Oxford Economic Papers, 70(4), 2018, 1084–1107
doi: 10.1093/oep/gpy016
Advance Access Publication Date: 8 June 2018

Prices vs quantities for international environmental agreements

By Ulrike Kornek[a] and Robert Marschinski[b]

[a]Mercator Research Institute on Global Commons and Climate Change, EUREF Campus 19, 10829 Berlin, Germany; Potsdam Institute for Climate Impact Research; and Department of Economics of Climate Change, Technische Universität Berlin; e-mail: kornek@mcc-berlin.net
[b]Joint Research Center, European Commission; Potsdam Institute for Climate Impact Research; and Department of Economics of Climate Change, Technische Universität Berlin

Abstract

Does the choice between price-based or quantity-based regulation matter for the formation of an international environmental agreement (IEA)? We introduce abatement cost uncertainty 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. However, uncertainty also aggravates free-riding whenever the endogenous preference of the coalition is to implement quantities, which implies lower equilibrium participation than in the benchmark case without uncertainty. A restriction to price-based agreements can lead to higher participation, but does not necessarily raise global welfare. Tradable quantities can both increase participation in the agreement and achieve higher 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.

JEL classifications: C72, D81, H41, Q54.

1. Introduction

It is one of the principle insights from the theory of self-enforcing international environmental agreements (IEA) that free-riding incentives undermine international cooperation in the presence of public goods (Barrett, 1994). In the field of climate change this is seen as one reason for the thus 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 KyotoProtocol was its specification of precise emission assignments for each eligible country, and also the new Paris agreement mandates developed countries to undertake 'economy-wide absolute emission reduction targets'

(UNFCCC, 2015, Article 4). The focus on 'targets and timetables' (Victor, 2011) follows the earlier examples of the Helsinki and Montreal Protocols and might be motivated by the desire to avoid extreme climate damages. However, in light of the US's 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. Formulating the obligations of members by means of other instruments, such as coordinated emission prices (Cooper, 1998), might be preferable. The present article analyses this question by modelling coalition formation in a global public goods game with uncertainty in countries' payoff function.

We combine two strands of literature. On the one hand, we base our payoff on the prices vs quantities literature. The seminal study of Weitzman (1974) showed that the symmetry between price and quantity regulation breaks down in the presence of uncertainty—a ubiquitous feature in most environmental policy areas, in particular climate change. Weitzman studied the policy choice of a single decision-maker in a general cost-benefit setting, demonstrating that in the presence of uncertain costs a price instrument is socially preferable if marginal costs are relatively steeper than marginal benefits. This finding has generally been reaffirmed in the subsequent literature, which extended the analysis to 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).

On the other hand, we build on the literature on IEAs by adopting a multi-country setting and analysing the two instruments' strategic effect on each countries' incentive to join an IEA, which—to our knowledge—has so far not been addressed in previous studies.[1] The IEA literature typically analyses these incentives by studying the stability of coalitions that provide a global public good (Hoel, 1992; Carraro and Siniscalco, 1993; Barrett, 1994). This literature assumes a payoff consisting of costs from private and benefits from global pollution abatement.

Relatively few authors have extended the basic model to let uncertainty influence the decision to join an IEA. Na and Shin (1998), Ulph (2004), Kolstad and Ulph (2011), and Finus and Pintassilgo (2013) all study the implications of uncertain abatement benefits or benefit-cost ratios, albeit under varying assumptions about the type of uncertainty and learning. Due to the idiosyncrasy of their models, no clear-cut conclusion emerges 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 in an IEA depends on the shape of the probability distribution of the uncertain benefit-parameter.

The present study considers the case when a global public good is supplied by many countries, each having its own sovereign regulator deciding on the regulatory instrument and the participation in an IEA. In contrast to the above cited literature, we investigate the implications of a Weitzman-type asymmetric form of uncertain abatement costs, which affects the regulator(s) at the time of the policy decision, but not the regulated entities at the time of policy implementation (see Section 3.1).

By combining the two strands of literature, we show that the problem of prices versus quantities goes beyond its relevance for the welfare of an individual or fixed group of cooperating countries and bears strategic implications for international cooperation. When an

1 Studies of the strategic implications of instrument choice—but without considering uncertainty—include Endres and Finus (1999) and Eichner and Pethig (2015). For a qualitative discussion see Hepburn (2006).

IEA falls short of implementing the Pareto optimum of full participation, the obtained level of global welfare depends on how much of the externality is internalized, which depends on the size of the coalition of cooperating countries. This paper's objective is to study how the choice between a quantity- or price-based formulation of a treaty influences the number of participating countries in a stable IEA, their internal welfare, and global welfare.

Our analysis indicates that instrument choice influences coalition formation by creating a potential trade-off between the coalition's internal welfare and its size (i.e. participation). Using an analytical approach to determine the equilibrium instrument choice, 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 members, and also globally. However, the preference for quantities at the same time undermines countries' willingness to actually participate in the cooperative effort. While we do not derive a closed-form solution of the equilibrium participation in the agreement, we show analytically that uncertainty aggravates free-riding whenever the coalition endogenously choses a quantity-based treaty. An *ex ante* restriction to only allow agreements based on prices can reduce free-riding incentives, but the welfare effects are ambiguous: as numerical simulations show, either treaty formulation may 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.

Due to the novel character of our model, no evident contradictions between our results and those of previous contributions arise. Within the overall ambiguous findings of the literature on uncertainty and IEA formation, our results on prices vs quantities align with those studies where participation is negatively affected by uncertainty. Our main conclusion echoes the effect of 'modesty' described in Finus and Maus (2008), i.e. the idea that less ambitious agreements—in our case price-based treaties—may be preferable as they counteract incentives to free-ride. In a similar vein, Hong and Karp (2014) find that IEAs with increasingly ambitious abatement—in their case driven by risk aversion to uncertain benefits of emissions—result in lower levels of participation.

The remainder of this article is structured as follows. The model is set up in Section 2. Section 3 analyses equilibrium IEAs under prices vs quantities. In Section 4 we extend our basic model by considering a treaty restricted to prices (4.1) and the option of tradable quantities (4.2). Numerical examples for all our theoretical results are presented in Section 4.3. The final section concludes.

2. The model

Our analysis of self-enforcing IEAs follows the predominant approach of modelling the decision to join a coalition as the first stage in a one-shot cartel-formation game. Following d'Aspremont and Gabszewicz (1986), an equilibrium is characterized by a single coalition that is both internally (no member wants to leave) and externally (no non-member wants to join) stable.[2]

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 (quantities) or an emission tax (prices). As often observed in reality, we assume that the agreement is based

2 Countries are assumed to be represented by a single regulator. We use these terms—countries and regulators—interchangeably.

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, in the latter case with the coalition as the 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.

3rd ('policy level') stage:
Members and non-members choose their level of regulation either simultaneously or sequentially, where in the latter case the coalition acts as a 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. In the following, we will analyse the subgame-perfect equilibria of the game using backwards induction.

This model 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 IEA (first stage)? Third, what are the implications for global welfare, which is a function of participation and instrument choice?

With regard to the payoff function, we choose a formalization as a global public 'bad' game in emissions among $N \geq 2$ *ex ante* symmetric countries, and use a total cost framework as previously employed by Chander and Tulkens (1995), for example. Different from abatement, a formulation in emissions allows for a straightforward formalization of two real-world policy instruments relevant for pollution problems like climate change, namely fixed emission targets and emissions taxes. We assume simple functional forms to derive analytical results that reveal the trade-offs in instrument choice under uncertainty (the generalizability of our results is addressed in the conclusion).

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 (list of symbols available in the online supplementary material):

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

Here, e_i represents emissions of country i, which are themselves a function of its implemented policy and its baseline emissions $\varepsilon_i > 0$. We have $e_i = \varepsilon_i$ if country i adopts no emission policy, and otherwise $e_i \leq \varepsilon_i$. Global emissions are given by $e = \sum_{j=1}^{N} e_j$, and $d_1 > 0$ and $d_2 \geq 0$ are normalized damage parameters.[3]

3 Without uncertainty, the total cost framework can be transformed into the quadratic cost-benefit framework of Barrett (1994) by setting $q_i = \varepsilon_i - e_i$ and relabelling parameters.

To study the question of instrument choice, we introduce uncertainty in the total cost function. Following Weitzman (1974), we only model abatement cost uncertainty, recalling his result that the asymmetry between price and quantity instrument is a consequence solely of uncertainty about abatement costs—at least as long as damage and abatement cost uncertainty are not correlated (Stavins, 1996).

More specifically, our model incorporates abatement cost uncertainty in the specific form of uncertain baseline emissions ε_i, which have been identified as a major driver of abatement cost uncertainty in the area of climate change (Edenhofer *et al.*, 2006; Rogelj *et al.*, 2013). This formalization turns out to be convenient for the analysis of coalition formation and is also fully equivalent to a modelling of abatement cost uncertainty by means of additive shocks on the marginal abatement cost function, as done in Weitzman (1974) and many subsequent studies (e.g. Krysiak, 2008; Ambec and Coria, 2013).

To see this, consider for the cost of abating a quantity q of emissions the following alternative formalization C^W directly taken from Weitzman (1974):

$$C^W(q,\theta) = a(\theta) + (C' + \alpha(\theta))(q - \widehat{q}) + \frac{C''}{2}(q - \widehat{q})^2, \tag{2}$$

where $\widehat{q}$ is the optimal *ex -ante* quantity-based regulation, constants C' and C'' specify, respectively, slope and curvature of the abatement cost function at the optimum, and $a(\theta)$ as well as $\alpha(\theta)$ are some functions of the uncertain parameter θ. If we generalize eq. (2) by moving to a multi-regulator setting, i.e. $q \to q_i$, and use the substitution $q_i = \varepsilon_i - e_i$ to switch to our notation, the formal equivalent of our abatement cost component in eq. (1) can be recovered (see online supplementary material for details).

However, Weitzman derived his specification of net costs from a more general payoff function by using a second-order approximation in the optimum's neighbourhood. Though, in this sense, our payoff function is less general than Weitzman's, it allows to analyse the general trade-offs involved in instrument choice under uncertain abatement costs and extend results from previous literature most directly.[4]

The uncertain baseline emissions ε_i could be distributed according to any of the standard probability density functions. Nevertheless, since total costs are quadratic and countries are *ex ante* symmetric, two parameters are sufficient for a full characterization. First, the standard deviation σ, which defines the common level of uncertainty faced by all countries. Accordingly, setting $\sigma = 0$ recovers the certainty case. Second, the coefficient of correlation $\rho \in [0, 1]$, representing a possible positive correlation of baseline emissions between countries.[5] The parameter ρ captures the relative strength of global factors moving baseline emissions of all countries in the same direction against idiosyncratic local stochasticity.

4 Our total cost payoff function with uncertainty can also be rewritten in the more common cost-benefit form of Barrett (1994). In particular, the payoff $b_1 Q - \frac{b_2}{2} Q^2 - \frac{1}{2} q_i^2 - \eta_i q_i$ with an uncertain parameter η_i leads to the same results as our total cost function when understanding the quantity policy as setting an expected abatement level $q_i = E[\varepsilon] - e_i$ and the price policy in the usual manner of determining *ex post* marginal abatement costs $q_i + \eta_i$.

5 We only consider positive values for ρ, following the argument of Weitzman (1974). However, the case $\rho < 0$ might still be relevant in particular cases, e.g. when modelling emission shocks from a relocation of industries within a two-country model.

In case of strong global interdependence (high ρ), uncontrolled emissions pose a greater risk of high damages because positive shocks will not cancel out, but rather reinforce each other.

3. Stable IEAs under prices vs quantities

This section derives the conditions characterizing an equilibrium of the game specified in the previous section, going from the third to the first stage. We refer to the fully unconstrained game as 'endogenous instrument choice', so as to distinguish it from the one with *ex ante* commitment to price-based treaties analysed in Section 4.1.

3.1 Third stage: expected total costs under prices vs quantities

In this section we solve the game's last stage, in which countries choose the policy level that minimizes their expected total costs. In what follows, we only consider interior solutions. Taken as given from the first stage of the game is the number of members in the agreement, which we denote as k. Also, from the second stage it is known how many non-members implement a price policy, denoted by ℓ_{-k}, and whether the coalition implements prices or quantities. If we denote the total number of countries with price-based policy as ℓ, we have $\ell = \ell_{-k}$ if the coalition implements quantities, and $\ell = \ell_{-k} + k$ if the coalition implements prices.

Formally, the symmetry break-up between prices and quantities arises because regulators have to choose under uncertainty, while firms are able to adjust their emission level after uncertainty is resolved at the end of the game (as in Weitzman, 1974). Firms are modelled implicitly by the abatement costs C_i. With an emission tax p_i imposed by the regulator in the last stage of the game, 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:

$$p_i = -C_i'(e_i) = \varepsilon_i - e_i \quad \Rightarrow \quad e_i = \varepsilon_i - p_i. \tag{3}$$

Equation (3) shows that shocks on baseline emissions are translated one-to-one into shocks on *ex post* emissions e_i in case of a price-based regulation. To fix *ex post* emissions the regulator can adopt a quantity target with an emissions assignment $\bar{e}_i$.[6] However, if at least one regulator chooses a price-based regulation, the amount of global emissions remains uncertain.

The expected total costs of each country i are:

$$\mathrm{E}[TC_i] = d_1 \cdot \mathrm{E}[e] + \frac{d_2}{2} \cdot \mathrm{E}\left[e^2\right] + \frac{1}{2}\mathrm{E}\left[(\varepsilon_i - e_i)^2\right]. \tag{4}$$

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

3.1.1 Emission choice of non-members We first derive the optimal level of regulation for non-members (with subscript 'nm'), who minimize their individual expected total costs. The first-order-condition (FOC) of a representative non-member i with a quantity-based regulation is found 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. (4). To obtain the Nash-equilibrium of the last stage, we take the derivative of

6 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.

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:

$$0 = d_1 + d_2 \cdot \bar{e}_{nm} + d_2 \cdot \mathrm{E}\left[\sum\nolimits_{j \neq i} e_j\right] + \bar{e}_{nm} - \mathrm{E}[\varepsilon], \tag{5}$$

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. (4) 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:

$$0 = -d_1 - d_2 \cdot (\mathrm{E}[\varepsilon] - p_{nm}) - d_2 \cdot \mathrm{E}\left[\sum\nolimits_{j \neq i} e_j\right] + p_{nm}. \tag{6}$$

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

$$\bar{e}_{nm} = \mathrm{E}[\varepsilon] - p_{nm} = \frac{\mathrm{E}[\varepsilon] - d_1 - d_2 \cdot \mathrm{E}\left[\sum_{j \neq i} e_j\right]}{1 + d_2}. \tag{7}$$

In addition, the optimal policy level results to be independent of the level of uncertainty (σ), reflecting the fact that baseline emissions and regulated emissions enter the FOC only linearly.

3.1.2 Emission choice of members Assuming that the coalition comprises the first $1..k$ countries, the members choose the policy level that minimizes the sum of their expected total costs, $\sum_{j=1}^{k} \mathrm{E}\left[TC_j\right]$. Under a quantity-based agreement the FOC of the coalition can be derived by setting the emissions of each member i to $e_i = \bar{e}_m\ \forall i \in \{1..k\}$. Assuming that non-members $(k+1)..\ell_{-k}$ implement prices, total emissions can be written as $e = \sum_{j=1}^{k} \bar{e}_m + \sum_{j=k+1}^{k+\ell_{-k}} \left(\varepsilon_j - p_j\right) + \sum_{j=k+\ell_{-k}+1}^{N} \bar{e}_j$, and be substituted in eq. (4). Taking the derivative with respect to $\bar{e}_m$ gives:

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

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 and derived by taking the derivative of eq. (7) for the case of sequential choices giving $\frac{d\bar{e}_j}{d\bar{e}_m} = -\frac{dp_j}{d\bar{e}_m} = -k\frac{d_2}{d_2(N-k)+1}$.

With a price-based agreement 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 in eq. (4) to

$e = \sum_{j=1}^{k}(\varepsilon_j - p_m) + \sum_{j=k+1}^{k+\ell_{-k}}(\varepsilon_j - p_j) + \sum_{j=k+\ell_{-k}+1}^{N}\bar{e}_j$. Taking the derivative with respect to p_m yields:

$$0 = kd_1 \cdot \left(-k - \sum_{j=k+1}^{k+\ell_{-k}} \frac{dp_j}{dp_m} + \sum_{j=k+\ell_{-k}+1}^{N} \frac{d\bar{e}_j}{dp_m}\right) + kd_2 \cdot \left(kp_m - k\mathrm{E}[\varepsilon] - \sum_{j=k+1}^{k+\ell_{-k}}(\mathrm{E}[\varepsilon] - p_j) - \sum_{j=k+\ell_{-k}+1}^{N}\bar{e}_j\right) \cdot \left(k + \sum_{j=k+1}^{k+\ell_{-k}} \frac{dp_j}{dp_m} - \sum_{j=k+\ell_{-k}+1}^{N} \frac{d\bar{e}_j}{dp_m}\right) + kp_m. \tag{9}$$

First, it can be observed that uncertainty about baseline emissions does not enter the FOC of either quantity or price regulation. Just as for non-members, the optimal policy level is thus independent of the level of uncertainty σ. Second, eqs (8) and (9) coincide when setting

$$\bar{e}_m = \mathrm{E}[\varepsilon] - p_m. \tag{10}$$

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. (7).

In conclusion, the general relationship between the optimal (indicated by '*') price and quantity policy levels follow from eqs (7) and (10) and represents the regulator's version of eq. (3):

$$\bar{e}_i^*(k) = \mathrm{E}[\varepsilon] - p_i^*(k), \ \forall \ i, \tag{11}$$

meaning that for all coalition sizes and both members and non-members, the optimal price and quantity instrument are equivalent in expected terms. The explicit optimal emission assignments $\bar{e}_i^*$ were derived by Barrett (1994) for the case of sequential choices[7] and by Finus and Ruebbelke (2008) for simultaneous choices.[8]

The optimal level of regulation is independent of uncertainty because baseline emissions ε_i do not influence how marginal abatement costs increase with the level of emission policy. Said differently, since marginal costs are linear with a certain slope, every unexpected unit of abatement has a specific cost above or below the mean with the same probability, and so uncertainty does not systematically influence the equalization of expected marginal costs and negative marginal damages. 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).

3.1.3 Equilibrium expected total costs in the last stage Although uncertainty about baseline emissions does not change optimal policy levels, it influences expected total costs and therefore leads to an asymmetry between the quantity- and price-based regulation. Inserting eqs (3) and (11) in the abatement cost function $C_i(e_i)$ shows that implementing a

7 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 $\bar{e}_{nm}^* = \frac{(\mathrm{E}[\varepsilon]-d_1)(d_2(N-k)+1) + d_2\mathrm{E}[\varepsilon]k[k-(d_2(N-k)+1)]}{(d_2(N-k)+1)^2 + k^2 d_2}$.

8 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 $\bar{e}_{nm}^* = \frac{\mathrm{E}[\varepsilon](1-d_2 k + d_2 k^2) - d_1}{d_2(N-k)+1+k^2 d_2}$.

quantity policy leads to an expected abatement costs mark-up of $1/2 \cdot \sigma^2$ vis-à-vis the price policy:

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

$$\mathrm{E}[C_i(\bar{e}^*)]|_{p^*} = \frac{1}{2}\{\mathrm{E}[\varepsilon] - \bar{e}_i^*(k)\}^2, \tag{13}$$

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 ℓ,

$$\mathrm{E}[D(e)] = \underbrace{d_1 \cdot e^*(k) + \frac{d_2}{2} \cdot [e^*(k)]^2}_{\text{certainty term}} + \underbrace{\sigma^2 \cdot \frac{d_2}{2} \cdot [\ell + \ell(\ell - 1)\rho]}_{\text{uncertainty term}}, \tag{14}$$

as can be confirmed by inserting eqs (3) and (11) 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. The first term in eq. (14) 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' uncertainty of baseline emissions is 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 (12), (13) and (14) one obtains:

$$\begin{aligned}\sum_{j=1}^{N} \mathrm{E}[TC_j(k)] = N \cdot \left\{ d_1 \cdot e^*(k) + \frac{d_2}{2} \cdot [e^*(k)]^2 + \sigma^2 \cdot \frac{d_2}{2} \cdot [\ell + \ell(\ell - 1)\rho] \right\} \\ + \sum_{j=1}^{N} \frac{1}{2}\left\{\mathrm{E}[\varepsilon] - \bar{e}_j^*(k)\right\}^2 + (N - \ell) \cdot \frac{1}{2}\sigma^2.\end{aligned} \tag{15}$$

The number of members k of the agreement determines expected global emissions and the certainty equivalent parts of eqs (12), (13), and (14), meaning that the externality will only be internalized and global total costs only approach the social optimum level if participation is sufficiently high. 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 the damage parameter d_2 and the correlation ρ.

If we set the number of countries N to one, eq. (15) 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 quantities to prices. In our model the slope of marginal abatement costs is normalized to one, and thus the price instrument is preferred whenever $d_2 < 1$, i.e. marginal damages are less steep than marginal abatement costs.

3.2 Second stage: 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.

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

$$\Delta = \mathrm{E}[TC_{nm}]|_{p^*} - \mathrm{E}[TC_{nm}]|_{\bar{e}^*} \tag{16}$$

is negative, it will adopt a price regulation. According to eq. (14), Δ depends on the total number of countries other than i with a price policy, which we denote by ℓ_{-i}. Using eqs (12), (13) and (14), and taking into account that the total number ℓ of countries with price policy differs by one in the two terms $\mathrm{E}[TC_{nm}]|_{p^*}$ and $\mathrm{E}[TC_{nm}]|_{\bar{e}^*}$ of eq. (16), the following holds:

Lemma 1 A non-member prefers a price over a quantity policy iff

$$\Delta = \frac{1}{2}\sigma^2\{d_2 \cdot (1 + 2\rho \cdot \ell_{-i}) - 1\} < 0. \tag{17}$$

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 ratio of the slopes of marginal damages and marginal abatement costs (recall that the latter is normalized to one). Third, there is a strategic interaction term shifting the relative advantage toward quantities whenever $\rho > 0$. While the first two reaffirm the well-known standard result, the last effect represents 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.

3.2.2 Instrument choice of the coalition The instrument choice of a coalition of size k is again determined by the difference in expected total costs, which in this case is computed by summing across all members:

$$\Delta^k = \mathrm{E}\left[\sum_{j=1}^{k} TC_m(k)\right]|_{p^*} - \mathrm{E}\left[\sum_{j=1}^{k} TC_m(k)\right]|_{\bar{e}^*}. \tag{18}$$

Considering the instrument choice of the non-members as given, eqs (12), (13), and (14) yield the following lemma:

Lemma 2 A coalition of size k prefers a price over a quantity policy iff

$$\Delta^k = k \cdot \frac{1}{2}\sigma^2\{d_2 \cdot [k + \rho(k^2 + 2k \cdot \ell_{-k} - k)] - 1\} < 0. \tag{19}$$

Again ℓ_{-k} denotes the number of non-members with a price policy. Notice that eqs (17) and (19) are the same if the coalition size is one $k = 1$. Hence the choice between prices and quantities is equal for non-members and a coalition of size one.

Several effects can be identified if $k \geq 2$: 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. 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 a first important insight: coalitions of at least two members 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.

3.2.3 Equilibrium of the second stage By eqs (17) and (19), 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}. To solve for the second stage's equilibrium we must 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. (17) 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 of endogenous instrument choice can support, which from now on we denote by ℓ_∞:

$$\ell_\infty = \begin{cases} \infty & \rho = 0 \text{ and } d_2 < 1 \\ 0 & \rho = 0 \text{ and } d_2 \geq 1 \\ \max\left(\left\lceil \frac{1}{2\rho}\left(\frac{1}{d_2} - 1\right)\right\rceil, 0\right) & \text{else} \end{cases} \quad (20)$$

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. Because the instrument choice for a member of a coalition of size one is governed by the same equation as for non-members, the equilibrium of the second stage is the same in this case, which is summarized in our first result:

***Result 1* (Nash equilibrium in instrument choice for $k \in \{0, 1\}$)** In the absence of a coalition or for a coalition of size one, i.e. $k \in \{0, 1\}$, the number of countries ℓ^* (member or non-members) that implement a price policy is $\ell^* = \min(N, \ell_\infty)$, while the remaining $N - \ell^*$ countries implement a quantity-based regulation.

As the result shows, if $\ell_\infty < N$ (implying $\rho > 0$), a mixed equilibrium is obtained, in which some countries choose prices and other quantities, even though countries are *ex ante* identical. This is a direct consquence of the interaction of policy instrument choice between countries, as implied by Lemma 1.

Next, Appendix 1 shows that the value of ℓ_∞ is also relevant for the equilibrium instrument choice in presence of a coalition of size $k \geq 2$, as summarized in the following result:

Result 2 **(Nash equilibrium of the policy instrument stage for** $k \geq 2$) In equilibrium of the second stage a coalition of size $k \geq 2$ implements quantities as a dominant strategy and $\ell^* = \min(\ell_\infty, N-k)$ non-members implement prices if the coalition's size k is greater than or equal to k_q, with:

1. $\ell_\infty > N$:

$$k_q = \begin{cases} \frac{1}{d_2}, & \rho = 0 \\ N + \frac{1}{2\rho} - \frac{1}{2} - \sqrt{\left(N + \frac{1}{2\rho} - \frac{1}{2}\right)^2 - \frac{1}{d_2\rho}} & \left(N + \frac{1}{2\rho} - \frac{1}{2}\right)^2 \geq \frac{1}{d_2\rho} \ \& \ \rho \neq 0 \\ \infty & \text{else} \end{cases}$$

2. $\ell_\infty \leq N$:

$$k_q = \begin{cases} 3 & \ell_\infty = 2 \ \& \ 2d_2(1+\rho) - 1 < 0 \\ 2 & \text{else} \end{cases}$$

The result shows that the coalition has an increasing incentive to implement quantities as its size grows. Whenever $\ell_\infty > N$, all non-members choose a price regulation, no matter how large the coalition is and what instrument it implements (eq. 17). The preference of the coalition, determined by Δ^k from eq. (19), 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.

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 to mitigate the risk of high damages present when marginal damages are steep (d_2 large) or uncertainties are highly correlated (ρ large).

If $k = 2$, the coalition will implement quantities as a dominant strategy if the combination of d_2 and ρ is high enough, except if ($\ell_\infty = 2$) and ($2d_2(1+\rho) - 1 < 0$) are both met, in which case there exists an equilibrium in which the coalition chooses prices and non-members implement quantities.

3.3 First stage: free-riding 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—or participation—stage of the game. In this section we analyse how the incentive to free-ride is influenced by the presence of uncertainty and endogenous instrument choice.

However, as a closed-form solution for the equilibrium participation in an IEA does not exist for our case of a quadratic payoff function, not even for the simpler case without uncertainty (Barrett, 1994; Finus, 2001), the analysis must be carried out without explicit expressions for the size of the stable coalition(s) and global total costs, and without formal proof of the uniqueness of the equilibrium. This caveat means that we will be able to rigorously determine the conditions under which equilibrium participation in an IEA tends to decrease (or increase) under uncertainty, but not by how much. We will resort to numerical solutions to illustrate the magnitude of effects in Section 4.3.

A stable IEA is characterized by a coalition of size k^* satisfying internal and external stability. Internal stability holds if a member of the k^* coalition has lower expected total costs than a non-member in presence of a coalition of size $k^* - 1$: $\mathrm{E}[TC_{nm}(k^* - 1)] - \mathrm{E}[TC_m(k^*)] \geq 0$. External stability holds if a non-member of the coalition of size k^* has lower (expected) total costs than a member of the enlarged $k^* + 1$ coalition: $\mathrm{E}[TC_m(k^* + 1)] - \mathrm{E}[TC_{nm}(k^*)] > 0$. Both conditions can be verified by the so-called stability function (see Hoel, 1992), defined as 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 $k - 1$:

$$\Phi(k) := \mathrm{E}[TC_{nm}(k-1)] - \mathrm{E}[TC_m(k)]. \tag{21}$$

Due to symmetry, all coalitions of size k with non-negative $\Phi(k)$ are internally stable and those with negative $\Phi(k)$ externally stable. Hence, if the stability function is non-negative at $k^* < N$ and negative at $k^* + 1$, the coalition of size k^* is stable. The grand coalition of all N countries is stable if $\Phi(N) \geq 0$.

In what follows, we use the function $\Phi(k)$ as a continuous indicator of equilibrium participation in an IEA. When higher values for, say, ρ or σ induce a downward shift of $\Phi(k)$, a coalition of size k^* that was stable before the shift loses internal stability if $\Phi(k^*)$ becomes negative. Therefore, whenever a downward shift of $\Phi(k)$ can be observed, we conclude that equilibrium participation in the IEA is weakly decreased or, synonymously, that free-riding is aggravated for all (under certainty) stable coalitions. For lack of a full analytical solution we cannot determine the minimum size of the shift required to decrease equilibrium participation by at least one member, and neither whether the number of stable coalitions changes, which we leave to numerical simulations in Section 4.3.

To compute eq. (21) 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, the following result, derived in Appendix 2, holds:

Result 3 **(Free-riding incentive under uncertainty)** Let $\bar{k}^*$ be the number of countries participating in a stable IEA in the absence of uncertainty. Uncertainty ($\sigma > 0$) increases the incentive to free-ride on this IEA if $k_q + 1 \leq \bar{k}^*$ and $\ell_\infty \geq 1$.

Recall that k_q, from Result 2, represents the minimum coalition size so that a coalition of size k_q or larger implements quantities and $min(\ell_\infty, N)$ the number of countries with prices if the coalition size is zero or one. Result 3 shows that endogenous instrument choice under uncertainty aggravates free-riding whenever three 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.[9] 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. This is the case for the conditions in Result 3.

The uncertainty-related increase of the free-riding incentive is stronger for larger coalition sizes if emission uncertainty is correlated.[10] 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.

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.[11] Our result thus echoes a known insight from studies of public good games (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.

4. Alternatives to quantities for IEAs: prices and tradable quantities

The last section showed that free-riding is enhanced under uncertainty and when the coalition's endogenous choice is to implement quantities. This section analyses two alternative options for formulating the IEA and their impact on free-riding: First, a restriction to price-based agreements (Section 4.1) and, second, quantity agreements with emission trading among the coalition members (Section 4.2). Finally, we also compare all options using numerical examples (Section 4.3).

9 Trivially, for $\ell_\infty > N$ and $3 \leq \bar{k}^* < k_q$ 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 can decrease because the leaving member may incur a loss from the coalition's collective instrument switch.

10 The different components of enhanced free-riding can be identified, e.g., in eq. (29) derived in the proof of Result 3, which quantifies the downward shift of the stability function: $\frac{1}{2}\sigma^2\{d_2[1 + 2\rho(N - k)] - 1\}$.

11 One could conceive that while participation is indeed weakly decreased for coalitions that are stable under certainty, uncertainty might at the same time lead to the emergence of new equilibria with larger coalition sizes. However, Appendix 2 shows that the stability function shifts downward for any $k \geq k^q + 1$. Hence, the size of the largest stable coalition weakly decreases under uncertainty if the conditions of Result 3 hold. As games with quadratic payoff functions, like ours, are generally not superadditive, we nevertheless refrain from depicting the coalition with the largest size as *the* equilibrium.

4.1 Free-riding incentive under an agreement restricted to prices

In the subgame perfect equilibrium derived in Section 3, members do not anticipate the effect of choosing a treaty based on quantities on equilibrium size and expected total costs of stable coalitions, because the instrument choice takes place after the participation decision. 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. As it turns out, this could in fact mitigate the additional free-riding incentive previously identified for the endogenous choice case:

Result 4 **(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 gains from instrument choice vanishes and the stable coalition(s) observed under certainty are restored.

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 4 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 4 suggests that an *ex ante* restriction to a price-based agreement could 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), indicating that this case could indeed be empirically relevant. Of course, without a full solution of the first stage we cannot assert whether the formal conditions of Result 4 are actually met by any stable coalition; however, Section 4.3 will provide a series of numerical examples confirming that the equilibrium participation under an *ex ante* restriction to prices indeed increases for a range of parameter values.

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. (15). 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. (15), 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.

4.2 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 consider an extension of the model addressing uncertain abatement costs: an agreement with tradable quantities. In this approach, each member country still has a quantitative emission assignment. However, firms of member countries can buy and sell emission rights—after uncertainty has been resolved—in an agreement-wide competitive market, allowing them to equalize marginal abatement costs.

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. (14) still holds, whereas expected abatement costs become lower than in the case for non-tradable quantities (eq. [12]), namely:[12]

$$\mathrm{E}[C_m(e)]|_{\tilde{e}_k^*} = \frac{1}{2}\left[\mathrm{E}[\varepsilon] - \bar{e}_m^*(k)\right]^2 + \frac{1}{2}\sigma^2 - \frac{1}{2}\sigma^2\frac{k-1}{k}(1-\rho). \tag{22}$$

The symbol $|_{\tilde{e}_k^*}$ refers to expected costs with tradable emission rights in presence of a k coalition. Hence, this approach is always at least as good as the non-tradable emission assignments analysed before. Based on the last equation, we derive the following result in Appendix 3:

Result 5 **(Free-riding incentive with agreement based on tradable quantities)** When $\rho < 1$ and $k \geq k_q$, tradable quantities lead to a lower free-riding incentive than non-tradable quantities. Uncertainty may result in an increase of equilibrium participation above a $\bar{k}^*$ 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 co-operation. 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.

12 See Williams (2002) for an equivalent equation and its derivation.

Table 1. Numerical solutions of the game specified in Section 2, in which the coalition acts as a Stackelberg leader in the last stage. $N = 100$, $d_1 = 0.1$ and values for d_2, σ and ρ as indicated. ℓ_∞ as defined in eq. (20), k^* is the unique size of the stable coalition, $\mathrm{E}[TC_m]$ the expected total costs of the members and $\sum_{i=1}^{N} \mathrm{E}[TC_i]$ the global expected total costs.

	Endogenous choice (always quantities)			Restriction to prices			Endogenous choice (always tradable quantities)		
σ	k^*	$\mathrm{E}[TC_m]$	$\sum_{i=1}^{N}\mathrm{E}[TC_i]$	k^*	$\mathrm{E}[TC_m]$	$\sum_{i=1}^{N}\mathrm{E}[TC_i]$	k^*	$\mathrm{E}[TC_m]$	$\sum_{i=1}^{N}\mathrm{E}[TC_i]$
$d_2 = 0.200,\ \rho = 0.00 \Rightarrow \ell_\infty = Inf$									
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	0.27	26.83
0.30	5	0.98	95.43	18	0.99	98.83	21	0.80	79.22
$d_2 = 0.077,\ \rho = 0.20 \Rightarrow \ell_\infty = 30.00$									
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	35.73
0.30	3	0.93	93.02	15	0.88	87.38	12	0.90	89.42
$d_2 = 0.077,\ \rho = 1.00 \Rightarrow \ell_\infty = 6.00$									
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	0.24	24.97	8	0.23	22.57
0.30	8	0.35	35.10	7	0.35	39.93	8	0.35	35.10

Source: Authors' calculations.

4.3 Numerical examples

Table 1 presents a series of numerically solved examples to illustrate the quantitative impact of endogenous instrument choice on coalition formation.[13] In this, k^* is the size of the identified stable coalition, i.e. the stability function $\Phi(k)$ in eq. (21) is non-negative for k^* (internal stability) and negative for $k^* + 1$ (external stability). All reported numerical equilibria are unique, i.e. for each set of parameters only one k^* fulfilling the above conditions was found.

$\mathrm{E}[TC_m]$ are the expected total costs of a coalition member, and $\sum_{j=1}^{N} \mathrm{E}[TC_j]$ global expected total costs. Equilibria are computed for various levels of uncertainty σ and values of d_2 and ρ (vertical dimension of table). However, the total set of parameters was in all cases chosen such that the k^* coalition's equilibrium instrument choice would always fall on quantities (as shown in the left column). Prices will then be adopted only 'artificially' if an IEA based on quantities is excluded *ex ante* (shown in the centre column). Conversely, the coalition's equilibrium instrument choice will become tradable quantities if this option becomes available (right column). For the third and last stage of the game, in which the policy level is chosen, we assume sequential emissions decisions with the coalition acting as 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 3.

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 the

13 A detailed description of the numerical algorithm can be found in the online supplementary material.

restriction to a price-based agreement, thus confirming Result 4. However, although the *ex ante* 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 4.1, the extra incentive to join a price-constrained agreement stems from the fact that entry into the coalition allows reducing 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.

It might seem counterintuitive that in the endogenous equilibrium coalition members prefer to implement a treaty based on quantities, despite the fact that with a price-based agreement the coalition would apparently be larger and the total costs of each member lower. However, it has to be kept in mind that the coalitions of 11 and 15 members shown in the centre column are only stable because they are 'artificially' restricted to prices. If one were to lift this restriction, the coalition would actually prefer switching to quantities because this would—when keeping the coalition size and the instrument choice of all other countries constant—decrease their expected costs. By definition of the Nash equilibrium, the coalition does not anticipate that such a switch will first prompt some non-members to switch from quantities to prices, and then some coalition members to drop out, eventually leading to the endogenous equilibria of eight and three member-coalitions shown in the left column.

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 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 imposed uncooperativeness, which creates an additional incentive not to join the agreement.

Finally, when comparing 'prices vs tradable quantities', and considering the case in which the coalition's equilibrium choice—as by our chosen parameter values—falls on the latter (right column), the outcome always improves upon the case without trade (left column), except if $\rho = 1$, when it is equal. In the most favourable case, i.e. with $\rho = 0$, participation becomes higher with than without uncertainty, confirming the theoretical possibility indicated by Result 5. A treaty implementing emissions trading is superior to the other design options both in terms of participation and the coalition's as well as global expected total costs.

If correlation increases to $\rho = 0.2$, the endogenously chosen tradable quantities still lead to a higher level of participation than in the certainty case, but lower than for a treaty restricted to prices. The latter is now also preferable from a total costs perspective, showing how the benefits of emissions trading rely quite strongly on a low level of correlation between emission shocks.

5. 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) for the emission reductions of participating countries. Also the Paris agreement resorts to this approach for developed countries. As shown by Weitzman (1974) and others, an alternative formulation based on emission taxes might enhance the welfare of cooperating countries when abatement costs 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: on the one hand a treaty based on quantities could be socially preferable to one based on prices, but on the other hand it could also increase the free-riding incentive.

The choice between a quantity or price instrument is generally 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 towards quantities.

Non-cooperating countries in principle face the same conflict between damages and abatement costs as the coalition. 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 generally alleviates the trade-off between instruments for free-riders when fixing its own emissions and reducing global damage uncertainty. Hence, coalitions that choose quantities can make free-riding more attractive.

We demonstrate that an agreement restricted to prices can decrease 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, and (ii) the uncertainty about damages faced by each country. If prices are individually preferred, a price-based agreement can ensure higher participation, which likely increases the internalization of the emission externality. However, by excluding quantities for cooperating countries, global emissions uncertainty can be higher than for a quantity-based agreement with lower participation, which increases each country’s expected damages. Any one of the two effects might dominate. Our analysis suggests that this trade-off could be alleviated if members create an emissions trading scheme, in which case uncertainty could even amplify the incentive to cooperate.

Although our results are derived within a specific theoretical model of reduced complexity, the basic mechanisms carry over 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. Therefore, an agreement formulated in quantities will benefit free-riders by reducing damage uncertainty (and hence expected damages) and increasing—relative to a price formulation—expected compliance costs for coalition members.

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 binding emission targets. Such targets could result in high abatement costs, which countries seek to avoid. 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).

Acknowledgements

The views expressed are purely those of the authors and may not in any circumstances be regarded as stating an official position of the European Commission. We would like to thank Jobst Heitzig, Patrick Doupe, Martin Weitzman and two anonymous reviewers for very valuable comments.

Funding

This work was supported by the German Federal Ministry for Education and Research (Bundesministerium für Bildung und Forschung) [grant number 01LA1121A].

Supplementary material

Supplementary material is available online at the OUP website. This material consists of an online appendix.

References

Ambec, S. and Coria, J. (2013) Prices vs quantities with multiple pollutants, *Journal of Environmental Economics and Management*, **66**, 123–40.

Barrett, S. (1994) Self-enforcing international environmental agreements, *Oxford Economic Papers*, **46**, 878–94.

Carraro, C. and Siniscalco, D. (1993) Strategies for the international protection of the environment, *Journal of Public Economics*, **52**, 309–28.

Chander, P. and Tulkens, H. (1995) A core-theoretic solution for the design of cooperative agreements on transfrontier pollution, *International Tax and Public Finance*, **2**, 279–93.

Cooper, R. (1998) Toward a real treaty on global warming, *Foreign Affairs*, **77**, 66–79.

d'Aspremont, C. and Gabszewicz, J.J. (1986). On the stability of collusion, in J.E. Stiglitz and G.F. Mathewson (eds) *New Developments in the Analysis of Market Structures*, Palgrave Macmillan, London, 243–64.

Edenhofer, O., Kemfert, C., Lessman, K., Grubb, M., and Koehler, J. (2006) Induced technological change: exploring its implications for the economics of atmospheric stabilization: synthesis report from the innovation modeling comparison project, *The Energy Journal*, **27**, 57–107.

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

Endres, A. and Finus, M. (1999) International environmental agreements: how the, policy instrument affects equilibrium emissions and welfare, *Journal of Institutional and Theoretical Economics*, **155**, 527–50.

Finus, M. (2001) *Game Theory and International Environmental Cooperation*, Edward Elgar, Cheltenham.

Finus, M. and Maus, S. (2008) Modesty may pay! *Journal of Public Economic Theory*, **10**, 801–26.

Finus, M. and Pintassilgo, P. (2013) The role of uncertainty and learning for the success of international climate agreements, *Journal of Public Economics*, **103**, 29–43.

Finus, M. and Ruebbelke, D. (2008) Coalition Formation and the Ancillary Benefits of Climate Policy. FEEM Working Paper, Nb. 62.2008, Milan.

Hepburn, C. (2006) Regulation by prices, quantities, or both: a review of instrument choice, *Oxford Review of Economic Policy*, **22**, 226–47.

Hoel, M. (1992) International environment conventions: the case of uniform reductions of emissions, *Environmental and Resource Economics*, **2**, 141–59.

Hong, F. and Karp, L. (2014) International environmental agreements with endogenous or exogenous risk, *Journal of the Association of Environmental and Resource Economists*, **1**, 365–94.

Karp, L. (2012) The effect of learning on the membership and welfare in an international environmental agreement, *Climatic Change*, **110**, 499–505.

Kolstad, C.D. and Ulph, A. (2011) Uncertainty, learning and heterogeneity in international environmental agreements, *Environmental and Resource Economics*, **50**, 389–403.

Krysiak, F.C. (2008) Prices vs quantities: the effects on technology choice, *Journal of Public Economics*, **92**, 1275–87.

Lessmann, K., Marschinski, R., Finus, M., Kornek, U., and Edenhofer, O. (2014) Emissions trading with non-signatories in a climate agreement- an analysis of coalition stability, *The Manchester School*, **82**, 82–109.

Na, S.-l. and Shin, H.S. (1998) International environmental agreements under uncertainty, *Oxford Economic Papers*, **50**, 173–85.

Newell, R.G. and Pizer, W.A. (2003) Regulating stock externalities under uncertainty, *Journal of Environmental Economics and Management*, **45**, 416–32.

Pizer, W.A. (2002) Combining price and quantity controls to mitigate global climate change, *Journal of Public Economics*, **85**, 409–34.

Rogelj, J., McCollum, D.L., Reisinger, A., Meinshausen, M., and Riahi, K. (2013) Probalistic cost estimates for climate change mitigation, *Nature*, **493**, 79–83.

Stavins, R.N. (1996) Correlated uncertainty and policy instrument choice, *Journal of Environmental Economics and Management*, **30**, 218–32.

Ulph, A. (2004) Stable international environmental agreements with a stock pollutant, uncertainty and learning, *The Journal of Risk and Uncertainty*, **29**, 53–73.

UNFCCC (2015) The Paris agreement, http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf (accessed 12 March 2018).

Victor, D. (2011). *Global Warming Gridlock: Creating More Effective Strategies for Protecting the Planet*, Cambridge University Press, Cambridge.

Weitzman, M. (1974) Prices vs. quantities, *The Review of Economic Studies*, **41**, 477–91.

Weitzman, M. (2015) Internalizing the climate externality: can a uniform price commitment help? *Economics of Energy & Environmental Policy*, **4**, 37–50.

Williams, R.C. (2002) Prices vs. Quantities vs. Tradable Quantities. NBER Working Paper No. 9283, Cambridge, MA.

Appendix 1: Derivation of Result 2

Consider first the behaviour of non-members in the second stage. A non-member will switch to a price-based regulation until the number of countries with prices reaches $\ell = \min[\ell_\infty, N]$, irrespective of whether the coalition or other non-members implement prices (Result 1). Hence, if the coalition implements quantities, $\ell_{-k} = \min[\ell_\infty, N-k]$ non-members will implement prices. If the coalition chooses prices, $\ell_{-k} = \max[0, \min[\ell_\infty - k, N-k]]$ non-members implement prices. These rules will be used when deriving the equilibrium behaviour of non-members and of the coalition.

In the first case of Result 2, $\ell_\infty > N$, non-members always implement prices. Hence, $\ell_{-k} = N-k$ non-members choose prices. Inserting this in eq. (19), the decision of the coalition to switch from prices to quantities as k increases hinges on Δ^k becoming positive:

$$\Delta^k = \frac{1}{2}\sigma^2\left\{d_2 \cdot \left[\underbrace{k + \rho\big(k^2 + 2k\cdot(N-k) - k\big)}_{=k+\rho(2Nk-k^2-k)}\right] - 1\right\}k \geq 0 \tag{23}$$

We determine when this inequality switches sign by setting it to zero and solving for its roots. For $\rho = 0$, the equality is linear and can be directly solved, thus concluding the first case. For $\rho > 0$, the equality becomes a quadratic equation, for which the lesser root identifies the switching point in case it is real. Solutions do not exist if the roots are not real, in which case the coalition will always choose prices (last case).

If $\ell_\infty \leq N$, it follows from eq. (17) that $d_2(1 + 2\rho\ell_\infty) - 1 \geq 0$ (*). We now check whether the coalition has an incentive to switch its instrument choice from prices to quantities, given the choices of non-members, thus defining the condition under which it will implement quantities in the second stage equilibrium. This holds because if the incentive to switch to quantities is positive under a price-based agreement (defined by eq. 19), then also the incentive to stay with quantities is positive under a quantity-based agreement, given that non-members weakly increase their number of price-based policies as a reaction to the coalition's switch to quantities.

Consider first the case with $k > \ell_\infty$. If we assume that all members would implement prices, none of the non-members will do so, i.e. $\ell_{-k} = 0$. Inserting this to determine the expected total cost difference Δ^k, eq. (19), gives:

$$\Delta^k = \frac{1}{2}\sigma^2\left\{d_2 \cdot \left[k + \rho \cdot \underbrace{(k^2 - k)}_{\geq(\ell_\infty+1)^2-(\ell_\infty+1)\geq 2\ell_\infty}\right] - 1\right\}k. \tag{24}$$

By (*) this is always greater than or equal to zero: the coalition switches to quantities.

Second, if $k \leq \ell_\infty$, $\ell_{-k} = \ell_\infty - k$ non-members would implement prices if the coalition implements prices. Again, inserting into eq. (19) and with the help of (*), the expected total cost difference:

$$\Delta^k = \frac{1}{2}\sigma^2\left\{d_2 \cdot \left[k + \rho\big(k^2 + 2k(\ell_\infty - k) - k\big)\right] - 1\right\}k \tag{25}$$

$$= \frac{1}{2}\sigma^2 \left\{ d_2 \cdot \left[k + 2\rho\ell_\infty + \rho \underbrace{\left(2\ell_\infty(k-1) - k^2 - k\right)}_{\geq 0 \forall \ell_\infty \geq 3 \text{or} k \geq 3} \right] - 1 \right\} k \tag{26}$$

is always greater than or equal to zero $\forall\, k \geq 2$ if $\ell_\infty \geq 3$. In case $\ell_\infty = 2$ we already know from above that a coalition of size $k = 3$ will implement quantities (case $k > \ell_\infty$).

The only exception we have to consider is $k = 2$ and $\ell_\infty = 2$. In this case the total expected cost difference is positive if:

$$\Delta^k = \frac{1}{2}\sigma^2\{d_2 \cdot [2 + 2\rho] - 1\}2 \geq 0 \tag{27}$$

$$\iff 2d_2(1+\rho) - 1 \geq 0 \tag{28}$$

Hence only if $2d_2(1+\rho) - 1 < 0$ does the coalition not switch to quantities when all non-members implement quantities, which is therefore an equilibrium of the game.

Appendix 2: Derivation of Result 3

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. (17). When both the coalitions of size k and $k - 1$ implement quantities ($k \geq k_q + 1$, such that any coalition at least as large as k will implement quantities), a member will increase the number of countries implementing prices by one when leaving the coalition. The stability function Φ in eq. (21) is:

$$\Phi(k) = \bar{\Phi}(k) + \underbrace{\frac{1}{2}\sigma^2\{d_2 \cdot [1 + 2\rho(N-k)] - 1\}}_{<0}, \tag{29}$$

where $\bar{\Phi}(k)$ is the stability function under certainty. Using (**), the uncertainty-related second term is negative and free-riding is aggravated.

For the case of $\ell_\infty \leq N$, eq. (17) implies $d_2[1 + 2\rho(\ell_\infty - 1)] - 1 < 0$ (***). Again, both coalitions of sizes k and $k - 1$ implement quantities. Hence, if $k + \ell_\infty \leq N$, ℓ_∞ non-members will implement prices. The total number of countries implementing prices then stays the same for coalitions of size k and $k - 1$. The stability function reflects the fact that the non-member has a chance to switch to prices when becoming a free-rider while damage increases are fixed:

$$\Phi(k) = \bar{\Phi}(k) - \frac{1}{2}\sigma^2\left\{\frac{\ell_\infty}{N - k + 1}\right\}. \tag{30}$$

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 stability function is:

$$\Phi(k) = \bar{\Phi}(k) + \underbrace{\frac{1}{2}\sigma^2\left\{ d_2 \cdot \left[1 + 2\rho \underbrace{(N-k)}_{\leq \ell_\infty - 1} \right] - 1 \right\}}_{<0}. \tag{31}$$

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

Appendix 3: Derivation of Result 5

The first part of this result was already derived non-formally in the corresponding section. The second part can be demonstrated when assuming $\rho = 0$. The stability function is equal to the difference in expected costs in eq. (29) plus the reduction in expected abatement costs due to trading in eq. (22), $\frac{\sigma^2}{2}\left(\frac{k-1}{k}\right)$, giving:

$$\Phi(k) = \bar{\Phi}(k) + \frac{\sigma^2}{2}\left(d_2 - 1 + \frac{k-1}{k}\right). \tag{32}$$

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