Prices vs. Quantities for International Environmental Agreements

Ulrike Kornek[a, b, c] Robert Marschinski [b, c]

20.10.2016

Abstract

Does the choice between price-based or quantity-based regulation matter for the formation of an international environmental agreement? We introduce uncertainty on baseline emissions in a standard coalition formation model and let countries choose their preferred regulatory instrument. It is shown that a coalition of cooperating countries is more likely to prefer a quantity regulation than non-cooperating countries. Compared to the case without uncertainty this implies stronger free-riding incentives and lower equilibrium participation in agreements that are based on quantities. A restriction to price-based agreements often leads to higher participation, but does not necessarily raise global welfare. Tradable quantities can both increase participation in the agreement and achieve higher levels of global welfare. Overall, our results suggest that free-riding incentives in global public good problems with uncertainty may be underestimated if the strategic implications of instrument choice are ignored.

Keywords: International Environmental Agreements; Uncertainty; Climate Change; Prices vs. Quantities; Global Public Goods

JEL Classifications: C72, D81, H41, Q54

[a]Corresponding author at: Mercator Research Institute on Global Commons and Climate Change, Torgauer Str. 12-15, 10829 Berlin, email: kornek@mcc-berlin.net

[b]Potsdam Institute for Climate Impact Research, PO Box 60 12 03, D-14412 Potsdam, Germany

[c]Technische Universität Berlin, Department 'Economics of Climate Change', Strasse des 17. Juni 145, 10623 Berlin, Germany

1 Introduction

It is one of the principle insights from the theory of self-enforcing international environmental agreements that free-riding incentives undermine international cooperation in the presence of public goods (Barrett, 1994; Karp and Simon, 2013). In the field of climate change this is seen as one reason for the so far inconclusive efforts of the international community to adopt a comprehensive global agreement. Consequently, the design characteristics of any such agreement ('treaty design') should as much as possible support the cooperation of countries who base their participation decision on self-interest.

A key design element of the Kyoto-Protocol is its specification of precise emission assignments for each eligible country, a choice likely to be retained by a possible successor agreement. The focus on "targets and timetables" (Victor, 2011) represents an intuitive approach and might be motivated by the desire to avoid extreme damages (WBGU, 2009). However, in light of the US' early and Canada's later withdrawal from the Kyoto Protocol, the question arises whether this choice of emission assignments might actually undermine the incentive of countries to join the agreement.

Other forms of treaty design have been forwarded and investigated (e.g. Barrett and Stavins, 2003). One proposal is a coordinated emission price (Cooper, 1998), which all participating countries would levy as a tax on the emissions generated on their territory. In an idealized economic setting this approach would be equivalent to Kyoto-style emission assignments, i.e. there would be a one-to-one correspondence between a given emission assignment and a tax level. However, there are many real-world characteristics under which the symmetry between the two approaches breaks down (Nordhaus, 2007). One particularly relevant example is the presence of uncertainty – a ubiquitous feature in most environmental policy areas, in particular climate change.

In terms of global efficiency – i.e. when considering the world as one region with a single regulator – many studies have investigated the ranking of price vs. quantity instruments under uncertainty (e.g. Weitzman, 1974; Stavins, 1996; Krysiak, 2008). However, the policy relevant multi-country setting and the strategic effect of each instrument on countries' incentive to join an international agreement have so far been mostly neglected, with the exception of Endres and Finus (1999), Eichner and Pethig (2015), and a limited number of qualitative discussions (Hepburn, 2006; McKibbin et al., 2012). This sharply contrasts with the fact that in global public good problems the global welfare resulting from an agreement typically depends on the number of participating countries.

For the formal analysis of international environmental agreements (IEAs) and the stability of coalitions, most studies adopt the approach first employed by Hoel (1992), Carraro and

Siniscalco (1993), and Barrett (1994). It assumes a payoff function for each country consisting of costs from private and benefits from global pollution abatement. Even though the assumption of payoff-functions with perfect information appears strong, only relatively few authors have investigated how uncertainty impacts the formation of self-enforcing agreements.

Na and Shin (1998), Ulph (2004), and Kolstad (2007) all study the implications of uncertain abatement benefits, albeit under different assumptions about the type of uncertainty and learning. Due to their idiosyncratic models, no clear-cut conclusions emerge from this group of contributions, with some finding uncertainty being conducive to cooperation, while others observe a negative impact. Karp (2012) shows that the effect of uncertainty on membership and global welfare of a self-enforcing agreement depends on the shape of the probability distribution of the uncertain benefit-parameter. Finus et al. (2013) find that the level of risk aversion plays a decisive role in determining how uncertainty influences coalition formation. Finus and Pintassilgo (2010) investigate uncertain benefits and costs of abatement, whereas Finus and Pintassilgo (2013) focus on uncertain benefits only. Since countries are modelled as ex-post heterogeneous and transfers are allowed, both contributions find that larger coalitions can be stabilized under certainty than with uncertainty.

Overall, the existing literature suggests that the presence of uncertainty could undermine cooperation (Kolstad and Ulph, 2008, 2011; Barrett, 2013). However, one option for mitigating the effects of uncertain abatement costs has been omitted: changing countries' decision variable from quantities to prices. This distinction between instruments under uncertainty goes back to the seminal work of Weitzman (1974), who studied the policy choice of a single decision-maker in a general cost-benefit setting. He demonstrated that expected welfare is higher under a price instrument whenever marginal costs are relatively steeper than marginal benefits. This finding has generally been reaffirmed in the subsequent literature, which incorporated correlation between cost and benefit uncertainty (Stavins, 1996), the possibility of tradable quantities (Williams, 2002), or the presence of a stock as opposed to a flow externality (Newell and Pizer, 2003).

What the above studies do not take into account, but arguably constitutes a crucial characteristic of global environmental problems (like climate change), is that the public good is supplied by many countries, each having its own sovereign regulator. The uncertainty about benefits and costs of abatement can be expected to influence countries' decision to participate or not in an environmental agreement. For instance, former US President George W. Bush apparently singled out this aspect as a main reason for the US' withdrawal from the Kyoto Protocol (Kolstad, 2007).

The problem of prices versus quantities may therefore go beyond its relevance for the welfare of a fixed group of cooperating countries and bear strategic implications for cooperation. When an IEA falls short of implementing the Pareto optimum of full participation, the level of global welfare depends on how much of the externality is internalized, which depends on the size of the coalition of cooperating countries (Barrett, 1994), and on how abatement costs are distributed. This paper's objective is to study how the choice between a quantity- or price-based formulation of a treaty influences the resulting number of participating countries, their internal welfare, and global welfare.

As our key results, we find that the larger the group of potentially cooperating countries becomes, the stronger is its preference for a treaty based on quantities. This choice would generate higher welfare for the treaty's participants, and also globally. However, it is counteracted by greater free-riding incentives resulting from this type of formulation, which undermines countries' willingness to actually participate in the cooperative effort. Formulating an agreement in terms of prices likely reduces free-riding incentives but does not necessarily lead to higher global welfare. Finally, we demonstrate that a coalition-wide market for tradable emission rights can create an additional incentive to join the agreement, and might thereby turn uncertainty into a facilitator of cooperation.

The remainder of this article is structured as follows. The general model is defined in Section 2. Section 3 derives the expected damages and abatement costs for the price and quantity policy. The resulting equilibrium in instrument choice for a given coalition size is characterized in Section 4. Section 5 analyzes the strategic implications of a treaty's formulation in prices or quantities on the free-riding incentive. The option of tradable quantities is discussed in Section 6. Numerical examples of the equilibrium outcome of the full game are presented in Section 7. The final section concludes.

2 The Model

Our analysis of self-enforcing IEAs follows the predominant approach of modelling the decision to join a single coalition as the first stage in a one-shot cartel-formation game. Following d'Aspremont and Gabszewicz (1986), the equilibrium is characterized by a single coalition that is both internally (no member has an incentive to leave) and externally (no non-member wants to join) stable.[1]

In the second stage of the game, we model instrument choice as a strategic variable, allowing members and non-members to either adopt an emission assignment (quantity policy) or an emission tax (price policy). As often observed in reality, we assume that the international agreement is based on one type of instrument only, applicable to all members. In the third and last stage, countries decide on the level of regulation. In both stages two and three we make the usual assumption of joint total cost minimization for the coalition. Non-members choose their strategy non-cooperatively so as to minimize their individual total costs. Choices within the last stage may be taken either simultaneously or sequentially (with the coalition as Stackelberg leader). In sum, the game's structure is

> 1st ("participation") stage:
> Countries choose whether to be members of the coalition.
>
> 2nd ("policy instrument") stage:
> The coalition (as a whole) and all non-members (individually) choose between price or quantity regulation.[2]
>
> 3rd ("policy level") stage:
> The coalition and the non-members choose their level of regulation either simultaneously or sequentially, where in the latter case the coalition acts as Stackelberg leader.

Uncertainty, to be formalized below, is only resolved after the last stage, forcing all countries to choose their strategies under incomplete information, based on expected total costs (abatement costs and damages, to be defined below). The game is solved by backwards induction.

Our approach allows studying the implications of endogenous instrument choice along three main questions. First, what determines the optimal choice between prices and quantities for a given coalition (second stage)? This generalizes Weitzman's 1974 analysis to the case of multiple regulators. Second, what is the effect on the free-riding incentive and, consequently, the participation in an international agreement (first stage)? Third and last, what are the implications for global welfare, which is a function of participation and instrument choice in the subgame-perfect equilibrium of the game?

With regard to the payoff function, we choose a formalization as a global public 'bad' game in emissions among the N countries, and use a total cost framework common in the literature on IEAs (Carraro and Siniscalco, 1993; Finus and Rundshagen, 1998). We assume simple functional forms in order to derive analytical results that reveal the trade-offs in instrument choice under uncertainty (see the concluding section for a discussion in how far our results hold in more general settings). The total costs TC_i for each country i are the sum of quadratic damages D from global emissions and quadratic abatement costs C_i for reducing individual emissions (see Appendix D for a complete list of symbols):

$$TC_i = D(e) + C_i(e_i) = d_1 \cdot e + \frac{d_2}{2} \cdot e^2 + \frac{1}{2}(\varepsilon_i - e_i)^2. \tag{1}$$

The symbol e_i represents emissions of country i, which are themselves a function of the implemented policy (discussed in detail in the next section). Global emissions are given by $e = \sum_{j=1}^{N} e_j$, d_1 and d_2 are damage parameters, and ε_i a parameter representing baseline emissions of country i, i.e. $e_i = \varepsilon_i$ if country i adopts no emission policy.

For the case of climate change all three parameters in the total costs (d_1, d_2, ε_i) are uncertain (Webster et al., 2002; Quirion, 2010; Dellink et al., 2008). For the question of instrument choice, Weitzman (1974) and Stavins (1996) show that the asymmetry between price and quantity instrument is solely driven by uncertainty about abatement costs, at least as long as damage uncertainty is not correlated with abatement cost uncertainty. We therefore only analyze uncertain abatement costs, in the specific form of uncertain baseline emissions, which have been identified as a major driver of abatement cost uncertainty (Edenhofer et al., 2006; Rogelj et al., 2013).[3]

With uncertain baseline emissions ε_i, the payoff function Eq. (1) becomes mathematically equivalent to the one of Weitzman (1974) and subsequent contributions, which all assumed an uncertain intercept of the marginal abatement cost function. However, Weitzman derived his specification from a more general payoff function using a second-order approximation in the optimum's neighborhood. This approach leads to the same uncertain payoff as in Eq. (1) if the additional assumption is made that the slope and the convexity of the payoff function are constant and independent of the level of emissions. Modeling a more general total cost function could make our analysis more general to some extent, but it would also make it less comprehensible as different parameters for the different optima at each coalition size would have to be introduced. By assuming that only baseline emissions are uncertain and total costs are quadratic, we are able to describe and analyze the general trade-offs involved in instrument choice under uncertain abatement costs and extend previous results most directly.

The uncertain ε_i could be distributed according to any of the standard probability density functions. However, because total costs are quadratic and countries are ex-ante symmetric, two parameters are sufficient to fully characterize the uncertainty of baseline emissions, which we derive in detail below. First, the standard deviation σ, which defines the common level of uncertainty about baseline emissions faced by all countries. Accordingly, setting $\sigma = 0$ recovers the certainty case. Second, the coefficient of correlation $\rho \in [0, 1]$, representing the possibly positive correlation of baseline emissions between countries. The ρ parameter captures the relative strength of local stochasticity of baseline emissions due to, for example, domestic growth shocks, against global factors affecting baseline emissions of all countries in the same direction, e.g. technology shocks that induce fuel switching or changes in productivity. In case of strong global interdependence (high ρ), uncontrolled emissions pose a greater risk of high damages because positive shocks in baseline emissions will not cancel out but rather reinforce each other.

Each country can regulate its contribution e_i to the public bad by either adopting an emission assignment $\bar{e}_i$, which fixes its amount of emissions, or by an emission tax p_i, which fixes the economy's ex-post marginal abatement costs, but not its emission level.[4]

3 Expected Total Costs under Prices vs. Quantities

In this section we solve the game's third and last stage, in which countries choose the policy level that minimizes their expected total costs. Formally, the symmetry break-up between price and quantity stems from the assumed ability of firms to react to baseline emissions after they have been observed (as in Weitzman 1974). Firms are modelled implicitly by the abatement costs C_i. Thus, with an emission tax p_i country i's economy reduces emissions until marginal abatement costs equal the tax level (emissions and marginal costs have the same dimension due to the normalization of Eq. 1), i.e. $p_i = -\frac{dC_i}{de_i}$, which by Eq. (1) implies:

$$p_i = -C_i'(e_i) = \varepsilon_i - e_i \quad \Rightarrow \quad e_i = \varepsilon_i - p_i. \tag{2}$$

Eq. (2) shows that shocks on baseline emissions are translated one-to-one into shocks on ex-post emissions e_i if a price-based regulation is chosen. To fix ex-post emissions the regulator can adopt a quantity target with an emissions assignment $\bar{e}_i$.[5] Hence, if one or more regulators choose a price-based regulation, the amount of global emissions is uncertain. The objective function to be miminized by each country i is then given by the expected total costs:

$$\mathrm{E}[TC_i] = d_1 \cdot \mathrm{E}[e] + \frac{d_2}{2} \cdot \mathrm{E}[e^2] + \frac{1}{2}\mathrm{E}[(\varepsilon_i - e_i)^2]. \tag{3}$$

where we used the E[·]-operator to denote the expected value.

Solving the last stage of the game, countries take the choice of instruments (prices or quantities) from the second stage of the game and the number of coalition members k from the first stage as given. We first derive the optimal level of regulation for non-members, who minimize their individual expected total costs.

The first-order-condition (FOC) of a respresentative non-member i with a quantity-based regulation (decision taken as given from stage 2) is derived by setting the emissions to $e_i = \bar{e}_{nm}$ and total emissions to $e = \bar{e}_{nm} + \sum_{j \neq i} e_j$ in Eq. (3). To obtain the Nashequilibrium of the last stage, we differentiate expected total costs with respect to $\bar{e}_{nm}$ taking the level of emission policy (emission assignments or taxes) of all other countries as given. The resulting FOC is:

$$0 = d_1 + d_2 \cdot \bar{e}_{nm} + d_2 \cdot \mathrm{E}[\sum_{j \neq i} e_j] + \bar{e}_{nm} - \mathrm{E}[\varepsilon], \tag{4}$$

where $\mathrm{E}[\varepsilon_i] = \mathrm{E}[\varepsilon] \; \forall \, i$ due to our assumption of ex-ante symmetric countries. In case the non-member implements a price-based regulation, the FOC is derived by taking Eq. (3) and setting the emission of a representative non-member i to $e_i = \varepsilon_i - p_{nm}$, total emissions to $e = \varepsilon_i - p_{nm} + \sum_{j \neq i} e_j$ and differentiating with respect to p_{nm}, yielding:

$$0 = -d_1 - d_2 \cdot (\mathrm{E}[\varepsilon] - p_{nm}) - d_2 \cdot \mathrm{E}[\sum_{j \neq i} e_j]) + p_{nm}. \tag{5}$$

Rearranging the last two equations, we observe that the optimal emissions level under quantities is equal to the expected level of emissions under optimal prices:

$$\bar{e}_{nm} = \mathrm{E}[\varepsilon] - p_{nm} = \frac{\mathrm{E}[\varepsilon] - d_1 - d_2 \cdot \mathrm{E}[\sum_{j \neq i} e_j]}{1 + d_2}. \tag{6}$$

The optimal level of the emission assignment and of the emission tax are hence equivalent in expected terms. In addition, the optimal policy level results to be independent of uncertainty, reflecting the fact that baseline emissions and regulated emissions enter the FOC only linearly.

As we show now, these two properties also hold for the FOC of the coalition members. Assuming without loss of generality that the coalition comprises the first $1..k$ countries, the coalition minimizes the sum of expected total costs of its members, $\sum_{j=1}^{k} \mathrm{E}[TC_j]$, by choosing the policy level of each of its members, where k is given from the first stage of the game. The coalition also takes as given the number of non-members that implement a price policy, which we denote by ℓ_{-k}. If the agreement specifies a quantity regulation, the FOC of the coalition can be derived by setting the emissions of each member i to $e_i = \bar{e}_m \; \forall i \in \{1..k\}$ and total emissions to $e = \sum_{j=1}^{k} \bar{e}_m + \sum_{j=k+1}^{k+\ell_{-k}} (\varepsilon_j - p_j) + \sum_{j=k+\ell_{-k}+1}^{N} \bar{e}_j$

in Eq. (3) and taking the derivative with respect to $\bar{e}_m$, giving:

$$
\begin{aligned}
0 \quad = \quad & kd_1 \cdot \left(k - \sum_{j=k+1}^{k+\ell_{-k}} \frac{dp_j}{d\bar{e}_m} + \sum_{j=k+\ell_{-k}+1}^{N} \frac{d\bar{e}_j}{d\bar{e}_m} \right) \\
& + kd_2 \cdot \left(k\bar{e}_m + \sum_{j=k+1}^{k+\ell_{-k}} (\mathrm{E}[\varepsilon] - p_j) + \sum_{j=k+\ell_{-k}+1}^{N} \bar{e}_j \right) \left(k - \sum_{j=k+1}^{k+\ell_{-k}} \frac{dp_j}{d\bar{e}_m} + \sum_{j=k+\ell_{-k}+1}^{N} \frac{d\bar{e}_j}{d\bar{e}_m} \right) \\
& - k\mathrm{E}[\varepsilon] + k\bar{e}_m. \qquad (7)
\end{aligned}
$$

The influence of the emission strategies of the coalition on the emissions of non-members, terms $\frac{dp_j}{d\bar{e}_m}$ and $\frac{d\bar{e}_j}{d\bar{e}_m}$, is equal to zero if emission choices are made simultaneously in the last stage, or otherwise derived by taking the derivative of Eq. (6), giving $\frac{d\bar{e}_j}{d\bar{e}_m} = -\frac{dp_j}{d\bar{e}_m} = -k\frac{d_2}{d_2(N-k)+1}$ for the case of sequential choices.

In case the coalition implements a price-based regulation, the FOCs are derived by setting the emissions of each member i to $e_i = \varepsilon_i - p_m \ \forall i \in \{1..k\}$ and the amount of total emissions to $e = \sum_{j=1}^{k}(\varepsilon_j - p_m) + \sum_{j=k+1}^{k+\ell_{-k}}(\varepsilon_j - p_j) + \sum_{j=k+\ell_{-k}+1}^{N} \bar{e}_j$ in Eq. (3) and taking the derivative with respect to p_m, yielding:

$$
\begin{aligned}
0 = \quad & kd_1 \cdot \left(-k - \sum_{j=k+1}^{k+\ell_{-k}} \frac{dp_j}{dp_m} + \sum_{j=k+\ell_{-k}+1}^{N} \frac{d\bar{e}_j}{dp_m} \right) \\
& + kd_2 \cdot \left(kp_m - k\mathrm{E}[\varepsilon] - \sum_{j=k+1}^{k+\ell_{-k}} (\mathrm{E}[\varepsilon] - p_j) - \sum_{j=k+\ell_{-k}+1}^{N} \bar{e}_j \right) \cdot \\
& \cdot \left(k + \sum_{j=k+1}^{k+\ell_{-k}} \frac{dp_j}{dp_m} - \sum_{j=k+\ell_{-k}+1}^{N} \frac{d\bar{e}_j}{dp_m} \right) \\
& + kp_m. \qquad (8)
\end{aligned}
$$

First, it can be observed that uncertainty about the level of unregulated emissions does not enter the FOC of either quantity or price regulation. As anticipated, the optimal policy level is thus independent of uncertainty. Second, Eqs. (7) and (8) coincide when setting

$$\bar{e}_m = \mathrm{E}[\varepsilon] - p_m. \qquad (9)$$

and recognizing that for the cross-dependency terms we obtain $\frac{dp_j}{de_m} = -\frac{dp_j}{dp_m}$ and $\frac{de_j}{de_m} = -\frac{de_j}{dp_m}$ from Eq. (6). In conclusion, the general relationship between the optimal (indicated by '*') price and quantity instrument levels is given by the regulator's analogue of Eq. (2)

and follows from Eqs. (6) and (9):

$$\bar{e}_i^*(k) = \mathrm{E}[\varepsilon] - p_i^*(k),\ \forall\, i. \tag{10}$$

In what follows, we only consider interior solutions. The explicit optimal emission assignments $\bar{e}_i^*$ were derived by Barrett (1994) for the case of a sequential choices[6] and by Finus and Ruebbelke (2008) for simultaneous choices[7].

The optimal level of regulation is independent of uncertainty because unregulated emissions ε_i do not affect how marginal abatement costs increase with the level of emission policy. Said differently, as marginal costs are linear with a certain slope, the last unit of abatement has a specific cost above or below the mean with the same probability. Analogously, correlation of emission uncertainty between countries also does not influence the optimal level of emission policy because uncertain global baseline emissions only shift the marginal damage curve (marginal damages are linear), and hence an increase of marginal damages above the mean is as likely as a decrease of the same level.

However, uncertainty about unregulated emissions does influence the expected total costs and therefore leads to an asymmetry between the quantity- and price-based regulation. Inserting Eqs. (2) and (10) in the abatement cost function $C_i(e_i)$ shows that implementing a quantity policy leads to an expected abatement costs mark-up of $1/2 \cdot \sigma^2$ vis-à-vis the price policy:

$$\mathrm{E}[C_i(\bar{e}^*)]|_{\bar{e}^*} = \frac{1}{2}\left\{\mathrm{E}[\varepsilon] - \bar{e}_i^*(k)\right\}^2 + \frac{1}{2}\sigma^2 \tag{11}$$

$$\mathrm{E}[C_i(\bar{e}^*)]|_{p^*} = \frac{1}{2}\left\{\mathrm{E}[\varepsilon] - \bar{e}_i^*(k)\right\}^2, \tag{12}$$

where the symbol $\cdot|_{\bar{e}^*/p^*}$ denotes the conditionality on the respective instrument.

Expected damages are, due to symmetry, only a function of the expected emission levels of members and non-members, and of the total number of countries implementing a price policy, hereafter denoted by $\ell \in [0, N]$,

$$\mathrm{E}[D\,(e)] = \underbrace{d_1 \cdot e^*(k) + \frac{d_2}{2} \cdot [e^*(k)]^2}_{\text{certainty term}} + \underbrace{\sigma^2 \cdot \frac{d_2}{2} \cdot [\ell + \ell(\ell-1)\rho]}_{\text{uncertainty term}} \tag{13}$$

as can be confirmed by inserting Eqs. (2) and (10) into the damage function. Here, $e^*(k) = \sum_{j=1}^{N} \bar{e}_j^*(k)$ denotes the *expected* amount of global emissions at the optimum in the presence of a coalition of size k. As argued earlier, the optimal expected emissions of each country are independent of the instrument choice, but depend on k and on whether the country is member or not of the agreement.

The first term in Eq. (13) is identical to the damage function obtained in the certainty case. Uncertainty ($\sigma > 0$) leads to an additional term that unambiguously increases expected damages if at least one country adopts a price regulation ($\ell \geq 1$). If countries' baseline emissions are uncorrelated ($\rho = 0$), total emissions uncertainty is the sum of individual uncertainties across all countries with price policy. In case of positive correlation ($\rho > 0$) expected damages are further amplified since emission shocks then tend to reinforce each other.

Expected global total costs therefore depend on the number of countries with price-based regulation ℓ and expected global emissions, which are a function of the size of the coalition k. Putting together Eqs. (11), (12), and (13) one obtains:

$$\sum_{j=1}^{N} \mathrm{E}[TC_j(k)] = N \cdot \left\{ d_1 \cdot e^*(k) + \frac{d_2}{2} \cdot [e^*(k)]^2 + \sigma^2 \cdot \frac{d_2}{2} \cdot [\ell + \ell(\ell-1)\rho] \right\} + \sum_{j=1}^{N} \frac{1}{2} \left\{ \mathrm{E}[\varepsilon] - \bar{e}_j^*(k) \right\}^2 + (N - \ell) \cdot \frac{1}{2} \sigma^2. \quad (14)$$

The number of members k of the agreement determines expected global emissions and the certainty equivalent parts of Eqs. (11), (12), and (13), meaning that global total costs will only approach the social optimum level if participation is sufficiently high (so as to enable the internalization of the externality). In other words, the rate of participation is an essential determinant of a treaty's environmental effectiveness. In addition, when the number ℓ of countries with price regulation increases, expected damages increase for all countries and expected abatement costs decrease, with the net effect on global total costs depending on d_2 and ρ.

By setting the number of countries N to one, Eq. (14) recovers the benchmark result of Weitzman (1974): the choice between price ($\ell = 1$) and quantity ($\ell = 0$) regulation is determined by the net effect of reducing expected abatement costs by $1/2 \cdot \sigma^2$ and increasing damage costs by $1/2 \cdot d_2 \cdot \sigma^2$ when switching from the quantity to the price instrument. In our case, in which the slope of marginal abatement costs is normalized to one, the price instrument would be preferred whenever $d_2 < 1$. The next section analyzes how this result is modified in a strategic multi-country setting.

4 Instrument Choice for a Given Coalition Size

This section derives the optimal instrument choice for non-members and members of the coalition, thus solving the second stage of the game. The size k of the coalition, being the outcome of the first stage, is taken as given.

4.1 Instrument Choice of Non-Members

A representative non-member country i prefers the instrument with the lower expected total costs, given the instrument choice of all other countries. Hence, if the total cost difference Δ

$$\Delta = \mathrm{E}[TC_{nm}]|_{p^*} - \mathrm{E}[TC_{nm}]|_{\bar{e}^*} \tag{15}$$

is negative, it will adopt a price regulation. According to Eq. (13), this difference depends on the number of countries (members and non-members) other than i with a price policy, which we denote by ℓ_{-i}. Using Eqs. (11), (12), and (13), and taking into account that the total number ℓ of countries with price policy differs by one in the two terms on the RHS of Eq. (15), the following holds:

Lemma 1

A non-member prefers a price over a quantity policy iff

$$\Delta = \frac{1}{2}\sigma^2 \{d_2 \cdot (1 + 2\rho \cdot \ell_{-i}) - 1\} < 0. \tag{16}$$

Three observations can be made. First, higher uncertainty amplifies the relative advantage of the preferred instrument but does not influence the sign of Δ. Second, the choice depends on the value of d_2, i.e. the relative slope of marginal damages and abatement costs. Third, there is a strategic interaction term shifting the relative advantage toward quantities whenever $\rho > 0$. While the first two are standard findings, the last effect is an extension of the single-regulator Weitzman (1974) rule. It reflects the increased risk of high damages when emission shocks are correlated, and may lead a non-member to prefer an emissions assignment even when marginal abatement costs are steeper than marginal benefits.

For $\rho > 0$, the interaction term also implies that in the case of no coalition ($k = 0$) a mix of countries with price and with quantity policy might be observed. Conversely, for $\rho = 0$ all countries would adopt the same instrument.

4.2 Instrument Choice of the Coalition

The instrument choice of a coalition of size k is again determined by the difference in expected total costs, but in this case computed as the sum across all members:

$$\Delta^k = \mathrm{E}[\sum_{j=1}^{k} TC_m(k)]|_{p^*} - \mathrm{E}[\sum_{j=1}^{k} TC_m(k)]|_{\bar{e}^*}. \tag{17}$$

Considering the instrument choice of the non-members as fixed, Eqs. (11), (12), and (13) yield the following lemma :

Lemma 2

A coalition of size k prefers a price over a quantity policy iff

$$\Delta^k = k \cdot \frac{1}{2}\sigma^2 \left\{ d_2 \cdot \left[k + \rho(k^2 + 2k \cdot \ell_{-k} - k) \right] - 1 \right\} < 0. \tag{18}$$

Again ℓ_{-k} denotes the number of non-members with a price policy. Several effects can be identified: First, because emissions are a public bad, marginal damages add up vertically and hence the effective ratio of marginal damages and marginal abatement costs is k-times higher for the coalition than for non-members, as captured by the first k in the term amplifying d_2. Second, if baseline emissions are correlated ($\rho > 0$), the incentive to choose quantities even grows with k^2. This reflects the fact that marginal abatement costs partially add up horizontally for the coalition.[8] Third, the strategic interaction effect observed for non-members also characterizes the coalition's choice: the higher the number of non-members with price policy, the stronger becomes its incentive to choose quantities. Because it affects all coalition members equally, this interaction effect is again k times stronger than for non-members. Finally, as all discussed effects apply to the total costs of all members, the entire cost difference scales with k.

This is the first important result: coalitions are more likely than non-members to prefer quantities over prices, especially if the considered coalition is large, uncertainty is correlated, and the number of non-members with price policy is high.

4.3 Equilibrium of the Second Stage

By Eqs. (16) and (18), the instrument choice of members and non-members of the coalition depends on the parameter values d_2 and ρ, the coalition size k, and the choice of the other countries, i.e. ℓ_{-i} and ℓ_{-k}. Therefore, in order to solve for the second stage's equilibrium it remains to determine the consistent combinations of instrument choice, with k taken as given from the first stage.

Consider first the simplest situation: the absence of a coalition, $k = 0$. The RHS of Eq. (16) can be solved for the value of ℓ_{-i} where it switches from a negative to a positive sign. The nearest higher integer of this number represents the maximum number of countries with a price policy any equilibrium can support, which from now on we denote by ℓ_∞. For $\rho \neq 0$ this parameter equals:[9]

$$\ell_\infty = \max\left(\left\lceil \frac{1}{2\rho}\left(\frac{1}{d_2} - 1\right)\right\rceil, 0\right), \tag{19}$$

where $\lceil\cdot\rceil$ is the ceiling-function, i.e. the function that returns the nearest higher integer number. In line with intuition, $d_2 < 1$ must hold for ℓ_∞ to be positive, i.e. the basic Weitzman criterion for choosing a price instrument must be fulfilled as a prerequisite. It follows that the equilibrium number of countries implementing a price policy is exactly ℓ_∞ if the total number of countries N is sufficiently large (for sure if $N \to \infty$, hence the notation) or, otherwise, equal to N.

Next, Appendix A shows that the distinction $\ell_\infty \lesseqqgtr N$ is relevant for the equilibrium instrument choice of a coalition of size k, as summarized in the following result:

Result 1 (*Equilibrium of the policy instrument stage*)

In equilibrium a coalition implements quantities if its size k is greater than or equal to k_q, with

1. $\ell_\infty > N$: $k_q = \begin{cases} N + \frac{1}{2\rho} - \frac{1}{2} - \sqrt{(N + \frac{1}{2\rho} - \frac{1}{2})^2 - \frac{1}{d_2\rho}} & \textit{if } \rho \neq 0 \\ \frac{1}{d_2}, & \textit{otherwise} \end{cases}$
2. $\ell_\infty \leq N$: $k_q = 3$.

Consider the first case $\ell_\infty > N$: by Eq. (16) all non-members choose a price regulation, no matter how large the coalition is and what instrument it implements. The preference of the coalition, determined by Δ^k from Eq. (18), increasingly moves toward quantities for growing k, with the eventual switch occurring at $k = k_q$ if it exists. In fact, the root may not exist or may be higher than N, in which case the coalition always chooses prices.

Second, for $\ell_\infty \leq N$ even the equilibrium in absence of a coalition would comprise $N - \ell_\infty$ countries implementing quantities, reflecting the potentially higher damages present in this case. As a consequence, all coalitions of size greater than two will regulate by quantities in order to mitigate the risk of high damages present when marginal damages are steep (d_2 large) or uncertainties are highly correlated (ρ large).[10] If the size of the coalition is smaller than the number of countries that implement quantities in the no-agreement case ($k + \ell_\infty \leq N$), ℓ_∞ non-members use a price policy. But as k increases and $k + \ell_\infty > N$, the total number of countries with price regulation will decrease.

5 Free-Riding and Participation under Prices vs. Quantities

In the previous section we showed that the more countries join a coalition, the more likely it becomes that it will adopt a treaty based on quantities. The choice of each country to either become a member of such a coalition or to free-ride constitutes the first stage of our game. In this section we analyze how the incentive to free-ride is influenced by the presence of uncertainty and endogenous instrument choice. However, since a closed-form solution of the game does not exist (Barrett, 1994), the analysis cannot provide explicit expressions for the size of the stable coalition and global total costs. We will resort to numerical solutions to illustrate the magnitude of effects in Section 7.

5.1 Free-Riding Incentive under Endogenous Instrument Choice of the Coalition

The free-riding incentive is commonly studied by considering the difference in welfare – in our model equivalent to negative expected total costs – between being a member of a coalition of size k and being a non-member of a coalition of size $k - 1$, the so-called stability function (see Hoel 1992):

$$\Phi(k) := \mathrm{E}[TC_{nm}(k-1)] - \mathrm{E}[TC_m(k)]. \tag{20}$$

If $\Phi(k)$ is negative, the coalition of size k is internally unstable because the k-th member has an incentive to leave in order to decrease its total costs. The largest coalition size k for which Eq. (20) is still positive represents the equilibrium outcome of the first stage, i.e. the endogenous *participation* in the agreement (external stability holds by symmetry). Here, we take the function $\Phi(k)$ as a continuous indicator of the free-riding incentive: if its value becomes lower – e.g. for higher values of ρ or σ – we conclude that free-riding is aggravated (and vice versa), even if the equilibrium participation may not change due to the integer formulation of the coalition model.

To compute Eq. (20) we have to determine which instrument a coalition member would implement when leaving the agreement. In case the resulting group of non-members comprises both countries with price and with quantity regulation, it cannot be inferred from the equilibrium conditions to which group the additional non-member would belong, since the conditions only allow to determine the total number of countries with prices, but not their identity. We assume that in this case all non-members have – and expect – the same probability for becoming a country with price regulation. Under this assumption, Appendix B derives the following result:

Result 2 (*Free-riding incentive under uncertainty*)

> *Let $\bar{k}^*$ be the equilibrium participation under certainty. Uncertainty ($\sigma > 0$) increases the free-riding incentive whenever $k_q + 1 \leq \bar{k}^*$ and $\ell_\infty \geq 1$.*

Endogenous instrument choice under uncertainty aggravates free-riding whenever three not very restrictive conditions are met: (i) the coalition's rational best choice is to adopt a quantity regulation, (ii) this remains to be the case even if the coalition's size is reduced by one, and (iii) quantities are not a dominant strategy.[11] As a noteworthy corollary, the free-riding incentive is always weakly increased if the coalition's agreement is exogenously constrained to quantities, e.g. for political reasons.

Intuitively, free-riding can be more attractive than in the certainty case because non-members not only benefit from the coalition's choice regarding emission reductions – the origin of free-riding in the standard case with certainty – but also with regard to the policy instrument. Free-riders can switch from the socially preferable quantity instrument to the individually preferable price-based regulation whenever the reduction in expected abatement costs outweighs their individual increase in expected damages. Result 2 determines when this condition is met.

The uncertainty-related increase of the free-riding incentive is stronger if emission uncertainty is correlated and the coalition is large. In this case the coalition's quantity agreement alleviates the individual trade-off between instruments for free-riders by reducing the risk of high damages associated with $\rho > 0$. If emission uncertainty is uncorrelated, the stronger free-riding incentive reduces to the basic Weitzman gain of the single regulator case, as there is no strategic interaction with the instrument choice of the coalition.[12]

Hence, due to uncertainty there exists an additional instrument-related dimension of free-riding, which aggravates the standard free-riding incentive. By using quantities the coalition's own ambitiousness undermines participation and therefore the environmental effectiveness of the treaty. Our result thus echoes a known insight from studies of public good games (Barrett, 2002; Lessmann et al., 2014): ambitious targets of a coalition decrease the incentive to join the treaty. "Modesty" (Finus and Maus, 2008) in the form of a commitment to prices could improve the situation.

5.2 Free-Riding Incentive under an Agreement Restricted to Prices

Because the coalition chooses its instrument after the participation decision is made, members do not anticipate the effect of their choice on the equilibrium size of the coalition and the resulting total costs. To illustrate how results might change if the effects of endogenous instrument choice on free-riding are taken into account, we study the situation in which the treaty instrument is restricted to prices.[13]

It turns out that restricting the treaty to prices would in fact mitigate the additional free-riding incentive observed in the endogenous choice case (Result 2):

Result 3 (*Free-riding incentive with price-based agreement*)

Compared to a treaty with endogenous instrument choice, a restriction to a price-based treaty decreases the free-riding incentive for all coalition sizes k satisfying $k_q + 1 \leq k \leq \ell_\infty$.

To derive this result, it is again instructive to distinguish the cases $\ell_\infty > N$ and $\ell_\infty \leq N$. In the former, non-members always implement prices, irrespective of the size of the coalition. The coalition members – now by definition – do the same, and hence the possibility of additional free-riding benefits from instrument choice, as described in Result 2, vanishes. Since free-riding is increased under endogenous choice, a treaty restricted to prices weakly enhances participation.

The case of $\ell_\infty \leq N$ is more complicated, as a large enough coalition – in which all members are obliged to implement prices – could lead to a situation in which the total number of countries with prices becomes greater than ℓ_∞. A cooperating country would be forced to implement prices when it would – as a non-member – choose quantities to reduce its individual total costs. In these circumstances the free-riding incentive in the presence of a (forcedly) price-based agreement could become higher than for a quantity-based agreement. To exclude this implausible case, Result 3 makes the restriction of $k \leq \ell_\infty$. Given that, it is clear that any new coalition member will benefit from eliminating the abatement cost mark-up associated with a quantity regulation, while uncertainty on damages does not change because the total number of countries implementing prices remains constant at ℓ_∞ for all coalition sizes (i.e. additional members of the coalition always 'crowd out' a non-member with price instrument). In such a setting, the presence of uncertainty actually leads to a lower free-riding incentive than in the certainty case, and therefore also with respect to endogenous instrument choice.

Result 3 shows that an ex-ante restriction to a price-based agreement would mitigate free-riding in a situation where marginal damages are moderately flat, such that non-members tend to prefer prices, while the coalition - at least at some critical size - prefers to implement a quantity policy. For climate change, marginal damages have been described as relatively flat in the short to medium term (Pizer, 2002), suggesting that this case could indeed be empirically relevant.

Comparing the global total costs (in equilibrium) of treaties based on either prices or quantities does not yield a clear-cut result. With quantities the resulting coalition may be smaller, which likely decreases the agreement's environmental effectiveness and thus increases the certainty terms in the global total cost expression Eq. (14). However, even if the price-based agreement increases the number of member countries, this is likely to come at the cost of higher global emission uncertainty, since the total number of countries with prices might be higher than under a quantity-based agreement. Hence, a price-based agreement tends to decrease the certainty terms in Eq. (14), but also to increase the uncertainty-related terms. The net effect depends on which contribution dominates, which in turn depends on the values of all parameters of the model. An exact quantification cannot be derived analytically due to the absence of a closed-form solution of the game.

6 Extension: Tradable Quantities

One driver of the potentially higher free-riding incentive under uncertainty is the abatement cost mark-up incurred by members of a quantity-based agreement. In this section, we therefore consider an extension of the model addressing uncertain abatement costs: an agreement with tradable quantities. Under this approach, each member country still has a quantitative emission assignment. However, actual emissions are determined only after uncertainty has been resolved, by means of a competitive market in which member countries can buy and sell emission rights. Trading allows to smooth the uncertain realizations of baseline emissions among the coalition members (unless $\rho = 1$), and hence decreases their individually expected abatement costs. The policy level of emissions is not affected by the decision to implement tradable quantities compared to fixed quantities because the first-order conditions are linear in emissions and the uncertain baseline emissions. Thus, Eq. (13) still holds, whereas expected abatement costs become lower than in Eq. (11), namely:[14]

$$\mathrm{E}[C_m(e)]|_{\bar{e}_k^*} = \frac{1}{2}\left[\mathrm{E}[\varepsilon] - \bar{e}_m^*(k)\right]^2 + \frac{1}{2}\sigma^2 - \frac{1}{2}\sigma^2\frac{k-1}{k}(1-\rho). \tag{21}$$

The symbol $|_{\tilde{e}_k^*}$ denotes expected costs with tradable emission rights. Hence, this approach is always at least as good as the non-tradable emission assignments analyzed before. Based on the last equation, Appendix C derives the following result:

Result 4 (*Free-riding incentive with agreement based on tradable quantities*)

Whenever $\rho < 1$ and $k \geq k_q$, tradable quantities lead to a lower free-riding incentive than non-tradable quantities. Uncertainty may even increase participation above the number $\bar{k}^*$ observed in the certainty case, if $\bar{k}^*$ is sufficiently large and ρ sufficiently small.

The first statement stems from the fact that the gains from emissions trading are kept exclusively within the coalition, while non-members do not experience any change at all. Clearly, this shifts the stability function upwards and makes free-riding less attractive than under the previously considered quantity approach.

The second part shows that emissions trading may turn uncertainty into an ally of cooperation. In fact, with increasing coalition size k, the mark-up on abatement costs associated with baseline emission uncertainty becomes smaller and smaller, reaching zero in the limit of $k \to \infty$ and $\rho = 0$. In other words, with $\rho < 1$ tradable quantities behave similar to prices for large k coalitions, allowing to combine the advantages of the two instruments. This is not possible for non-members, for whom instrument choice always entails a trade-off. As a consequence, the incentive to join such an agreement can become higher than in the certainty case.

7 Numerical Examples

So far we used an analytical approach to identify the different effects that determine participation in and expected total costs of a treaty in the presence of uncertainty. As a consequence, our analysis lacks information on the relative importance of the individual effects. This section presents a series of numerical simulations to illustrate the quantitative impact of endogenous instrument choice on coalition formation. Table 1 lists numerically derived equilibria for various levels of uncertainty and for three different possibilities of instrument choice (horizontal dimension) and sets of parameters (vertical dimension). The parameters are such that the coalition's equilibrium choice always falls on quantities (left column) or respectively tradable quantities (right column). For the last stage of the game, in which the policy level is chosen, we assume sequential emissions decisions with the coalition acting as a Stackelberg leader.

The first column of Table 1, the case in which the coalition endogenously chooses to implement quantities, shows that for all three parameter sets a higher level of uncertainty leads to a lower level of equilibrium participation in the agreement, in line with Result 2.

The next column demonstrates how – at zero ($\rho = 0$) or moderate ($\rho = 0.2$) correlation – the increase in free-riding due to uncertainty can be mitigated by restriction to a price-based agreement, thus confirming Result 3. However, although the commitment to prices leads to higher participation, it does not guarantee lower expected total costs for coalition members or globally: for the case of $\rho = 0$, a five-countries coalition implementing quantities still realizes a lower level of expected total costs for all countries than a much larger coalition of size 18 that implements prices. Hence, there is no conflict between a price-based or a quantity-based agreement in this case, since the coalition's endogenous choice of quantities induces the equilibrium that is preferable from all perspectives.

This situation changes when increasing the coefficient of correlation to $\rho = 0.2$: the restriction to prices lifts equilibrium participation even above the certainty level (i.e. uncertainty becomes an ally of cooperation), and now also decreases both the coalition members' and global expected total costs. As discussed in Section 5.2, the extra incentive to join a price-constrained agreement stems from the fact that entry into the coalition allows to reduce expected abatement costs (no uncertainty cost mark-up for countries with prices) without increasing the level of global damage uncertainty, since – as long as $k \leq \ell_\infty$ – the total number of countries with quantities remains constant.

If ρ is increased to the extreme case of perfect correlation, the restriction to prices turns again into a disadvantage: as the bottom section of Table 1 shows, participation is lower and total costs are higher (or equal) than for the endogenous choice equilibrium with quantities. Because of the low $\ell_\infty = 6$ value, the restriction to prices now has the effect of pushing the total number of countries with prices above the number observed in the no-agreement equilibrium. Thus, committing to prices no longer represents "modesty" (Finus and Maus, 2008) but rather self-imposed uncooperativeness, which creates an additional incentive not to join the agreement. Total costs of the remaining members are lower than in the endogenous choice case, but globally the world is worse off.

Finally, when changing to the 'prices vs. tradable quantities' setting and the coalition's equilibrium choice falls on the latter (right column), the outcome always improves upon the endogenous choice case without trade, except if $\rho = 1$, when it is equal. In the most favorable case, i.e. with $\rho = 0$, participation becomes higher with than without uncertainty, confirming Result 4. A treaty implementing emissions trading is superior to the other design options with respect to participation, the coalition's and global expected total costs.

If correlation increases to $\rho = 0.2$, the endogenous equilibrium choice of tradable quantities still leads to a higher level of participation than in the certainty case, but remains lower than for a treaty restricted to prices. The latter is now also preferable from a total costs perspective, showing how the benefits of trade rely quite strongly on a low level of uncertainty correlation.

[Table 1 about here.]

8 Discussion and Conclusion

To limit the impacts of transboundary environmental pollutants, treaties like the Helsinki-Protocol, the Montreal-Protocol, or the Kyoto-Protocol established "targets and timetables" (Victor, 2011) of the emission reductions of participating countries. As shown by Weitzman (1974) and others, an alternative formulation based on emission taxes might enhance the welfare of cooperating countries when costs of abatement are uncertain. Our study shows – to our knowledge for the first time – that in the presence of global public goods, for which cost uncertainty is ubiquitous, the choice of the policy instrument also bears a strategic dimension influencing cooperation: on the one hand a treaty based on quantities could be socially preferable to one based on prices, but on the other hand it also increases the free-riding incentive.

The intuition of our main result follows the one first described by Weitzman (1974), who considers instrument choice in the absence of international strategic effects. The choice between a quantity or price instrument is characterized by the conflicting objectives of avoiding unexpectedly high damages and avoiding unexpectedly high abatement costs: in case of a relatively higher importance of damages the quantity instrument should be chosen. A coalition of countries by definition internalizes all damages of its members, while non-cooperating countries take only their own individual damages into account. As a consequence, avoiding unexpectedly high damages is relatively more important for a coalition than for a same size group of non-cooperating countries (all else equal). Hence we find that cooperation shifts instrument preference in the direction of quantities.

The choice of non-cooperating countries is in principle characterized by the same conflict between damages and abatement costs as the coalition's. But because free-riders internalize only their own damages, they gain from adopting the individually preferred price policy instead of the socially preferred quantity regulation. A coalition that has adopted a quantity regulation generally alleviates the trade-off between instruments for free-riders, because by fixing its own emissions the coalition also reduces the damage uncertainty faced by the non-cooperating countries. Hence, the coalition's instrument choice – if it falls on quantities – can make free-riding more attractive.

We demonstrate that an alternative treaty formulation based on prices often decreases the free-riding incentive, but may not necessarily enhance global welfare. The reason is that each country's welfare is influenced by instrument choice in a two-fold way: a treaty's instrument determines (i) the participation rate in the agreement and hence the expected environmental effectiveness of the treaty and (ii) the uncertainty about damages faced by each country. If prices are individually preferred, a price-based treaty ensures a higher number of cooperating countries, which likely increases the internalization of the emission externality. However, because a treaty based on prices excludes quantities for cooperating countries, there could be a higher level of uncertainty about emissions than for a quantity-based treaty with lower participation, which increases each country's expected damages. Any one of the two effects might dominate. Furthermore, our analysis suggests that the conflict resulting from these opposing effects could be alleviated if the treaty includes emissions trading among its members, in which case uncertainty could even amplify the incentive to cooperate.

Our model shows how free-riding incentives can be shaped differently according to whether an international agreement adopts a price or quantity formulation, thus addressing one of the major obstacles to cooperation in the presence of public goods (Barrett, 2003). Although our results are derived within a special theoretical model of reduced complexity to derive analytical insights, the basic mechanisms extend to more general settings. Any environmental problem with a transboundary pollution externality, unless linear in costs or damages, implies a trade-off between prices and quantities under abatement cost uncertainty (see the discussion following Eq. 1). Therefore, a treaty formulated in quantities will benefit free-riders by reducing uncertainty about damages (and hence expected damages) and increase expected compliance costs for coalition members as compared to a price-based regulation.

From a policy perspective, our results can contribute to a better understanding of the international community's ongoing struggle to agree on a comprehensive climate treaty with Kyoto-style emission targets. Such targets could result in high abatement costs, which countries seek to avoid, while uncertainty about high damages is moderate as marginal damages are rather flat in the short to medium term (Pizer, 2002). Our analysis suggests that cooperation might be facilitated if countries negotiated over carbon prices instead of emission targets, as also recently discussed by Weitzman (2015).

A Derivation of Result 1

In the first case of Result 1, $\ell_\infty > N$, non-members always implement prices. Hence, the number of non-members choosing prices is $\ell_{-k} = N - k$. Inserting this in Eq. (18), the decision to switch from prices to quantities as k increases is determined by Δ^k becoming positive:

$$\Delta^k = \frac{1}{2}\sigma^2 \left\{ d_2 \cdot \left[\underbrace{k + \rho(k^2 + 2k \cdot (N-k) - k)}_{=k+\rho(2Nk-k^2-k)} \right] - 1 \right\} k \geq 0 \tag{A.1}$$

Solving for the smaller root in the size of the coalition k in the case of equality leads to the first condition of Result 1. When setting $\rho = 0$, Eq. (18) is linear and can be directly solved for its root, deriving the second condition.

In the second case, $\ell_\infty \leq N$, the equilibrium in the absence of an agreement comprises some countries that choose a quantity-based regulation. It follows from Eq. (16) that $d_2(1+2\rho\ell_\infty)-1 \geq 0$ (*). We now test whether the coalition has an incentive to switch its instrument choice from prices to quantities given the choices of non-members, such that the coalition will implement quantities in the equilibrium of the second stage of the game.

Consider first the case where $k > \ell_\infty$. If we assume that all coalition members would implement prices, none of the outsiders will find it optimal to choose a price-based regulation: $\ell_{-k} = 0$. Inserting this to determine the expected total cost difference Δ^k, Eq. (18), gives:

$$\Delta^k = \frac{1}{2}\sigma^2 \left\{ d_2 \cdot \left[k + \rho \cdot \underbrace{(k^2 - k)}_{\geq(\ell_\infty+1)^2-(\ell_\infty+1)\geq 2\ell_\infty} \right] - 1 \right\} k. \tag{A.2}$$

By (*) this is always greater than or equal to zero: the coalition has an incentive to switch to quantities. Hence, the coalition would choose quantities in equilibrium.

Second, if $k \leq \ell_\infty$, $\ell_{-k} = \ell_\infty - k$ non-members would implement prices if the coalition implements prices collectively. Again inserting to Eq. (18) and with the help of (*), the expected total cost difference:

$$\Delta^k = \frac{1}{2}\sigma^2 \left\{ d_2 \cdot \left[k + \rho(k^2 + 2k(\ell_\infty - k) - k) \right] - 1 \right\} k \tag{A.3}$$

$$= \frac{1}{2}\sigma^2 \left\{ d_2 \cdot \left[k + 2\rho\ell_\infty + \rho \underbrace{(2\ell_\infty(k-1) - k^2 - k)}_{\geq 2k(k-1)-k^2-k\geq 0\,\forall\, k\geq 3} \right] - 1 \right\} k \tag{A.4}$$

is always greater than or equal to zero $\forall\ k \geq 3$.

Hence, a coalition of size $k \geq 3$ would switch to quantities. In conclusion, for all parameters of the model and coalition sizes $k \geq 3$ members would choose a quantity-based regulation in the equilibrium of the seond stage.

B Derivation of Result 2

We prove the result by considering the two cases $\ell_\infty \gtrless N$. For $\ell_\infty > N$ the equilibrium in the absence of an agreement consists only of prices, meaning $d_2\,[1 + 2\rho(N-1)] - 1 < 0$ (**) from Eq. (16). Within the regime in which both the coalitions of size k and at $k-1$ implement quantities ($k_q + 1 \geq k^*$), a non-member will increase the number of countries implementing prices by one when leaving. The stability function Φ , i.e. the change in expected total costs from free-riding in Eq. (20), is:

$$\Phi(k) = \bar{\Phi}(k) + \underbrace{\frac{1}{2}\sigma^2\,\{d_2 \cdot [1 + 2\rho(N-k)] - 1\}}_{<0}, \tag{B.1}$$

where $\bar{\Phi}(k)$ is the difference in expected total costs under certainty. Using (**), the uncertainty-related second term is always negative and hence the free-riding incentive is increased when $\sigma > 0$.

For the case of $\ell_\infty \leq N$, Eq. (16) implies $d_2\,[1 + 2\rho(\ell_\infty - 1)] - 1 < 0$ (***). If $k > 3$, Result 1 implies that both coalitions of sizes $(k, k-1)$ will implement quantities. If $k + \ell_\infty \leq N$, ℓ_∞ non-members will implement prices. Hence the total number of countries implementing prices stays the same for both coalitions of size k and $k-1$. The difference in expected total costs from free-riding, Eq. (20), then reflects the fact that the non-member has a certain chance to switch to prices when becoming a free-rider while damage increases are fixed:

$$\Phi(k) = \bar{\Phi}(k) - \frac{1}{2}\sigma^2 \left\{\frac{\ell_\infty}{N-k+1}\right\}. \tag{B.2}$$

Since $\ell_\infty \geq 1$, free-riding is aggravated. If $k + \ell_\infty > N$, all non-members will implement prices. A member country will therefore increase the number of countries implementing prices by one when becoming a free-rider. The change in expected total costs from free-riding is:

$$\Phi(k) = \bar{\Phi}(k) + \underbrace{\frac{1}{2}\sigma^2 \left\{d_2 \cdot \left[1 + 2\rho\underbrace{(N-k)}_{\leq \ell_\infty - 1}\right] - 1\right\}}_{<0}. \tag{B.3}$$

As per (***), the uncertainty-related summand is again always negative meaning that free-riding is aggravated.

C Derivation of Result 4

The first part of this result was already derived non-formally in the corresponding section. The second part can easily be demonstrated when we assume $\rho = 0$. The difference in expected total costs from free-riding is

$$\Phi(k) = \bar{\Phi}(k) + \frac{\sigma^2}{2}\left(d_2 - 1 + \frac{k-1}{k}\right). \tag{C.1}$$

If we let k approach infinity, the second summand reduces to $1/2 \cdot d_2 \cdot \sigma^2$, which is always positive. Therefore, the free-riding incentive is decreased compared to certainty.

D List of symbols

[Table 2 about here.]

Notes

[1]Countries are assumed to be represented by a single regulator. We use these terms - countries and regulators - interchangeably.

[2]The case of Stackelberg-leadership of the coalition in instrument choice is not considered here, but we discuss in Section 5.2 the case when only a price-based agreement is available.

[3]For a discussion of the influence of uncertain damages on the policy level and participation decision see for example Na and Shin (1998); Kolstad (2007); Finus and Pintassilgo (2013).

[4]One may oppose that the regulators can adjust the policy level once new information arrives but before all abatement decisions have been taken, which would add a dynamic dimension to our model. However, our static model still reveals basic characteristics of instrument choice under uncertainty because an adjustment of the policy level will possibly not (i) remove sunk costs or sunk emissions that occurred in the previous legislation period or (ii) influence the fundamental trade-off between prices and quantities for the next legislation period, for which uncertainty prevails.

[5]We assume that a quantity target is always binding, i.e. $e_i = \bar{e}_i$. For this to be reasonable the abatement implied by target $\bar{e}_i$ must be sufficiently larger than the typical baseline fluctuation σ.

[6]The optimal levels of expected emission are $\bar{e}^*_m = \frac{\mathrm{E}[\varepsilon](d_2(N-k)+1)^2 - d_1 k - d_2 k(N-k)\mathrm{E}[\varepsilon]}{(d_2(N-k)+1)^2 + k^2 d_2}$ and $\bar{e}^*_{nm} = \frac{(\mathrm{E}[\varepsilon]-d_1)(d_2(N-k)+1) + d_2\mathrm{E}[\varepsilon]k[k-(d_2(N-k)+1)]}{(d_2(N-k)+1)^2 + k^2 d_2}$.

[7] The optimal levels of expected emissions are $\bar{e}^*_m = \frac{\mathrm{E}[\varepsilon](d_2(N-k)+1-kd_2(N-k))-kd_1}{d_2(N-k)+1+k^2 d_2}$ and $\bar{e}^*_{nm} = \frac{\mathrm{E}[\varepsilon](1-d_2k+d_2k^2)-d_1}{d_2(N-k)+1+k^2 d_2}$

[8]Eq. (18) reduces to Eq. (28) in Weitzman (1974) if $\ell_{-k} = 0$ and when aggregating the coalition into a single regulator with k firms.

[9]For $\rho = 0$ equilibria without a coalition are characterized by dominant strategies in instrument choice: $d_2 \leq 1$ implies $\ell_\infty = \infty$, i.e. all countries choose prices, while $d_2 > 1$ implies $\ell_\infty = 0$, i.e. all countries choose quantities.

[10]For $k = 2$ equilibria with the coalition choosing prices or quantities exist.

[11]Trivially, for $\ell_\infty > N$ and $\bar{k}^* < k_q$ the free-riding incentive is not changed under uncertainty because both coalition and non-members implement prices. In the special case of $\ell_\infty > N$ and $k_q + 1 > \bar{k}^* \geq k_q$, the free-riding incentive is decreased because the leaving member would incur a loss from the coalition's collective instrument switch.

[12] The components of enhanced free-riding can be illustrated by discussing Eq. (B.1) derived in the proof of Result 2, which represents the downward shift of the stability function: $\frac{1}{2}\sigma^2 \{\cdot d_2[1 + 2\rho(N - k)] - 1\}$. In the case of uncorrelated uncertainties, $\rho = 0$, the expression simplifies to $\frac{1}{2}\sigma^2 \{\cdot d_2 - 1\}$, which represents the single regulator gain from instrument switch as in Weitzman (1974). If uncertainties are correlated, $\rho > 0$, the presence of k members with quantity regulations reduces the individual increase in expected damages (when switching to quantities) by decreasing the risk of high damages, which is represented by the term $2\rho(N - k)$ that adds to the uncorrelated damage increase. Thus, the trade-off between instruments for free-riders is alleviated by the coalition and free-riding is more attractive.

[13]In principle, one could introduce a pre-stage to the game in Section 2 in which countries can vote whether they prefer a treaty based on quantities or prices without knowing whether they will be a member of the agreement or not (as in Finus and Maus, 2008). However, no closed-form solution of our model exists and we cannot solve for the resulting total costs in equilibrium. The following analysis nevertheless identifies the driving forces that would govern such a choice.

[14] See Williams (2002) for an equivalent equation and its derivation.

References

Scott Barrett. Self-Enforcing International Environmental Agreements. *Oxford Economic Papers*, 46:878–894, 1994.

Scott Barrett. Consensus Treaties. *Journal of Institutional and Theoretical Economics*, 158(4):529–547, 2002.

Scott Barrett. *Environment and Statecraft: The Strategy of Environmental Treaty-Making*. Oxford University Press, 2003.

Scott Barrett. Climate treaties and approaching catastrophes. *Journal of Environmental Economics and Management*, 66:235–250, 2013.

Scott Barrett and Robert Stavins. Increasing Participation and Compliance in International Climate Change Agreements. *International Environmental Agreements: Politics, Law and Economics*, 3:349–376, 2003.

Carlo Carraro and Domenico Siniscalco. Strategies For The International Protection Of The Environment. *Journal of Public Economics*, 52(3):309–328, 1993.

Richard Cooper. Toward a Real Treaty on Global Warming. *Foreign Affairs*, 77:66–79, 1998.

C. d'Aspremont and J. J. Gabszewicz. *New developments in the analysis of market structures*, chapter On the stability of collusion, pages 243–64. Macmillan, New York, 1986.

Rob Dellink, Michael Finus, and Niels Olieman. The stability likelihood of an international climate agreement. *Environmental and Resource Economics*, 39:357–377, 2008.

Ottmar Edenhofer, Claudia Kemfert, Kai Lessman, Michael Grubb, and Jonathan Koehler. Induced Technological Change: Exploring its Implications for the Economics of Atmospheric Stabilization: Synthesis Report from the Innovation Modeling Comparison Project. *The Energy Journal*, Endogenous Technological Change and the Economics of Atmospheric Stabilisation:57–107, 2006.

Thomas Eichner and Rüdiger Pethig. Self-enforcing international environmental agreements and trade: taxes versus caps. *Oxford Economic Papers*, pages 1–21, 2015. doi: doi:10.1093/oep/gpv037.

Alfred Endres and Michael Finus. International Environmental Agreements: How the Policy Instrument Affects Equilibrium Emissions and Welfare. *Journal of Institutional and Theoretical Economics*, 155:527–550, 1999.

Michael Finus and Stefan Maus. Modesty may pay. *Journal of Public Economic Theory*, 10:801–826, 2008.

Michael Finus and Pedro Pintassilgo. International Environmental Agreements under Uncertainty: Does the Veil of Uncertainty Help? *FEEM working paper*, 2010.

Michael Finus and Pedro Pintassilgo. The Role of Uncertainty and Learning for the Success of International Climate Agreements. *Journal of Public Economics*, 103: 29–43, 2013.

Michael Finus and Dirk Ruebbelke. Coalition Formation and the Ancillary Benefits of Climate Policy. *FEEM working paper*, (62), 2008.

Michael Finus and Bianca Rundshagen. Renegotiation-Proof Equilibria in a Global Emission Game When Players Are Impatient. *Environmental and Resource Economics*, 12:275–306, 1998.

Michael Finus, Pedro Pintassilgo, and Alistair Ulph. International Environmental Agreements with Uncertainty, Learning and Risk Aversion. *The University of Manchester Economics Discussion Paper Series*, (EDP-1329), 2013.

Cameron Hepburn. Regulation by Prices, Quantities, or both: a Review of Instrument Choice. *Oxford Review of Economic Policy*, 22(2):226–247, 2006.

Michael Hoel. International environment conventions: The case of uniform reductions of emissions. *Environmental and Resource Economics*, 2(2):141–159, 1992.

Larry Karp. The effect of learning on the membership and welfare in an International Environmental Agreement. *Climatic Change*, 110:499–505, 2012.

Larry Karp and Leo Simon. Participation games and international environmental agreements: A non-parametric model. *Journal of Environmental Economics and Management*, 65:326–344, 2013.

Charles D. Kolstad. Systematic Uncertainty in Self-enforcing International Environmental Agreements. *Journal of Environmental Economics and Management*, 53(1):68–79, 2007.

Charles D. Kolstad and Alistair Ulph. Learning and International Environmental Agreements. *Climatic Change*, 89(1-2):125–141, 2008.

Charles D. Kolstad and Alistair Ulph. Uncertainty, Learning and Heterogeneity in International Environmental Agreements. *Environmental and Resource Economics*, 50: 389–403, 2011.

Frank C. Krysiak. Prices vs quantities: The effects on technology choice. *Journal of Public Economics*, 92:1275–1287, 2008.

Kai Lessmann, Robert Marschinski, Michael Finus, Ulrike Kornek, and Ottmar Edenhofer. Emissions Trading with Non-signatories in a Climate Agreement- an Analysis of Coalition Stability. *The Manchester School*, 82:82–109, 2014. doi: 10.1111/manc.12045.

Warwick McKibbin, Adele Morris, and Peter Wilcoxen. Bridging the gap: Intergrating price mechanisms into international climate negotiations. *Brookings: Climate and Energy Economics Discussion Paper*, 2012.

Seong-lin Na and Hyun Song Shin. International environmental agreements under uncertainty. *Oxford Economic Papers*, 50:173–185, 1998.

Richard G. Newell and William A. Pizer. Regulating stock externalities under uncertainty. *Journal of Environmental Economics and Management*, 45:416–432, 2003.

William D. Nordhaus. To Tax or Not to Tax: Alternative Approaches to Slowing Global Warming. *Review of Environmental Economics and Policy*, 1:26–44, 2007.

William A. Pizer. Combining price and quantity controls to mitigate global climate change. *Journal of Public Economics*, 85:409–434, 2002.

Philippe Quirion. Complying with the Kyoto Protocol under uncertainty: Taxes or tradable permits? *Energy Policy*, 38:5166–5173, 2010.

Joeri Rogelj, David L. McCollum, Andy Reisinger, Malte Meinshausen, and Keywan Riahi. Probalistic cost estimates for climate change mitigation. *Nature*, 493:79–83, 2013. doi: doi:10.1038/nature11787.

Robert N. Stavins. Correlated Uncertainty and Policy Instrument Choice. *Journal of Environmental Economics and Management*, 30:218–232, 1996.

Alistair Ulph. Stable International Environmental Agreements with a Stock Pollutant, Uncertainty and Learning. *The Journal of Risk and Uncertainty*, 29(1):53–73, 2004.

David Victor. *Global warming gridlock: creating more effective strategies for protecting the planet*. Cambridge University Press, 2011.

WBGU. *Solving the Climate Dilemma - the Budget Approach*. German Advisory Council on Global Change (WBGU), 2009.

M. Webster, M. Babiker, M. Mayer, J. M. Reilly, J. Harnisch, R. Hyman, M. C. Sarofim, and C. Wang. Uncertainty in emissions projections for climate models. *Atmospheric Environment*, 36:3659–3670, 2002.

Martin Weitzman. Prices vs. Quantities. *The Review of Economic Studies*, 41(4):477–491, 1974.

Martin Weitzman. Internalizing the Climate Externality: Can a Uniform Price Commitment Help? *Economics of Energy & Environmental Policy*, 4(2):37–50, 2015. Symposium on 'International Climate Negotiations'.

Roberton C. Williams. Prices vs. Quantities vs. Tradable Quantities. *NBER Working Paper No. 9283*, 2002.

Table 1: Numerical solutions of the game specified in Section 2, in which the coalition acts as a Stackelberg leader in emission choices in the last stage. $N = 100$, $d_1 = 0.1$ and values for d_2, σ and ρ as indicated. ℓ_∞ follows as defined in Eq. (19), k^* is the resulting size of the stable coalition, E[TC_m] the expected total costs of the members and $\sum_{j=1}^{N}$ E[TC_j] the global expected total costs

	Endogenous choice (always quantities)			Restriction to prices			Endogenous choice (always tradable quantities)		
σ	k^*	E[TC_m]	$\sum_{j=1}^{N}$ E[TC_j]	k^*	E[TC_m]	$\sum_{j=1}^{N}$ E[TC_j]	k^*	E[TC_m]	$\sum_{j=1}^{N}$ E[TC_j]
			$d_2 = 0.200$, $\rho = 0.00 \Rightarrow \ell_\infty = Inf$						
0.00	18	0.09	8.83	18	0.09	8.83	18	0.09	8.83
0.15	14	0.29	29.06	18	0.31	31.33	19	0.28	28.88
0.30	5	0.98	95.43	18	0.99	98.83	21	0.94	95.08
			$d_2 = 0.077$, $\rho = 0.20 \Rightarrow \ell_\infty = 30.00$						
0.00	9	0.18	17.92	9	0.18	17.92	9	0.18	17.92
0.15	8	0.37	36.85	11	0.36	35.16	10	0.36	36.78
0.30	3	0.93	93.02	15	0.88	87.38	12	0.90	92.71
			$d_2 = 0.077$, $\rho = 1.00 \Rightarrow \ell_\infty = 6.00$						
0.00	9	0.18	17.92	9	0.18	17.92	9	0.18	17.92
0.15	8	0.23	22.57	8	0.21	22.57	8	0.23	22.57
0.30	8	0.35	35.10	7	0.31	35.47	8	0.35	35.10

Table 2: List of symbols with explanation

$\mathrm{E}[\cdot]$	Expectation value operator
i, j	Indices running over all countries
m, nm	Subscript indicating a member and non-member, respectively
TC	Total costs, sum of damages and abatement costs
D	Damages
C	Abatement costs
d_1, d_2	Damage parameters
ε_i, $\mathrm{E}[\varepsilon]$	Country i's uncertain and expected baseline emissions, respectively
σ, ρ	Parameters characterizing uncertainty in baseline emissions: standard deviation and coefficient of correlation between two distinct countries, respectively
N	Total number of countries
$k, k^*, \bar{k}^*$	Size of the coalition: given at second stage, in equilibrium of the game, in equilibrium of the game in the absence of uncertainty, respectively
e_i	Individual ex-post emissions
e, $e^*(k)$	Level of global emissions and level at optimum for a k-sized coalition, respectively
p_i, p_i^*	Price and optimal price set by regulator, respectively
$\bar{e}_i$, $\bar{e}_i^*$	Quantity and optimal level of quantity set by regulator, respectively
$\tilde{e}_k^*$	Optimal total emission allowances of the coalition in an emissions trading regime
$\mathrm{E}[\cdot]\|_{\bar{e}^*/p^*}$	Expected value under either instrument
Δ, Δ^k	Difference in expected total costs when switching the instrument from prices to quantities for: a single regulator; a coalition of size k, respectively
ℓ, ℓ_{-i}, ℓ_{-k}	Number of: countries with prices; countries but i with prices; non-members to coalition of size k with prices, respectively
ℓ_∞	Maximum number of countries with prices in the no-agreement equilibrium (and exact number for $N \to \infty$)
k_q	For all $k > k_q$ the k-coalition implements quantities in equilibrium

$\Phi(k)$	Stability function: difference in expected total costs of a non-member in the presence of a $(k-1)$-coalition and a member in presence of a (k)-coalition
$\bar{\Phi}(k)$	Stability function under certainty
$\frac{d\bar{e}_j}{d\bar{e}_\mathrm{m}}$, $\frac{d\bar{e}_j}{dp_\mathrm{m}}$	Derivative of emission policy level of non-members with respect to emission policy levels of a member
$\frac{dp_j}{d\bar{e}_\mathrm{m}}$, $\frac{dp_j}{dp_\mathrm{m}}$	Derivative of emission tax level of non-members with respect to emission policy levels of a member