{\partial NE_A^e}{\partial e_B} = 1 \tag{3.34}$$

and similarly,

$$\frac{\partial NE_B^e}{\partial e_B} - \frac{\partial NE_B^e}{\partial e_A} = 1. \tag{3.35}$$

In this case, marginal abatement costs are equalised, $C'_A = C'_B$, capturing technological efficiency, but exchange efficiency is not reached. Technological efficiency occurs as both countries equalise their marginal cost of abatement. That is, the ratio of marginal abatement costs of the two countries is equal to 1 from eq. (3.22), given our assumption that damages are the same across both countries. The effect of a change in a country's emissions on its own net emissions is positive. The cross effect, the effect of a change in a country's emissions on the other country's net emissions, is negative, i.e., an increase in levels of pollution of country type A (B) will decrease the level of net emissions of country B (A). From eqs. (3.34) and (3.35), the total effect of the adjustment of e_A and e_B on NE_i^e is 1. That means the total effect is a weighted sum of the individual effects. When the two countries cooperate under model (M3), the countries collectively choose their net emission levels in the first stage and their levels of emissions individually in the second stage, given the first stage agreed upon net emission levels, where model (M3) captures equality in marginal abatement costs of both countries.

3.3 An Illustrative Example

We consider the following quadratic benefit function for country i, $B_i(e_i) =$

$b_i\left(e_i - \frac{1}{2}e_i^2\right)$, where b_i is a positive parameter for country i. We assume quadratic abatement cost function, $C_i(x_i) = \frac{1}{2}c_i x_i^2$, where $i \in \{A, B\}$, and c_i is a type specific positive parameter. Each country i suffers from damages that depend on aggregate net emissions, $NE = E - X$, where $E = \sum_{i\in\{A,B\}} e_i$. That is, aggregate net emissions are defined as the difference between global emissions and global abatement. The quadratic damage function considered for each country is given by $D(NE) = \frac{1}{2}d(NE)^2 = \frac{1}{2}d\left[\sum_{i\in A,B}(e_i - x_i)\right]^2$, where d is a positive parameter. As mentioned earlier, to simplify the analysis and focus on differences in emission benefits and abatement technologies, we assume d is common to both countries, which is similar to the model utilised by Diamantoudi, Sartzetakis, and Strantza (2022) in the absence of adaptation.

If we substitute the specific functional forms mentioned above for benefit, damage and abatement cost functions, the social welfare of country of i is defined as,

$$W_i = b_i\left(e_i - \frac{1}{2}(e_i)^2\right) - \frac{1}{2}d\left(\sum_{i\in A,B}(e_i - x_i)\right)^2 - \frac{1}{2}c_i(x_i)^2. \tag{3.36}$$

The next section presents the solutions of the four cases given the specific functional forms in eq. (3.36).

M0. The non-cooperative pure Nash case

Each country simultaneously chooses its own emission and abatement levels taking the other country's emission and abatement levels as given. It is assumed that the two countries act independently. Country i chooses its optimal levels of emissions and abatement by maximising its welfare given in (3.36). Each country determines its level of emissions by equating its marginal benefit from its emissions to the marginal damage inflicted on it when global emissions increases as a result of its own emissions increase. Each country also determines its levels of abatement by equating its marginal abatement cost to the decreased marginal damage as net emissions decrease due to its increase in abatement. The solution of the two countries' first-order conditions yield;

$$e_A^{nc} = \frac{\sigma_A + \sigma_B + \sigma_A\sigma_B\left(\gamma_B - \gamma_A + 1\right)}{\sigma_A + \sigma_B + \sigma_A\sigma_B\left(\gamma_A + \gamma_B + 1\right)}, \tag{3.37}$$

$$e_B^{nc} = \frac{\sigma_A + \sigma_B + \sigma_A\sigma_B\left(\gamma_A - \gamma_B + 1\right)}{\sigma_A + \sigma_B + \sigma_A\sigma_B\left(\gamma_A + \gamma_B + 1\right)}, \tag{3.38}$$

$$x_A^{nc} = \frac{2\sigma_B}{\sigma_A + \sigma_B + \sigma_A\sigma_B\left(\gamma_A + \gamma_B + 1\right)}, \tag{3.39}$$

$$x_B^{nc} = \frac{2\sigma_A}{\sigma_A + \sigma_B + \sigma_A\sigma_B\left(\gamma_A + \gamma_B + 1\right)}, \tag{3.40}$$

where $\gamma_i = \frac{d}{b_i}$ and $\sigma_i = \frac{c_i}{d}$. Therefore, each country i's net emissions, $NE_i^{nc} = e_i^{nc} - x_i^{nc}$, are,

$$NE_A^{nc} = \frac{\sigma_A - \sigma_B + \sigma_A\sigma_B\left(\gamma_B - \gamma_A + 1\right)}{\sigma_A + \sigma_B + \sigma_A\sigma_B\left(\gamma_A + \gamma_B + 1\right)}, \tag{3.41}$$

$$NE_B^{nc} = \frac{\sigma_B - \sigma_A + \sigma_A\sigma_B\left(\gamma_A - \gamma_B + 1\right)}{\sigma_A + \sigma_B + \sigma_A\sigma_B\left(\gamma_A + \gamma_B + 1\right)}. \tag{3.42}$$

The aggregate net emissions level, $NE^{nc} = NE_A^{nc} + NE_B^{nc}$, is,

$$NE^{nc} = \frac{2\sigma_A\sigma_B}{\sigma_A + \sigma_B + \sigma_A\sigma_B\left(\gamma_A + \gamma_B + 1\right)}. \tag{3.43}$$

M1. Cooperation on abatement and emissions

In this case, both countries collectively choose optimal levels of emissions and abatement by solving the following maximisation problem,

$$\max_{e_i,x_i} \sum_{i\in A,B} \left[b_i \left(e_i - \frac{1}{2}\left(e_i\right)^2 \right) - \frac{1}{2}d \left(\sum_{i\in A,B} \left(e_i - x_i\right) \right)^2 - \frac{1}{2}c_i\left(x_i\right)^2 \right]. \tag{3.44}$$

Given the specified functional forms, the solution of the two countries' reaction functions yields,

$$e_A^c = \frac{\sigma_A + \sigma_B + \sigma_A\sigma_B\left(\frac{1}{2} + \gamma_B - \gamma_A\right)}{\sigma_A + 2\sigma_B + \sigma_A\sigma_B\left(\frac{1}{2} + \gamma_B + \gamma_A\right)}, \tag{3.45}$$

$$e_B^c = \frac{\sigma_A + \sigma_B + \sigma_A\sigma_B\left(\frac{1}{2} - \gamma_B + \gamma_A\right)}{\sigma_A + \sigma_B + \sigma_A\sigma_B\left(\frac{1}{2} + \gamma_B + \gamma_A\right)}, \tag{3.46}$$

$$x_A^c = \frac{2\sigma_B}{2\sigma_A + 2\sigma_B + \sigma_A\sigma_B\left(\frac{1}{2} + 2\gamma_B + 2\gamma_A\right)}, \tag{3.47}$$

$$x_B^c = \frac{2\sigma_A}{\sigma_A + \sigma_B + \sigma_A\sigma_B\left(\frac{1}{2} + \gamma_B + \gamma_A\right)}. \tag{3.48}$$

Therefore, each country i's net emissions, $NE_i^c = e_i^c - x_i^c$, are,

$$NE_A^c = \frac{\sigma_A - \sigma_B + \sigma_A\sigma_B\left(\frac{1}{2} + \gamma_B - \gamma_A\right)}{\sigma_A + \sigma_B + \sigma_A\sigma_B(\frac{1}{2} + \gamma_B + \gamma_A)}, \quad (3.49)$$

$$NE_B^c = \frac{\sigma_B - \sigma_A + \sigma_A\sigma_B\left(\frac{1}{2} - \gamma_B + \gamma_A\right)}{\sigma_A + \sigma_B + \sigma_A\sigma_B(\frac{1}{2} + \gamma_B + \gamma_A)}. \quad (3.50)$$

Aggregate net emissions levels, NE^c, are

$$NE^c = \frac{\sigma_A\sigma_B}{\sigma_A + \sigma_B + \sigma_A\sigma_B(\frac{1}{2} + \gamma_B + \gamma_A)}. \quad (3.51)$$

Direct comparison of total net emissions in the two benchmark cases, given in (3.51) to (3.41), reveal that $NE^{nc} > NE^c$: aggregate net emissions are reduced under the cooperative case. The next section covers the cases of two countries collectively choosing their net emission levels in the first stage, then individually maximising their welfare in the second stage to find their levels of abatement and emissions, given the level of net emissions in the first stage.

M2. Information on preferences

The joint maximisation in the first stage given the specific functional forms for benefit, damage and abatement cost functions is given by,

$$\max_{NE_A, NE_B} \sum_{i \in A,B} \left[b_i \left(e_i - \frac{1}{2}(e_i)^2 \right) - \frac{1}{2} d \left(\sum_i NE_i \right)^2 - \frac{1}{2} c_i (x_i)^2 \right] \quad (3.52)$$

where $e_i = NE_i + x_i$. The maximisation problem in the first stage can be rewritten as,

$$\max_{NE_A, NE_B} \quad b_A\left(e_A - \tfrac{1}{2}\left(e_A\right)^2\right) + b_B\left(e_B - \tfrac{1}{2}\left(e_B\right)^2\right)$$
$$-\tfrac{1}{2}c_A\left(x_A\right)^2 - \tfrac{1}{2}c_B\left(x_B\right)^2 - d\left(NE_A + NE_B\right)^2. \tag{3.53}$$

The first-order conditions provide the solution to countries' net emissions, as functions of both countries' choice of abatement,

$$NE_A^x\left(x_A, x_B\right) = \frac{2d\left(b_A - b_B\right) + b_A b_B + 2d b_B x_B - x_A(b_A b_B + 2d b_A)}{2d\left(b_A + b_B\right) + b_A b_B}, \tag{3.54}$$

and

$$NE_B^x\left(x_A, x_B\right) = \frac{2d\left(b_B - b_A\right) + b_A b_B + 2d b_A x_A - x_B(b_A b_B + 2d b_B)}{2d\left(b_A + b_B\right) + b_A b_B}. \tag{3.55}$$

Our example with quadratic functions verifies the results we derived using general functions. First, the effect on its own net emissions is negative but less than -1, given by,

$$\frac{\partial NE_i^x}{\partial x_i} = -\left(\frac{b_i b_j + 2d b_i}{2d(b_i + b_j) + b_i b_j}\right) \tag{3.56}$$

We can also derive the value of the cross effect, which is positive, given by,

$$\frac{\partial NE_j^x}{\partial x_i} \Rightarrow \frac{2d b_i}{2d\left(b_i + b_j\right) + b_i b_j} > 0 \tag{3.57}$$

Thus, an increase in the level of abatement of country type A (B) will increase the level of net emissions of country B (A). For given net emissions, this also indicates that country B (A) will pollute more. We can see that the denominator in both $\frac{\partial NE_i^x}{\partial x_i}$ and $\frac{\partial NE_i^x}{\partial x_j}$ is greater than the numerator, $2d\left(b_i + b_j\right) + b_i b_j > b_i b_j + 2db_i$ and $2d\left(b_i + b_j\right) + b_i b_j > b_i b_j + 2db_j$. Thus, we conclude that $-1 < \frac{\partial NE_i^x}{\partial x_i} < 0$ and $0 < \frac{\partial NE_i^x}{\partial x_j} < 1$. Note that the sum of their absolute values equals to unity, that is, $|\frac{\partial NE_i^x}{\partial x_i}| + |\frac{\partial NE_i^x}{\partial x_j}| = 1$. Thus, in our setup the total effect of changes in both countries' abatement choices on the net emissions of one of them is the weighted sum of the individual effects.

Given net emission levels in the first stage, each country simultaneously chooses emission levels in the second stage by maximizing its individual welfare given by,

$$\max_{x_i} b_i \left(e_i - \frac{1}{2}\left(e_i\right)^2\right) - \frac{1}{2}d\left(NE_i^x + NE_j^x\right)^2 - \frac{1}{2}c_i\left(x_i\right)^2, \tag{3.58}$$

where $i \neq j$. After computing emission levels, each country deduces its corresponding level of abatement satisfying the net emissions constraint, $NE_i = e_i - x_i$. Thus, the equilibrium levels computed under (M2) are denoted by e_i^x and x_i^x and given by,

$$e_A^x = \frac{\sigma_A + \sigma_B + 4(\sigma_B\gamma_A + \gamma_B\sigma_A) + \sigma_A\sigma_B\left(2\gamma_A - 2\gamma_B + 1\right)\left(2\gamma_A + 2\gamma_B + 1\right)}{\sigma_A + \sigma_B + 4(\sigma_B\gamma_A + \gamma_B\sigma_A) + \sigma_A\sigma_B\left(2\gamma_A + 2\gamma_B + 1\right)^2}, \tag{3.59}$$

$$e_B^x = \frac{\sigma_A + \sigma_B + 4(\sigma_B\gamma_A + \gamma_B\sigma_A) + \sigma_A\sigma_B\left(2\gamma_B - 2\gamma_A + 1\right)\left(2\gamma_A + 2\gamma_B + 1\right)}{\sigma_A + \sigma_B + 4(\sigma_B\gamma_A + \gamma_B\sigma_A) + \sigma_A\sigma_B\left(2\gamma_A + 2\gamma_B + 1\right)^2}, \tag{3.60}$$

$$x_A^x = \frac{2\sigma_B\left(4\gamma_A + 1\right)}{\sigma_A + \sigma_B + 4(\sigma_B\gamma_A + \gamma_B\sigma_A) + \sigma_A\sigma_B\left(2\gamma_A + 2\gamma_B + 1\right)^2}, \tag{3.61}$$

$$x_B^x = \frac{2\sigma_A\left(4\gamma_B + 1\right)}{\sigma_A + \sigma_B + 4(\sigma_B\gamma_A + \gamma_B\sigma_A) + \sigma_A\sigma_B\left(2\gamma_A + 2\gamma_B + 1\right)^2}. \tag{3.62}$$

In this case, net emissions for each country, $NE_i^x = e_i^x - x_i^x$, are

$$NE_A^x = \frac{\sigma_A - \sigma_B + 4(\gamma_B\sigma_A - \sigma_B\gamma_A) + \sigma_A\sigma_B\left(2\gamma_A - 2\gamma_B + 1\right)\left(2\gamma_A + 2\gamma_B + 1\right)}{\sigma_A + \sigma_B + 4(\sigma_B\gamma_A + \gamma_B\sigma_A) + \sigma_A\sigma_B\left(2\gamma_A + 2\gamma_B + 1\right)^2}, \tag{3.63}$$

$$NE_B^x = \frac{\sigma_B - \sigma_A + 4(\sigma_B\gamma_A - \gamma_B\sigma_A) + \sigma_A\sigma_B\left(2\gamma_B - 2\gamma_A + 1\right)\left(2\gamma_A + 2\gamma_B + 1\right)}{\sigma_A + \sigma_B + 4(\sigma_B\gamma_A + \gamma_B\sigma_A) + \sigma_A\sigma_B\left(2\gamma_A + 2\gamma_B + 1\right)^2}. \tag{3.64}$$

M3. Information on abatement technologies

The maximisation problem in the first stage can be written as,

$$\max_{NE_A, NE_B} \quad b_A\left(e_A - \tfrac{1}{2}\left(e_A\right)^2\right) + b_B\left(e_B - \tfrac{1}{2}\left(e_B\right)^2\right) - \tfrac{1}{2}c_A\left(e_A - NE_A\right)^2 - \tfrac{1}{2}c_B\left(e_B - NE_B\right)^2 - d\left(NE_A + NE_B\right)^2. \tag{3.65}$$

The first order conditions provide the solution to countries' net emissions, as a function of both countries' emission level e_A and e_B,

$$NE_A^e\left(e_A, e_B\right) = \frac{c_A e_A\left(2d + c_B\right) - 2dc_B e_B}{2d\left(c_A + c_B\right) + c_A c_B} \tag{3.66}$$

and

$$NE_B^e\left(e_A, e_B\right) = \frac{c_B e_B\left(2d + c_A\right) - 2dc_A e_A}{2d\left(c_A + c_B\right) + c_A c_B} \tag{3.67}$$

In this case, we can deduce that second term in both FOCs (3.66) and (3.67) is negative, since

$$\frac{\partial NE_i^e}{\partial e_j} \Rightarrow -\frac{2dc_j}{2dc_j + 2dc_i + c_i c_j} < 0 \tag{3.68}$$

Thus, an increase in levels of pollution of country type A (B) will decrease the level of net emissions of country B (A). We also have that:

$$\frac{\partial NE_i^e}{\partial e_i} \Rightarrow \frac{c_i\left(2d + c_j\right)}{2dc_i + 2dc_j + c_i c_j} > 0 \tag{3.69}$$

The denominator in both $\frac{\partial NE_i^e}{\partial e_i}$ and $\frac{\partial NE_i^e}{\partial e_j}$ is greater than the numerator, $2dc_i + 2dc_j + c_i c_j > c_i\left(2d + c_j\right)$ and $2dc_i + 2dc_j + c_i c_j > c_j\left(2d + c_i\right)$. Thus, we conclude that $|\frac{\partial NE_i^e}{\partial e_i}| + |\frac{\partial NE_i^e}{\partial e_j}| = 1$. Thus, the total effect in this case of changes in both countries' emission choices on the net emissions of one of them is the weighted sum of the individual effects.

Given net emission levels in the first stage, each country simultaneously chooses emission levels in the second stage by maximizing its individual welfare given by,

$$\max_{e_i} b_i\left(e_i - \frac{1}{2}\left(e_i\right)^2\right) - \frac{1}{2}d\left(NE_i^e + NE_j^e\right)^2 - \frac{1}{2}c_i\left(e_i - NE_i^e\right)^2 \tag{3.70}$$

where $i \neq j$. After computing emission levels, each country deduces its corresponding level of abatement satisfying the net emissions constraint, $NE_i = e_i - x_i$. Thus, the equilibrium levels computed under (M3) are denoted by e_i^e and x_i^e and given by,

$$e_A^e = \frac{\sigma_A^2 + \sigma_B^2 + \sigma_A\sigma_B\left[\sigma_A\left(1+\gamma_B\right) + \sigma_B\left(1-\gamma_A\right) + \frac{1}{4}\sigma_A\sigma_B(1-\gamma_A+\gamma_B) + 2\right]}{\sigma_A^2 + \sigma_B^2 + \sigma_A\sigma_B\left[\sigma_A\left(1+\gamma_B\right) + \sigma_B\left(1+\gamma_A\right) + \frac{1}{4}\sigma_A\sigma_B(1+\gamma_A+\gamma_B) + 2\right]} \tag{3.71}$$

$$e_B^e = \frac{\sigma_A^2 + \sigma_B^2 + \sigma_A\sigma_B\left[\sigma_A\left(1-\gamma_B\right) + \sigma_B\left(1+\gamma_A\right) + \frac{1}{4}\sigma_A\sigma_B(1+\gamma_A-\gamma_B) + 2\right]}{\sigma_A^2 + \sigma_B^2 + \sigma_A\sigma_B\left[\sigma_A\left(1+\gamma_B\right) + \sigma_B\left(1+\gamma_A\right) + \frac{1}{4}\sigma_A\sigma_B(1+\gamma_A+\gamma_B) + 2\right]} \tag{3.72}$$

$$x_A^e = \frac{\sigma_B\left[\sigma_A\sigma_B + 2\left(\sigma_A+\sigma_B\right)\right]}{\sigma_A^2 + \sigma_B^2 + \sigma_A\sigma_B\left[\sigma_A\left(1+\gamma_B\right) + \sigma_B\left(1+\gamma_A\right) + \frac{1}{4}\sigma_A\sigma_B(1+\gamma_A+\gamma_B) + 2\right]} \tag{3.73}$$

$$x_B^e = \frac{\sigma_A\left[\sigma_A\sigma_B + 2\left(\sigma_A+\sigma_B\right)\right]}{\sigma_A^2 + \sigma_B^2 + \sigma_A\sigma_B\left[\sigma_A\left(1+\gamma_B\right) + \sigma_B\left(1+\gamma_A\right) + \frac{1}{4}\sigma_A\sigma_B(1+\gamma_A+\gamma_B) + 2\right]} \tag{3.74}$$

In this case, net emissions for each country, $NE_i = e_i^{ne} - x_i^{ne}$, are computed to be,

$$NE_A^e = \frac{\left[\begin{array}{c}\sigma_A^2 + \left(1-\sigma_A\right)\sigma_B^2 - 2\sigma_B\left(\sigma_A+\sigma_B\right) \\ +\sigma_A\sigma_B\left[\sigma_A\left(1+\gamma_B\right) + \sigma_B\left(1-\gamma_A\right) + \frac{1}{4}\sigma_A\sigma_B(1-\gamma_A+\gamma_B) + 2\right]\end{array}\right]}{\sigma_A^2 + \sigma_B^2 + \sigma_A\sigma_B\left[\sigma_A\left(1+\gamma_B\right) + \sigma_B\left(1+\gamma_A\right) + \frac{1}{4}\sigma_A\sigma_B(1+\gamma_A+\gamma_B) + 2\right]} \tag{3.75}$$

$$NE_B^e = \frac{\left[\begin{array}{c}(1-\sigma_B)\,\sigma_A^2+\sigma_B^2-2\sigma_A\,(\sigma_A+\sigma_B)\\ +\sigma_A\sigma_B\left[\sigma_A\,(1-\gamma_B)+\sigma_B\,(1+\gamma_A)+\frac{1}{4}\sigma_A\sigma_B(1+\gamma_A-\gamma_B)+2\right]\end{array}\right]}{\sigma_A^2+\sigma_B^2+\sigma_A\sigma_B\left[\sigma_A\,(1+\gamma_B)+\sigma_B\,(1+\gamma_A)+\frac{1}{4}\sigma_A\sigma_B(1+\gamma_A+\gamma_B)+2\right]} \tag{3.76}$$

The next section presents numerical simulations to compare all four cases.

3.4 Total Welfare and Stability Analysis

The total social welfare, ω, is the sum of indirect welfare of countries A and B, ω_A and ω_B, given by,

$$\omega = \omega_A + \omega_B. \tag{3.77}$$

In order to examine the stability of any of the two types of agreements examined above, we compare the welfare each country achieves under the agreement to the welfare they get under the non-cooperative Nash case assuming that if the agreement break down, they revert to the Nash outcome. That is, the solution in M2 and M3 is stable for country A and country B if,

$$\omega_i\,(M2) > \omega_i\,(M0)\,, \tag{3.78}$$

and

$$\omega_i\,(M3) > \omega_i\,(M0)\,. \tag{3.79}$$

It is clear that if none of the two countries wants to leave the coalition under either of the two setups, the coalition increases aggregate welfare relative to the non-cooperative scenario. That is,

$$\omega_A(M2) + \omega_B(M2) > \omega_A(M0) + \omega_B(M0) \tag{3.80}$$

and

$$\omega_A(M3) + \omega_B(M3) > \omega_A(M0) + \omega_B(M0). \tag{3.81}$$

Given the complexity of the expressions, we resort to simulations. To illustrate the benefits of cooperation on net emissions, the analysis starts with the assumption of homogeneity. This allows clarity in visualising emission, abatement, and net emission strategies before heterogeneity is introduced. The simulations that follow present three different case studies under the heterogeneity assumption: (1) in the first case, the two countries are identical in their emission benefits but have different abatement technologies. (2) In the second case, countries have identical abatement technologies but different emission benefits. The final case assumes both different emission benefits and abatement technologies across the two countries and presents two simulations where in one, (M3) dominates (M2) and in the other, (M2) dominates (M3). The results are compared to the case of the pure non-cooperative Nash case (M0) and the full cooperative case (M1).

Starting with the special case of homogeneity, i.e., both countries are identical in their benefits and costs, we assume the following values for the parameters: $b_A =$

$b_B = 10$, $d = 0.4$, and $c_A = c_B = 3$, hence $\gamma_A = \gamma_B = 0.04$ and $\sigma_A = \sigma_B = 7.5$. Table 3.1 illustrates the levels of emissions, abatement, individual and total net emissions, and individual and total indirect welfare under the four cases (M0), (M1), (M2) and (M3).

Table 3.1 Homogenous case
b_A=b_B = 10, d = 0.4, c_A =c_B = 3

	M0	M1	M2	M3
e_A	0.94	0.9055	0.8878	0.9504
e_B	0.94	0.9055	0.8878	0.9504
x_A	0.198	0.315	0.186	0.3305
x_B	0.198	0.315	0.186	0.3305
NE_A	0.742	0.59	0.7	0.6198
NE_B	0.742	0.59	0.7	0.6198
NE	1.485	1.18	1.4	1.239
ω_A	4.48	4.527	4.491	4.516
ω_B	4.48	4.527	4.491	4.516
ω	8.964	9.054	8.982	9.032

It is clear that countries' net emissions get their higher value under the pure Nash case. Given that each country is free-riding on the other country's emissions reduction, the aggregate net emissions, in this case, are $NE^{nc} = 1.485$. This leads to an aggregate welfare level, $\omega^{nc} = 8.964$. In the case of full cooperation under (M1), both countries emit less, $e^c = 0.9055$, and abate more, $x^c = 0.315$, compared to (M0), thus net emissions are lowest under (M1) with $NE^c = 1.181$. This generates an aggregate welfare, $\omega^c = 9.054$. This is straightforward as countries reduce their emissions and abate more when they cooperate. Global damages are reduced, given the reduction in net emissions.

In the case of cooperation on net emissions with revealed information on emissions (M2), countries agree on a level of net emissions closer to values of the Nash case, $NE^x = 0.7$, both countries reduce their emissions, $e^x = 0.8878$, but choose lower abatement levels, $x^x = 0.186$. In the case of cooperation on net emissions with revealed information on abatement (M3), both countries increase their emissions further, $e^e = 0.9504$, but this is accompanied with higher abatement levels, $x^e = 0.3305$. Thus, the level of net emissions under (M3) produces a lower aggregate level of net emissions compared to (M2), $NE^e = 0.619$, a net emissions level closer to the (M1) case. These results are particular to the choice of the parameter values. The result that model M3 welfare dominates model M2 is reversed if the cost of abatement is higher or benefits from emissions are lower. A numerical simulation showing M2 dominating M3 is presented in the final heterogeneous case. Given that both countries' net emissions are lower, the welfare of both countries and total welfare are higher compared to (M0). With symmetry, (M3) achieves 99.76 % of the (M1) total welfare, and (M2) achieves 99.2 % of the (M1) total welfare. Countries abate more under cooperation on net emissions in (M3), leading to lower aggregate net emissions compared to (M0) and generate welfare levels close to (M1).

Table 3.2 Heterogenous case in abatement technology
b_A=b_B = 10, d = 0.4, c_A =2, c_B = 4

	M0	M1	M2	M3
e_A	0.942	0.897	0.89	0.946
e_B	0.942	0.897	0.89	0.958
x_A	0.289	0.341	0.273	0.47
x_B	0.144	0.17	0.136	0.23
NE_A	0.652	0.454	0.616	0.47
NE_B	0.797	0.681	0.753	0.72
NE	1.449	1.135	1.369	1.19
ω_A	4.478	4.493	4.489	4.475
ω_B	4.521	4.597	4.527	4.594
ω	9.0	9.09	9.016	9.069

Table 3.2 illustrates the case of heterogeneity in abatement technology. We use the same values for the benefit and damage, $b_A = b_B = 10$, $d = 0.4$, and we differentiate the cost parameters such that their average value is the same as in the homogeneous case $c_A = 2$, and $c_B = 4$. Hence $\gamma_A = \gamma_B = \frac{d}{b} = 0.04$, $\sigma_A = \frac{c_A}{d} = 5$ and $\sigma_B = \frac{c_B}{d} = 10$. The aggregate net emissions in the non-cooperative Nash case is computed to be $NE^{nc} = 1.449$. This leads to an aggregate welfare level, $\omega^{nc} = 9.0$. Given that the two countries have identical benefits from emissions, under models (M0), (M1), and (M2), emissions of both countries are equal ($e_A^{nc} = e_B^{nc} = 0.942$, $e_A^c = e_B^c = 0.897$, and $e_A = e_B = 0.89$), which is not the case under (M3). In the model (M3), countries, in determining their agreement on net emissions, take into account the difference in their abatement technologies; thus, their emissions are different. Given that country B has higher abatement costs, $c_B = 4$, it not only abates less than country A but also

emits slightly higher than country A. With asymmetry in abatement cost, the (M3) model achieves higher benefits since it also sets marginal cost of abatement equal across countries. In the case of cooperation on net emissions (M2), aggregate welfare also improves relative to the Nash equilibrium. However, the welfare improvement achieved is not as high as that achieved by model (M3) since the latter leads also to equalisation of marginal abatement cost across countries. While both models (M2) and (M3) bring total welfare gains to both countries, this specific example shows that (M3) is not stable as country A would deviate to the Nash equilibrium.

Table 3.3 Heterogenous case in emission benefits
b_A=14, b_B = 6, d = 0.4, c_A =c_B = 3

	M0	M1	M2	M3
e_A	0.958	0.933	0.921	0.965
e_B	0.902	0.845	0.816	0.918
x_A	0.195	0.309	0.171	0.327
x_B	0.195	0.309	0.194	0.327
NE_A	0.762	0.624	0.75	0.637
NE_B	0.706	0.535	0.62	0.59
NE	1.468	1.159	1.37	1.227
ω_A	6.498	6.55	6.536	6.52
ω_B	2.482	2.515	2.465	2.517
ω	8.981	9.071	9.002	9.046

Table 3.3 illustrates the case of heterogeneity in emission benefits; the parameters utilised are as follows: $c_A = c_B = 3$, $d = 0.4$, and $b_A = 14$, and $b_B = 6$, keeping the same average as in the homogenous case. Hence, $\sigma_A = \sigma_B = \frac{c}{d} = 7.5$, $\gamma_A = \frac{d}{b_A} =$

0.0285 and $\gamma_B = \frac{d}{b_B} = 0.0666$. The aggregate net emissions computed in this case is $NE^{nc} = 1.468$. This leads to an aggregate welfare level, $\omega^{nc} = 8.981$. Comparing (M1) to (M0), net emissions are lower under the full cooperative case, $NE^c = 1.159$. This generates an aggregate welfare, $\omega^c = 9.071$. Given that the two countries have identical abatement technologies, under models (M0), (M1), and (M3), abatement levels for both countries are equal ($x_A^{nc} = x_B^{nc} = 0.195$, $x_A^c = x_B^c = 0.309$, and $x_A = x_B = 0.327$). This is not the case under (M2), since countries' benefits from emissions are different and, thus, when collectively maximising their joint welfare, country B abates, $x_B^{ne} = 0.194$, which is higher than country A, $x_A^{ne} = 0.171$. Emissions levels for country A are higher due to the large benefits from emissions, thus attains a much higher welfare than country B.

In the case of cooperation on net emissions (M2), compared to (M1), both countries reduce their emissions and choose lower levels of abatement. Compared to (M0), aggregate welfare under (M2) is $\omega^x = 9.03$, this identifies an improvement in total welfare compared to the non-cooperative case with countries emitting less and abating more, thus achieving lower levels of global net emissions. In the case of cooperation on net emission under (M3), $\omega^e = 9.046$ reports a total welfare 99.73 % that of the full cooperative case (M1) as both countries' marginal costs of abatement are equalised and abatement levels are higher under (M3) compared to (M2) in such scenario. In this example, given the large benefits from emissions for country A, model (M2) is not stable as country B wants to deviate.

Table 3.4 Heterogenous case in emission benefit and abatement technology (M3>M2)
b_A=14, b_B = 6, d = 0.4, c_A =2, c_B = 4

	M0	M1	M2	M3
e_A	0.959	0.936	0.922	0.961
e_B	0.904	0.851	0.82	0.93
x_A	0.286	0.446	0.252	0.473
x_B	0.143	0.223	0.143	0.236
NE_A	0.672	0.489	0.67	0.488
NE_B	0.761	0.627	0.676	0.694
NE	1.433	1.116	1.346	1.182
ω_A	6.495	6.522	6.53	6.486
ω_B	2.52	2.58	2.499	2.593
ω	9.015	9.106	9.03	9.079

Table 3.4 illustrates the case of heterogeneity in both emissions benefits and abatement technology, the parameters utilised are as follows: $b_A = 14$, $b_B = 6$, $c_A = 2$, $c_B = 4$ and $d = 0.4$, hence $\gamma_A = 0.0285$, $\gamma_B = 0.0666$, $\sigma_A = 5$, and $\sigma_B = 10$. Aggregate net emissions get their highest value under (M0) at $NE^{nc} = 1.433$. This leads to an aggregate welfare level, $\omega^{nc} = 9.015$. Net emissions get their lowest value under the cooperative case, $NE^c = 1.116$. This generates an aggregate welfare $\omega^c = 9.106$. In the case of cooperation on net emissions (M2), both countries reduce their emissions and choose lower abatement levels compared to (M1). The level of total welfare under (M2) is computed to be $\omega^x = 9.03$, and that of (M3) to be $\omega^e = 9.079$. Country A has higher benefits, so when the agreement is based on equalizing marginal benefits, it emits more than country B. Country A also has lower abatement cost, so (M3), by equalizing marginal abatement cost between countries, pushes country A to do much

more abatement and thus can emit more. In such a scenario, when one country has the advantage in both benefits from emissions and abatement costs, cooperation with a less technologically developed country can be welfare improving. In addition, given low abatement costs, model (M3) equalizes the marginal abatement costs of both countries and this leads to higher abatement by both countries. That is, even when country B has a higher abatement cost, country B cleans up more under (M3) compared to (M0), (M1), and (M2). Given the heterogeneity in both emission benefits and abatement technology, the gains from cooperation are present for both countries, and (M3) generates a higher welfare than (M2).

We now present a simulation in which (M2) dominates (M3), given lower benefits from emissions and higher abatement costs for each country.

Table 3.5 Heterogenous case in emission benefit and abatement technology (M2>M3)
b_A=4, b_B = 2, d = 0.4, c_A =8, c_B = 12

	M0	M1	M2	M3
e_A	0.855	0.773	0.762	0.859
e_B	0.71	0.547	0.524	0.734
x_A	0.072	0.113	0.052	0.136
x_B	0.0481	0.075	0.044	0.091
NE_A	0.783	0.66	0.71	0.722
NE_B	0.662	0.471	0.479	0.643
NE	1.445	1.131	1.2	1.365
ω_A	1.519	1.589	1.592	1.512
ω_B	0.484	0.504	0.478	0.506
ω	2.003	2.094	2.071	2.018

Table 3.5 illustrates a second case of heterogeneity in both emissions benefits and abatement technology, the parameters utilised are as follows: $b_A = 4$, $b_B = 2$, $c_A = 8$, $c_B = 12$ and $d = 0.4$, hence $\gamma_A = 0.1$, $\gamma_B = 0.2$, $\sigma_A = 20$, and $\sigma_B = 30$. Note that in this example, the benefit parameters are much lower while the cost parameters are much higher relative to the example we considered up to now. In particular, the average benefit parameter is 3 instead of 10, while the average abatement cost parameter is now 10 instead of 3. Aggregate net emissions under (M0) are highest at $NE^{nc} = 1.445$. This leads to an aggregate welfare level, $\omega^{nc} = 2.003$. Given the higher cost of abatement for both countries, net emissions' commitments of both countries under (M2), $NE = 1.2$ are lower than that of (M3), $NE = 1.365$. In the case of cooperation on net emissions (M2), this generates an aggregate welfare $\omega = 2.071$, where both countries reduce their emissions but choose lower abatement levels compared to (M1). The level of total welfare under (M1) is computed to be $\omega^c = 2.094$, and that of (M3) to be $\omega = 2.018$. In such a case, (M2) dominates (M3) and generates a welfare closer to (M1).

Model (M2) equalizes the marginal benefits from emissions of both countries, thus, given both countries' lower benefits from emissions, the agreement under (M2) achieves a welfare level closer to (M1), specifically 98.9% of (M1)'s welfare. Country A has a lower abatement cost, so model (M3) by equalizing marginal abatement cost between countries, pushes country A to do much more abatement. Given that country A also has a higher benefit from emissions, it also emits more than country B. Given the high abatement costs present, model (M2) equalizes the marginal benefits of both countries, and this leads to a lower level of emissions chosen by both countries.

That is, even when country B has a lower benefit from emissions, country B emits less under (M2) compared to (M0), (M1), and (M3), and thus abates less. Given higher abatement costs, model (M2) captures lower levels of net emissions and thus outperforms model (M3). Given the heterogeneity in both emission benefits and abatement technology, the gains from cooperation are present for both countries under both (M2) and (M3), and with the parameters values chosen, we show that (M2) dominates (M3).The last two examples show that country A wants to deviate under model (M3) and country B wants to deviate under model (M2).

Given the simulations above, in the symmetric case, depending on the values of b and c, one could have (M2) or (M3) closer to (M1). Let's assume that (M3) is closer to (M1), then if there is asymmetry only in c's, (M3) performs even better since it equates the marginal cost of abatement between countries. If asymmetry is in b's, then the performance of (M2) improves. The simulations identify welfare levels close to (M1), with asymmetry in emission benefits and cost of abatement, (M3) achieves 99.94% of the (M1) total welfare, and (M2) performs better if b is low or c is high. If c is low, the total welfare of (M3) outweighs that of (M2). The results of (M2) and (M3) illustrate that when benefits from emissions are higher than the cost of abatement, (M3) equalizes the marginal costs, and model (M3) generates a higher welfare than (M2). This is due to the lower level of net emissions chosen. With a similar intuition, when benefits from emissions are lower than the cost of abatement, (M2) generates a higher welfare than (M3) as the level of net emissions chosen under (M2) are lower.

3.5 Conclusion

In this essay, we consider abatement effort as a separate choice variable from emissions, and utilise the choice of net emissions as a coalition's policy instrument to examine the case of heterogeneity in abatement and emission technologies across two countries. Given that countries have two choice variables, we examine agreements on net emissions. If we try to follow the commitments made by countries in the Paris Agreement, each country commits to one net target collectively in the first stage choosing net emissions, and then go on to individually decide on emission and abatement levels. For countries to collectively choose certain variables by maximizing total welfare, the welfare functions must be shared among countries so they become common information. However, which part of the welfare function countries share depends on the choice variables they choose.

We present four cases: the pure Nash non-cooperative case, the case of cooperation with a commitment to both abatement and emissions, and two cases where countries cooperate on net emissions. Results show that cooperation on net emissions is possible even with a high degree of heterogeneity among countries. Comparing our results to the pure Nash non-cooperative benchmark case, we show that the model achieves lower aggregate net emissions and allows gains from cooperation for both countries.

As each country free rides on the other country's net emission reduction efforts, it increases its emissions as the other country reduces its net emissions. It also increases its abatement efforts given the other country's reduction in net emissions. Cooperation with policy coordination on net emissions leads to lower global net emissions compared to the non-cooperative case. With both choice variables, emissions and abatement, the analysis on the choice of net emissions in international environmental agreements can be extended to study coalition formation as well as contingency plans that deter free-riding incentives.

Appendix A. List of Notations

i	index of a country
n	total number of countries
N	set of countries
S	a coalition set
s	size of a coalition
e_i	level of emissions generated by country i
E	aggregate pollution
x_i	abatement of country i
X	aggregate abatement
NE_i	net emissions of country i
W_i	welfare of country i
a, b, c, d	positive parameters
B_i	emission benefit function of country i
$D(E)$	damage function of aggregate pollution
$D(NE)$	damage function of global net emissions
C_i	abatement cost function of country i

BIBLIOGRAPHY

AUMANN, R. (1959). "Acceptable points in general cooperative n-person games." In "Contributions to the Theory of Games IV," Princeton University Press, Princeton, N.J..

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

BARRETT, S. (1997). "Towards a theory of international environmental cooperation." In "New directions in the economic theory of the environment," 239-280.

BARRETT, S. (2006). "Climate Treaties and "Breakthrough" Technologies." *American Economic Review*, **96**(2), 22-25.

BERNHEIM, B. D., PELEG, B., WHINSTON, M. D. (1987). "Coalition-proof Nash equilibria i. concepts." *Journal of Economic Theory*, **42**(1), 1-12.

CARRARO, C., D. SINISCALCO (1992). "The International Dimension of Environmental Policy." *European Economic Review*, vol. 36, 379-387.

CARRARO, C., SINISCALCO, D. (1993). "Strategies for the international protection of the environment." *Journal of Public Economics*, **52**(3), 309-328.

CARRARO, C., SINISCALCO, D. (1997). "New directions in the economic theory of the environment." Cambridge University Press.

CARRARO, C., MORICONI, F. (1998). "International games on climate change control." FEEM Working Paper, 56-98.

CHANDER, P., TULKENS, H. (1995). "A core-theoretic solution for the design of cooperative agreements on trans-frontier pollution." 176-193.

CHANDER, P., TULKENS, H.(1992). "Theoretical foundations of negotiations and cost-sharing in trans-frontier pollution problems." Eur. Econ. Rev., **36**, 388-398.

CHANDER, P., TULKENS, H. (1997). "The core of an economy with multilateral environmental externalities." *International Journal of Game Theory*, **26**, 379-401.

CHANDER, P. (2007). "The gamma-core and coalition formation." *International Journal of Game Theory*, **35**, 539-556.

CHWE, M. S.Y. (1994). "Farsighted coalitional stability." *Journal of Economic Theory*, **63**(2), 299-325.

D'ASPREMONT, C.A., JACQUEMIN, J., GABSZEWICZ, J., WEYMARK, J.A. (1983). "On the stability of collusive price leadership." *Canadian Journal of Economics*, **16**, 17-25.

DIAMANTOUDI, E., SARTZETAKIS, E. S. (2006). "Stable international environmental agreements: An analytical approach,." *Journal of Public Economic Theory*, **8**(2), 247-263.

DIAMANTOUDI, E., SARTZETAKIS, E. S. (2015). "International environmental agreements: coordinated action under foresight." *Economic Theory*, **59**(3), 527-546.

DIAMANTOUDI, E., SARTZETAKIS, E. S. (2018). "International environmental agreements- the role of foresight." Environmental and Resource Economics: The Official Journal of the European Association of Environmental and Resource Economists, **71**(1), 241-257.

DIAMANTOUDI, E., SARTZETAKIS, E. S., STRANTZA, S. (2018). "International environmental agreements the impact of heterogeneity among countries on stability." FEEM working paper No. 22.

DIAMANTOUDI, E., SARTZETAKIS, E. S., STRANTZA, S. (2022). "International environmental agreements with emission, abatement and adaptation choices." Preliminary version (April 2022), Working paper.

EYCKMANS, J. (2001). "On the farsighted stability of the Kyoto protocol." Working Paper Series, Faculty of Economics and Applied Economic Sciences, University of Leuven, No. 2001-03.

FINUS, M., RUNDSHAGEN, B. (1998). "Toward a positive theory of coalition formation and endogenous instrumental choice in global pollution control." public choice, **96**(1), 145-186.

FINUS, M., MCGINTY, M. (2019). "The anti-paradox of cooperation: Diversity may pay!" *Journal of Economic Behavior & Organization*, **157**, 541-559.

HOEL, M. (1991). "Global Environmental Problems: The Effects of Unilateral Actions Taken by One Country." *Journal of Environmental Economics and Management*, vol. 20, 55-70.

HOEL, M. (1992). "International Environment Conventions: The Case of Uniform Reductions of Emissions." *Environmental and Resource Economics*, vol. 2, 141-159.

HOEL, M. and DE ZEEUW A. (2010). "Can a Focus on Breakthrough Technologies Improve the Performance of International Environmental Agreements?" *Environmental and Resource Economics*, **47**(3), 395-406.

MAO, L. (2018). "A note on stable cartels." *Economics Bulletin*, **38**(3), 1338-1342.

MASOUDI, N. (2022). "Designed to be stable: international environmental agreements revisited." *International Environmental Agreements: Politics, Law and Economics*, 1-14.

NORDHAUS, W. (2019). "Climate change: The ultimate challenge for economics." *American Economic Review*, **109**(6), 1991-2014.

RUBIO, S. J., ULPH, A. (2006). "Self-enforcing international environmental agreements revisited." *Oxford Economic Papers*, **58**(2), 233-263.

SAMUELSON, P. A. (1954). "The Pure Theory of Public Expenditure." *The Review of Economics and Statistics*, **36**(4), 387–389.

SARTZETAKIS, E.S., STRANTZA, S. (2013). "International Environmental Agreements: An emission choice model with abatement technology." Discussion Paper Series, Paper No 5, Department of Economics, University of Macedonia, ISSN 1791-3144.

VELDERS, G. J., ANDERSEN, S. O., DANIEL, J. S., FAHEY, D. W., MCFARLAND, M. (2007). "The importance of the Montreal Protocol in protecting climate." *Proceedings of the National Academy of Sciences*, **104**(12), 4814-4819.

UNFCCC (2015). "Adoption of the Paris Agreement." FCCC/CP/2015/L.9/Rev.1. Available at unfccc.int.