A model of the (incomplete) Paris Agreement on climate change

Alejandro Caparrós
Institute for Public Goods and Policies (IPP)
Consejo Superior de Investigaciones Científicas (CSIC)
Albasanz 26, 28037 Madrid, Spain
alejandro.caparros@csic.es

Abstract: The paper analyses a two-tier climate policy: a long-term agreement followed by a sequence of short-term agreements. Long-term climate agreements can be incomplete in a twofold sense. In the first sense, a long-term agreement is complete, 'comprehensive', when countries can agree on abatement and investments, while it is 'non-comprehensive' if they can just agree on abatement. In the second sense, the agreement is complete if it is 'contingent' on a stochastic variable that summarizes climate uncertainty. Short-term agreements written after the state of nature is realized follow the long-term agreement if profitable for all parts, and are renegotiated otherwise. Results show that a series of short-term agreements written after a contingent but non-comprehensive long-term agreement can implement the first best. Short-term agreements following a non-contingent and non-comprehensive long-term agreement cannot yield the first best, but can improve upon the situation without a long-term agreement. The Paris Agreement is an example of a two-tier climate policy with a non-contingent and non-comprehensive long-term agreement. The analysis provides a rationale for this treaty and guidelines for its amendment and improvement.

JEL classifications: C73, C78, H41, Q54.

Key words: Bargaining, renegotiation, contract theory, climate change, international environmental agreements.

1 Introduction

1.1 Motivation and preview

Most countries of the world have signed and ratified the Paris Agreement (henceforth, 'the Agreement'). According to this long-term agreement, countries aim at "holding the increase in the global average temperature to well below 2 °C above pre-industrial levels and [..] pursue efforts to limit the temperature increase to 1.5 °C above pre-industrial levels[,. . .] so as to achieve a balance between anthropogenic emissions by sources and removals by sinks of greenhouse gases in the second half of this century" (UNFCCC, 2015). The Agreement does not have short-term commitments, but rather it urges countries to send their voluntary commitments, their (Intended) Nationally Determined Contributions[1], (I)NDCs, which are not legally binding. In addition, as any other International Environmental Agreement, or any international agreement for that matter, countries can denounce the agreement at any moment and abandon it.

Although initial attention focused mainly on the role of the (I)NDCs, the Agreement has a second key feature: the mechanism to foster cooperation provided by Articles 6 and 9 of the Agreement, which may become the most important articles in the long term (Stavins and Stowe, 2017). Article 6 states that countries may cooperate voluntarily to increase the reach of their abatement targets. Cooperation between countries can significantly reduce costs and allow for more ambitious goals. Agreements based on this Article can be legally binding and will probably have a shorter time horizon. Article 9 is also relevant as it considers monetary transfers from developed countries to assist developing countries in their efforts to mitigate and adapt to climate change. Developed countries have agreed to mobilize a substantial amount of money: "prior to 2025 the Conference of the Parties [. . .] shall set a new collective quantified goal from a floor of USD 100 billion a year".

Rogelj et al. (2016) recently published a review of analyses determining the contribution

[1]INDCs were submitted before the Paris summit. Under the Paris Agreement, future mitigation contributions will be referred to as NDCs, without the 'intended'.

of currently submitted INDCs to the 2 °C target. The analysis shows that current climate policies cover about 25% of the required effort, that current INDCs would add another 25%, so that roughly 50% of the required effort is missing. In addition, about one third of the INDCs are conditional, requiring, for example, the provision of international funds, and an additional 45% of the INDCs came with both conditional and unconditional components. Thus, only about 20% of INDCs are purely unconditional. In other words, even if countries meet their (unconditional) INDCs there is a need for substantial additional negotiations, possibly under Articles 6 and 9, to reach the 2 °C target (let alone the 1.5 °C target).

Thus, instead of focusing on (I)NDCs, I will focus on the role of the long-term term Paris Agreement on shaping the subsequent short-term agreements to be signed under Articles 6 and 9. To model the process described above, I assume that countries can sign, or not, a long-term agreement (such as the Paris Agreement), that sets out some general rules that govern future short-term interactions. This long-term agreement is not a burden-sharing agreement, as countries did not negotiate targets for the different countries in the Paris Agreement. On the contrary, it is simply a long-term Nash equilibrium. In addition, countries can decide at any moment in time to abandon the long-term agreement. Countries can also negotiate short-term agreements, which are burden sharing agreements and are legally binding. Countries can negotiate these short term agreement within the general rules set out in the long-term agreement, but they can also abandon the agreement and negotiate these short term agreements freely.

The long-term agreement (the Paris Agreement) is incomplete in a double sense[2]. First, countries can agree on abatement levels and associated transfers, but they cannot negotiate and sign agreements over investments. To avoid confusion, I call agreements which are incomplete in this sense 'non-comprehensive'. Second, the long-term agreement on abatement levels and associated transfers is complete if it is contingent on a stochastic variable that summarizes climate uncertainty, and is incomplete when it cannot depend upon the

[2] Barrett (2002) uses the term 'incomplete' in a third sense, in the context of International Environmental Agreements, referring to agreements signed only by a subset of all the countries.

stochastic component. Real-world treaties are unlikely to include obligations that depend on the realization of a stochastic variable. Even over an issue such as climate change, where uncertainty is a key ingredient of the problem, countries are unlikely to agree on transfers and abatement efforts that depend on the severity of climate change. Moreover, the problem is so complex and multidimensional that it is probably not possible to write a meaningful agreement as a function of the state of nature. I call agreements which are incomplete in this sense 'non-contingent'.

The analysis shows that a sequence of short-term burden sharing agreements cannot implement the first-best, explaining the interest in writing a long-term agreement such as the Paris Agreement. The reason for this is the standard hold-up problem, as countries will underinvest to improve their bargaining power in the sequence of negotiations determining the short-term burden sharing agreements. However, if it were possible to write a long-term agreement contingent on the state of nature the first-best would be attaint. This holds even if the long-term agreement does nothing more than 'put on paper' a Nash equilibrium and countries are able to abandon at any time the terms agreed upon in this long-term agreement. Thus, if the Paris Agreement could be made contingent on the state of nature it would be able to solve the hold-up problem, despite its weakness. However, as it is probably not possible to write an Agreement that is truly contingent on the state of nature, I analyze also the role that a non-comprehensive and non-contingent agreement can have. Not surprisingly the results show that the first-best can no longer be attaint. Nevertheless, even this form of agreement brings the world closer to the first-best, thus justifying the interest in the Paris Agreement. Furthermore, the analysis shows that the problem lies not so much in the non-comprehensiveness of the agreement but more on the fact that the agreement is non-contingent. Thus, the more dependent the terms of the agreement are on the state of nature finally realized (e.g. on the severity of the climate change observed), the closer the agreement will bring us to the first best.

1.2 Relation to the literature

A large body of literature on game-theoretic analyses of International Environmental Agreements (IEA) has emerged over the last three decades. A recent survey and a selection of the most relevant papers can be found in Finus and Caparrós (2015). The literature on IEA can be divided into several broad categories. First, we have literature based on the concept of the internal and external stability of a coalition (agreement), in the sense that no player wants to leave or join the coalition (Carraro and Siniscalco 1993; Barrett 1994). Results show that for identical players, only a very small number of players form a coalition, while asymmetries among players allow for larger coalitions. The second approach is based on cooperative game theory and, more precisely, on the core (Chander and Tulkens 1997). The objective is to set up a burden-sharing rule that is able to favor the cooperation of all, ensuring that the rule prevents any individual country, but also any sub-group of countries, from being interested in leaving the IEA. The two approaches just mentioned are static but there is also a large set of papers that have focused on the dynamic nature of the climate change problem (see the survey in Calvo and Rubio 2013). The papers in this third approach are characterized by using dynamic state-space games to analyze IEA, combined with one of the stability concepts discussed above. The fourth approach focuses on the negotiation process, a part from which the previous approaches abstract. Here the goal is to investigate the impact of different bargaining protocols on the final agreement (for a survey, see Caparrós 2016).

None of these papers are appropriate to model the two-tier process described above: a long-term non-binding agreement followed but a series of short-term burden sharing agreements. In all these papers the focus was on a long-term burden sharing agreement, such as the Kyoto Protocol. In the long list of papers pioneered by Carraro and Siniscalco (1993) and Barrett (1994), countries determine their emissions jointly once they sign the agreement (all signatories essentially become one player). The papers following Chander and Tulkens (1997) assume that a supranational authority imposes abatement efforts and transfers, ensuring that all players, and all groups of players, are better off with the proposed burden sharing rule. The agreement is then enforced by a threat to go back to all-singleton behavior. Papers in the bargaining tradition (Caparrós et al., 2004) also assume that the long-term agreements are binding or, alternatively, that defection of one player will bring all countries back to the all-singleton behavior.

Although less related in terms of modeling approach, there is also a growing literature applying experimental economics methods to analyze alternative IEA designs. Within this group, Barrett and Dannenberg (2016) present a laboratory experiment that analyzes the second key feature of the Paris Agreement mentioned above, i.e. the submission of INDCs by countries followed by a review, with no legal consequences, of the degree in which the pledges have been met (the so-called 'pledge and review' process). As already stated, this is not the focus of my model.

The closest precedent to my analysis can be found in Harstad (2016), and to a lesser extend in Harstad (2012), Battaglini and Harstad (2016) and Beccherle and Tirole (2011). Although in a different framework, Harstad (2012 and 2016) analyze incomplete contracts in the first sense, i.e. in the sense that countries can negotiate and sign agreements on emissions and not on investment (what I have called non-comprehensive). Battaglini and Harstad (2016) extend the analysis to include the role of participation and duration of the agreements. Harstad (2012) and Battaglini and Harstad (2016) analyze a deterministic framework, while Harstad (2016) includes a stochastic component, and is therefore more relevant for my analysis. However, the agreements that he analyzes are complete in the sense that they depend on the stochastic variable (contingent in the terminology introduced above). As discussed above, contingent agreements are difficult to implement for a problem as complex as climate change. Furthermore, in this paper his analysis is based on a quadratic specification and, although he considers long-term agreements, he does not consider the two-tier procedure described above (one long-term agreement followed by a sequence of short-term agreements). Other relevant precedents can be found in Dutta and Radner (2004, 2009, 2012) and in Barrett (2002). The former focus on equilibrium selection in analyses that combine theory and simulations and the latter focuses on the role of alternative assumptions when analyzing the possibility to renegotiate agreements.

The paper is also closely related to the vast literature on incomplete contracts (e.g. Hart and Moore 1988, Hart and Holstrom 1987 or Crawford 1988), and several of the results shown below are well known in that literature (see Segal and Whinston 2010 for a survey). My model is basically a hold-up model, although it does not assume self-investments as most of these models do, using the terminology used, e.g., in Segal and Whinston (2010). On the contrary, the model has imperfect coalition investments, despite the fact that investments only reduce the cost function of the player investing. The reason is that, due to the public good nature of the abatement done in the last stage, when one player increases its investment it improves the disagreement point for the remaining players. The model also differs from most of this literature by the fact that it considers a larger number of players, and not only two as this literature typically does.

The rest of the paper is organized as follows. Section 2 presents the model and discusses the results. To provide two extreme benchmarks, the first sub-section analyzes the first-best and the second a sequence of short-term agreements without a long-term agreement. The following sub-section analyzes the role of a non-comprehensive but contingent long-term agreement. The last sub-section analyzes a non-comprehensive and non-contingent long term agreement, what I consider the best stylized representation of the Paris Agreement. Section 3 concludes.

2 The model

There are N countries which can provide abatement efforts to mitigate a public bad problem, such as climate change (i.e., abatement is a public good). Through investments, countries can also reduce the costs of the abatement efforts. The model has four stages, as shown in Figure 1. At stage 1, countries decide to sign, or not, a long term 'non-comprehensive' agreement, in the sense that the agreement cannot specify investment levels. This agreement can be complete ('contingent'), if it is written as a function of the state of nature described below, or incomplete, if it is not contingent on the state of nature. The long-term agreement specifies the additional abatement that countries will perform at stage 4, and the associated transfers, as detailed below.

Whether or not a long-term agreement is signed, countries decide at stage 2 their investment levels, denote by $\mathbf{I} = (I_1, ...I_N)$. Investing I_i units has a cost equal to I_i for country i. This investment is private and benefits directly only the country investing, by reducing its abatement cost in the subsequent stages.

At stage 3, the state of nature θ is randomly realized, out of the finite set of possible states of nature Θ. The probability of each state of nature is known, and is denoted $\gamma(\theta)$. After the value of θ is observed at stage 3, it becomes common knowledge.

At stage 4, countries bargain repeatedly over abatement and associated transfer levels, potentially signing short-term agreements. I assume that there is an infinite sequence of time periods and that in each one countries sign a short-term agreement that covers only one period. If no long-term agreement was signed at stage 1, countries negotiate freely. If a long-term agreement was signed, they follow the terms of the long-term agreement if they are profitable for all parts, taking into account the value of θ observed at stage 3. Otherwise, it is renegotiated, as detailed below.

[Figure 1 about here]

Gross payoffs per period at stage 4 are given by the function

$$v_i(\mathbf{q}, I_i, \theta) = B_i(Q, \theta) - C_i(q_i, I_i), \ \forall i \in N$$

where $\mathbf{q} = (q_1, ...q_N)$ shows the abatement effort undertaken by each country, B_i summarizes benefits from climate change abatement and C_i shows the cost incurred through the abatement efforts. Benefits B_i depend on aggregate abatement ($Q = \sum_{i \in N} q_i$) and on the realization of a random variable θ. Costs C_i depend on the amount of abatement performed by county i at stage 4 (abatement stage) and on the investment done by county i at stage 1 (investment stage). Investments is a private good, and only the country investing benefits from the costs reduction. Note that none of the results shown below depend on the separable nature of benefits and costs, as I could work directly with the function v. Nevertheless, I use functions B_i and C_i to facilitate the interpretation and the comparability with the rest of the literature on IEA. The discount rate is given by $r > 0$, with $\delta \equiv e^{-r\tau}$. Regarding the payoff function mentioned above, I make the following assumptions:

Assumption 1: *For all* $i \in N$: $\frac{\partial B_i}{\partial q_i} > 0$, $\frac{\partial^2 B_i}{\partial q_i^2} \leq 0$, $\frac{\partial C_i}{\partial q_i} > 0$, $\frac{\partial^2 C_i}{\partial q_i^2} > 0$, $\frac{\partial C_i}{\partial I_i} < 0$, $\frac{\partial^2 C_i}{\partial I_i^2} > 0$, $\frac{\partial^2 C_i}{\partial q_i \partial I} < 0$, $\lim_{Q \to 0} \frac{\partial B_i}{\partial q_i} > \lim_{q \to 0} \frac{\partial C_i}{\partial q_i} > 0$.

Before continuing, I define formally self-investments and cooperative investments for future reference.

Definition 1 *Country i's investment is a "self-investment" if country j's disagreement payoff,* $v_j(\mathbf{q}^A, I_j, \theta)$*, is independent of* I_i *for all* $j \neq i$*, at all* I_{-i} *and* θ*. Country i's investment is "cooperative" if country j's disagreement payoff is non decreasing in* I_i *for all* $j \neq i$*, at all* I_{-i} *and* θ*.*

As already mentioned in the introduction, although investment is a private good in the sense that it benefits directly only the country that undertakes the investment, using the terminology commonly used in the literature on incomplete contracts (Segal and Whinston, 2010) it is a cooperative investment, in the sense that it affects the disagreement payoff of other countries.

2.1 First best

To provide a benchmark, let us analyze the first best outcome.

Lemma 1 *The first best equilibrium in terms of abatement and investment is defined by the following equations:*

$$E_\theta\left[\sum_{i\in N}\frac{\partial B_i}{\partial q_i}-\frac{\partial C_i}{\partial q_i}\right] = 0 \Longrightarrow q_i^*=q_i^*(I_i,\theta) \quad \forall i\in N \tag{1}$$

$$E_\theta\left[-\left.\frac{\partial C_i}{\partial I_i}\right|_{q_i^*(I_i,\theta)}\right] = 1-\delta \Longrightarrow I_i^* \quad \forall i\in N \tag{2}$$

Proof. As all short term interactions are identical, the central planner has the following objective function:

$$f(\mathbf{q},I_i,\theta)=\arg\max_{\mathbf{I},\mathbf{q}}\left(E_\theta\left[\frac{\sum_{i\in N}v_i(\mathbf{q},I_i,\theta)}{1-\delta}\right]-\sum_{i\in\mathbf{I}}I_i\right) \tag{3}$$

The FOC for the abatement efforts are given by:

$$\frac{\partial f}{\partial q_i}=E_\theta\left[\frac{1}{(1-\delta)}\left(\sum_{i\in N}\frac{\partial B_i}{\partial q_i}-\frac{\partial C_i}{\partial q_i}\right)\right]=0 \quad\Longrightarrow\quad q_i^*=q_i^*(I_i,\theta) \quad \forall i\in N, \tag{4}$$

yielding equation (1). Define the maximum value function:

$$F(\mathbf{q},\mathbf{I},\theta)=f(\mathbf{q}^*(\mathbf{I},\theta),\mathbf{I},\theta)=f(q_1^*(I_1,\theta),...,q_N^*(I_N,\theta),I_1,...,I_N,\theta)$$

where $q_i^*(I_i,\theta)$ are the optimal levels of abatement stage 4. From the envelope theorem, using (4):

$$\frac{\partial F(\mathbf{q},\mathbf{I},\theta)}{\partial I_1}=\frac{\partial f}{\partial q_1}\frac{\partial q_1}{\partial I_1}+...+\frac{\partial f}{\partial q_N}\frac{\partial q_N}{\partial I_1}+\frac{\partial f}{\partial I_1}=\frac{\partial f}{\partial I_1}$$

and the FOC in terms of investment is

$$E_\theta\left[\frac{\sum_{i\in N}\frac{\partial v_i(q_i^*(I_i,\theta),I_i,\theta)}{\partial I_i}}{1-\delta}\right]-1=E_\theta\left[\frac{-\left.\frac{\partial C_i}{\partial I_i}\right|_{q_i^*(I_i,\theta)}}{1-\delta}\right]-1=0 \quad \forall i\in N,$$

or equation (2). ■

The first order condition in (1) is the standard equation defining optimal abatement, only taking into account that abatement benefits have a random component. Condition (2) yields the optimal investment level. The central planner takes into account the impact of the investment of player i on the infinite stream of future costs of abatement. However, as one only needs to take into account direct effects, due to the envelope theorem, and investments are a private good and only benefit directly the country that undertakes the investment, the FOC in terms of investment are independent from each other. Thus, each FOC yields the optimal investment for one country.

2.2 Nash equilibrium at the investment stage with no long-term agreement

Consider now the outcome if countries act independently and do not sign any long-term agreement at stage 1. At the abatement stage (stage 4), after all players have invested and θ is known, if there are no short-term agreements either all countries abate at their Nash equilibrium (NE). I denote the abatement that countries perform in this case by $\mathbf{q}^A = (q_1^A, ...q_N^A)$. Note that the value chosen by each country i would depend both on I_i and θ. However, at each period of stage 4 countries have the opportunity to agree on the abatement effort that maximizes the additional surplus created by signing the short-term agreements. To simplify this stage, I assume that countries use the asymmetric Nash Bargaining Solution for these short-term agreements[3]. Thus, countries agree on the abatement level that maximizes the additional surplus to be shared, and each country obtains a share φ_i of the additional surplus created, plus its disagreement payoff (if all countries are symmetric and have the same bargaining power, the split the difference rule yields $\varphi_i = \varphi = 1/n$). The following proposition summarizes the main results of this section (I_i^S denotes investment without a long term agreement).

[3]One would obtain similar results with the alternative assumption that negotiations follow a multilateral offer-counter-offer bargaining protocol à la Rubinstein, as long as one assumes that countries use stationary strategies and the discount factor tends to one (Osborne and Rubinstein, 1994, p131). If countries do not use stationary strategies (note that the assumption of stationary strategies is different, and stronger, than the standard assumption that the game is stationary) the multilateral case would yield multiple equilibria. There are different alternatives to overcome this difficulty, such as the "exit game" in Krishna and Serrano (1993) or the model in Huang (2001), which allows for conditional and unconditional offers and has been extended to games with externalities in Caparrós and Péreau (2017). Although the proof would be more complicated, dropping the assumption of stationary strategies and using Huang's (2001) model would also yield similar results, as long as one assumes that the discount rate tends to one and each country has the right to talk once following a cyclical protocol.

Proposition 1 *Countries will not implement the first best through a series of short term agreements without a long-term agreement. For a given level of investment, abatement at the abatement stage will be first best, as it will be defined by (1), i.e.* $q_i^S(I_i,\theta) = q_i^*(I_i,\theta)$. *However, investment,* I_i^S*, will be defined,* $\forall i \in N$ *and* $j \neq i$*,by*

$$\underbrace{\underbrace{\varphi_i E_\theta\left[-\left.\frac{\partial C_i}{\partial I_i}\right|_{q_i^*(I_i,\theta)}\right]}_{Agreement}+\underbrace{(1-\varphi_i)E_\theta\left[-\left.\frac{\partial C_i}{\partial I_i}\right|_{q_i^A(I_i,\theta)}\right]}_{Disagreement}}_{Self\text{-}investment\ terms}\underbrace{-\varphi_i\sum_{j\in N/i}E_\theta\left[\left.\frac{\partial B_j}{\partial q_i}\right|_{q_i^A(I_i,\theta)}\frac{q_i^A(I_i,\theta)}{\partial I_i}\right]}_{Cooperative\ investment\ term}=1-\delta \tag{5}$$

Thus, there will be underinvestment as compared to the first best and as a consequence abatement will also be below the optimal level. Overall underinvestment will be more severe the larger the number countries, and the larger the benefit obtained by other countries from the increased abatement in case of disagreement induced by larger investments. For a particular country, a larger share of the surplus generated by the short-term agreement, φ_i*, reduces underinvestment if increasing investment has a stronger impact on payoffs in case of agreement than in case of disagreement.*

Proof. At stage 4 countries agree on the abatement level that maximizes the additional surplus to be shared, i.e. $q_i^S = q_i^S(I_i,\theta) = q_i^*(I_i,\theta)$ $\forall i \in N$ is given by (1). Taking this into account, at stage 2 (investment stage), the expected payoff for each player i is:

$$E_\theta\left[\frac{v_i(\mathbf{q}^A, I_i,\theta)+\varphi_i\sum_{j\in N}\left[v_j(\mathbf{q}^S, I_j,\theta)-v_j(\mathbf{q}^A, I_j,\theta)\right]}{1-\delta}\right]-I_i \tag{6}$$

i.e., its disagreement payoff in each period plus a share of the additional surplus created through the short-term agreement in each period, minus the cost associated with the investment. And the NE at stage 2 is given by:

$$E_\theta\left[\frac{\partial v_i(\mathbf{q}^A, I_i,\theta)}{\partial I_i}\right]+\varphi_i\sum_{j\in N}E_\theta\left[\frac{v_j(\mathbf{q}^S, I_j,\theta)}{\partial I_i}-\frac{\partial v_j(\mathbf{q}^A, I_j,\theta)}{\partial I_i}\right]=1-\delta$$

Applying again the envelope theorem, but taking into account that $\frac{\partial v_j(\mathbf{q}^A,I)}{\partial q_i^A} \neq 0$, we obtain (5). Overall underinvestment will be more severe the larger the number countries because the last term on the LHS of (5) becomes larger, for a given φ_i. Larger φ_i reduces underinvestment if $E_\theta\left[-\left.\frac{\partial C_i}{\partial I_i}\right|_{q_i^S(I_i,\theta)}\right] > E_\theta\left[-\left.\frac{\partial C_i}{\partial I_i}\right|_{q_i^A(I_i,\theta)} + \sum_{j\in N/i}\left.\frac{\partial B_j}{\partial q_i}\right|_{q_i^A(I_i,\theta)}\frac{q_i^A(I_i,\theta)}{\partial I_i}\right]$, yielding the last statement in the proposition, ■

As the proposition shows, for a given country, increasing its investment at stage 2 has three effects. First, it decreases its costs in the event of an agreement at stage 4. This is the first term on the LHS of (5). The larger the share of the surplus created by the agreement that the country appropriates, φ_i, the more important this term becomes. The second effect is that it reduces its costs in the event of a disagreement at stage 4. This is the second term on the LHS in (5). The larger the share obtained by country i the less important this term becomes. The two terms just discussed appear if investment is a self-investment and if it is a cooperative investment (see Definition 2), as they refer to the impact that investment has on the agreement and disagreement payoffs of the country itself. The last term on the LHS appears only because investment is cooperative in this model. The investment done by country i in stage 2 impacts the disagreement point of all the other countries. The reason is that, although investment only impacts the cost function of the investing country, the investment implies more abatement by country i at stage 4 and this improves the disagreement point of all the other countries. This term becomes more important the larger the share of the surplus obtained by the country, and reduces the investment level of country i, as improving the disagreement point of the other countries reduces the bargaining power of the country investing.

Figure 2 illustrates equations (2) and (5). The thick horizontal line is the RHS of (5). The dashed line is the LHS of (2). The first-best investment level, I^*, is given by the intersection of these two lines. The thin solid line is the first term on the LHS of (5). Note that if investment would have no impact on the disagreement point, the second and third terms of (5) would vanish, and I^S would be given by the intersection of both solid lines. In fact, this is the case in the standard model analyzed in Hart and Moore (1988) or Muthoo(1995), where the no-agreement situation implies no-trade between a buyer and seller. The reduction would be explained exclusively by the share φ_i that the player receives from the additional surplus created, as the individual country only cares about the share of the additional surplus created through the investment that it appropriates. The line with full dots includes the first two terms on the LHS of (5). That is, the situation that would prevail if there is abatement at the disagreement point, and hence investment affects this disagreement point, but investment were self-investments and would not impact the disagreement point of other players. As shown in the Figure, the impact on the own disagreement point brings the investment level closer to the first-best level. The reason is that there is now a second incentive to invest, to reduce the costs in the event of a disagreement. Finally, the line with empty dots includes all the terms on the LHS of (5). Including the 'cooperative investment' term moves the investment level again further away from to the first-best situation. As already mentioned, the reason is that increasing the investment level increases the benefits of all other countries at the disagreement point, and this weakens the bargaining position of the country investing. That is, although the hold-up problem still exists, it is less severe as in the canonical model where the no-agreement situation implies no-trade between a buyer and seller.

[**Figure 2 about here**]

The following example illustrates the model using particular functions that are similar to those usually used in the literature on IEA (Barrett, 1994).

Example 1 *Consider $N \geq 3$ identical players, implying $\varphi = 1/N$. Payoffs are $B(Q, \theta) = b\theta Q$ and $C(q, I) = \frac{c}{2I}\left(q^2 + F\right)$, where b, c and F are non-negative parameters, with $E_\theta\left[N^2 b^2 \theta^2\right] < 2(1-\delta)c$. The states of nature in Φ are $\{0.5, 1, 1.5\}$, with equal probability. Linear benefits are used frequently in this literature, even in papers where they are assumed to capture the infinite stream of future benefits. In the context analyzed here, $B(Q, \theta)$ only captures benefits in one period, and a linear specification is therefore more appropriate. The abatement cost function is also similar to the one frequently used in this literature (identical for $I = 1$), although with an additional fix cost F. However, in this model investment at stage 1 can* *reduce abatement costs (for $I > 0$). Dropping subscripts, optimal abatement and investment levels at the first best are*

$$
\begin{aligned}
q^*(I,\theta) &= \frac{Nb\theta I^*}{c} \\
I^*(\theta) &= \left(\frac{Fc^2}{2c(1-\delta) - N^2b^2\theta^2}\right)^{\frac{1}{2}}.
\end{aligned}
$$

For the case of a series of short-term agreements but a NE at the investment stage with no long-term agreement, we have:

$$
\begin{aligned}
q^S(I,\theta) &= \frac{Nb\theta I^S}{c} \\
I^S(\theta) &= \left(\frac{Fc^2}{2c(1-\delta) + b^2\theta^2(N-3+1/N)}\right)^{\frac{1}{2}}.
\end{aligned}
$$

As predicted, in this case we have underinvestment: $I^S < I^*$. To see that $I^S < I^*$, subtract the denominator from the expression for I^S from the denominator of the expression for I^*. This yields $\frac{1}{N}b^2\theta^2(N-1)(N^2+2N-1)$, which is positive. Note also that $I^S > 0$ for $N \geq 3$ and that we also have $q^S < q^*$. Figure 2 is drawn using the functional forms in this example, with $(N, b, \theta, c, F, \delta) = (3, 1, 1, 50, 3, 0.9)$. For these numbers, in terms of investment we have $I^* = 87$ and $I^S = 31$, and in terms of abatement $q^* = 5.1962$ and $q^S = 1.8766$. That is, although the expressions for abatement are similar in both cases, underinvestment implies a significantly lower level of abatement if there is no long-term agreement. If there is no long-term agreement and countries are not able to reach an agreement at stage 4, abatement would be $q^A(I^S, 1) = 0.6255$ while abatement at the disagreement point with the first-best investment level would be $q^A(I^*, 1) = 1.7321$.

2.3 Equilibrium with a non-comprehensive but contingent long term agreement

I now assume that countries can write a long-term agreement as a function of θ. This long-term agreement, written at stage 1, before θ is know, specifies the transfers that each player will receive and the abatement that it will perform in each short-term agreement written at stage 4. This long-term agreement is not binding in the sense that countries can break it. To simplify the analysis I assume that a new long-term agreement is not possible. Note, however, that this is not really a relevant restriction as a new long-term agreement after investment was made at stage 2 and uncertainty was revealed at stage 3 would not add any strategic component to the sequence of short-term agreements.

Let us fix an arbitrary long-term agreement $\langle\tau(\theta)\rangle_{\theta\in\Theta}$ such that for any θ in each short interaction at stage 4:

$$v_i(\mathbf{q}^A, I_i, \theta) + \left[v_i(\mathbf{q}^L, I_i, \theta) - v_i(\mathbf{q}^A, I_i, \theta) + \tau_i(\theta)\right] \geq v_i(\mathbf{q}^A, I_i, \theta) \quad \forall i \in N$$

or

$$v_i(\mathbf{q}^L, I_i, \theta) + \tau_i(\theta) \geq v_i(\mathbf{q}^A, I_i, \theta) \qquad \forall i \in N \tag{7}$$

As in the previous sub-section, abatement in case of agreement is optimal for a given level of investment, as countries have no interest in leaving unexhausted surplus, i.e. $q^L(I_i, \theta) = q^*(I_i, \theta)$, where L in superscript stands for contingent long-term agreement. Given my assumptions on v_i there are many long-term agreements that satisfy the restrictions in (7). Let us further assume that the terms of the long-term agreement will only be abandoned if it is not profitable for all parts, in the sense that at least one player can credibly threat to prefer a no short-term agreement situation (i.e. no renegotiation to improve the terms of the agreement). This implies that the contingent long-term agreement $\langle\tau(\theta)\rangle_{\theta\in\Theta}$ defined above will never be denounced, as it ensures that (7) holds for any realization of θ. Thus, taking into account that (7) holds for any value of θ, the expected payoff for player i for any $\mathbf{I}$ at stage 2 is:

$$E_\theta\left[\frac{v_i(\mathbf{q}^A, I_i, \theta) + \left(v_i(\mathbf{q}^*, I_i, \theta) - v_i(\mathbf{q}^A, I_i, \theta) + \tau_i(\theta)\right)}{1-\delta}\right] - I_i$$

or

$$E_\theta\left[\frac{v_i(\mathbf{q}^*, I_i, \theta) + \tau_i(\theta)}{1-\delta}\right] - I_i \quad \forall i \in N \tag{8}$$

Hence, the NE at the investment stage is given by

$$E_\theta\left[\frac{\partial v_i(\mathbf{q}^*, I_i, \theta)}{\partial I_i}\right] = 1 - \delta \quad \forall i \in N$$

or

$$E_\theta\left[-\left.\frac{\partial C_i}{\partial I_i}\right|_{q_i^*(I_i,\theta)}\right] = 1 - \delta \;\; \forall i \in N \tag{9}$$

which are the same as first-best FOCs (see equation (2)).

Thus, we can write:

Proposition 2 *Countries will implement the first best, in terms of abatement and in terms of investment, through a series of short term agreements regulated by a non-comprehensive but contingent long-term agreement. That is, even if countries cannot sign an agreement on investments the first best will be attaint if the long-term agreement defines transfers that are contingent on the state of nature.*

The second and the third terms in (5) disappear because now the share obtained by each country of the additional surplus created through an agreement on stage 4 is not anymore under discussion at that stage, because transfers were already determined at stage 1, when the long-term agreement was signed. This implies that, despite the weakness of the long-term agreement, which can be abandoned at any time, the first best is attaint even if countries cannot agree on investments.

As discussed in the introduction, the Paris Agreement is incomplete in a double sense. First, it is non-comprehensive, as it does not determine investment levels. What Proposition 5 tells us is that this incompleteness is not problematic, as the world could attain the first-best by writing a long-term agreement that is contingent on the state of nature. In theory, a detailed development of Article 6 and 9 could make the agreement contingent on the state of nature. However, it is no secret that the practical challenges of this task would be enormous. This is the reason why I now move on to analyze agreements that are also incomplete in this second sense, i.e. non-contingent agreements.

2.4 Equilibrium with a non-comprehensive and non-contingent long-term agreement

Assume now that θ is so complex and multidimensional that it is not possible to write a meaningful agreement as a function of the state of nature. Thus, countries cannot write an agreement in which transfers are a function of the state of nature finally realized. This is clearly the case for any long-term agreement on climate change abatement efforts, as the number of variables is extremely large and the uncertainties surrounding them are also large.

Let us therefore assume that countries write a long-term agreement where transfers are specified independently from the state of nature. As before, these transfers are specified for the case where the players move to the abatement level that maximizes the surplus to be shared, for a given level of investment, as countries have no interest in leaving unexhausted surplus. These agreements state, e.g., that country j has to pay country i a given amount of money for increasing its abatement efforts. The difference with the previous section is that the amount paid is determined a priori, and that it does not depend on the state of nature finally realized.

Under this assumption, let us fix an arbitrary incomplete long-term agreement $\boldsymbol{\tau} = (\tau_1, ..., \tau_N)$. At stage 4, investment levels $\mathbf{I}$ are sunk and the state of nature θ is known. Once θ is observed by the parties, all countries are happy with the terms of this agreement if

$$v_i(\mathbf{q}^l, I_i, \theta) + \tau_i \geq v_i(\mathbf{q}^A, I_i, \theta) \quad \forall i \in N, \tag{10}$$

where l in superscript stands for long-term non-comprehensive and non-contingent agreement. If these conditions hold for all players, I assume that they honour the agreement, as in the previous sub-section (i.e., no renegotiation just to improve agreement). If (10) does not hold for at least one player (an eventuality that was not possible in the previous sub-section), the long-term agreement will not be honoured and players will renegotiate each short-term agreement. As before, I assume that players cannot write a new long term agreement (this assumption has again no strategic bite as the state of nature is known and investments are sunk). That is, in each short-term interaction countries will renegotiate and agree on a transfer $\boldsymbol{\tau}'$ such that

$$v_i(\mathbf{q}^l, I_i, \theta) + \boldsymbol{\tau}_i' \geq v_i(\mathbf{q}^A, I_i, \theta) \text{ for } i = 1, ..N.$$

Note that, as there is no new long-term agreement, this new sequence of short-term agreements has an expected payoff identical to (6). Hence, at stage 4 equilibrium expected payoffs are given by the terms of the long-term agreement if they are profitable for all parts, and by the expected payoffs associated with the sequence of short-term interactions without any agreement otherwise. That is, the payoff is given by:

$$\begin{cases} \frac{v_i(\mathbf{q}^l, I_i, \theta) + \tau_i}{1-\delta} & \text{if } \tau_i \geq v_i(\mathbf{q}^A, I_i, \theta) - v_i(\mathbf{q}^l, I_i, \theta) \text{ for } i = 1, ..N \\ \frac{v_i(\mathbf{q}^A, I_i, \theta) + \varphi_i \sum_{i \in N} \left[v_i(\mathbf{q}^S, I_i, \theta) - v_i(\mathbf{q}^A, I_i, \theta) \right]}{1-\delta} & \text{otherwise.} \end{cases}$$

Define now, for each vector of investment $\mathbf{I}$, the set

$$\Psi(\mathbf{I}) = \left\{ \theta \in \Theta : \ \tau_i \geq v_i(\mathbf{q}^A, I_i, \theta) - v_i(\mathbf{q}^l, I_i, \theta) \right\}.$$

Set $\Psi(\mathbf{I})$ includes all the states of nature, given the investment vector decided at stage 2, for which (10) holds. Hence, given the arbitrary incomplete agreement $\boldsymbol{\tau}$, the expected payoff at stage 2 for player i is:

$$E_{\theta \in \Psi(\mathbf{I})} \left[\frac{v_i(\mathbf{q}^l, I_i, \theta) + \tau_i}{1 - \delta} \right] + E_{\theta \in \Theta / \Psi(\mathbf{I})} \left[\frac{v_i(\mathbf{q}^A, I_i, \theta) + \varphi_i \sum_{i \in N} \left[v_i(\mathbf{q}^S, I_i, \theta) - v_i(\mathbf{q}^A, I_i, \theta) \right]}{1 - \delta} \right] - I_i \tag{11}$$

For abatement the FOC will again be given by (1), implying $q_i^l(I_i, \theta) = q_i^S(I_i, \theta) = q_i^*(I_i, \theta)$. For investment, applying again the envelope theorem, but taking into account that

$\frac{\partial v_j(\mathbf{q}^A,I)}{\partial q_i^A} \neq 0$, the NE investment vector $\mathbf{I}^l$ is given by:

$$\begin{aligned} & E_{\theta\in\Psi(I)}\left[-\left.\frac{\partial C_i}{\partial I_i}\right|_{q_i^*(I_i,\theta)}\right] \\ & +E_{\theta\in\Theta/\Psi(I)}\left[-\varphi_i\left.\frac{\partial C_i}{\partial I_i}\right|_{q_i^*(I_i,\theta)}-(1-\varphi_i)\left.\frac{\partial C_i}{\partial I_i}\right|_{q_i^A(I_i,\theta)}-\sum_{j\in N/i}\varphi_i\left.\frac{\partial B_j}{\partial q_i}\right|_{q_i^A(I_i,\theta)}\frac{q_i^A(I_i)}{\partial I_i}\right] \\ = & \; 1-\delta \quad \text{for } i=1,..N \text{ and } j\neq i \end{aligned} \tag{12}$$

Note that the LHS of (12) is a linear combination of (2) and (5), implying that the non-comprehensive and incomplete agreement brings the investment level closer to the first best. Thus, we can write the following proposition:

Proposition 3 *Countries will not implement the first best through a series of short term agreements regulated by a non-comprehensive and non-contingent long term agreement. Given a level of investment, abatement will be optimal, as it will be defined by (1), i.e.* $q_i^l(I_i,\theta) = q_i^S(I_i,\theta) = q_i^*(I_i,\theta)$*. However, there will be underinvestment, although it will be less severe than in the absence of the long-term agreement. As a consequence, abatement will also be below the optimal level but closer to the first best than without the long-term agreement. Furthermore, each additional state of nature included in the long-term agreement, i.e. each additional state of nature included in* $\Psi(I)$*, weakly approaches the solution to the first best. If* $\Psi(I) \rightarrow \Theta$ *a series of short term agreements following the incomplete long term agreement tends to the first best.*

As described above, the Paris Agreement is best described as a non-comprehensive and non-contingent long-term agreement that simply puts on paper the Nash equilibrium. However, despite this weakness, Proposition 3 shows that it has a role to play, although it will not be able to bring investment to the optimal level, it will nevertheless bring investments closer to the first best. Propositions 2 and 3 also tell us that being non-comprehensive is not the main weakness of the agreement, as investments would be optimal even if countries cannot agree on them if they were able to write a contingent agreement, in the sense that all possible states of nature are considered. Furthermore, the more complete the agreement becomes in this sense, the closer it moves us to the first best. Thus, the Paris Agreement should define as clearly as possible the conditions under which transfers will occur under Articles 6 and 9 and these conditions should depend on the severity of the climate change finally observed (and on other stochastic variables).

3 Conclusion

This paper presents a stylized model to analyze the two two-tier process initiated by the Paris Agreement, where a long-term non-binding agreement is followed by a series of short-term burden sharing agreements. To this end, the paper analyzes long-term climate agreements that are incomplete in a twofold sense. In the first one, they are non-comprehensive because countries do not agree on investments, only on some general rules governing future short-term agreements on abatement levels and associated transfers. In the second sense, long-term agreements on abatement are incomplete because the climate change problem is so complex and multidimensional that it is not possible to write a meaningful agreement as a function of the state of nature. Hence, transfers are non-contingent and cannot depend on the state of nature finally observed. Short-term agreements written after the realization of the stochastic variable are consistent with the long-term agreement if profitable for all parts, and are renegotiated otherwise. These agreements are burden sharing agreements that are binding for the period in which they are negotiated.

Results show that a series of short-term agreements written after a contingent but non-comprehensive long-term agreement can implement the first best. Short-term agreements following a non-contingent and non-comprehensive long-term agreement, as the Paris Agreement, cannot render the first best solution. Nevertheless, this process can improve upon the situation without a long-term agreement, therefore providing a rationale for the Agreement. Furthermore, being non-comprehensive is not the main weakness of the agreement, as investments would be optimal even if countries cannot agree on them if countries were able to write a contingent agreement, in the sense that all possible states of nature are considered.

References

[1] Barrett, S. 1994. Self-enforcing international environmental agreements. *Oxford Econ Pap* 46: 878-894.

[2] Barrett, S., 2002. Consensus Treaties. *J Inst Theor Econ* 158: 529-547.

[3] Barrett, S. and Dannenberg, A., 2016. An experimental investigation into 'pledge and review' in climate negotiations. *Clim change* 138: 339-351.

[4] Battaglini, M. and Harstad, B., 2016. Participation and Duration of Environmental Agreements. *J Political Econ* 124(1):160-204.

[5] Beccherle, J. and Tirole, J., 2011. Regional Initiatives and the Cost of Delaying Binding Climate Change Agreements. *J Public Econ* 95 (11-12): 1339-1348.

[6] Calvo, E. and Rubio, S.J., 2013. Dynamic Models of International Environmental Agreements: A Differential Game Approach. *Int Rev Environ Resour Econ* 6: 289–339.

[7] Caparrós, A., 2016. Bargaining and International Environmental Agreements. *Environ Resour Econ* 65(1): 5–31.

[8] Caparrós, A. and Péreau, J.C., 2017. Multilateral versus sequential negotiations over climate change. *Oxford Econ Pap* 69(2): 365–387.

[9] Caparrós, A., Péreau, J.-C. and Tazdaït, T., 2004. North-South Climate Change Negotiations: a Sequential Game with Asymmetric Information. *Public Choice* 121 (3-4): 455-480.

[10] Carraro, C. and Siniscalco, D., 1993. Strategies for the international protection of the environment. *J Public Econ* 52(3): 309-328.

[11] Chander, P. and Tulkens, H., 1997. The Core of an Economy with Multilateral Environmental Externalities. *Int J Game Theory* 26: 379-401.

[12] Crawford, V., 1988. Long-Term Relationships Governed by Short-Term Contracts. *Am Econ Rev* 78: 485-499

[13] Dutta, P.K. and Radner, R. 2004. Self-enforcing Climate-Change Treaties. *Proceedings of the National Academy of Sciences* 101: 4746–4751.

[14] Dutta, P.K. and Radner, R., 2009. A Strategic Analysis of Global Warming: Theory and Some Numbers. *J Econ Behavr Organ* 71: 187–209.

[15] Dutta, P.K. and Radner, R. 2012. Capital Growth in a Global Warming Model: Will China and India Sign a Climate Treaty? *Econ Theory* 49: 411–443.

[16] Finus, M. and Caparrós, A., 2015. *Game Theory and International Environmental Cooperation: Essential Readings*. Edward Elgar, Cheltenham.

[17] Harstad, B., 2012. Climate Contracts: A Game of Emissions, Investments, Negotiations, and Renegotiations. *Rev Econ Stud* 79(4): 1527-57.

[18] Harstad, B., 2015. The Dynamics of Climate Agreements. *J Euro Econ Assoc* 14(3): 719–752.

[19] Hart, O. and Holstrom, B., 1987. The Theory of Contract. In T. Bewley (ed.) *Advances in Economic Theory*. Cambridge University Press.

[20] Hart, O. and Moore., 1988. Incomplete Contracts and Renegotiation. *Econometrica* 56: 755-786.

[21] Huang, C.-Y., 2002. Multilateral bargaining: conditional and unconditional offers. *Econ Theory* 20:401–412.

[22] Krishna, V. and Serrano, R., 1996. Multilateral bargaining. *Rev Econ Stud* 63:61–80.

[23] Muthoo, A., 1995. Bargaining in a Long-Term Relationship with Endogenous Termination. *J Econ Theory* 66: 590-598.

[24] Osborne, M. and Rubinstein, A., 1994. *A course in Game Theory*, MIT Press, Cambridge, Massachusetts.

[25] Rogelj, J., Elzen, M., Höhne, N., Fransen, T., Fekete, H., Winkler, H., Schaeffer, R., Sha, F., Riahi, K. and Meinshausen, M., 2016. Paris Agreement climate proposals need a boost to keep warming well below 2 °C. *Nature* 534: 631-639.

[26] Segal, I. and Whinston, M. D., 2010. Property Rights. In Gibbons, R. and Roberts, J. (eds) *Handbook of Organizational Economics*. Princeton University Press.

[27] Stavins, R.N. and . Stowe, R.C. (editors), 2017. *Market Mechanisms and the Paris Agreement*. Working Paper of the Harvard Project on Climate Agreements. Cambridge, Mass.

[28] UNFCCC, 1992. *United Nations Framework Convention on Climate Change*. Available at unfcc.int.

[29] UNFCCC, 2015. *Adoption of the Paris Agreement* (FCCC/CP/2015/L.9/Rev. 1). Available at unfcc.int.

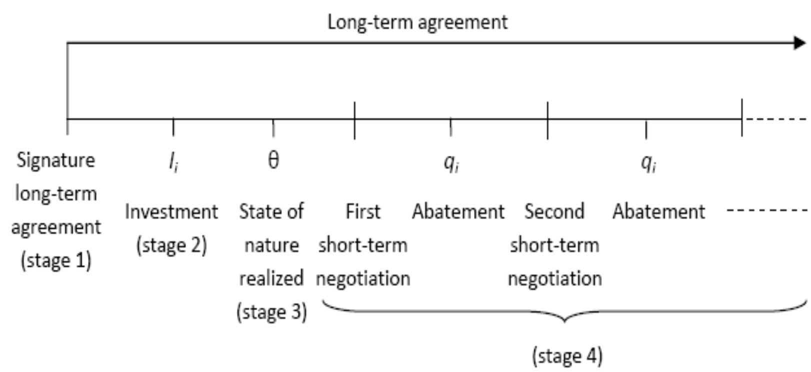

Figure 1. Stages of the game

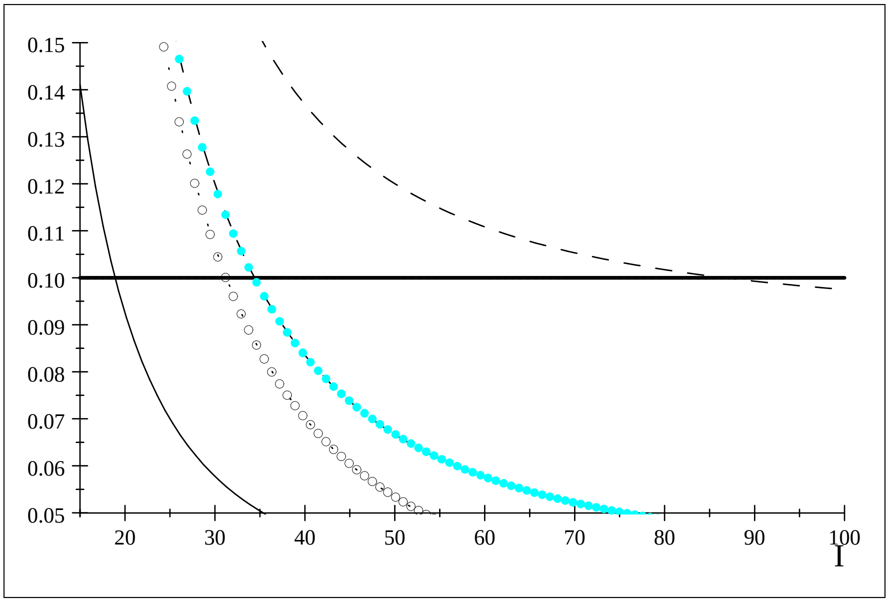

Figure 2. Investment in the first-best and with a sequence of short-term agreements with no long-term agreement.