Designing International Environmental Agreements immune to deviations

Effrosyni Diamantoudi, Concordia University
Fadi Harb, Concordia University   Eftichios S. Sartzetakis *

Abstract

The paper contributes to the literature that examines the formation of international environmental agreements as a two-stage game, where in the first, countries decide whether or not to sign the agreement, and in the second, they choose either emissions or abatement. Within this framework it has been shown that when the second stage is modelled as a Nash equilibrium, only small coalitions are stable, while when modelled as a Stackelberg equilibrium, larger coalitions become stable but attain little welfare improvement, a result coined as the paradox of cooperation. In the present paper we propose amending the agreement with a contingency plan to avert possible deviations from either a single country or a group of countries. We develop this contingency plan by assuming that coalition members when faced with a deviation do not adjust their emission levels by maximizing their collective welfare, but instead they calculate the level of their emissions that will inflict on the deviator a damage large enough to outweigh the possible benefits from reducing its emission efforts by exiting the coalition. We calculate the contingency plan using a model in which emissions is countries' choice variable and the coalition behaves as a leader. Our work suggests that coalition members realizing the potential gains of cooperation, give more substance to their leadership role, and instead of reacting to potential deviations by just choosing emissions through joint welfare maximization, they commit to emission levels that can sustain the coalition and their welfare level.

Keywords: International Environmental Agreements, Immune to deviations, Stable coalitions

JEL: D6, Q5, C7

*Corresponding author: Department of Economics, University of Macedonia, Email: esartz@uom.edu.gr.

1 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, generated mainly by human activities, are responsible for "global public bad" problems, the solution of which requires IEAs. According to William Nordhaus, "... 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, regulating the emissions of ozone-depleting substances at a global level (Velders et al., 2007). IEAs aiming at mitigating climate change are not yet as successful. Despite all effort made, including the Kyoto Protocol and the more recent the Paris Agreement, expectations are that we will fail to attain the goal of limiting global warming to well below 2 degrees Celsius (UNFCCC, 2015). This is mainly due to the fact that the high costs of mitigation create large free riding incentives, rendering a self-enforcing global climate agreement with high emission reduction targets unstable.

An IEA is considered stable if none of the countries that join the agreement, hereafter called signatories, has an incentive to pull out of the agreement (internal stability) and none of the countries that do not join the agreement, hereafter called non-signatories, has an incentive to participate in the agreement (external stability). This coalition stability notion was originally introduced by D'Aspremont, Jaquemin, Gabszewicz and Weymark (1983) in a price leadership model examining cartel stability, and has been has been employed in examining IEAs' stability by Hoel (1992), and Carraro and Siniscalco (1993). Barrett (1994) and Diamantoudi and Sartzetakis (2006). The rich literature on the stability of IEAs that has developed yields the pessimistic result that only very small coalitions are immune to deviations. This result can be attributed to the assumptions regarding the reaction of coalition's remaining members to potential deviations. In particular, it is assumed that when one country contemplates deviating, it expects that no other country will follow and the remaining coalition members will adjust

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

their choice so as to maximize their joint welfare. 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 is because the costs avoided by reducing mitigation efforts outweigh any additional environmental damages caused from its own reduction in mitigation efforts, especially taking into account the remaining members' adjustment of mitigation effort.

There are some differences in the size of stable coalition, depending on the choice variable used[2] and whether the coalition is assumed to behave as a Stackelber leader or a simultaneous move game is employed. However, these differences are small, with the exception of corner solutions (see Rubio and Ulph, 2006), and easily attributed to different models' assumptions. Furthermore, if under the assumption of leadership, larger coalitions form, they achieve very little welfare improvements, a result coined by Barret (1994) as the paradox of cooperation.[3] Large stable coalitions that achieve large welfare improvements become possible only when the notion of stability changes. For example, assuming that potential deviators expect the coalition to collapse in response of their deviators, Chander and Tulkens (1995) and (1997) show that any large coalition, including the grand coalition is stable. Furthermore, Diamantoudi and Sartzetakis (2015) show that, when endogenizing the reaction of the coalitions' remaining members to a deviation by a single country or a group of countries, large farsighted stable coalitions become possible, including the grand coalition.

In the present paper we propose a change in the design of an IEA that could make it immune to deviations. We retain the first part of the assumptions regarding the reaction of the remaining coalition's members to

[2]Some papers use countries' abatement level as a choice variable (see for example Barrett, 1994), while other use countries' emissions as a choice variable (see for example Diamantoudi and Sartzetakis, 2015). More recently Diamantoudi (2022) develop a model with emissions and abatement as separate choice variables.

[3]Recently, there is some interesting discussion regarding the intuition of the paradox of cooperation. McGinty (2020) is the first to explain the paradox of cooperation by identifying the damage internalization and the leadership effects. However, it does so using specific functional forms in a model with abatement as the only choice variable, not mentioning positivity constraints. Finus et al. (2021) extend the analysis using general functional forms in a model that also used abatement as the only choice variable. Diamantoudi et al. (2022), provide a clear intuition for the paradox of cooperation, using general functional forms in a model that treats emissions and abatement as separate choice variables.

a potential deviation. That, is we assume that when a country (group of countries) contemplates deviating, it (they) assumes that no other countries will follow. However, the remaining members instead of adjusting their choice variable in order to maximize anew their collective welfare, they adjust their choice variable so as to penalize the potential deviator(s) and prevent its (their) exit. In particular, we employ a model similar to Diamantoudi and Sartzetakis (2006), in which the choice variable is emissions and the coalition acts as a leader. We start with a coalition of a given size and we derive the level of emissions to which the coalition has to adjust to prevent any possible deviation by either one member or a group of members. These emission reaction levels are build as an integral part to the initial agreement to which all members commit. Drafting an agreement that contains a contingency plan prescribing the response of coalition members when faced with a potential deviation, will make it immune to any potential deviation. Our work suggests that coalition members realizing the potential gains of cooperation, give more substance to their leadership role, and instead of reacting to potential deviations by just choosing emissions through joint welfare maximization, they commit to emission levels that can sustain the coalition and their welfare level.

The remainder of the essay is structured as follows. Section 2 develops the model and solves for countries' choice of emissions under the contingency plan. Section 3 presents a numerical example of the contingency plan in the case of unilateral deviations. Section 4 examines contingency plans in the case of multiple deviations. The last Section concludes the paper.

2 The Model

The model in Diamantoudi and Sartzetakis (2006) is utilized 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, $N = \{1, \dots, n\}$, where all countries incur the same costs and attain the same 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. Country i's damages from total emissions are $D(E) = \frac{1}{2}cE^2$. Given the above benefit and damage functions, country i's 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 maximizes its welfare. 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,

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

unilateral member's deviation from the grand coalition, which nullifies any gains to the deviator. The contingency plan built by the coalition allows members to adjust their level of emissions in case of such unilateral deviation. An example with ten countries is provided that illustrates the contingency plan levels relative to the optimal solutions of the benchmark model.

The above literature builds on a framework that highlights the existence of strong free-riding incentives, which are moderated only when forming coalitions is not relevant, in the sense that they do not provide any welfare benefits over non-cooperation. However, this pessimistic result depends on a number of assumptions: Karp and Simon (2013), using an abatement model and dropping the assumption of strictly convex abatement costs, show with specific examples that larger coalitions are stable and yield significant welfare benefits. Furthermore, if R&D investments that reduce abatement costs exhibit increasing returns to scale, Barret (2006) shows that even the grand coalition will be stable. In general, technology agreements and R&D cooperation are considered club goods whose attractiveness may outweigh the incentive to free-ride,[4] the same as linking trade to environmental agreements.[5] Optimistic results are also obtained if we assume that countries are

[4] See for example Hoel and de Zeeuw (2010) and Goeschl and Perino (2012).

[5] See for example, Eichner and Pethig (2013) and (2015) and Diamantoudi et. al.

farsighted, that is, when a country contemplating joining or leaving the coalition takes into account all other countries' participation decisions.[6] More recently there is an interesting debate as to whether introducing adaptation as a choice could lead to larger stable coalitions.[7]

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

In the case of full cooperation, the grand coalition maximizes its joint welfare, which yields a per country emissions, e_c,

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

Following the joint welfare maximization and choice of emissions e_c, the grand coalition is not stable since 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 only adjust their emissions so as to maximize their joint welfare. As it has been shown in the literature, under this 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 could incorporate into their agreement a contingency plan, under which they all commit to adjust their emission level to a level nullifying any benefit from exiting, this agreement could sustain the initial number of signatories. In the present paper we retain the assumption of the main model about the perception of deviators regarding the membership choice of the remaining coalition's members. That is, we assume that when one (or a number of) member(s) consider exiting the coalition, it (they) assumes no other member(s) will follow. However, we change the assumption regarding the emission choice of the remaining coalition's members. That is, instead of adjusting their emission level so as to maximize their collective welfare, they choose a relatively larger level of emissions, capable of penalizing the deviator. This level of emissions it is build as an integral part of the initial agreement, to which coalition members commit. Through this contingency plan, i.e. the new rule of emissions'

(2020).

[6] Diamantoudi and Sartzetakis (2015) and (2018), de Zeeuw (2008), and Osmani and Tol (2009). However, Benchekroun and Chaudhuri (2012), using a farsighted stability concept, find that eco-innovations can reduce the stability of IEAs.

[7] Breton and Sbragia (2017), Benchekroun et al. (2017), Bayramoglu et al. (2018), Rubio (2018) and Finus et al. (2021).

adjustment, in the face of deviations, coalition members can eliminate single or multiple countries' incentives to free ride on the remaining coalition members' effort.

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. We assume that the initial size of the coalition is given and we want to derive the coalition members' contingency plan in case of deviations by single countries.

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 wants 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. Non-signatory countries choose their emissions level by maximizing their individual welfare level. The solution of the welfare maximization problem yields the level of emissions by non-signatories e_{ns} a function of signatories' emission level e_s. Each non-signatory's emissions best response function $e_{ns}(e_s)$ is,

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

Signatories choose their emission level to account for potential deviations, that is, 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 maximize 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,

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

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,

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

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

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

where $t = n - (s-1)$. Substituting equations (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 $\widehat{e}_s\left(s-1\right)$ adopted by the coalition at $s-1$,

$$\begin{aligned}&a\frac{[a-(s-1)\gamma\widehat{e}_s(s-1)]}{1+\gamma t} - \frac{[a-(s-1)\gamma\widehat{e}_s(s-1)]^2}{2(1+\gamma t)^2} - \\ &\frac{\gamma}{2}\left[(s-1)\widehat{e}_s(s-1) + t\frac{a-\gamma(s-1)\widehat{e}_s(s-1)}{1+\gamma t}\right]^2 \\ &\leq a^2[\frac{1}{2} - \frac{n^2\gamma}{2\psi}]\end{aligned}$$

The solution to the maximisation uses the binding constraint, $\omega_{ns}(e_{ns}(\widehat{e}_s(s-1))) = \omega_s(e_s^*(s))$.[8] The level of emissions resulting from the solution of the equality is presented in the following Proposition.

[8]The proofs and derivations of the model are found in Appendix 1.6 at the end of the essay.

Proposition 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 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 favorable 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, $\widehat{e}_s(s-1) > e^*_s(s-1)$.

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 immunise the grand coalition from deviation.

3 A numerical example

To facilitate direct comparison, we employ the same parameter values used in Diamantoudi and Sartzetakis (2006). 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 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 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 members adjust its emissions to $\widehat{e}_s(s = 9)$.

1.png

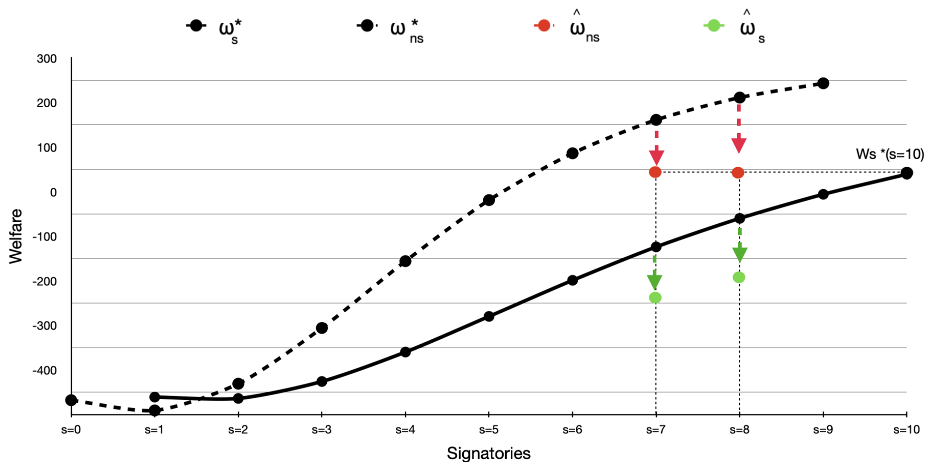

Figure 1: Optimal emission levels and contingency plan at $s = 9$

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 $\widehat{E}(9) > E^*(9)$.[9] 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 2 illustrates the indirect welfare of signatories and non-signatories facilitating welfare comparisons. The solid (dashed) curve in Figure 2 illustrates the welfare level of signatories W_s^*(non-signatories W_{ns}^*). The figure also presents the levels of welfare that correspond to $(\widehat{e}_s(9), \widehat{e}_{ns}(9))$ for signatories $\widehat{W_s}(9)$ and non-signatories $\widehat{W_{ns}}(9)$, respectively. At the grand coalition, $s = 10$, the optimal level of emissions is $e_{s=10}^* = 1.3044$ from the

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

joint profit maximisation. In case a signatory contemplates 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 $\widehat{W}_s(9)$ and $\widehat{W}_{ns}(9)$. 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, $W_s^*(10)=\widehat{W}_{ns}(9)$.

1.png

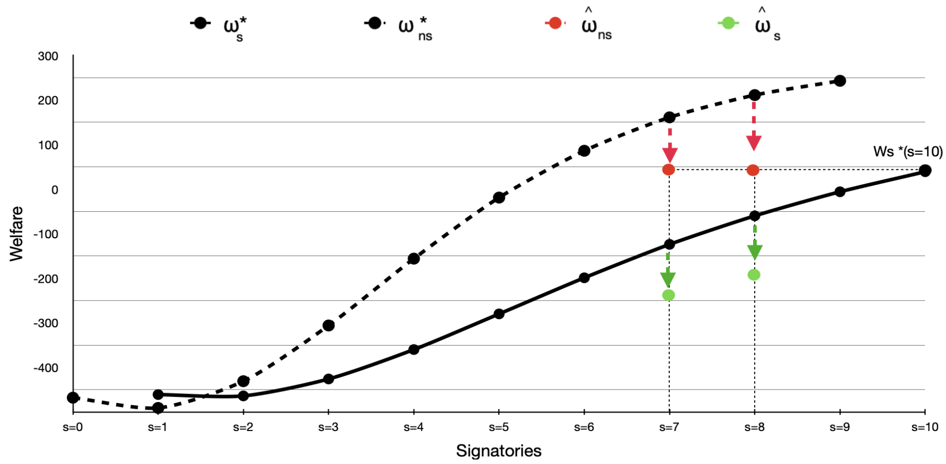

Figure 2: Welfare levels given optimal level of emissions and contingency plan at $s=9$

It is clear from Figure 2 that $\widehat{e}_s(9)$ is not the level of emissions that maximise the nine members' aggregate welfare, that is, $W_s^*(9) > \widehat{W}_s(9)$. The coalition members commit to an action that will result in lower welfare in order to immunize 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 behavior; 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.

4 Agreements immune to multiple deviations

This section studies the case of multiple simultaneous deviations from the coalition 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 an 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 exiting the agreement. 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

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

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 $s - j$,

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

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 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,*

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

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 created by coalition members drafts the emission adjustment necessary to eliminate the incentives of the coalition's 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 two examples of developing a contingency plan in case two or three countries simultaneously deviate from the grand coalition.

Using the same parameter values as in the previous section, the solid and dashed lines in Figure 3 depict the optimal emission levels of signatory e_s^* and non-signatory e_{ns}^* countries, respectively, as in Figure 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. We examine cases where two or three of the grand coalition's members contemplate exiting simultaneously. Figure 3 also depicts the levels of emission $\widehat{e}_s(8)$ and $\widehat{e}_s(7)$ to which all signatories commit to adjusting if two or three countries potentially decide to simultaneously free ride. The Figure also depicts the emission response of non-signatories at $\widehat{e}_{ns}(8)$ and $\widehat{e}_{ns}(7)$ when the remaining member adjust to $\widehat{e}_s(8)$ and $\widehat{e}_s(7)$, respectively.

1.png

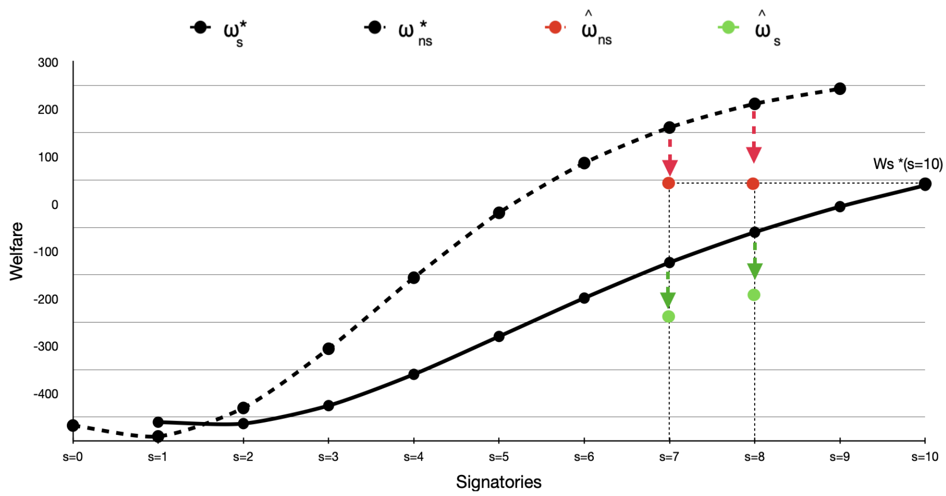

Figure 3: Optimal level of emissions and contingency plans for $s = 8$ and $s = 7$

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 $\widehat{e}_s(8) = 2.454$. Similarly, in the case that three countries contemplate exit, the remaining coalition's members, a coalition of seven, raise their emission level to $\widehat{e}_s(7) = 1.709$.

The two curves in Figure 4 illustrate the level of indirect welfare of signatories W_s^* and non-signatories W_{ns}^*, same as in Figure 1.2 above. It also presents the indirect levels of welfare that correspond to $(\widehat{e}_s(8), \widehat{e}_{ns}(8))$, for signatories $\widehat{W}_s(8)$ and non-signatories $\widehat{W}_{ns}(8)$, 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 $\widehat{W}_s(7)$ and non-signatories $\widehat{W}_{ns}(7)$, when three countries contemplate exiting the agreement simultaneously. 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, $W_s^*(10) = \widehat{W}_{ns}(8) = 39.1$. 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 strategies for multiple deviations until non-signatory welfare levels drop lower than full cooperation.

1.png

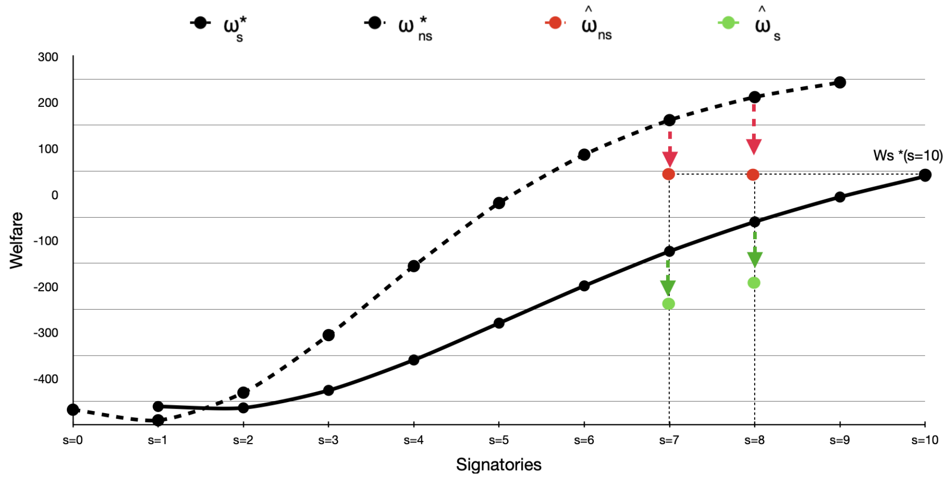

Figure 4: Welfare levels given optimal level of emissions and contingency plans for s=8 and s=7

When three countries contemplate exiting, 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 $\widehat{W}_{ns}(7) = 39.1$ instead of $W_{ns}^*(7) = 161$. Given the adjustment taken by the remaining coalition, this disincentivizes contemplating signatories from exiting the agreement.

5 Conclusion

The present paper proposes amending International Environmental Agreements with a contingency plan of emission adjustments in response to potential unilateral or group deviations, in order to make the agreement immune to deviations. We develop a framework assuming 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. The assumption that, following a potential deviation, remaining coalition members will choose their emission levels in order to maximize their joint welfare is revised. Assuming one-step deviations, i.e., simultaneous deviations (either unilateral or multiple), 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. We show that such contingency plans involve increasing rather than decreasing coalition members' emission levels so as to impose a cost on deviators, 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 coalition members, that is, it assumes that the membership status of coalition members does not change. A natural extension of our work is to loosen this assumption and derive the contingency plans under farsighted behavior, taking into account possible subsequent deviation by members.

6 References

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

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

Barrett, S. (2006) "Climate treaties and breakthrough technologies." *American Economic Review,* **96**, 22–25.

Bayramoglu, B., M. Finus, and J.-F. Jacques (2018). "Climate Agreements in a Mitigation-Adaptation Game." *Journal of Public Economics*, **165**, 101-113.

Benchekroun, H., W. Marrouch, and A. Ray Chaudhuri (2017). "Adaptation Technology and Free-Riding Incentives in International Environmental Agreements." In *Economics of International Environmental Agreements*, edited by M. Ozgur Kayalica, Selim Cagatay, and Hakan Mihci, Chapter 11, Routledge.

Benchekroun H. and Chaudhuri A.R., (2012). "Cleaner technologies and the stability of international environmental agreements." CentER Discussion Paper, 2012-051.

Breton, M. and L. Sbragia (2017). “Adaptation to Climate Change: Commitment and Timing Issues.” *Environmental and Resource Economics*, **68**, 975-995.

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

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

Diamantoudi, E. and Sartzetakis, E. (2006). “Stable International Environmental Agreements: An Analytical Approach.” *Journal of Public Economic Theory* ***8**, 247–263.*

Diamantoudi, E. and E. Sartzetakis (2015) "International environmental agreements: coordinated action under foresight." *Economic Theory*, **59**(3), 527-546.

Diamantoudi, E. and E. Sartzetakis (2018) "International Environmental Agreements-The Role of Foresight." *Environmental & Resource Economics*, **71**(1), 241-257.

Diamantoudi, E., E. Sartzetakis and S. Strantza (2020) "International Environmental Agreements and Trading Blocks: Can Issue Linkage Enhance Cooperation?" *Strategic Behavior and the Environment*, **8**(3), 269-310.

Eichner, T. and Pethig, R. (2013) “Self-enforcing environmental agreements and international trade.” *Journal of Public Economics*, **102**, 37–50.

Eichner, T. and Pethig, R. (2015) “Self-enforcing international environmental agreements and trade: taxes versus caps.” *Oxford Economic Papers*, **67(4)**, 897–917.

Finus, M. and Rundshagen B. (2001), “Endogenous coalition formation global pollution control.” *working paper, FEEM,* Nota di Lavoro 43.2001.

Finus, M., F. Furini and A. V. Rohrer (2021) "The efficacy of international environmental agreements when adaptation matters: Nash-Cournot vs Stackelberg leadership." *Journal of Environmental Economics and Management*, **109**, doi.org/10.1016/j.jeem.2021.102461.

Frondel, M., Horbach, J. and Rennings, K. (2007) "End-of-pipe or cleaner production? An empirical comparison of environmental innovation decisions across OECD countries." *Bussiness Strategy and the Environment*, **16**, 571-584.

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.

Karp, L., Simon, L. (2013) "Participation games and international environmental agreements: a non-parametric model." *Journal of Environmental Economics and Management*, **65** (2), 326–344.

McGinty, M. (2020) "Leadership and Free-Riding: Decomposing and Explaining the Paradox of Cooperation in International Environmental Agreements." *Emvironmental and Resource Economics*, **77**, 449-474.

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

Osmani, D. and R. Tol (2009) "Toward farsightedly stable international environmental agreements." *Journal of Public Economic Theory*, **11**, 455–492

Rubio, J. S. and Casino, B. (2001), "International Cooperation in Pollution Control" *mimeo.*

Rubio J.S. (2018). "Self-enforcing international environmental agreements: adaptation and complementarity." FEEM Working Paper 0 29.2018.

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

Ulph, A & Rubio, S (2006) "Self-enforcing Environmental Agreements Revisited", *Oxford Economic Papers*, **58(2)**, 233-263.

Velders G. J., Aandersen S. O., Daniel J. S., Fahey D. W. and 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

7 Appendices

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

7.1 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 $\widehat{e}_s\,(s-1)$ 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 non-signatories at $s - 1$ is given by

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

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

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

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

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

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

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

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

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

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

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

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

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.

7.2 Appendix B. Proofs

Proposition 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 $t = n-(s-1)$

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

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

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

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

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

Proposition 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*:

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

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

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

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

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

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

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