Regional Initiatives and the Cost of Delaying Binding Climate Change Agreements*

Julien Beccherle†
Jean Tirole ‡

May 17, 2010

Abstract

The Kyoto and Copenhagen Protocols on climate change mitigation postponed the specification of binding commitments to a future negotiation. This paper analyzes the strategic implications of delayed negotiations. While, as is well-understood, the incentive to free ride leads to excessive emissions prior to a binding agreement, the cost of delay is magnified by players' attempt to secure a favorable bargaining position in the future negotiation. A "brinkmanship", an "effort rebalancing", and a "raising rival's cost" effects all concur to generate high post-agreement emissions. The paper applies this general insight to a variety of policy instruments, from the issuance of forward or bankable permits to standards and green investment policies.

Keywords: International negotiations, climate change, cap and trade, bankable permits, standards.

JEL numbers: D62, F51, H23, Q52.

*The authors are grateful to Dominique Bureau, Dominik Grafenhofer and especially Bard Harstad for very helpful comments on an earlier draft.

†Ministère de l'Ecologie, de l'Energie, du Développement Durable, et de la Mer [julien.beccherle@developpement-durable.gouv.fr]

‡Toulouse School of Economics [jean.tirole@tse-fr.eu]

1 Introduction

Climate change is a global issue in need of a global answer. The first attempt at an integrated approach to mitigation, the 1997 Kyoto Protocol, did not specify firm commitments from ratifying countries to cut their emissions. The 2009 Copenhagen conference was meant to define the contours of the post-Kyoto world. Despite widespread agreement on the urgency to act, Copenhagen did not deliver an agreement with binding commitments to emissions cuts, and left the definition of objectives and their verification to future negotiations.

The paper argues that extending the waiting game until Kyoto 3 (say, 2020) would have serious consequences, that go well beyond the celebrated free-riding incentive. Namely, not only will countries engage in suboptimal efforts to reduce their emissions in the next ten years, but they will also consider how their behavior will impact the outcome of negotiations in 2020.[1]

We consider a two-period framework. In period 1, each region of the world chooses a public policy, anticipating the negotiation of a global agreement at date 2. In the generic version of the model, this policy refers to any instrument that impacts the region's date-2 welfare: It may determine its date-2 technological feasibility set, its installed base of polluting equipments, or, in a key application of our theory, the domestic allocation of property rights on pollution allowances. The key feature of the date-1 policy choice is that it affects the region's marginal cost of date-2 abatement, which in turn implies that the region's date-1 choice of public policy is made with an eye on the future negotiation.

The paper's first contribution is to investigate the exact nature of the resulting commitment effect. A natural benchmark is regional optimization or cost minimization. Because the date-1 policy affects the date-2 incentives for abatement, regional optimization aims at minimizing the total (intertemporal) cost to the region. Cost minimization obtains both in the first best, in which a binding agreement is reached at date 1, and in the complete absence of negotiation.

Under delayed negotiation, two new effects concur to push date-2 emissions up. A decrease in one's marginal incentive to abate first implies that the region would pol-

[1]Of course there will be some progress. Carbon permits markets exist or may be created in Europe, the US, Japan and some other developed economies. Emerging countries are taking some action as well. A mixture of collateral damages (the emission of SO_2, a local pollutant, jointly with that of CO_2 by coal plants), the direct impact of own production of CO_2 for large countries like China, and the desire to placate domestic opinion and avoid international pressure will all lead to some carbon control.

lute more, were the negotiation to fail. The "brinkmanship effect" works through a reduction in the *other* region's payoff when the negotiation fails; it changes the region's threat point in the negotiation and enables it to extract more of the surplus; it is particularly potent when the region has substantial bargaining power in the negotiation. Second, the outcome in the negotiation depends on *one's own* welfare, were the negotiation to fail. Thus a region's date-1 policy is in part guided by its own welfare when negotiations break down at date 2. Because emissions are higher when negotiations break down, the country can afford a lower date-1 investment in pollution control; this "effort rebalancing effect" by contrast is most potent when the country's bargaining power is weak. These two effects combined are shown to imply that a delay in negotiating a global agreement increases post-negotiation pollution and not only the pre-negotiation one.

When environmental damage costs are convex, a third strategic effect arises, that reinforces the other two: By committing to a higher pollution level, a region raises the marginal damage cost of all regions and therefore induces others to cut down on their emissions. This is the "raising the other's marginal environmental cost" effect, or "raising rival's cost" effect for short.

We show that delaying negotiation always raises *date-2* emissions compared to the first-best. Indeed, it may even be the case that a delayed negotiation induces more date-2 emissions than in the complete absence of negotiation, and this despite the reduction in emissions brought about by the date-2 agreement.

These results are robust and apply even to small countries, for which all three effects remain non-negligible. In particular the brinkmanship effect applies just the same as the vanishing bargaining power is offset by an almost complete lack of internalization. The date-1 effort is always lower than in the complete absence of negotiation.

The first application of the generic model is to policies that change the installed bases of green equipments or the regions' technological frontiers. Such policies include green investments and standards. We show that delayed negotiations again lead to more post-negotiation emissions than in the first best; put differently, standards are too lenient, investments too brown, and revamping of emitting sources too slow.

The paper's second contribution is to apply these generic insights to the issuance of future allowances and to the bankability of pollution allowances (as embedded in the Waxman-Markey bill). In particular, we predict that regions will issue too many forward or bankable allowances.

Proponents of regional cap-and-trade systems, such as the one existing in Europe or those that are/were under consideration in the United States and a number of other developed countries, take the view that regional markets will jump-start climate change mitigation and later lead to convergence to a single, worldwide climate treaty. We investigate the strategic implications of a delayed agreement in the context of the issuance of forward allowances; we also show that similar insights apply to the issuance of bankable permits. In either case, the region puts today into private hands allowances that can be used to cover future emissions.

Under linear environmental damage costs, we first show that, following the logic of the Coase conjecture, regions issue forward allowances neither in the complete absence of negotiation nor in the first best (in which negotiation takes place at date 1). By contrast, in the intermediate situation in which negotiations are delayed, the brinkmanship and effort rebalancing effects both imply that regions issue forward allowances whenever date-2 emissions increase with the number of such allowances. The latter property holds for example if the regions auction off date-2 new (spot) allowances and face a shadow cost of public funds; because regions put some weight on revenue, they have a tendency to "over issue" spot permits at date 2, thereby lowering the price of carbon. Alternatively, the government could internalize only partially the welfare of the holders of forward permits. Or else, the new permits could be distributed for free at date 2, but the government might at date 2 negotiate domestically with a powerful industrial lobby, whose status-quo welfare is stronger, the larger the number of allowances it received at date 1.

Either way, delayed negotiations lead to high future emissions through an excessive issuance of forward or bankable permits even though they can be retired or fewer spot permits issued. A political implication of our analysis is that if an ambitious climate treaty is impossible today, countries should at least agree to limit banking and forward-selling.[2] Finally, we show that markets are merged under symmetrical conditions, but that the agreement may otherwise content itself with a specification of the volume of emissions in maintained regional markets.

Finally, we allow for leakage. Provided that a large enough fraction of industries are mobile, regional carbon markets de facto merge into a single one even when negotiations break down. A region's issuing forward or bankable permits as earlier is a com-

[2]Of course bankability has the (well-known) benefit of smoothing the carbon price when transitory shocks to economic activity or technological progress would make this price very volatile.

mitment to pollute more in the absence of agreement. We show that the brinkmanship and effort rebalancing effects obtain, and that a new raising rival's cost effect arises: even in the absence of convex damage costs, region j's marginal benefit from issuing permits in the absence of agreement is decreased by the abundance of permits worldwide.

The paper is organized as follows: Section 2 sets up the generic model. Section 3 identifies the brinkmanship, effort rebalancing, and raising rival's cost effects, derives the excess pollution result and gives first applications. Section 4 develops the application to forward allowances and bankable pollution permits. Section 5 concludes.

Literature review and contribution

This paper contributes to several literatures. First, the forward-market application relates to, among others, Allaz and Vila (1993) and Mahenc and Salanié (2004), who investigate the idea that forward markets can be used to influence rivals. Allaz and Vila find that if competing firms set quantities of forward rights over time, then they will sell more than they would in a spot market. Selling one unit today reduces the competitors' future marginal benefits and therefore increases the firm's profit. In much of the paper, we abstract from the Allaz-Vila (raising rival's cost) effect by assuming that marginal pollution damages are linear in total pollution and focus on the impact of negotiation, a question that is moot in Allaz and Vila's oligopoly framework. Laffont and Tirole (1996a,b) study cap-and-trade policies with spot and forward markets and analyze how regulatory commitment and flexibility to news can be made consistent. Like Allaz-Vila, these papers however do not consider negotiations among multiple regulators/countries, which is the focus of the current paper.[3]

Second, by taking the view that negotiations take time, we implicitly study the role of incomplete contracts and the importance of property rights allocations in a context of global externalities. Our paper thereby builds on the incomplete contracts and hold up literature (Grossman and Hart (1986), Hart and Moore (1990), Williamson (1985)). The common thread with that literature is the idea that ex-ante investments influence the outcome of ex-post negotiations.

Third, this paper contributes to the growing literature on climate change agreements[4] and casts some light on the effects of delayed negotiations on international

[3]There is also a growing literature on closed-economy, optimal dynamic multi-instrument policies (for example, carbon price and R&D subsidies, as in Acemoglu et al (2009) and Grimaud and Rouge (2008)).

[4]See e.g., Aldy and Stavins (2007) and Barrett (2005). The paper also builds on the extensive literature

climate change agreements. Our contribution is to include dynamics into the analysis and to put forward the potential cost of delayed negotiations. An early paper emphasizing the potential of R&D reduction when investments can be held up in future renegotiations is Buchholz and Konrad (1994). The most closely related paper is Harstad (2009) who develops a very interesting analysis of the dynamics of climate change agreements. Harstad studies an economy where sovereign countries repeatedly make investment and emission decisions. The cost of pollution depends on the total stock of emissions, accumulated over time. Harstad demonstrates, as we do, that under some conditions short-lasting agreements lead to higher pollution than no agreement at all. Notably, he establishes this result in an infinite horizon model while we have only two periods.

Under some assumptions,[5] Harstad's state space, which a priori has $(n+1)$ dimensions (the existing stock of pollution, together with the technological state of each the n countries), collapses to a two-dimensional recursive one, in which outputs are fully determined by the existing stock of pollution and the latter's evolution depends on these outputs and the global stock of knowledge. We chose to be more general in terms of technologies and to allow for asymmetries (in particular, in our model, which country is more technology advanced impacts the continuation equilibrium) at the expense of a two-period analysis. This added generality allows us to investigate interesting policy instruments such as forward and bankable allowances, which the Harstad assumptions do not capture, as well as the consequences of asymmetric preferences and asymmetric bargaining powers. We unveil the three effects at stake and check that strategic incentives apply equally well to small and large regions.

2 A generic model

2.1 Timing and utility functions

The model has two periods, $t = 1,\ 2$ and two regions of the world, $i = A,\ B$.[6] For notational simplicity, and without loss of generality, we normalize date-2 payoffs so that regions do not discount the future. In period 1, regional authorities (regulators) non-cooperatively and simultaneously choose their levels of pollution control, a_1^i for

on coalition formation in international agreements (e.g., Carraro and Siniscalco (1993)).

[5]See Section 3.5 for a statement of these assumptions.

[6]The analysis is generalized to an arbitrary number of countries in Section 3.6.

region i. At this stage $a_1^i \in \mathbb{R}$ may stand for any instrument[7] available to the regulator to control the pollution of her region. We normalize a_1^i such that a higher a_1^i is *less* environment-friendly. For example, a high a_1^i may correspond to a lax pollution standard, a low investment in green technology or a high level of issuance of forward allowances.

The negotiation with the other region takes place at the beginning of period 2. Thus, the game is a stylized representation of a situation in which regions delay a binding agreement, be it for a lack of adequate preparation on the framework of governance, a delay in setting up a measurement system, non-synchronized political agendas [8] or for another reason [9].

The absence of date-1 agreement implies that the choices of date-1 emissions are driven by the familiar free-riding incentive. Without loss of generality and unless otherwise stated,[10] we omit the corresponding analysis for clarity of exposition. At date 2, the regulators will agree on second-period pollution levels $a_2^i \in \mathbb{R}^+$ in both regions and on a side payment, for example through an allocation of pollution allowances or direct cash transfers. While a_1^i can stand for any (environment-unfriendly) policy that changes region i's date-2 emissions incentives, a_2^i denotes the actual level of date-2 emissions.[11]

Regions' date-1 and date-2 welfares, gross of pollution damages, are denoted $U_1^i(a_1^i)$ and $U_2^i(a_1^i, a_2^i)$. These functions are increasing in a_1^i and a_2^i respectively, and twice differentiable. U_2^i is twice differentiable and strictly concave in a_2^i. Region i's pollution damage depends linearly[12] on the total amount of emissions at date 2; if c^i is region i's marginal damage, then its environmental cost is $c^i a_2$ where $a_2 \equiv a_2^A + a_2^B$. Let $c \equiv c^A + c^B$ denote the total marginal damage cost.

[7]The results carry over to a multi-dimensional date-1 action space, provided that the standard supermodularity assumptions are satisfied (see e.g., Milgrom and Roberts (1990)).

[8]If a region is governed by a political party opposed to international negotiations on climate change, the other region must wait for this government to be defeated and replaced by a more favorable government before it can enter real negotiations.

[9]Yet another possible explanation is an initial asymmetry of information between regions, leading to a bargaining breakdown in early stages.

[10]Date-1 and date-2 emissions are interdependent in the case of bankable allowances (Section 4.2). There, we will formally introduce date-1 pollution.

[11]As we will later discuss, this is unrestrictive since any date-2 pollution-control policy corresponds to a unique level of emissions.

[12]We make the assumption that the environmental costs depend linearly on the total amount of pollution for simplicity, but we show in Section 3 that the results obtained in the linear case can be extended to a general environmental cost function. The benefit of assuming linear damage functions is that we rule out Allaz-Vila style effects.

Assumption 1 *(i) The marginal utility of polluting in the second period is decreasing with the first-period pollution control:*

$$\frac{\partial^2 U_2^i}{\partial a_1^i \partial a_2^i} > 0.$$

(ii) Let

$$\Gamma^i(a_1^i) \equiv \frac{dU_1^i}{da_1^i}(a_1^i) + \frac{\partial U_2^i}{\partial a_1^i}\Big(a_1^i, \widehat{a}_2^i(a_1^i)\Big) \quad \textit{for all } a_1^i$$

where

$$\widehat{a}_2^i(a_1^i) \equiv \arg\max\{U_2^i(a_1^i, a_2^i) - ca_2^i\}.$$

There exists a_1^{iFB} such that $\Gamma^i(a_1^{iFB}) = 0$, $\Gamma^i(a_1^i) > 0$ if $a_1^i < a_1^{iFB}$ and $\Gamma^i(a_1^i) < 0$ if $a_1^i > a_1^{iFB}$.

In the following, "hats" refer to the values of the parameters that emerge from an efficient date-2 agreement, while "star" superscripts correspond to the Nash outcome that would result from a failure to agree. Thus, $\widehat{a}_2^i$ denotes region i's second-period pollution control after the agreement. Were the negotiation to break down at date 2, regional regulators would choose non-cooperatively second-period pollution levels, ${a_2^i}^*$ for region i. This benchmark defines the outside options in the date-2 negotiation. Both $\widehat{a}_2^i$ and ${a_2^i}^*$ are functions of prior policies.

2.2 Welfare functions and date-2 bargaining

Letting T^i denote the transfer received by region i as part of the date-2 agreement ($T^A + T^B = 0$), the second-period welfares of the world and of region i, W_2 and W_2^i, respectively, can be written as:

$$W_2(a_2^A, a_2^B, a_1^A, a_1^B) \equiv U_2^A(a_1^A, a_2^A) + U_2^B(a_1^B, a_2^B) - c(a_2^A + a_2^B) \tag{1}$$

and

$$W_2^i(a_2^i, a_2^j, a_1^i) + T^i \equiv U_2^i(a_1^i, a_2^i) - c^i(a_2^i + a_2^j) + T^i. \tag{2}$$

Let, for given date-1 actions,

$$\widehat{W}_2(a_1^A, a_1^B) \equiv W_2(\widehat{a}_2^A(a_1^A), \widehat{a}_2^B(a_1^B), a_1^A, a_1^B),$$

and

$${W_2^i}^*(a_1^i, a_1^j) \equiv W_2^i({a_2^i}^*(a_1^i), {a_2^j}^*(a_1^j), a_1^i),$$

where

$$a_2^{i*}(a_1^i) \equiv \arg\max\{U_2^i(a_1^i, a_2^i) - c^i a_2^i\}$$

and

$$\widehat{a}_2^i(a_1^i) \equiv \arg\max\{U_2^i(a_1^i, a_2^i) - c a_2^i\} < a_2^{i*}(a_1^i),$$

using revealed preference: A successful negotiation reduces emissions.[13] Revealed preference also implies that the functions $\widehat{a}_2^i$ and a_2^{i*} are monotonic in the first-period effort:

Lemma 1. *$\widehat{a}_2^i$ and a_2^{i*} are non-decreasing (strictly increasing if $c^i > 0$ for all i) functions of a_1^i. Furthermore, for all a_1^i, $a_2^{i*}(a_1^i) > \widehat{a}_2^i(a_1^i)$.*

Thus, the more stringent the first-period pollution control policy, the lower the second-period pollution. In particular, applied to the case in which negotiations are delayed, by adopting loose pollution control policies in the first-period a region can credibly commit to high date-2 pollution, were the negotiation to break down. This commitment is a key element of our analysis.

Date-2 bargaining: At date 2 the two regulators agree on their pollution levels and on a side payment. We model the bargaining outcome by the Nash bargaining solution.[14] Calling α^A and α^B the bargaining powers of region A and B (with $\alpha^A + \alpha^B = 1$), region $i \in \{A,\ B\}$'s intertemporal payoff after negotiation, W_{neg}^i, is given by:

$$W_{neg}^i(a_1^i, a_1^j) = U_1^i(a_1^i) + W_2^{i*}(a_1^i, a_1^j) + \alpha^i\left(\widehat{W}_2(a_1^i, a_1^j) - \left(W_2^{A*}(a_1^i, a_1^j) + W_2^{B*}(a_1^i, a_1^j)\right)\right) \tag{3}$$

The term $U_1^i(a_1^i)$ in (3) is the date-1 utility associated with choosing policy a_1^i. This term is fixed (sunk) at date 2. $W_2^{i*}(a_1^i, a_1^j)$ is the outside option of region i, while

$$\alpha^i\left(\widehat{W}_2(a_1^i, a_1^j) - \left(W_2^{A*}(a_1^i, a_1^j) + W_2^{B*}(a_1^i, a_1^j)\right)\right)$$

is the share of the total surplus extracted by region i during the negotiation.

[13]Interestingly, the European Union ETS price fell by 21% in 2009 as it became more and more unlikely that a satisfactory agreement would be drawn. While some of this decrease may be due to news about the economic recession, the 9% drop in price immediately after the Copenhagen Accord is a clear sign.

[14]They are other approaches to modeling bargaining. It seems reasonable to assume that the outcome of a negotiation depends at least partially on the outside options available to the parties.

3 The cost of delaying negotiations

3.1 Benchmarks: first best and autarky

Before tackling the case of delayed negotiations, let us look at the two polar cases in which negotiations take place at both dates (first best)[15] or never take place (autarky).

First best: The pollution levels agreed upon during the negotiation $(\widehat{a}_2^A,\ \widehat{a}_2^B)$ are optimal given the first-period choices $(a_1^A,\ a_1^B)$. From the envelope theorem, the first best can be obtained by differentiating

$$U_1^A(a_1^A) + U_1^B(a_1^B) + U_2^A(a_1^A, a_2^A) + U_2^B(a_1^B, a_2^B) - c(a_2^A + a_2^B)$$

with respect to $\{a_1^A, a_1^B\}$, and evaluating it at $\{\widehat{a}_2^A,\ \widehat{a}_2^B\}$. This gives, for $i \in \{A, B\}$:

$$\Gamma^i(a_1^i) \equiv \frac{dU_1^i}{da_1^i}(a_1^i) + \frac{\partial U_2^i}{\partial a_1^i}(a_1^i, \widehat{a}_2^i(a_1^i)) = 0 \tag{4}$$

As announced in Assumption 1, let a_1^{iFB} denote the unique value of a_1^i satisfying this equation.

Autarky: The outcome when there is never any negotiation is given by a similar regional optimization equation:

$$\frac{dU_1^i}{da_1^i}(a_1^i) + \frac{\partial U_2^i}{\partial a_1^i}(a_1^i, a_2^{i*}(a_1^i)) = 0, \tag{5}$$

Let a_1^{i*} denote the solution to equation (5), which we will assume is unique.

3.2 Delayed negotiation: the effort rebalancing and brinkmanship effects

Suppose now that regions negotiate only at date 2. We assume that the game with payoff functions $\{W_{neg}^i(a_1^i, a_1^j),\ W_{neg}^i(a_1^i, a_1^j)\}$ admits a unique Nash equilibrium, $\{a_1^{iDN},\ a_1^{jDN}\}$, where "$DN$" stands for "Delayed Negotiation". Before proving our main result, we identify more precisely the effects mentioned above. Region i's regulator chooses a_1^i so as to maximize W_{neg}^i as defined by equation (3). Note that, while W_{neg}^i depends on a_1^j,

[15] It does not matter whether the two regions negotiate at date 1 a long-term agreement for the two periods, or negotiate at each date t an agreement specifying the countries' short-term policies.

its derivative with respect to a_1^i does not. From the envelope theorem, the first-order condition is:

$$\begin{aligned}\frac{\partial W_{neg}^i}{\partial a_1^i}(a_1^i) = & \left(\frac{dU_1^i}{da_1^i}(a_1^i) + \frac{\partial U_2^i}{\partial a_1^i}(a_1^i, \widehat{a}_2^i(a_1^i))\right) \\ & + (1-\alpha^i)\left(\frac{\partial U_2^i}{\partial a_1^i}(a_1^i, {a_2^i}^*(a_1^i)) - \frac{\partial U_2^i}{\partial a_1^i}(a_1^i, \widehat{a}_2^i(a_1^i))\right) \\ & + \alpha^i c^j \frac{d{a_2^i}^*}{da_1^i}(a_1^i) = 0.\end{aligned} \tag{6}$$

The regional optimization effect is

$$\left(\frac{dU_1^i(a_1^i)}{da_1^i} + \frac{\partial U_2^i}{\partial a_1^i}(a_1^i, \widehat{a}_2^i(a_1^i))\right)$$

and is nil at $a_1^i = a_1^{iFB}$.

The effort rebalancing effect is given by

$$(1-\alpha^i)\left(\frac{\partial U_2^i}{\partial a_1^i}(a_1^i, {a_2^i}^*(a_1^i)) - \frac{\partial U_2^i}{\partial a_1^i}(a_1^i, \widehat{a}_2^i(a_1^i))\right).$$

It is positive since ${a_2^i}^*(a_1^i) > \widehat{a}_2^i(a_1^i)$ for all a_1^i (the negotiation leads to less pollution) and

$$\frac{\partial^2 U_2^i(a_1^i, a_2^i)}{\partial a_1^i \partial a_2^i} > 0.$$

Furthermore, it is larger, the smaller is α^i. A region with a low bargaining power has a strong incentive to free-ride, in the expectation that an agreement will induce a stricter pollution control in the second period.

The brinkmanship effect is given by

$$\alpha^i c^j \frac{d{a_2^i}^*}{da_1^i}(a_1^i).$$

Because ${a_2^i}^*$ increases with a_1^i, this effect is always positive. It is larger when α^i is large: A country with substantial bargaining power is able to extract most of the surplus created by an agreement and therefore benefits from lowering the other party's outside option. Also the brinkmanship effect is more important, the larger the other region's environmental cost. Indeed, if region j is sensitive to pollution (c^j is large), then region i's threat is all the more effective.

Proposition 1. *Delayed negotiations raise the post-negotiation pollution by encouraging lax environmental policies* a_1^{iDN} *prior to the negotiation: For all i,* $a_1^{iDN} > a_1^{iFB}$ *and so* $a_2^{iDN} \equiv \widehat{a}_2^i(a_1^{iDN}) > \widehat{a}_2^i(a_1^{iFB})$.

Proof. Country i's first-order condition is:

$$\frac{\partial W_{neg}^i}{\partial a_1^i}(a_1^i) = \left(\frac{dU_1^i(a_1^i)}{da_1^i} + \frac{\partial U_2^i}{\partial a_1^i}(a_1^i, \widehat{a}_2^i(a_1^i)) \right)$$

$$+ (1-\alpha^i) \left(\frac{\partial U_2^i}{\partial a_1^i}(a_1^i, a_2^{i*}(a_1^i)) - \frac{\partial U_2^i}{\partial a_1^i}(a_1^i, \widehat{a}_2^i(a_1^i)) \right)$$

$$+ \alpha^i c^j \frac{da_2^{i*}}{da_1^i}(a_1^i) = 0$$

Because, for all a_1^i, $a_2^{i*}(a_1^i) > \widehat{a}_2^i(a_1^i)$ and $\dfrac{\partial^2 U_2^i(a_1^i, a_2^i)}{\partial a_1^i \partial a_2^i} > 0$,

$$(1-\alpha^i) \left(\frac{\partial U_2^i}{\partial a_1^i}(a_1^i, a_2^{i*}(a_1^{iDN})) - \frac{\partial U_2^i}{\partial a_1^i}(a_1^i, \widehat{a}_2^i(a_1^{iDN})) \right) \geq 0.$$

Assumption 1 *(i)* implies that

$$\alpha^i c^j \frac{da_2^{i*}}{da_1^i}(a_1^{iDN}) > 0.$$

And so

$$\Gamma^i(a_1^{iDN}) < 0$$

Assumption 1 (ii) then implies that $a_1^{iDN} > a_1^{iFB}$. In equilibrium there is an overprovision of a_1^i relative to the first-best level. □

3.3 Comparison with autarky

More assumptions are needed in order to compare the autarky level ${a_1^i}^*$ with the equilibrium value ${a_1^i}^{DN}$ under period-2-only negotiation. The outcome under autarky is given by condition (5). Note that condition (5) is a special case of condition (6) with, instead of $\alpha^A + \alpha^B = 1$,

$$\alpha^i = 0 \quad \text{for all } i.$$

Thus a reinterpretation of the autarky case is that the brinkmanship effect vanishes while the effort-rebalancing effect is strongest.

For symmetric, quadratic payoff functions, these two considerations exactly offset and so the date-1 policies are the same with and without date-2 negotiation.[16] Further results on this comparison will be provided in specialized versions of the generic model.

3.4 Non-linear pollution damages: raising the rival's cost

The linear environmental-costs assumption allowed us to cleanly focus on the effect of the date-2 negotiation on the date-1 incentive. Yet the effects of greenhouse gases (GHG) seem to be convex rather than linear. The nonlinearity in the environmental damage function actually magnifies the strategic incentive. The intuition resembles that developed in Allaz and Vila (1993).

Let $C^i(a_2)$ denote the environmental cost function of region i. Let $C(a_2) \equiv C^A(a_2) + C^B(a_2)$ denote the total environmental cost. We assume that C^i is twice differentiable and is increasing and convex (that is $\frac{dC^i}{da_2}(a_2) > 0$, and $\frac{d^2C^i}{d{a_2}^2}(a_2) > 0$).

The second-period welfare functions are now:

$$W_2(a_2^A, a_2^B, a_1^A, a_1^B) = U_2^A(a_1^A, a_2^A) + U_2^B(a_1^B, a_2^B) - C(a_2) \tag{7}$$

$$W_2^i(a_2^i, a_2^j, a_1^i) = U_2^i(a_1^i, a_2^i) - C^i(a_2) \tag{8}$$

[16] Let $U_2^i(a_1^i, a_2^i) = -\frac{\beta}{2}[a_2^i]^2 + \gamma a_2^i + \delta[a_1^i a_2^i]$ and $U_1^i(a_1^i) = \phi a_1^i - \frac{\Phi}{2}[a_1^i]^2$. Then

$$a_1^{iDN} = {a_1^i}^* = \frac{\phi\beta + \delta(\gamma - c^i)}{\beta\Phi - \delta^2}.$$

Therefore $a_2^{FB} < a_2^{DN} < {a_2}^*$.

Under *autarky*, a region's second-period emissions impact the other region's marginal damage cost. Regions play a Nash equilibrium of the game with payoff functions

$$W_2^i(a_2^i, a_2^j, a_1^i).$$

Let $\{a_2^{i*}(a_1^i, a_1^j)\}_{i\in\{A,B\}}$ denote the Nash equilibrium of the game which, we assume, is unique and stable. Similarly, $\{\widehat{a}_2^i(a_1^i, a_1^j)\}_{i\in\{A,B\}}$ maximizes $W_2(a_2^A, a_2^B, a_1^A, a_1^B)$ and is assumed to be unique.

For given date-1 actions, let

$$\widehat{W}_2(a_1^A, a_1^B) \equiv W_2(\widehat{a}_2^A(a_1^A, a_1^B), \widehat{a}_2^B(a_1^A, a_1^B), a_1^A, a_1^B),$$

and

$$W_2^{i*}(a_1^i, a_1^j) \equiv W_2^i(a_2^{i*}(a_1^A, a_1^B), a_2^{j*}(a_1^A, a_1^B), a_1^i).$$

Region i's intertemporal payoff after negotiation, W_{neg}^i is:

$$W_{neg}^i(a_1^i, a_1^j) = U_1^i(a_1^i) + W_2^{i*}(a_1^i, a_1^j) + \alpha^i \left(\widehat{W}_2(a_1^A, a_1^B) - \left(W_2^{i*}(a_1^i, a_1^j) + W_2^{j*}(a_1^i, a_1^j)\right)\right).$$

From the envelope theorem, the first-order condition is:

$$\begin{aligned}
\frac{\partial W_{neg}^i}{\partial a_1^i}(a_1^i, a_1^j) = & \left(\frac{dU_1^i(a_1^i)}{da_1^i} + \frac{\partial U_2^i}{\partial a_1^i}(a_1^i, \widehat{a}_2^i(a_1^i, a_1^j))\right) \\
& + (1-\alpha^i)\left(\frac{\partial U_2^i}{\partial a_1^i}(a_1^i, a_2^{i*}(a_1^i, a_1^j)) - \frac{\partial U_2^i}{\partial a_1^i}(a_1^i, \widehat{a}_2^i(a_1^i, a_1^j))\right) \\
& + \alpha^i C^{j\prime}\big(a_2^*(a_1^A, a_1^B)\big)\frac{\partial a_2^{i*}}{\partial a_1^i}(a_1^i, a_1^j) \\
& + (1-\alpha^i)C^{i\prime}\big(a_2^*(a_1^A, a_1^B)\big)\big(-\frac{\partial a_2^{j*}}{\partial a_1^i}(a_1^i, a_1^j)\big) = 0
\end{aligned} \tag{9}$$

Condition (9) is very similar to equation (6); the three effects identified earlier (regional optimization, effort rebalancing and brinkmanship) are still at work. But introducing non-linear environmental costs adds a fourth term:

$$(1-\alpha^i)C^{i\prime}\big(a_2^*(a_1^A,a_1^B)\big)\Big(-\frac{\partial a_2^{j\,*}}{\partial a_1^i}(a_1^i,a_1^j)\Big).$$

An increase in a_1^i makes date-2 abatement more costly to region i and so commits that region to high emissions in the absence of agreement. This raises region j's marginal damage cost and thus induces region j to reduce its own emissions under autarky. Region j faces a higher marginal cost of pollution and reduces its second-period pollution: Because we assumed that the Nash equilibrium is stable, $a_2^{j\,*}$ is decreasing in a_1^i (and increasing in a_1^j). This is the *raising rival's cost effect*.

This effect is larger, the larger is $1-\alpha^i$, that is, the smaller the bargaining power of the region. The new effect reinforces the effort rebalancing and brinkmanship effects. We thus conclude that Proposition 1 holds a fortiori.

3.5 The symmetric additive specification

Harstad (2009) assumes that region i derives an increasing and concave benefit $B(y_2^i)$ from date-2 consumption y_2^i.[17] Symmetry obtains, as $B(\cdot)$ as well as the damage function $C(\cdot)$ are the same for both regions. Finally, emissions take an additive form:

$$a_2^i = y_2^i + a_1^i.$$

Thus

$$B^{'}\big(a_2^{i*}-a_1^i\big) = C^{'}\big(a_2^*\big), \text{ where } a_2^* \equiv \Sigma_i a_2^{i*}$$

and

$$B^{'}\big(\widehat{a}_2^i-a_1^i\big) = 2C^{'}\big(\widehat{a}_2\big), \text{ where } \widehat{a}_2 \equiv \Sigma_i \widehat{a}_2^i.$$

And so, letting $a_1 \equiv \Sigma_i a_1^i$, $\psi(x) \equiv (B^{'})^{-1}\big(C^{'}(x)\big)$, and $\widehat{\psi}(x) \equiv 2(B^{'})^{-1}\big(2C^{'}(x)\big)$ (ψ and $\widehat{\psi}$ are decreasing),

$$a_2^* = a_1 + \psi(a_2^*)$$

and

$$\widehat{a}_2 = a_1 + \widehat{\psi}(\widehat{a}_2).$$

The linear additive assumption of Harstad's model implies that knowledge is a pure public good, on par with environmental quality. The outcomes in the presence or absence of negotiation are independent of who made the prior investments in green technologies (only a_1 matters).

[17] Recall that his model has an infinite horizon. We recast it in our two-period framework for the sake of comparison.

We can push the analysis a bit further and allow for asymmetric bargaining weights in the symmetric additive specification. The effort rebalancing, brinkmanship and raising rival's costs effects are given by the following three terms, respectively:

$$\Delta(a_1) \equiv (1-\alpha^i)\Big[2C^{'}(\widehat{a}_2(a_1)) - C^{'}(a_2^*(a_1))\Big]$$

$$+\alpha^i C^{'}(a_2^*(a_1))\left(\frac{1-\psi^{'}(a_2^*(a_1))}{1-2\psi^{'}(a_2^*(a_1))}\right)$$

$$+(1-\alpha^i)C^{'}(a_2^*(a_1))\left(\frac{-\psi^{'}(a_2^*(a_1))}{1-2\psi^{'}(a_2^*(a_1))}\right).$$

For symmetric bargaining powers ($\alpha^i = 1/2$ as is assumed in Harstad), the total strategic effect is $\Delta = C^{'}(\widehat{a}_2)$.

Another simple case arises when damage costs are linear ($\psi^{'} = 0$). Then Δ is independent of the bargaining powers and $da_2^{i*}/da_1^i = d\widehat{a}_2^i/da_1^i = 1$.

The following result is a direct application of these formulas:

Proposition 2. *Consider the symmetric additive specification and assume that the damage function $C(a_2)$ is linear. Then date-1 efforts are the same whether or not negotiation occurs at date 2. Put differently, the brinkmanship and effort rebalancing effects cancel out: $a_1^{iDN} = a_1^{i\,*}$. And so $a_2^{iDN} < a_2^{i\,*}$ for all i.*

Remark: An application of the symmetric additive specification is to the choice at date 1 of a pollution standard[18] that defines the environmental quality of equipments that will not be revamped at date 2. Letting a_1^i denote the resulting *date-2* pollution and y_2^i denote the new pollution at date 2, then $a_2^i = y_2^i + a_1^i$.

3.6 n-country version

We have so far assumed that there are only two regions. While it is technically straightforward to extend the analysis to an arbitrary number of countries, it may not be obvious how the various effects play out for small countries. This section's key insight

[18] Examples include CO_2 emission standards for automobiles in Europe, and in the United States, a minimum mileage legislation for cars and trucks (miles traveled per gallon of gasoline). In France, an environmental law package called "Grenelle de l'Environnement" introduced upper limits on housing consumption (by 2012 every new building will have to consume less than 50 kWh/m^2/year).

is that even small countries do play strategically. Namely the brinkmanship effect remains strong despite of the fact that small countries' bargaining power is low. The intuition can be grasped from looking at the expression of the brinkmanship effect: This effect is proportional to the country's bargaining power, suggesting that it should vanish for small countries. However it is also proportional to the damage cost of the "exported pollution". When the country is small, the exported pollution becomes infinitely large relative to internalized pollution. This offsetting effect guarantees that the brinkmanship effect never vanishes.

Let $\vec{a}_1 = (a_1^1, a_1^2, \ldots a_1^n)$ and α^i denote region i's bargaining power. The first-order condition is:

$$\begin{aligned}\frac{\partial W^i_{neg}}{\partial a^i_1}(\vec{a}_1) = &\left(\frac{dU^i_1(a^i_1)}{da^i_1} + \frac{\partial U^i_2}{\partial a^i_1}(a^i_1, \hat{a}^i_2(\vec{a}_1))\right) \\ &+ (1-\alpha^i)\left(\frac{\partial U^i_2}{\partial a^i_1}(a^i_1, {a^i_2}^*(\vec{a}_1)) - \frac{\partial U^i_2}{\partial a^i_1}(a^i_1, \hat{a}^i_2(\vec{a}_1))\right) \\ &+ \alpha^i\left[\sum_{j\neq i} C^{j\prime}\big(a^*_2(\vec{a}_1)\big)\right]\frac{\partial {a^i_2}^*}{\partial a^i_1}(\vec{a}_1) \\ &+ (1-\alpha^i)C^{i\prime}\big(a^*_2(\vec{a}_1)\big)\Big[-\sum_{j\neq i}\frac{\partial {a^j_2}^*}{\partial a^i_1}(\vec{a}_1)\Big] \\ &+ \alpha^i\sum_{j\neq i}\sum_{k\neq j,i}\left[C^{j\prime}\big(a^*_2(\vec{a}_1)\big)\right]\frac{\partial {a^k_2}^*}{\partial a^i_1}(\vec{a}_1) = 0\end{aligned} \tag{10}$$

The second term captures the effort rebalancing effect and the third the brinkmanship effect. For linear damage costs, the fourth and fifth terms vanish.

Proposition 3. *Let environmental damage costs be linear. Suppose that one increases the number n of countries so that $\alpha^i c = O(1)$, where $c = \sum_j c^j$ is the total marginal pollution damage. Then, as n tends to infinity:*

(i) because pollution is more and more exported when the country becomes small, the brinkmanship effect does not collapse despite vanishing bargaining power;

(ii) for all i, the date-1 effort is always lower under delayed negotiation than under autarky.

3.7 An application to investment in green technologies

Even though the bulk of investment in green technologies is made by the private sector, governments have substantial control over it through tax incentives, feed-in tariffs and so forth. Furthermore, public R&D may be geared toward green technologies, and the government can decide on the green investment level of public sector companies.[19] Let us show how the level of investment in green technologies can also be used as a strategic instrument to influence the negotiation outcome.

Strategic actions and utility functions.

In each period, regulators (directly or indirectly) choose their investment levels: in period 1, region i's regulator makes a fraction X^i of its industry pollution-free for date 2 (we neglect date-1 pollution. For example, X^i might be the fraction of green buildings or equipments built at date 1 and becoming operational at date 2). This investment costs $\Psi_1^i(X^i)$. The function Ψ_1^i is increasing and convex, with $\Psi_1^{i'}(0) = 0$ and $\Psi_1^{i'}(1) = +\infty$. At date 1, only the remaining fraction $1 - X^i$ of the domestic firms pollute. Similarly, the regions invest in pollution abatement in the second period. Let Y^i denote the fraction of remaining polluting firms for which abatement is decided at date 2 (so Y^i can be viewed as the fraction of "brown" buildings or equipments that will need to be revamped in order to achieve green performance). This investment costs $(1 - X^i)\Phi_2^i(Y^i) \equiv \Psi_2^i(X^i, Y^i)$, where Φ_2 is increasing and convex.

Let

$$a_2^i \equiv (1 - X^i)(1 - Y^i),$$

and

$$a_1^i \equiv 1 - X^i.$$

The first-period utility U_1^i is (minus) the investment cost a_1^i:

$$U_1^i(a_1^i) = -\Psi_1^i(1 - a_1^i).$$

The second-period utility is also (minus) the investment cost for a pollution level a_2^i:

[19]The French government holds an 85 % share in the country's largest utility EDF; the Chinese government owns most of the large emitters.

$$U_2^i(a_1^i, a_2^i) = -a_1^i \Phi_2^i\Big(1 - \frac{a_2^i}{a_1^i}\Big).$$

Checking Assumption 1

We can check that Assumption 1 is satisfied. First, differentiating U_2^i with respect to a_1^i and a_2^i

$$\frac{\partial^2 U_2^i}{\partial a_2^i \partial a_1^i} = \Phi_2^{i''} \frac{a_2^i}{(a_1^i)^2} > 0$$

since Φ_2^i is convex. Second, from the first-order condition in the negotiated case

$$\Phi_2^{i\prime}\Big(1 - \frac{\widehat{a}_2^i(a_1^i)}{a_1^i}\Big) = c.$$

Thus $\widehat{a}_2^i(a_1^i)/a_1^i$ is constant, and so the term in brackets in

$$\Gamma^i(a_1^i) = \Psi_1^{i\prime}(1 - a_1^i) + \Big[- \Phi_2^i\Big(1 - \frac{\widehat{a}_2^i(a_1^i)}{a_1^i}\Big) - \frac{\widehat{a}_2^i(a_1^i)}{a_1^i}\Phi_2^{i\prime}(1 - \frac{\widehat{a}_2^i(a_1^i)}{a_1^i})\Big]$$

is also constant. The function Γ^i is decreasing and admits a zero. Thus Assumption 1 is satisfied.

Policy implications.

Investments in green technologies fit our framework. Therefore, the second-period pollution level a_2^i is increasing in the first-period investment in brown equipment a_1^i. More importantly, a delayed negotiation provides countries with an incentive to under-invest in the first period (X^i is smaller than its first-best value).

Under delayed negotiation regions under-invest in the first-period compared to the autarky situation.

Although no general result can be drawn concerning second-period emissions, it can be shown that for some quadratic investment cost functions, $a_2^{iDN} > {a_2^i}^*$.[20]

[20]Take $\Psi_1^i(x) = \beta\frac{x^2}{2}$ and $\Phi_2^i(x) = \frac{x^2}{2}$. We assume $c < 1$. Then $\widehat{k}^i = 1 - c$ and ${k^i}^* = 1 - c^i$. And so ${a_1^i}^* = \text{Max}\,\{0, 1 - \frac{c^i}{\beta} + \frac{(c^i)^2}{2\beta}\}$ and $a_1^{iDN} = {a_1^i}^* + \alpha^i\frac{(c^j)^2}{2\beta}$. So, if $1 - \frac{c^i}{\beta} + \frac{(c^i)^2}{2\beta} > 0$

$$a_2^{iDN} = {a_2^i}^* + (1-c)\alpha^i\frac{(c^j)^2}{2\beta} - c^j(1 - \frac{c^i}{\beta} + \frac{(c^i)^2}{2\beta}).$$

We get $a_2^{iDN} = {a_2^i}^* - c^j{a_1^i}^* + (1-c)\alpha^i\frac{(c^j)^2}{2\beta}$. For $\beta \to +\infty$, $a_2^{iDN} = {a_2^i}^* - c^j$. Also, ${a_1^i}^*$ is increasing in β, and approaches 1 for $\beta \to +\infty$. $\alpha^i\frac{(c^j)^2}{2\beta}$ is decreasing with β. So there exists $\bar{\beta}$ such that $c^j{a_1^i}^* = \alpha^i\frac{(c^j)^2}{2\bar{\beta}}$. Therefore for $\beta < \bar{\beta}$: $a_2^{iDN} > {a_2^i}^*$.

Proposition 4. *A delayed negotiation decreases a region's first-period investment in green technologies compared to the autarky situation:* $a_1^{iDN} > a_1^{i\,*}$. *Furthermore, for some cost functions,* $a_2^{iDN} > a_2^{i\,*}$.

4 Market consolidation under forward or bankable allowances

4.1 Forward allowances

4.1.1 Overview

As the United States, the emerging countries and actually most of the world resist binding agreements, many experts place their hopes in the emergence of regional pollution permit markets such as the already existing European Union Emission Trading System (ETS) for CO_2 emission permits.

Proponents of regional ETS initiatives take the view that regional markets will jump-start climate change mitigation and will pave the way for an international binding agreement over region's emissions. However a deal will not happen overnight. We investigate the strategic implications of a delayed agreement. As we previously did, we ensure strategic independence under autarky by assuming that the damages depend linearly on total emissions. We first focus on strategic choices regarding forward allowances; we later show that similar insights apply to the issuance of bankable permits. In either case, the region puts today into private hands allowances that can be used to cover future emissions. We also assume, as we did before, that the date-2 agreement defines both regions' emissions. The resulting outcome is equivalent to a merger of the two ETS systems only in a symmetric configuration.

The brinkmanship effect applies provided that country i's date-2 pollution under autarky increases with the number of forward allowances distributed or sold at date 1. The effort rebalancing effect further requires that domestic pollution be reduced by an international agreement. We posit that the government faces a shadow cost of public funds and auctions off date-2 allowances,[21] and show that these properties are indeed satisfied.

[21]Existing or planned ETS schemes, while distributing most allowances for free, plan to move to a full auctioning approach (of course there is some uncertainty as to the credibility of such commitments).

But the properties may hold even if permits are distributed for free at date 2. For example, the government might at date 2 negotiate domestically with a powerful industrial lobby, whose status-quo welfare is stronger, the larger the number of allowances it received at date 1.

Let us first consider a situation in which the two regions sell forward allowances at date 1 in a non-coordinated way. In the absence of future negotiation, we show that it is suboptimal for a region to sell forward. This *regional optimization effect* is a reinterpretation of the Coase conjecture. If a region sells some allowances at date 1, at date 2 it will not fully internalize the decrease in value of these forward allowances when issuing spot allowances. Anticipating this incentive, the buyers of forward allowances buy the allowances below the price that would prevail if only a spot market existed. A region that chooses to sell allowances forward ends up selling more allowances than it would in a date-2 spot market.

By contrast, it becomes profitable for a region to sell some allowances forward when regions negotiate over emissions at date 2. As earlier, the outside options in the date-2 negotiation consist in signing no agreement and choosing noncooperatively the number of spot allowances.

To understand the two effects at work it is useful to consider the polar case in which one region has all the bargaining power. Because fewer allowances are issued than in the absence of an agreement, the date-2 price of allowances (P^i) is higher than in the noncooperative case (p^i). The buyers of forward allowances, rationally anticipating the negotiation outcome, are willing to purchase the forward allowances at a price P^i greater than p^i. For a region with no bargaining power, issuing some allowances forward increases its immediate profit by P^i per unit, while only decreasing its outside option by p^i. So the region has an incentive to free-ride on the negotiated outcome by selling at date 1 forward allowances at the post-agreement price. The *effort rebalancing effect* thus creates an incentive for the region without bargaining power to sell forward allowances.

The region with full bargaining power also has an incentive to sell forward allowances albeit for a different reason: It benefits from lowering the other region's outside option. The lower the other region's outside option, the larger the surplus that it will be able to extract in the negotiation. By selling allowances forward, the region credibly commits to increase *ex post* the total number of allowances it will sell in the absence of agreement. This *brinkmanship effect* gives an incentive for the region with bargaining power to sell allowances forward.

4.1.2 Description of the forward allowance game

In the case of forward trading, the first-period strategic action a_1^i is the number of forward allowances region i sells at date 1. Similarly, the second-period strategic action a_2^i, that is the level of emissions produced by region i in period 2, is the total number of allowances issued in region i (a_2^i is equal to a_1^i *plus* the number of spot allowances issued at date 2).

Each region hosts captive firms[22] on its soil. Section 4.3 extends the analysis to leakages. In each region, the firms take the price of allowances in their region as given when choosing their output. We call $p^i(a_2^i)$ region i's inverse demand function for allowances. p^i is non-increasing in a_2^i.

The region's second-period utility, U_2^i, is made of two terms. First, the regulator values the pollution of firms producing on its soil to the extent that this economizes on abatement costs, generating more profits, employment or taxes. We stay as general as possible and assume that the value associated with domestic firms emitting a_2^i units of pollution is $V_2^i(a_2^i)$. The function V_2^i is increasing in a_2^i and concave. In the absence of taxes or other externalities on the rest of society, one can assume that $V_2^i(a_2^i)$ is equal to the domestic firms' profit $\pi^i(a_2^i)$, so that $V_2^{i\prime}(a_2^i) = \pi_2^{i\prime}(a_2^i) = p^i(a_2^i)$ but we do not impose such a restriction for the moment.

Second, the regulator internalizes part of the cash generated by the sale of allowances; indeed we assume that the regions face a shadow cost of public funds λ, that is that raising \$1 of public money costs society $\$(1+\lambda)$. And so:

$$U_2^i(a_2^i, a_1^i) \equiv \lambda[a_2^i - a_1^i]p^i(a_2^i) + V_2^i(a_2^i) \tag{11}$$

The first-period utility associated with the sale of forward allowances is λ times the revenues from the sale.[23] More precisely, the forward allowances a_1^i are sold at price $p^i(\widehat{a}_2^i(a_1^i))$ as agents rationally anticipate that the date-2 agreement will lead to a price $p^i(\widehat{a}_2^i)$. Let

[22] That is, we assume that outsourcing costs are important, so that firms are fixed and therefore obliged to buy permits from the region where they were initially located. This is typically the case for utilities: since electricity cannot be transported over long distances, electricity producers must have power plants located relatively close to their end consumers.

[23] Whether forward allowances are sold or distributed for free at date 1 is irrelevant for the argument.

$$U_1^i(a_1^i) \equiv \lambda a_1^i p^i(\widehat{a}_2^i(a_1^i)) \tag{12}$$

Checking Assumption 1

These two functions satisfy the assumptions of the generic model.[24] In particular

$$\Gamma^i(a_1^i) = \frac{dU_1^i}{da_1^i}(a_1^i) + \frac{\partial U_2^i}{\partial a_1^i}\Big(a_1^i, \widehat{a}_2^i(a_1^i)\Big) = \lambda p^{i\prime}(\widehat{a}_2^i(a_1^i))\frac{d\widehat{a}_2^i}{da_1^i}a_1^i$$

has the same sign as $(-a_1^i)$. Assumption 1 *(ii)* is satisfied with $a_1^{iFB} = 0$, and Proposition 1 applies.

The first-order condition takes the specific form:

$$\begin{aligned}\frac{\partial W_{neg}^i}{\partial a_1^i}(a_1^i) = &\Big[\lambda p^{i\prime}(\widehat{a}_2^i(a_1^i))\,\frac{d\widehat{a}_2^i}{da_1^i}a_1^i\Big] \\ &+ (1-\alpha^i)\lambda\Big[p^i(\widehat{a}_2^i(a_1^i)) - p^i(a_2^{i*}(a_1^i))\Big] \\ &+ \alpha^i c^j \frac{da_2^{i*}}{da_1^i}(a_1^i).\end{aligned}$$

Policy implications

Applied to forward trading, Lemma 1 implies that the level of date-2 emissions in region i (under autarky or negotiation) is an increasing function of the number of its forward allowances. Thus, issuing forward allowances is a credible commitment to emit more in the future. Proposition 1 in turn implies that forward allowances are issued while none would be in the first best, and that delayed negotiation results in more pollution at date 2.

Comparison with autarky

---

[24] First, we have

$$\frac{\partial^2 U_2^i}{\partial a_1^i \partial a_2^i}(a_2^i, a_1^i) = -\lambda p^{i\prime}(a_2^i) > 0.$$

Second,

$$\frac{dU_1^i}{da_1^i}(a_1^i) + \frac{\partial U_2^i}{\partial a_1^i}(a_1^i, \widehat{a}_2^i(a_1^i)) = \lambda p^{i\prime}(\widehat{a}_2^i(a_1^i))\frac{d\widehat{a}_2^i(a_1^i)}{da_1^i}a_1^i.$$

Under autarky, rational agents anticipate second-period price of allowances $p^i({a_2^i}^*(a_1^i))$. Therefore ${U_1^i}^*(a_1^i) = \lambda a_1^i p^i({a_2^i}^*(a_1^i))$[25]. The intertemporal welfare function of region i in the autarky situation is

$$\begin{aligned} W^i_{aut}(a_1^i, a_1^j) &= {U_1^i}^*(a_1^i) + U_2^i({a_2^i}^*(a_1^i), a_1^i) - c^i[{a_2^i}^*(a_1^i) + {a_2^j}^*(a_1^j)] \\ &= \lambda p^i({a_2^i}^*(a_1^i)){a_2^i}^*(a_1^i) + V_2^i({a_2^i}^*(a_1^i)) - c^i[{a_2^i}^*(a_1^i) + {a_2^j}^*(a_1^j)]. \end{aligned} \tag{13}$$

So W^i_{aut} can be written as a function of ${a_2^i}^*$. So since ${a_2^i}^*(a_1^i)$ is non-decreasing, then ${a_1^i}^* = 0$. The effect of introducing a date-2 negotiation is to induce regions to sell allowances forward, while in autarky they would have refrained from selling forward. The following proposition is proved in the Appendix.

Proposition 5. *For linear demand functions and when V_2^i is equal to profit, the strategic incentive to be in a stronger bargaining position and the reduction of pollution achieved through date-2 negotiation cancel out: $a_2^{DN} = a_2^*$.*

4.1.3 Merger of emission trading systems

Without loss of generality, we have assumed that at date 2 the two regions agree on a vector of emissions, one for each region. When regions are completely symmetric ($c^A = c^B = c/2$, $\alpha^A = \alpha^B = 1/2$, $V^A(x) = V^B(x)$ and $p^A(x) = p^B(x)$ for all x) bargaining over regional allowances amounts (on the equilibrium path) to merging the regional markets and setting a total number of allowances.

The optimality of merging markets requires that the optimal agreement leads to equal prices of CO_2 in both regions. This condition however is not in general satisfied. It does not hold in the symmetric case off the equilibrium path (that is, when one of the regions deviates from the equilibrium number of forward permits) or in the asymmetric case. Intuitively, the date-2 carbon price determines the rent enjoyed by those private actors who acquired forward permits at date 1. Because such rents are costly to the regions, the latter at date 2 cooperatively set regional allowances and prices with an eye on limiting them. This incentive for price discrimination implies

[25] ${U_1^i}^*(a_1^i)$ differs from the first-period utility under delayed negotiation: Under autarky forward allowances sell at price $p^i({a_2^i}^*(a_1^i))$, while they sell at price $p^i(\hat{a}_2^i(a_1^i)) > p^i({a_2^i}^*(a_1^i))$ under delayed negotiation.

that setting equal prices for the two regions – a corollary of market merger – is not optimal in general.

With a few adaptations, though, our model can deal with the case where the negotiation is constrained to focus on merging the two systems and on setting the total allowances world-wide. We here content ourselves with a sketch of the main insights.

Under carbon-price equalization, the regional optimization and the brinkmanship effects remain. The effort rebalancing effect however can go in the opposite direction. This effect is driven by the price difference between the autarky outcome (the threat point) and the negotiated outcome. If regions have sufficiently different industry structures (and thus abatement costs), the autarky price in one region may be larger than the negotiated price. In that case, the effort rebalancing effect becomes negative for this region. By contrast, if both autarky prices are below the negotiated price, the analysis of the general model remains valid. The following proposition is proved in the Appendix.

Proposition 6. *For linear inverse demand functions, and when the negotiation bears on merging the two allowance markets:*

(i) When regions have the same technology (i.e. $p^i(x) = p^j(x)$ for all x), then both regions sell forward.

(ii) When regions have different technologies, then provided that technologies are differentiated enough (alternatively that environmental costs are low enough), only the region with the most performing technology (region i such that $p^i(x) < p^j(x)$ for all x) sells forward.

4.2 Banking of pollution allowances

Another policy through which the authorities can put future permits into private hands is the banking clause in Emission Trading Systems. For example, the Waxman-Markey bill which passed the House of Representatives in June 2009 had built in such a scheme. The bill adopted a cap-and-trade approach, set the quantity of permits to be issued over the next 20 years, and allowed permit holders to bank their permits for future use.

Strategic actions and utility functions

Let n_1^i denote region i's number of allowances issued at date 1. These allowances can be used in period 1 or banked for use at time 2. The amount of allowances banked by the private sector is a_1^i. So $n_1^i \equiv a_1^i + q_1^i$ where q_1^i is the number of allowances used in period 1. As earlier, a_2^i denote emissions in region i in period 2.

Firms' net benefit functions in periods 1 and 2, $\pi^i(q_1^i)$ and $\pi^i(a_2^i)$, respectively, are increasing and convex. They anticipate a post-negotiation second-period allowances price $p^i(\widehat{a}_2^i(a_1^i))$. So, assuming that there is banking in equilibrium, they will bank until their marginal benefit to produce is equal to that price:

$$\pi^{i\prime}(n_1^i - a_1^i) = p^i(\widehat{a}_2^i(a_1^i)) = \pi^{i\prime}(\widehat{a}_2^i(a_1^i))$$

and so

$$n_1^i - a_1^i = \widehat{a}_2^i(a_1^i) \tag{14}$$

Because the benefit is convex and the inverse demand function of region i is decreasing, condition (14) defines an increasing function $a_1^i(n_1^i)$. Regions therefore maximize indifferently over a_1^i or over n_1^i. We choose a_1^i as the first-period strategic action.

The pollution at date 1 has a direct cost $c_1^i q_1^i$ to society.[26] The regulator values a share λ of the proceedings from the sale of the allowances. We therefore have

$$U_1^i(a_1^i, a_1^j) = \lambda p^i(\widehat{a}_2^i(a_1^i))(\widehat{a}_2^i(a_1^i) + a_1^i) - c_1^i(\widehat{a}_2^i(a_1^i) + \widehat{a}_2^j(a_1^j)) + \pi^i(\widehat{a}_2^i(a_1^i)).$$

Region i's second-period utility is:

$$U_2^i(a_1^i, a_2^i) = \lambda p^i(\widehat{a}_2^i(a_1^i))(\widehat{a}_2^i(a_1^i) - a_1^i) + \pi^i(\widehat{a}_2^i(a_1^i)).$$

Checking Assumption 1.

We verify that $\dfrac{\partial^2 U_2^i}{\partial a_1^i \partial a_2^i} = -\lambda p^{i\prime}(\widehat{a}_2^i(a_1^i)) > 0$. So assumption 1 (i) is satisfied. We will assume that Assumption 1 (ii) is also satisfied. We have checked that this is indeed the case for linear demand functions. Although U_1^i depends on both a_1^i and a_1^j, the general result applies since $U_1^i(a_1^i, a_1^j)$ is separable in a_1^i and a_1^j.

Proposition 7. *A delayed negotiation increases the quantity of banked allowances and yields an overprovision of allowances in the first period.*

[26] If all effects of pollution are delayed and there is no regeneration, then $c_1^i = c^i$ (the second-period cost). In general c_1^i can be greater or smaller than c^i.

4.3 Leakage: A new Raising Rival's Cost effect

Patchworks of regional markets for pollution rights promote fiscal competition between regions and keep the price of allowances low. Indeed, if the regulators (at least) partly internalize the profits made by firms that operate on their soil, and if (at least a fraction of) the firms in a given country can outsource their production, then regulators have an incentive to issue a generous amount of allowances so as to attract more firms.[27]

In the model of Section 4.1, firms were captive and could not outsource their production. To account for industrial leakage, we now assume that there is a worldwide continuum of firms of mass 1, that firms all have the same technology (with non-decreasing profit function $\pi_2^i(q) = \pi_2^j(q)$ for any amount of allowances q). We further assume that an exogenous fraction θ of *mobile firms* can costlessly choose where to produce. The remaining $1-\theta$ firms cannot outsource their production to the other region. We call these firms *captive firms*. For simplicity we focus on the symmetric situation where region A and region B each have $\frac{1-\theta}{2}$ captive firms on their soil.

We will assume for conciseness (and also because it provides the starkest illustration of the impact of leakage) that θ is "large enough" (larger than some $\underline{\theta}$), so that allowance prices are equalized across markets.

The profit of captive and mobile firms producing at home is fully internalized by the local regulator. This profit depends only on the worldwide price $p(a_2^A + a_2^B) \equiv p^i(\frac{a_2^A+a_2^B}{2})$. For notational simplicity let $\pi_2(a_2^A + a_2^B) \equiv \pi_2^i(\frac{a_2^A+a_2^B}{2}) = \pi_2^j(\frac{a_2^A+a_2^B}{2})$.

Calling x^i the share of mobile firms that choose region i, then the second-period autarky welfare function of region i, W_2^i, is

$$W_2^i = \lambda p(a_2^i + a_2^j)[a_2^i - a_1^i] + \frac{1-\theta}{2}\pi_2(a_2^i + a_2^j) + x^i\theta\pi_2(a_2^i + a_2^j) - c^i[a_2^i + a_2^j].$$

But $\dfrac{1-\theta}{2} + x^i\theta = \dfrac{a_2^i}{a_2^i + a_2^j}$ and so we can rewrite the previous equation as:

$$W_2^i = \lambda p(a_2^i + a_2^j)[a_2^i - a_1^i] + \frac{a_2^i}{a_2^i + a_2^j}\pi_2(a_2^i + a_2^j) - c^i[a_2^i + a_2^j]. \tag{15}$$

[27] There is an extensive literature on strategic trade with environmental leakages (see e.g., Hoel (2001), Babiker (2005) and Felder and Rutherford (1993)). Harstad (2010) shows that if signatories can coordinate, and if markets for the ownership of polluting natural resources in non-signatory countries exist, then inefficiencies induced by leakage concerns disappear. In contrast with this literature, we focus on the interaction between leakages and negotiation.

If negotiations break down at date 2, regions choose a_2^A and a_2^B noncooperatively in a Cournot duopoly like manner. Let a_2^{A*} and a_2^{B*}, and $a_2^* \equiv a_2^{A*} + a_2^{B*}$ denote the equilibrium of this game. These are functions of $\left(a_1^A, a_1^B\right)$.

In the first period, the autarky objective function of region i, W^i_{aut}, is

$$W^i_{aut}(a_1^i, a_1^j) = \lambda p(a_2^*(a_1^A, a_1^B))a_1^i + W_2^{i*}(a_1^i, a_1^j).$$

Using the envelope theorem, the first-order condition for a_1^i for the autarky case is:

$$\begin{aligned}\frac{\partial W^i_{aut}}{\partial a_1^i}(a_1^i, a_1^j) = {} & \lambda p'\left(a_2^*(a_1^A, a_1^B)\right)\frac{\partial a_2^*}{\partial a_1^i}(a_1^i, a_1^j)a_1^i \\ & - \left(\frac{\pi_2(a_2^*(a_1^A, a_1^B))}{a_2^*(a_1^A, a_1^B)} + \lambda p(a_2^*(a_1^A, a_1^B))\right)\frac{\partial a_2^{j*}}{\partial a_1^i}(a_1^i, a_1^j) \\ = {} & 0\end{aligned} \tag{16}$$

Mobility has no effect on the outcome of the negotiation $\left(\widehat{a}_2^i(a_1^i, a_1^j), \widehat{a}_2^i(a_1^i, a_1^j)\right)$, but it affects the threat points in the negotiation. Formally, the objective function of region i in the negotiation case is:

$$\begin{aligned}W^i_{neg}(a_1^i, a_1^j) = {} & \lambda p(\widehat{a}_2(a_1^i, a_1^j))a_1^i + W_2^{i*}(a_1^i, a_1^j) \\ & + \alpha^i\left(\widehat{W}_2(a_1^i, a_1^j) - \left(W_2^{i*}(a_1^i, a_1^j) + W_2^{j*}(a_1^i, a_1^j)\right)\right)\end{aligned} \tag{17}$$

The first-order condition on W^i_{neg} is given by:

$$\begin{aligned}\frac{\partial W^i_{neg}}{\partial a_1^i} = {} & \lambda p'(\widehat{a}_2(a_1^A, a_1^B))\frac{\partial \widehat{a}_2(a_1^A, a_1^B)}{\partial a_1^i}a_1^i + (1-\alpha^i)\lambda(p(\widehat{a}_2(a_1^A, a_1^B)) - p(a_2^*(a_1^i, a_1^j)))) \\ & + \left(\frac{\pi_2(a_2^*(a_1^i, a_1^j))}{a_2^*(a_1^i, a_1^j)} + \lambda p(a_2^*(a_1^i, a_1^j))\right)\left(\alpha^i\frac{\partial a_2^{i*}}{\partial a_1^i}(a_1^i, a_1^j)\right) \\ & + \left(\frac{\pi_2(a_2^*(a_1^A, a_1^B))}{a_2^*(a_1^A, a_1^B)} + \lambda p(a_2^*(a_1^i, a_1^j))\right)\left(-(1-\alpha^i)\frac{\partial a_2^{j*}}{\partial a_1^i}(a_1^i, a_1^j)\right) = 0\end{aligned} \tag{18}$$

The first two terms of equation (18) correspond to the regional optimization and effort rebalancing effects. The additional terms $\left(\frac{\pi_2(a_2^*(a_1^i,a_1^j))}{a_2^*(a_1^i,a_1^j)} + \lambda p(a_2^*(a_1^i,a_1^j))\right)$ $\left(\alpha^i \frac{\partial a_2^{i*}}{\partial a_1^i}(a_1^i,a_1^j)\right)$ and $\left(\frac{\pi_2(a_2^*(a_1^i,a_1^j))}{a_2^*(a_1^i,a_1^j)} + \lambda p(a_2^*(a_1^i,a_1^j))\right)$ $\left(-(1-\alpha^i)\frac{\partial a_2^{j*}}{\partial a_1^i}(a_1^i,a_1^j)\right)$ can be thought of as the equivalents of the brinkmanship and the raising rival's cost effects, respectively. Indeed, since the mobility of firms creates *de facto* a single market for allowances, the brinkmanship effect consists in stealing some of the firms and their associated revenues from the other region. The marginal revenue of attracting firms from one region to the other is equal to $\left(\frac{\pi_2(a_2^{A*}+a_2^{B*})}{a_2^{A*}+a_2^{B*}} + \lambda p\right)$. Furthermore, in response to an increase in the total number of allowances by region A, region B decreases its total allowances. This raising rival's cost effect represents the net gain by the region to have the other reduce its total number of allowances.

Interestingly, the brinkmanship and raising rival's cost effects do not directly depend on the region's environmental cost, but on the total number of allowances in both markets ($a_2^{A*} + a_2^{A*}$) which in turn depend on c^A and c^B.

Note also that the raising rival's cost effect is not a negotiation-related effect; on the contrary, the prospect of a negotiation dampens its magnitude by a factor $(1-\alpha^i) \leq 1$. A country with substantial bargaining power has little interest in raising its rival's cost (or in this case in stealing its rival's business), since it already earns most of the surplus from the negotiation.

Proposition 8. *Leakage reinforces the conclusions of Proposition 1 applied to the allowance futures game: Not only do regions sell allowances forward when negotiation is delayed, but leakage creates an incentive to sell forward even under autarky.*

5 Alleys for future research

The introduction has already summarized the main insights of the analysis. These concluding notes rather focus on how it could be enriched. Besides the obvious extension to a longer horizon, several alleys for research seem particularly promising.

First, while our model already covers a wide range of instruments, its generality could be further enhanced. Consider for instance the clean development mechanism (CDM) set up in the aftermath of Kyoto. This mechanism allows countries that have committed to emission abatement targets to implement these in part through projects in countries that have ratified the Kyoto protocol but are not subject to such targets.

The developed countries' willingness to go along with the CDM impacts not only their own effort prior to the negotiation of a binding agreement (through the earned credits), but also the effort of the emerging countries' region. The basic model could be enriched to account for such interdependencies.

A second promising alley is to consider asymmetric information between regions regarding, say, their political resolve to combat climate change. As we noted, such asymmetries may be one of the causes of delay, as regions are engaged in a form of "war of attrition". Furthermore, this extension could generate interesting insights regarding signaling strategies. For example, before negotiating in Copenhagen, Europe made a commitment to a 20% emission reduction relative to 1990 (30% in case of a "satisfactory agreement") and several countries added a carbon tax for those economic agents who are not covered by the ETS system. Some observers argued that European politicians were thereby putting themselves in a weak bargaining position because they made concessions before negotiating and over-signalled their eagerness to reach an agreement; others disagreed and viewed this move as a way to signal good intentions and to jump-start real negotiations.

We have considered only a global agreement between the regions. While we feel that negotiating a global agreement is the most reasonable way to go, many advocate a sectoral agreement approach. The study of multiple negotiations could be fascinating, as interest group politics would then play a major role.

Finally, a common political argument in favor of tough pollution control at home is that strict anti-pollution policies give the country's industry a technological edge by stimulating R&D in green technologies and help create tomorrow's "green growth". This strategic trade effect might alleviate the impact of the strategic effects of delayed negotiation.

We leave these open questions and other exciting issues on the climate negotiation research agenda to future work.

References

Acemoglu, D., Aghion, P., Bursztyn, L., and D. Hemous (2009) "The Environment and Directed Technical Change," NBER Working Paper 15451 (October).

Aldy, J., and R. Stavins (ed.) (2007) *Architectures for Agreement: Addressing Global Climate Change in the Post-Kyoto World*. Cambridge University Press.

Allaz, B., and J.L. Vila (1993) "Cournot Competition, Forward Markets and Efficiency," *Journal of Economic Theory*, 59(1): 1–16.

Babiker, M. (2005) "Climate Change Policy, Market Structure, and Carbon Leakage," *Journal of International Economics*, 65(2): 421–445.

Barrett, S. (2005) "The Theory of International Environmental Agreements," in K.-G. Mäler and J.R. Vincent, eds., *Handbook of Environmental Economics*, 3, Elsevier.

Buchholz, W., and K. Konrad (1994) "Global Environmental Problems and the Strategic Choice of Technology," *Journal of Economics*, 60(3): 299–321.

Carraro, C., and D. Siniscalco (1993) "Strategies for the International Protection of the Environment," *Journal of Public Economics*, 52: 309–328.

Felder, S. and Rutherford, T. (1993) "Unilateral CO2 Reductions and Carbon Leakage: The Consequences of International Trade in Oil and Basic Materials," *Journal of Environmental Economics and Management*, 25(2): 162–176.

Grimaud, A., and L. Rouge (2008) "Environment, Directed Technical Change and Economic Policy," *Environmental Resource Economics*, 41: 439–463.

Grossman, S., and O. Hart (1986) "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 94(4): 691–719.

Harstad, B. (2009) "The Dynamics of Climate Change Agreements," Working Paper, Northwestern University, Center for Mathematical Studies in Economics and Management Science.

——– (2010) "Buy Coal! Deposit Markets Prevent Carbon Leakage," mimeo, Northwestern University.

Hart, O., and J. Moore (1990) "Property Rights and the Nature of the Firm," *Journal of Political Economy*, 98(6): 1119–1158.

Hoel, M. (2001) "International Trade and the Environment: How to Handle Carbon Leakage," *Frontiers of Environmental Economics*, 2001: 176–191.

Laffont, J.J., and J. Tirole (1996a) "Pollution Permits and Compliance Strategies," *Journal of Public Economics*, 62(1-2): 85–125.

—— (1996b) "Pollution Permits and Environmental Innovation," *Journal of Public Economics*, 62(1-2): 127–140.

Mahenc, P., and F. Salanié (2004) "Softening Competition Through Forward Trading," *Journal of Economic Theory*, 116(2): 282–293.

Milgrom P., and J. Roberts (1990) "Rationalizability, Learning and Equilibrium in Games with Strategic Complementarities," *Econometrica*, 58: 1255–1278.

Salop, S., and D. Sheffman (1983) "Raising Rivals' Costs," *American Economic Review, Papers & Proceedings*, 73: 267–271.

Willamson, O. (1985) *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*. New York: Free Press.

Appendix

Proof of Proposition 3

For linear environmental cost, country i's date-2 emission level (with or without negotiation) depends only on its date-1 choice. The brinkmanship effect term writes $\alpha^i \sum_{j\neq i} c^j \frac{da_2^{i\,*}}{da_1^i}(a_1^i)$. Under the assumption of Proposition 3, the brinkmanship effect, which is proportional to $\alpha^i(c - c^i)$, does not tend to zero as n tends to infinity, which proves (i).

The limit of equation (10) as α^i tends to zero is

$$\begin{aligned}\frac{\partial W^i_{neg}}{\partial a_1^i}(a_1^i) = &\left(\frac{dU_1^i(a_1^i)}{da_1^i} + \frac{\partial U_2^i}{\partial a_1^i}(a_1^i, a_2^{i\,*}(a_1^i))\right)\\ &+ \alpha^i\left(\sum_{j\neq i} c^j\right)\frac{da_2^{i\,*}}{da_1^i}(a_1^i) = 0\end{aligned} \tag{19}$$

Evaluating it at $a_1^{i\,*}$ we have $\frac{\partial W^i_{neg}}{\partial a_1^i}(a_1^{i\,*}) = \alpha^i \sum_{j\neq i} c^j \dfrac{da_2^{i\,*}}{da_1^i}(a_1^{i\,*}) > 0$. This proves (ii).

□

Proof of Proposition 4

The first-order conditions with respect to $\widehat{a}_2^i$ and $a_2^{i\,*}$ are

$$\Phi_2^{i\prime}\left(1 - \frac{\widehat{a}_2^i(a_1^i)}{a_1^i}\right) = c$$

and

$$\Phi_2^{i\prime}\left(1 - \frac{a_2^{i\,*}(a_1^i)}{a_1^i}\right) = c^i.$$

Define k^{i*} as the unique solution to $\Phi_2^{i\prime}(1 - k^{i*}) = c^i$. Similarly, define $\widehat{k}^i$ as the unique solution to $\Phi_2^{i\prime}(1 - \widehat{k}^i) = c$. Then

$$a_2^{i\,*}(a_1^i) = k^{i*} a_1^i$$

and

$$\widehat{a}_2^i(a_1^i) = \widehat{k}^i a_1^i.$$

Call

$$\Delta^i(k^i) \equiv \Phi_2^i(1-k^i) + k^i \Phi_2^{i\prime}(1-k^i) = -\frac{\partial U_2^i}{\partial a_1^i}(a_1^i, k a_1^i).$$

The first-order condition with respect to ${a_1^i}^*$ is:

$$\Psi_1^{i\prime}(1-{a_1^i}^*) - \Delta^i({k^i}^*) = 0.$$

Similarly, the first-order condition with respect to a_1^{iDN} is:

$$\Psi_1^{i\prime}(1-\widehat{a}_1^i) - \Delta^i(\widehat{k}^i) + (1-\alpha^i)\Big(\Delta^i(\widehat{k}^i) - \Delta^i({k^i}^*)\Big) + \alpha^i c^j {k^i}^* = 0$$

Substracting the latter equation from the former and rearranging, we get:

$$\Psi_1^{i\prime}(1-{a_1^i}^*) - \Psi_1^{i\prime}(1-\widehat{a}_1^i) = \alpha^i\Big[\Phi_2^i(1-{k^i}^*) - \Phi_2^i(1-\widehat{k}^i) + ({k^i}^* - \widehat{k}^i)\Phi_2^{i\prime}(1-\widehat{k}^i)\Big].$$

Since Φ_2^i is convex and ${k^i}^* \geq \widehat{k}^i$, the term in bracket is positive. Therefore $\Psi_1^{i\prime}(1-{a_1^i}^*) \geq \Psi_1^{i\prime}(1-\widehat{a}_1^i)$. The convexity of Ψ_1^i then implies that $a_1^{iDN} \geq {a_1^i}^*$.

□

Proof of Proposition 5

Define $p^i(a_2^i) = \beta^i - \gamma^i a_2^i$. Then

$${a_2^i}^*(a_1^i) = \frac{(1+\lambda)\beta^i - c^i + \lambda\gamma^i a_1^i}{(1+2\lambda)\gamma^i}.$$

Since ${a_1^i}^* = 0$, the autarky level of emission is $\frac{(1+\lambda)\beta^i - c^i}{(1+2\lambda)\gamma^i}$.

Similarly

$$\widehat{a}_2^i(a_1^i) = \frac{(1+\lambda)\beta^i - c + \lambda\gamma^i a_1^i}{(1+2\lambda)\gamma^i}.$$

□

Using the first-order condition with respect to a_1^i, $a_1^{iDN} = \frac{c^j}{\lambda\gamma^i}$, and so $\widehat{a}_2^i(a_1^{iDN}) = {a_2^i}^*({a_1^i}^*)$.

Proof of Proposition 6

Take again $p^i(a_2^i) = \beta^i - \gamma^i a_2^i$. We define the merged-market price $P(a_2)$ by the implicit equation: $p^{A^{-1}}(P(a_2)) + p^{B^{-1}}(P(a_2)) = a_2$.

So $P(a_2) = \dfrac{\beta^i\gamma^j + \beta^j\gamma^i - \gamma^i\gamma^j a_2}{\gamma^i + \gamma^j}$. Define $\beta \equiv \frac{\beta^i\gamma^j+\beta^j\gamma^i}{\gamma^i+\gamma^j}$ and $\gamma \equiv \frac{\gamma^i\gamma^j}{\gamma^i+\gamma^j}$. Then $P(a_2) = \beta - \gamma a_2$.

The first-order condition with respect to a_1^i writes

$$\frac{\partial W^i_{neg}}{\partial a_1^i}(a_1^i, a_1^j) = \lambda P'(\widehat{a}_2(a_1))\frac{d\widehat{a}_2}{da_1^i}a_1^i + (1-\alpha^i)\lambda\Big(P(\widehat{a}_2(a_1)) - p^i(a_2^{i*}(a_1^i))\Big) + \alpha^i c^j \frac{da_2^{i*}}{da_1^i}(a_1^i) = 0$$

Hence:

$$\frac{\partial W^i_{neg}}{\partial a_1^i}(a_1^i, a_1^j) = \frac{\lambda}{1+2\lambda}\Big[(1-\alpha^i)\lambda(\beta-\beta^i) + c^j - (1-\alpha^i)\lambda\gamma a_1^j - \lambda\Big(\gamma - (1-\alpha^i)(\gamma^i-\gamma)\Big)a_1^i\Big],$$

and:

$$\frac{\partial W^i_{neg}}{\partial a_1^i}(0,0) = \frac{\lambda}{1+2\lambda}\Big(\lambda(1-\alpha^i)\frac{\gamma^i}{\gamma^i+\gamma^j}(\beta^j-\beta^i) + c^j\Big) \equiv \Delta^i.$$

First note that for $c^i = c^j = 0$ then $\Delta^i > 0$ if and only if $\beta^j > \beta^i$. For simplicity we will assume that $\gamma^i = \gamma^j = 2\gamma$. We verify that

$$\frac{\partial^2 W^i_{neg}}{\partial a_1^i \partial a_1^j} \leq 0$$

Furthermore, since $\gamma^i = \gamma^j = 2\gamma$ then $\gamma - (1-\alpha^i)(\gamma^i - \gamma) \geq 0$ and so W^i_{neg} is concave in a_1^i for any a_1^j.

We conclude that for $(1-\alpha^i)\lambda\frac{\gamma^i}{\gamma^i+\gamma^j}(\beta^j-\beta^i) + c^j \geq 0$ and $(1-\alpha^j)\lambda\frac{\gamma^j}{\gamma^j+\gamma^i}(\beta^i-\beta^j) + c^i \leq 0$ then $a_1^i > 0$ and $a_1^j = 0$.

Indeed, if the conditions above hold: $\frac{\partial W^j_{neg}}{\partial a_1^j}(0, a_1^i) < 0$ for all $a_1^i \geq 0$ and so by quasi-concavity of W^j_{neg} in a_1^j, $a_1^j = 0$ for all $a_1^i \geq 0$. Then the quasi-concavity of W^i_{neg}, together with $\frac{\partial W^i_{neg}}{\partial a_1^i}(0,0) > 0$, implies that $a_1^i > 0$. $\square$