Helm, Carsten; Wirl, Franz

Working Paper
Climate policies with private information: The case for unilateral action

Oldenburg Discussion Papers in Economics, No. V-378-15

Provided in Cooperation with:
University of Oldenburg, Department of Economics

Suggested Citation: Helm, Carsten; Wirl, Franz (2015) : Climate policies with private information: The case for unilateral action, Oldenburg Discussion Papers in Economics, No. V-378-15, University of Oldenburg, Department of Economics, Oldenburg

This Version is available at:
http://hdl.handle.net/10419/113269

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.
You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.
If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
Climate policies with private information: The case for unilateral action

Carsten Helm
Franz Wirl

V – 378– 15

April 2015

Department of Economics
University of Oldenburg, D-26111 Oldenburg
Climate policies with private information: The case for unilateral action

Carsten Helm* and Franz Wirl†

April 28, 2015

Abstract

Countries often have private information about their willingness to pay for protecting the climate system and their cost of emission reductions. We use a principal-agent model to re-examine the economic case for unilateral action by individual countries, in our case of the principal. We find that the incentive structure that arises in an incomplete information framework can motivate (i) unilateral action before contract negotiations, (ii) optimal contracts in which the principal accepts higher marginal abatement costs for herself, as well as (iii) overcompliance by the principal after the contract has been negotiated. Multilateral externalities and type-dependent outside options, which are characteristic for climate policies, play a crucial role to explain these results.

Keywords: unilateral action, voluntary action, unilateral commitment, private information, multilateral externalities, international environmental agreements, type-dependent outside options.

JEL: D82, Q54, H87

*Corresponding author: University of Oldenburg, Department of Economics and Law, 26111 Oldenburg, Germany, phone +49 441 798-4113, fax +49 441 798-4116, carsten.helm@uni.oldenburg.de.
†University of Vienna, Faculty of Business, Economics and Statistics, Oskar Morgenstern Platz 1, Room 4.635, A- 1090 Wien, AUSTRIA. Email: franz.wirl@univie.ac.at, Tel: +43 1 4277 38101.
1 Introduction

Some policy makers and, in particular, many environmentalist groups advocate unilateral actions by developed countries in the context of global warming. This idea has also been codified in Article 3 of the United Nations Framework Convention of Climate Change (UNFCCC), which stipulates that "the developed country Parties should take the lead in combating climate change". Often this recommendation is based on equity considerations, referring to countries' "common but differentiated responsibilities and respective capabilities" (Art. 3, UNFCCC). By contrast, economists have usually been reserved, arguing that unilateral action by some countries may trigger less action by others so that aggregate emissions may even increase. This paper abstracts from equity motives for unilateral action. Instead, we re-examine the economic case against it for an environment where parties have private information about their costs and benefits of emission abatement.

It is widely accepted that climate change causes substantial impacts on natural and human systems, and that successful climate policies require a major overhaul of existing energy systems (IPCC 2014). However, the associated costs depend on regional patterns of climate change and technology developments, both of which are still uncertain. The perception of these costs by relevant societal groups and by voters, as well as their willingness to invest in such a long-term project, is of major importance for politicians that negotiate a climate contract. Moreover, politicians are usually better informed about these aspects in their home country, and they may also have their own agenda. As a consequence, negotiators are often uncertain about the objective function of their counterparts. This relevance of private information for negotiating a climate contract has been emphasized in some recent contributions (e.g., Martimort and Sand-Zantman (2015), Martimort and Sand-Zantman (2013), Konrad and Thum (2014), Helm and Wirl (2014)), and it motivates our methodological approach to analyze the consequences unilateral action.

In a seminal contribution, Hoel (1991) models unilateral action as a commitment to do more than in the non-cooperative solution if contract negotiations – modelled as Nash-bargaining without side-payments – fail. In this framework, unilateral action by player 1 (the North) worsens its disagreement point and, conversely, improves the bargaining position of the other player (the South). Consequently, the South can capture a larger share of the bargaining rent. In the absence of transfer payments, this is implemented by higher abatement of the North and less abatement of the South. If marginal abatement cost of the North are relatively high, then its additional abatement cannot compensate the lower abatement of the South and aggregate emissions rise due to unilateral action.

In the first part of this paper, we re-examine these effects of unilateral commitment for an environment with private information about the damages of climate change, respectively about the willingness to pay (WTP) to avoid them. We assume take-it-or-leave-it offers in order to avoid some of the complexities that arise in bargaining with incomplete information (see, e.g., Kennan and Wilson (1993)). In addition, we assume that only one of the parties – the South – has private information. While this is certainly a simplification, it reflects that information about the costs and benefits of emission abatement, and also about policy preferences, are probably better for industrialized countries because more scientific studies exist and political processes are often more transparent. This allows us to use the principal-agent model, where we assign the North to be the principal (she) that proposes a climate
change contract to the South (the agent, he).\footnote{1}

By the revelation principle, we can restrict the analysis to contracts in which it is incentive compatible for the agent to truthfully reveal his type. The agent has an incentive to underreport his WTP for emission abatement so as to pretend a better outside option.\footnote{2} To counteract this, types with a high WTP receive an information rent, and the principal distorts the emissions scheme so as to reduce this rent. Specifically, the principal raises her own and the agent’s emissions above the efficient level if a low type is reported. This makes underreporting less attractive, but has the effect that aggregate contract emissions for low types may exceed out-of-contract emissions, i.e. those at the disagreement point (see Helm and Wirl (2014)). For such types the principal offers a boundary contract for which aggregate emissions are the same as in the out-of-contract solution. These are reduced by the principal’s unilateral commitment. Hence it has a positive effect on aggregate emission reductions, in contrast to Hoel (1991).

In the second part of the paper, we abstract from the principal’s unilateral commitment. Instead, we assume that the agent has not only private information about his WTP for emissions abatement, but also about the associated abatement costs. We will show that this leads to optimal contracts with unilateral action in the sense that the principal bears higher marginal abatement costs than the agent. This result obtains even if both parties have the same abatement cost functions and the same WTP for emissions abatement; hence it is not driven by differences in the parties’ preferences. Rather, unilateral action is a consequence of the optimal incentive structure in the incomplete information framework with multilateral externalities. The principal offers to shoulder a higher abatement burden in order to incentivize the agent to honestly reveal that he has a high WTP for abatement and/or low abatement costs. Since the optimal contract specifies higher emission reductions for such types, one may equivalently interpret the principal’s unilateral action as an instrument for convincing the agent to accept higher emission reductions.

In the third part of the paper, we show that the principal may have a further motive for unilateral action; this time after the contract has been signed. A crucial aspect in which the application of the principal-agent model to climate change differs from standard applications is the presence of multilateral externalities. This enables the principal to use not only the usual instrument of subsidies to incentivize the agent, but also her own emissions. Specifically, the principal distorts her emissions upwards – as compared to the efficient solution – in order to make it less attractive for agents with a high WTP for abatement and/or with low abatement costs to misrepresent their type. Often this leads to emissions of the principal that even exceed their out-of-contract level, which implies marginal abatement cost below her marginal WTP for abatement. Therefore, the principal has an incentive to overfulfill the terms of the contract, i.e. to undertake unilateral action. Importantly, this action does not depend on the fact that the agent has revealed his type through the contract, but simply arises from the principal’s comparison of her own marginal abatement and damage costs.

Starting with Hoel (1991), several papers have analyzed the relation between unilateral action and international environmental agreements. Of these, Konrad and Thum (2014)
is probably most closely related to our work because it also uses an incomplete information framework, pointing out that "one of the main reasons for the breakdown of efficient bargaining is asymmetric information" (p. 244). In Konrad and Thum (2014) the offering party makes a single take-it-or-leave-it offer that specifies emissions of the other country and transfer payments. Since the offering party does not know the other country’s abatement costs (only the distribution), it faces a trade-off: it prefers to pay less, but this raises the probability that the offer will be rejected. Now, when the offering party unilaterally commits to reduce own emissions to an efficient level, the stakes of reaching an agreement are lower. Hence, the offering party puts less weight on the probability that the offer will be accepted, and contract negotiations fail more often as a consequence of unilateral action. The main conceptual difference in our paper is that the offering party is not restricted to a single offer, which is of course a strong simplification of negotiations about a climate contract. Instead, the principal proposes a whole menu consisting of emission profiles and associated transfer payments. What might seem a minor detail, reverses the result in Konrad and Thum (2014) on unilateral action; mainly because optimal contracts are designed such that the participation constraint is always satisfied.

Likewise assuming an incomplete information framework, some studies have analyzed unilateral action as a signaling device (Brandt 2004; Espinola-Arredondo and Munoz-Garcia 2011; Brandt and Nannerup 2013). For example, Brandt (2004) considers a situation where one country obtains private information that its abatement costs are low. If costs are highly correlated, then also the other country would have an interest to abate more. However, before doing so it would need to be convinced that costs are indeed low. Unilateral emission reductions may constitute a credible signal for this.

Aside from the above contributions, most papers on the effects of unilateral action have assumed complete information and focus on carbon leakage. In line with Hoel (1991), this literature warns that unilateral emission reductions may trigger higher emissions elsewhere (or in other periods), caused by lower energy prices, terms-of-trade effects and intertemporal changes in extraction paths (Eichner and Pethig 2011; Böhringer, Lange, and Rutherford 2014). Naturally, such effects are absent in our static, partial equilibrium model. Nevertheless, if our principal-agent model were extended in this direction, the effects are likely to be small because the principal would account for carbon leakage when setting the optimal contract terms. A more positive assessment of unilateral action has been obtained by papers that focus on R&D and technologies for greenhouse gas mitigation. If these are characterized by substantial spillovers, then technology driven abatement by some countries reduces abatement costs of other countries as well and, therefore, may induce them to raise their abatement level (Golombek and Hoel 2004; Bosetti and Cian 2013).

The remainder of the text is structured as follows. In Section 2, we generalize the principal-agent model of Helm and Wirl (2014) by taking into account that countries may have private information about both, costs and benefits of emissions. This serves as the starting point from which we discuss different reasons for unilateral action that arise in an environment with asymmetric information. In Section 3, we analyze the effects of unilateral action before contract negotiations. Next, we show that optimal contracts will often be characterized by unilateral action of the principal, meaning that she accepts higher marginal abatement costs (Section 4). Finally, we consider unilateral action after contract negotiations (Section 5).
2 Framework

Consider a principal (indexed 1) and an agent (indexed 2). Reflecting our focus on global environmental problems like climate change, we often refer to the principal as the group of industrialized countries and to the agent as the group of developing countries. For both, payoffs depend on a benefit term, $B_i(x_i)$, that is a function of own emissions, $x_i \in \mathbb{R}_+$, and on a cost term, $D(X)$, that is a function of aggregate emissions, $X := x_1 + x_2$. We assume that $B_i(x_i)$ is an increasing and strictly concave function that satisfies the Inada conditions, while $D(X)$ is strictly increasing and linear in $X$ with $D(0) = 0$.

In addition, there is a region-specific parameter $\theta_i > 0$ that affects benefits and costs of emissions. Specifically, payoffs are given by

$$V_i = \alpha [B_i(x_i) - \theta_i D(X)] + (1 - \alpha) \left[ \frac{B_i(x_i)}{\theta_i} - D(X) \right], \quad i = 1, 2,$$

where $\alpha \in [0, 1]$. Accordingly, the payoff function is a convex combination of the cases where $\theta_i$ affects the benefits of emissions (as in Helm and Wirl (2014)), or the costs (as in Martimort and Sand-Zantman (2015)). Observe that a higher $\theta_i$ always reduces the payoff, independent of $\alpha$.

In line with the principal-agent framework, we assume that $\theta_1$ is common knowledge, while $\theta_2$ is the agent’s private information. Specifically, $\theta_2$ is a random variable with a known distribution: $f$ denotes the density with support $[\theta_2, \theta_2]$, $F$ the cumulative distribution function.

Helm and Wirl (2014) have analyzed the case $\alpha = 1$, for which payoffs are

$$\tilde{V}_i := B_i(x_i) - \theta_i D(X).$$

Hence private information are restricted to climate damages or, more generally, the willingness to pay (WTP) for greenhouse gas emission abatement. Specification (1) generalizes this by allowing different allocations of private information to the benefit and damage component of the agent’s payoff function. Rearranging terms, (1) can be written equivalently as

$$V_i = \frac{1 - \alpha (1 - \theta_i)}{\theta_i} \tilde{V}_i.$$  

Accordingly, $V_i$ is a positive affine transformation of $\tilde{V}_i$. It is well known that the Nash equilibrium is not affected by this transformation and, therefore, does not depend on $\alpha$.

---

3The assumption that $D(X)$ is linear is widely used in the literature (e.g., Barrett (2006)) and substantially simplifies the analysis. It is sometimes criticized on the grounds that linear damage cost functions neglect incentives to free ride on the environmental benefits of cooperation. However, this problem is less severe in our set-up with binding contracts. Moreover, Finus, Ierland, and Dellink (2006) show that a linear specification can also be justified for substantive reasons since discounted climate damages that are linear in emissions are a good approximation of the figures in the RICE model (Nordhaus and Yang 1996), although damages are non-linear in temperature change.
Accordingly, emissions in this solution (indicated by superscript 0) are

\[ x_1^0(\theta_1) = \arg \max_{x_1} B_1(x_1) - \theta_1 \int_{\theta_2} D(x_1 + x_2^0(\theta_2)) \, dF(\theta_2), \]  

\[ x_2^0(\theta_2) = \arg \max_{x_2} B_2(x_2) - \theta_2 D(x_1^0(\theta_1) + x_2). \]  

The choice of \( x_1^0(\theta_1) \) reflects that the principal does not know the agent’s valuation and, therefore, maximizes her (expected) payoff over all possible realizations of \( \theta_2 \), while the agent has complete information.

First-best emissions (indicated by superscript 1) maximize \( V_1 + V_2 \) and, thus, solve

\[ \frac{1 - \alpha (1 - \theta_i)}{\theta_i} B_i'(x_i) - [2 - \alpha (2 - \theta_1 - \theta_2)] D'(X) = 0, \quad i = 1, 2. \]  

Next, consider the situation where the principal can present to the agent a contract \( \{ x_1(\theta_2), x_2(\theta_2), s(\theta_2), \theta_2 \in [\theta_2, \bar{\theta}_2] \} \), specifying emissions and transfers \( s(\theta_2) \in \mathbb{R} \) from the principal to the agent. Denote by

\[ U(\hat{\theta}_2, \theta_2) := \big[ 1 - \alpha (1 - \theta_2) \big] \left[ \frac{1}{\theta_2} B_2(x_2(\hat{\theta}_2)) - D(X(\hat{\theta}_2)) \right] + s(\hat{\theta}_2) \]  

the payoff of a type \( \theta_2 \) who pretends to be of type \( \hat{\theta}_2 \), and by

\[ R(\theta_2) := \max_{x_2} \big[ 1 - \alpha (1 - \theta_2) \big] \left[ \frac{1}{\theta_2} B_2(x_2) - D(x_1^0 + x_2) \right] \]  

the agent’s out-of-contract payoff, i.e. the Nash equilibrium outcome (see above). Using the revelation principle, the principal’s problem can be stated as choosing a contract \( \{ x_1(\theta_2), x_2(\theta_2), s(\theta_2), \theta_2 \in [\theta_2, \bar{\theta}_2] \} \) that maximizes her expected payoff, subject to the agent’s incentive and participation constraint:

\[ \max_{\theta_2} \int_{\theta_2} \left[ 1 - \alpha (1 - \theta_1) \right] \left( \frac{B_1(x_1(\theta_2))}{\theta_1} - D(X(\theta_2)) \right) - s(\theta_2) \, dF(\theta_2), \text{ s.t.} \]  

\[ \forall \theta_2 \in [\theta_2, \bar{\theta}_2] : \quad \theta_2 = \arg \max_{\theta_2} U(\hat{\theta}_2, \theta_2), \]  

\[ \forall \theta_2 \in [\theta_2, \bar{\theta}_2] : \quad U(\theta_2) := U(\theta_2, \theta_2) \geq R(\theta_2). \]  

Observe that the principal can choose her own emissions besides subsidies to incentivize the agent, which is a consequence of the multilateral externalities. Moreover, the agent’s outside option depends on his type (Maggi and Rodriguez-Clare 1995). One consequence of this is that the agent has an incentive to understate his type so as to pretend a better outside option (remember that payoffs are decreasing in \( \theta_1 \)). Therefore, the contract binds from below. Further consequences for the optimal contract are discussed in Helm and Wirz (2014), which analyzed basically the same problem, but only for private information about
damages ($\alpha = 1$). Extending this to the case of private information about benefits ($\alpha < 1$), we obtain the following lemma.

**Lemma 1** The optimal contract has the following properties:

(i) The contract binds at the lowest type, i.e. $U(\theta_2) = R(\theta_2)$.

(ii) Emissions of the principal and the agent in the optimal contract satisfy (arguments are dropped for ease of notation)

$$\frac{1 - \alpha (1 - \theta_1)}{\theta_1} B_1' - [2 - \alpha (2 - \theta_1 - \theta_2)] D' = \frac{F - 1 + \mu}{f} \alpha D', \quad (12)$$

$$\frac{1 - \alpha (1 - \theta_2)}{\theta_2} B_2' - [2 - \alpha (2 - \theta_1 - \theta_2)] D' = \frac{F - 1 + \mu}{f} \left[ \frac{1 - \alpha}{\theta_2^2} B_2' + \alpha D' \right], \quad (13)$$

where $\mu(\theta_2)$ is the Lagrangian multiplier of the participation constraint (11).

(iii) Except for the highest type, $\bar{\theta}_2$, emissions of the principal and of the agent are below their efficient level.

**Proof.** See appendix. ■

### 3 Unilateral action before contracting

We now examine different motives for unilateral action and its implications. In this section, we follow the approach in Hoel (1991). He used the Nash bargaining solution to model the outcome of contract negotiations and analyzed the effects of a unilateral commitment to reduce emissions by more than in the Nash equilibrium should negotiations fail. Thus, the unilateral commitment worsens the disagreement point. We extend this analysis by skipping the assumption that cost and benefits of climate change are common knowledge. However, the introduction of private information precludes the Nash bargaining solution (NBS). Therefore, we assume that one party holds the complete bargaining power. This allows us to use the principal-agent model as introduced above.

The next subsection determines the disagreement points. It also shows that the main arguments in Hoel (1991) are not affected by the extreme allocation of bargaining in the principal-agent model. By contrast, private information fundamentally changes the results, as we show in subsection 3.2.

#### 3.1 Unilateral commitment and the disagreement point

Consider two countries with payoff functions as specified in Section 2. Nash equilibrium emissions of the agent are given by (5) and satisfy

$$B_2'(x_2) - \theta_2 D'(x_1 + x_2) = 0. \quad (14)$$

Some suggestions for extending the NBS to situations with private information have been made; such as the generalized Nash solution (Harsanyi and Selten 1972) and the neutral bargaining solution (Myerson 1984). However, none of them has reached a level of support that resembles the NBS. Moreover, they do not carry over straightforwardly to our model; e.g., because the disagreement point is not fixed but type-dependent.
For the principal, we assume that she commits to emit less than in the Nash equilibrium. Specifically, following Hoel (1991, 59) we assume that the principal chooses her emissions so as to maximize

\[ B_1(x_1) - \theta_1 \int_{\theta_2} D(x_1 + x_2(\theta_2)) \, dF(\theta_2) - mX, \quad m > 0, \]  

(15)

while Nash equilibrium emissions would follow from (4) (or, equivalently, from (15) for \( m = 0 \)). Thus the principal commits to choose emissions according to

\[ B_1'(x_1) - \theta_1 \int_{\theta_2} D'(x_1 + x_2(\theta_2)) \, dF(\theta_2) - m = 0 \]  

(16)

should contract negotiations fail. Accordingly, the principal’s emissions in this solution will be a function of her inclination to undertake unilateral action, \( m \), and we denote them by \( x_1^0(m) \). Given the assumption that \( D(X) \) is linear, countries’ emissions follow independently from (14) and (16). Implicit differentiation then yields

\[ \frac{dx_1^0(m)}{dm} = \frac{1}{B_1''(x_1)} \quad \text{and} \quad \frac{dx_2^0(m)}{dm} = 0. \]  

(17)

Using the curvature assumptions and payoff functions, we obtain the following result.\(^5\)

**Lemma 2** The principal’s unilateral commitment – i.e., a higher \( m \) – has no effect on the agent’s out-of-contract emissions, but those of the principal and aggregate emissions, \( X^0 = x_1^0 + x_2^0 \), fall. Consequently, a higher \( m \) reduces the principal’s payoff and raises the agent’s payoff at the disagreement point.

The solution without private information can be seen as a special case of the above where all probability mass lies on a single type \( \theta_2 \). For this specification, an increase in \( m \) has similar effects independent of whether one models contract negotiations using Nash bargaining (as in Hoel (1991)) or take-it-or-leave it offers (as in this paper). Specifically, with a better outside option the agent requests a more attractive offer from the principal. If no transfer payments are feasible as in Hoel (1991), this can be achieved by higher emission reductions of the principal and by lower reductions of the agent. The extent to which these two instruments will be used depends on the responsiveness of countries’ marginal abatement costs to changes in emissions. For example, if \( |B_1''(x_1)| > |B_2''(x_2)| \), then less weight will be put on emission reductions by the principal because this would lead to a rapid increase in marginal benefits (respectively marginal abatement cost). Accordingly, aggregate emissions would increase in consequence of the agent’s better outside option that is caused by the

---

\( ^5 \)The last statement in the lemma is obvious: The agent benefits from the lower emissions of the principal without adjusting his own emissions. The principal does not choose her emissions as best response, but according to condition (16).
principal’s commitment to unilateral action. This is the case emphasized by Hoel (1991) in his warning that unilateral emission reductions may well imply higher total emissions than if both countries act selfishly. By contrast, if $|B_2'(x_2)| > |B_1'(x_1)|$, then marginal benefits of the agent decrease rapidly as it is allowed to emit more. Accordingly, less weight will be put on emission increases by the agent and aggregate emissions fall.

Finally, let us consider the case if transfer payments were allowed (under full information). Then, Nash bargaining and take-it-or-leave-it offers would both implement efficient emission levels and rely on transfers to allocate rents that correspond to the parties’ bargaining power. Therefore, unilateral action would have no effect on the outcome of negotiations. We now show that private information fundamentally changes this result.

3.2 Effects of unilateral commitment on optimal incentive contract

The optimal contract is determined along the same line as in section 2. However, we restrict the analysis to the case of private information about damages, i.e. we set $\alpha = 1$. This setup is sufficient to examine the effects of the principal’s unilateral commitment, and it keeps the analysis focused on this issue. Moreover, in contrast to section 2 we assume that the principal is committed to unilateral emission reductions that affect the disagreement points as summarized in Lemma 2. Therefore, the agent’s outside option is a function of the (exogenous) parameter $m$, and his participation constraint becomes

$$ U(\theta_2) \geq R(\theta_2, m) := \max_{x_2} B_2(x_2) - \theta_2 D(x_1^0(m) + x_2) \quad \forall \theta_2 \in [\underline{\theta_2}, \bar{\theta_2}] . $$

(18)

Accordingly, the optimal contract follows from (9) and (10), evaluated at $\alpha = 1$, as well as (18). The unilateral commitment parameter $m$ enters the program only via the participation constraint so that the associated Lagrangian multiplier, $\mu(\theta_2, m)$, is now also a function of $m$. However, $\mu(\theta_2, m) = 0$ for interior solutions, i.e. if the participation constraint does not bind. Using this, it follows immediately from the conditions that determine emissions in the optimal contract (see Lemma 1) that these are independent of $m$ for interior solutions.

By contrast, the principal’s unilateral commitment does affect emissions for boundary solutions ($\mu(\theta_2, m) > 0$). Differentiation of the agent’s outside option $R(\theta_2, m)$ as given in (18), and of his payoff from telling the truth (i.e., (7) evaluated at $\hat{\theta}_2 = \theta_2$ and $\alpha = 1$) yields (using the envelope theorem and dots to denote derivatives with respect to the type $\theta_2$)

$$ U(\theta_2) - R(\theta_2, m) = D(X^0(\theta_2, m)) - D(X(\theta_2)) . $$

(19)

Moreover, the agent’s participation constraint (18) can be written alternatively as $\bar{U}(\theta_2) - \bar{R}(\theta_2, m) \geq 0$ whenever $U(\theta_2) = R(\theta_2, m)$ (see proof of Lemma 1). Intuitively, this assures that the contract payoff does not fall below the out-of-contract payoff. By complementary slackness, it follows that the conditions for the optimal contract include $\mu(\theta_2, m) \geq 0$.
0, \mu(\theta_2, m) \left[ \dot{U}(\theta_2) - \dot{R}(\theta_2, m) \right] = 0 \text{ (see 42). Using (19), this condition can be stated alternatively as}
\begin{align*}
\mu(\theta_2, m) &\geq 0, \\
\mu(\theta_2, m) &\left[ X^0(\theta_2, m) - X(\theta_2) \right] = 0. \tag{20}
\end{align*}

It follows immediately that $X(\theta_2) = X^0(\theta_2, m)$ for boundary solutions. From Lemma 2, $dX^0/dm < 0$. Accordingly, the principal’s unilateral commitment reduces aggregate emissions of boundary solutions.

In addition, unilateral commitment affects the range of types that receive a boundary contract. Intuitively, if the principal excludes low types from a (interior) contract, it becomes less attractive for agents with a high WTP to undertake their true type. Often this exclusion is beneficial for the principal because it reduces the information rent that she has to pay. From Proposition 2 in Helm and Wirl (2014), it is known that the boundary and interior parts of the optimal contract are joined at the highest type at which aggregate relaxed program emissions – i.e., emissions that solve (12) and (13) for $\mu = 0$ – cross aggregate out-of-contract emissions from above. Given that $m$ reduces out-of-contract emissions but leaves relaxed program emissions unaffected, it follows immediately that more types receive a boundary contract as $m$ increases. Moreover, for these additional boundary types emissions equal those in the out-of-contract solutions (from 20) and, therefore, are below relaxed program emissions (due to the crossing from above). Accordingly, also this effect of the principal’s unilateral commitment reduces aggregate emissions (below we provide a graphical illustration).

More formally, define with $\theta_{IR} := \max\{\theta_2 \in (\bar{\theta}_2, \bar{\theta}_2]: X^e(\theta_2) = X^0(\theta_2, m)\}$ the highest type at which aggregate relaxed program emissions, denoted $X^e(\theta_2)$, cross aggregate out-of-contract emissions, i.e., $X^e(\theta_{IR}) = X^0(\theta_{IR}, m)$. If $\theta_{IR}$ exists, then an interior solution obtains for types $\theta_2 > \theta_{IR}$ and a boundary solution for types $\theta_2 \leq \theta_{IR}$ (see Proposition 2 in Helm and Wirl (2014)). From the above, $X^0(\theta_2, m)$ is decreasing in $\theta_2$ and $m$. Moreover, by definition $X^e(\theta_2) < X^0(\theta_2, m)$ for all $\theta_2 > \theta_{IR}$. Therefore, $\theta_{IR}$ is increasing in $m$. We summarize the above results in the following proposition.

**Proposition 1** If the principal is committed to reduce emissions by more than in the Nash equilibrium should contract negotiations fail, then this has the following effects on the optimal incentive contract: aggregate emissions of boundary types are lower, and more types receive a boundary contract. Both effects reduce aggregate emissions.

Figure 1 illustrates the effects of unilateral commitment. It is based on the following specification:

\begin{equation}
B_1 = \ln x_1, B_2 = 0.5 \ln x_2, D = X, \theta_1 = 1, \theta_2 \in [0.2, 1], f(\theta_2) = 1.25. \tag{21}
\end{equation}

The upper solid line depicts aggregate emissions without unilateral commitment ($m = 0$) for which $\theta_{IR} = 0.25$. The segment to the right of this threshold are (relaxed program) emissions of interior types, $X^e$, and the segment to the left are emissions of boundary types, denoted $X^b_{m=0}$. In line with Proposition 1, there are two effects of raising $m$. First, $\theta_{IR}$ rises from roughly 0.25 to 0.5 so that more types receive a boundary contract. Second, aggregate out-of-contract emissions for $m = 1$ are lower than for $m = 0$ (segments that do not belong to the contract are dotted and denoted $X^0_{m=0}$ and $X^0_{m=1}$, respectively). This leads to aggregate emissions in the optimal contract with unilateral commitment ($m = 1$).
as depicted by the lower solid line, where the right segment are emissions of interior types, $X_r$, and the left segment ($\theta_2 \leq 0.5$) emissions of boundary types, $X^b_{\theta_2 = 1}$. Thus, emission reductions due to the principal’s unilateral commitment are given by the area between the two solid lines to the left of $\theta_2 = 0.5$.

While we have shown that unilateral commitment reduces aggregate emissions, a different question is whether it raises the principal’s payoff. Remember that $R_2 (\theta_2, m)$ increases in $m$ from Lemma 2, and the relaxed program solution is independent of $m$. Accordingly, all contracts that are implementable (i.e. satisfy the agent’s incentive and participation constraint) under $m > 0$ are also implementable under $m = 0$. If the principal still chooses a different contract, then the contract chosen for $m = 0$ must yield a higher payoff than the one for $m > 0$. Accordingly, the principal’s unilateral commitment reduces her (expected) payoff (as in Hoel (1991)), which simply reflects that her emissions are lower and the agent’s higher along the boundary part of the contract (see Lemma 2).

4 Optimal contracts with unilateral action

In the previous section, we characterized unilateral action as a situation where the principal commits to additional emission reductions should contract negotiations fail. In this section, we say that the principal undertakes unilateral action if she reduces emissions up to a level
where she has higher marginal benefits – respectively, higher marginal abatement costs – than the agent, even if both face the same marginal damage. For equal benefit functions, unilateral action is then equivalent to lower emissions.

Specifically, denote differences in marginal abatement costs as

$$
\Delta := \frac{1 - \alpha (1 - \theta_1)}{\theta_1} B'_1 (x_1 (\theta_2)) - \frac{1 - \alpha (1 - \theta_2)}{\theta_2} B'_2 (x_2 (\theta_2)) .
$$

Observe from (6) that there is no unilateral action in the first-best solution. From (4) and (5), this is also the case in the Nash equilibrium if and only if $\theta_1 = \theta_2$, i.e. if both players have the same damage functions. This changes if we account for private information. Specifically, remember that emissions in the optimal contract must satisfy conditions (12) and (13) in Lemma 1. Subtracting these conditions yields

$$
\Delta = \left( \frac{1 - F(\theta_2) - \mu (\theta_2) 1 - \alpha}{f(\theta_2)} \right) B'_2 (x_2 (\theta_2)) .
$$

For an interior solution of the contract, we have $\mu (\theta_2) = 0$ as well as $1 - F(\theta_2) > 0$ for all $\theta_2 < \bar{\theta}_2$ and $1 - F(\bar{\theta}_2) = 0$. Moreover, for a boundary solution it has been shown in the proof of Lemma 1 that $1 - F(\theta_2) > \mu (\theta_2)$. Hence we obtain the following result.

**Proposition 2** For all but the highest type, $\bar{\theta}_2$, the principal chooses a contract where she has higher marginal abatement cost than the agent and, thus, undertakes unilateral action if and only if $\alpha < 1$ (i.e. if the agent has at least some private information about his abatement cost function).

To understand the intuition for this result, consider first the situation where the agent has only private information about his benefits, respectively abatement costs. Thus, $\alpha = 0$ and differentiation of the agent’s outside option (8) and of his payoff from telling the truth (i.e., (7) evaluated at $\theta_2 = \theta_2$) yields

$$
\dot{U} (\theta_2) - \dot{R} (\theta_2) = - \frac{1}{\theta_2^2} \left[ B_2 (x_2 (\theta_2)) - B_2 (x_2^0 (\theta_2)) \right] .
$$

For the moment, suppose that all types had the same outside option, denoted $\dot{R}$. Remember that the agent’s payoff (3) is falling in $\theta_2$. Hence he would have an incentive to overstate his true type as this would require a higher payment from the principal to shift his payoff above $\dot{R}$. To counteract this incentive, the principal would have to pay an information rent, $U (\theta_2) - R (\theta_2)$, that is higher for low types. This is the outcome that would obtain from (24) after setting the second term on both sides equal to zero so as to represent a constant outside option.

However, the type-dependence of the outside option provides a countervailing incentive. Specifically, high types have a worse outside option, which reduces the payment that is needed to assure their participation. Hence the agent has an incentive to understate his true type. This effect dominates because the agent’s contract emissions are below their out-of-contract level. Consequently, the contract binds from below (see Lemma 1) and high types receive a larger information rent.
The agent’s strategic considerations do not depend on the principal’s emissions; hence these are set at the first-best level.\(^9\) Efficiency would require to do this also for the agent’s emissions. However, the principal also wants to reduce the information rent that she has to pay and distorts the agent’s emissions upwards, especially for low types. This reduces the difference between contract and out-of-contract emissions and therefore, makes it less attractive to pretend a better outside option by underreporting of types. As a consequence, the principal has higher marginal abatement cost, i.e. she undertakes unilateral action. Put differently, unilateral action is the price that the principal has to pay in order to induce agents with low abatement costs to reveal their type.

Now consider the other extreme of \(\alpha = 1\), which does not lead to unilateral action according to Proposition 2. In this case, payoff functions are

\[
V_i = B_i(x_i) - \theta_i D(X), \quad i = 1, 2
\]  

(25)

so that the agent has private information about his damages only. These depend only on aggregate emissions, and so does the information rent. Accordingly, the principal has no reason to sacrifice cost-efficiency, and she chooses a contract that equalizes marginal abatement cost.

## 5 Unilateral action after contracting

By choosing a contract offer, the agent reveals his type. Having learned the type, the principal could design an alternative contract that is efficient and shares the resulting rent between the two parties. In our example, this would require to adjust the agent’s and the principal’s emissions as well as transfer payments. However, despite being mutually beneficial ex post, such a contract renegotiation turns out detrimental from an ex ante perspective for the principal because it is anticipated by the agent (see, e.g., Bolton and Dewatripont (2005, ch. 9)).\(^{10}\)

This problem is associated with a lack of commitment and well known. Therefore, it is not addressed in the following. However, the multilateral externalities add a further aspect to ex post deviations from the contract. The principal may find it beneficial to unilaterally reduce her emissions below the contract level.\(^{11}\) This does not require a renegotiation of the contract and, therefore, it will be more difficult to commit ex-ante against such a unilateral deviation, which is also to the benefit of the agent. Hence he is unlikely to oppose against this breach of contract. In fact, it is not uncommon that individual countries’ overfulfill their obligations under international environmental agreements, and other countries rarely

---

\(^{9}\)Efficiency of the principal’s emissions follows straightforwardly from noting that the first-order conditions for the first-best emissions (6) and contract emissions (12) are the same for \(\alpha = 0\) and \(D'\) constant.

\(^{10}\)For example, consider a buyer with either high or low valuation. It is often optimal for the principal to set a price such that only the high valuation type buys. Observing that the agent did not buy, the principal has an incentive to lower the price ex post. However, anticipating this behavior a high type agent has an incentive not to buy in stage 1 so as to benefit from the lower price in stage 2.

\(^{11}\)Observe that in the standard example of footnote 10, not only the principal but also the agent changes his behavior after the contract, namely from not buying to buying. A unilateral deviation by the principal would be to demand a lower than the contracted price from the agent that bought the product (or to deliver a better quality). Obviously, this is not in her interest.
follows immediately that the added constraint (weakly) reduces the principal’s (expected) payoff. The problem with the added constraint is also a feasible solution of the problem without this constraint. It reduced its SO$_2$ emissions ex post. However, if the agent anticipates this, then the “original” contract is no longer incentive compatible; hence it needs to be amended.

Consequently, this contract design allows the principal to reduce the information rent, but it may also bring her emissions above the out-of-contract level. Given linear damage costs, it then follows immediately that the principal would benefit from unilaterally reducing her emissions ex post. However, if the agent anticipates this, then the “original” contract is no longer incentive compatible; hence it needs to be amended.

To further investigate this, denote by \{x^c_1(\theta_2), x^c_2(\theta_2), s^c(\theta_2)\} contract emissions and subsidies that result from the “original” contracting problem in Section 2, i.e., from equations (9), (10) and (11). The principal’s payoff as given by (3) increases in unilateral emission reductions if $B'_1(x^c_1(\theta_2)) < \theta_1 D'(X^c(\theta_2))$. The out-of-contract solution follows from (4) and – using the assumption of linear damage cost – satisfies $B'_1(x^c_1) = \theta_1 D'(X^c(\theta_2))$. Therefore, the principal has an interest to unilaterally reduce her emissions ex post if and only if $x^c_1(\theta_2) > x_1^0$.

For specification (21) and $\alpha = 1$, on which Figure 1 was based, this happens for all types $\theta_2 < 0.5$. The dashed lines in Figure 2 depicts this. Moreover, the difference may be substantial; for the lowest type the principal’s emissions in the original contract, $x^c_1(\theta_2)$, is four times her out-of-contract emissions, $x_1^0$. By contrast, for $\alpha = 0$, the principal’s emissions of the original (unconstrained) solution are first-best (see 6 and 12). Therefore, they are always below the out-of-contract level, i.e., $x^c_1(\theta_2) < x_1^0$ for $\alpha = 0$. Intuitively, with private information about benefits only, the principal cannot use her own emissions to incentivize the agent, simply because the incentive compatibility constraint (10) does not depend on it (see equation (31) in the appendix). Thus the multilateral externality has no effect on incentive compatibility if $\alpha = 0$, and we obtain the standard result that the principal has no interest to unilaterally deviate from the contract ex post.

To analyze this more generally, we now assume that the principal unilaterally reduces her contract emissions ex post whenever this raises her payoff, and that the agent anticipates this. Effectively, this constrains the principal’s emissions in a contract to $x_1(\theta_2) \leq x_1^0$, and we add this as a control constraint to the original contracting problem (9) to (11). Doing so (see the proof of Proposition 3 below) yields that emissions of the principal and the agent in the “amended” contract satisfy (13) and

$$\frac{1 - \alpha (1 - \theta_1)}{\theta_1} B'_1 - [2 - \alpha (2 - \theta_1 - \theta_2)] D' = \frac{F - 1 + \mu}{f} \alpha D' + \frac{\nu}{f},$$

where $\nu(\theta_2) \geq 0$ is the Lagrangian multiplier of the new constraint. Comparing this with

\[\text{(26)}\]

\[\text{(13)}\]

\[\text{(14)}\]
the conditions for the original problem, (12) and (13), it follows that the only change is the added term $\nu(\theta_2)/f(\theta_2)$ in (26).

Subtracting the new equilibrium conditions (13) and (26) yields differences in marginal abatement costs

$$
\Delta = \left(1 - \frac{F(\theta_2) - \mu(\theta_2)}{f(\theta_2)} \frac{1 - \alpha}{\theta_2^2} \right) B_2'(x_2(\theta_2)) + \frac{\nu(\theta_2)}{f(\theta_2)},
$$

(27)

which is positive for all $\alpha \in [0, 1]$ if $\nu(\theta_2) > 0$. Thus the optimal amended contract is characterized by unilateral action of the principal as defined in Section 4, even in situations where this would not be the case for the original contract (i.e., for $\alpha = 1$; see Proposition 2). In addition, a binding constraint ($\nu(\theta_2) > 0$) strengthens the extent of unilateral action due to the positive last term in (27). This reflects that the principal undertakes more abatement if her emissions in the amended contract are constrained from above by her out-of-contract emissions.

For $\alpha > 0$ and interior solutions ($\mu(\theta_2) = 0$), comparison of the conditions that determine emissions in the original problem (12 and 13) and in the amended problem (26 and 13)

---

15In the final paragraph of the proof of Lemma 1, we have shown that $1 - F(\theta_2) - \mu(\theta_2) > 0$ for the original problem. If one replaces the original first-order conditions for the principal’s emissions (12) by the new one (26), then applying exactly the same steps of the proof show that $1 - F(\theta_2) - \mu(\theta_2) > 0$ also for the amended problem.
shows that emissions of the agent are unchanged, but those of the principal and, therefore, aggregate emissions are lower in the amended problem. Moreover, differentiation of the agent’s payoff from telling the truth (i.e., (7) evaluated at $\theta_2 = \theta_2$) yields (using the envelope theorem)

$$
\dot{U}(\theta_2) = - \left( \frac{1 - \alpha}{\theta_2^2} \right) B_2(x_2(\theta_2)) - \alpha D(X(\theta_2)) < 0. \quad (28)
$$

From the above, $x_2^a(\theta_2) = x_2^c(\theta_2)$ and $X^c(\theta_2) - X^a(\theta_2) > 0$ along an interior solution, where superscript $a$ denotes emissions in the amended contract. Therefore,

$$
\dot{U}^a(\theta_2) - \dot{U}^c(\theta_2) = \alpha [D(X^c(\theta_2)) - D(X^a(\theta_2))] > 0, \quad (29)
$$

i.e. the agents payoff decreases less steeply in the amended contract. Moreover, the out-of-contract payoff does not change. Since the contract binds from below (see Lemma 1), this means that the agent’s information rent is higher in the amended problem (for $\alpha > 0$) if the boundary type is weakly lower, i.e. if we can exclude the case that there exists a $\theta^a_{IR} > \theta^c_{IR}$ (which we show below).\(^{17}\) The reason is that the principal cannot use her own emissions as effectively as before to deter the agent from underreporting his type.

For boundary solutions $(\mu(\theta_2) > 0)$, the analysis has to account for possible effects of the added constraint on $\mu(\theta_2)$. In addition, the range of types that receives a boundary solution may change. In the proof of Proposition 3 (see Appendix A2) we show that along a boundary solution aggregate emissions of the amended contract are the same as in the out-of-contract solution (for the relevant case of $\alpha > 0$). Moreover, the boundary and interior parts of the optimal contract are joined at the highest type at which aggregate relaxed program emissions – i.e., emissions that solve (12) and (13) for $\mu = 0$ – cross aggregate out-of-contract emissions from above (see Proposition 2 in Helm and Wirl (2014)).\(^{18}\) The former are lower in the amended contract, and the latter do not change. Hence, if such a crossing exists, it must take place for a lower type so that $\theta^a_{IR} \leq \theta^c_{IR}$. Moreover, aggregate emissions for the additional types that receive an interior contract are below their level in the original boundary contract. The following proposition summarizes the effect that unilateral action after contracting (weakly) reduces the principal’s and aggregate emissions.

**Proposition 3** Ex post unilateral action has no effects on the agent’s emissions, but often it reduces the principal’s and, therefore, aggregate emissions. Specifically:

(i) For $\alpha = 0$, i.e., with private information about benefits only, there are no effects.

(ii) For $\alpha > 0$ and types that receive an interior contract $(\mu(\theta_2) = 0)$: aggregate emissions are lower for all types $\theta_2$ for which the principal’s emissions in the original contract exceed her out-of-contract emissions; i.e. for which $x_1^c(\theta_2) > x_0^1$.

(iii) Less types receive a boundary contract; i.e. either $\theta^a_{IR} \leq \theta^c_{IR}$ or, if $\theta^c_{IR}$ does not exist, all types receive an interior contract. For types that switch from a boundary to an interior contract, aggregate emissions are lower. For types $\theta_2 \leq \theta^a_{IR}$ that continue

\(^{17}\) Remember that if $\theta_{IR}$ exists, then it separates the solution such that types $\theta_2 \leq \theta_{IR}$ receive a boundary contract and types $\theta_2 > \theta_{IR}$ an interior contract.

\(^{18}\) An adaption of the corresponding proof to the model in this section is available upon request.
to receive a boundary contract, aggregate emissions of the original and the amended contract are the same.

**Proof.** See appendix. □

Figure 2, which has already been used to illustrate the effects of unilateral action after contracting on the principal’s emissions (dashed lines), also depicts the effects on aggregate emissions (solid lines). For types $\theta_2 < 0.5$, the principal’s emissions in the original contract exceed her out-of-contract emissions so that $X^u < X^c$. The difference may be substantial; for the lowest type, $\theta_2 = 0.1$, aggregate emissions in the amended contract are only half as large as in the original contract (not depicted). Moreover, in this original contract, types $\theta_2 \leq \theta_{IR} = 0.25$ receive a boundary contract because aggregate relaxed program emissions exceed out-of-contract emissions, $X^0$. In the amended contract this never happens due to the principal’s unilateral reductions; hence $\theta_{IR}^u$ does not exist and all types now receive an interior contract.

### 6 Concluding remarks

Most of the literature on climate agreements is based on the concept of coalitional stability, which was originally introduced by d’Aspremont, Jacquemin, Gabszewicz, and Weymark (1983) to analyze cartel formation. The inefficiency in these models arises from the parties’ inability to write binding contracts about coalition membership. This assumption is certainly appropriate for cartels, simply because they are illegal. It is less obvious for countries that, in fact, often write contracts, and many of these are widely considered as binding (especially in the area of trade agreements). This paper has used a different approach. The source of inefficiency is not countries’ inability to write contracts, but private information about their abatement cost function and their WTP for emissions abatement.

We have used this setup to analyze motives for unilateral action that result from the incentive structure in an incomplete information framework with multilateral externalities and type-dependent outside options. First, we have found that a unilateral commitment to emission reductions that is made before contract negotiations always reduces aggregate emissions, in contrast to the results in the seminal contribution by Hoel (1991). Second, optimal contracts are often characterized by unilateral action in the sense that the principal accepts higher marginal abatement costs than the agent, even if both have the same preferences for emission abatement. Finally, we have shown that the principal often has an interest to unilaterally reduce emissions below the level to which she is obliged to by the contract. Thus, the principal-agent model leads to a more positive assessment of unilateral action than many other economic papers on this issue (see introduction).

Nevertheless, regarding policy implications our results should certainly not be read as a naive endorsement of unilateral action. Global problems like climate change require global solutions, and unilateral action can only be useful if it contributes to such a global solution. In our analysis this is the case, provided that the unilateral action is carefully designed and within the limitations of our model. For example, detrimental general equilibrium effects – in particular, carbon leakage – that may result from unilateral action have not been considered simply because they fall out of the scope of our partial equilibrium model. There
is also no inter-temporal carbon leakage because the model is static. Moreover, the usual caveat applies that our model is a rather stylized description of complex climate change negotiations.

As Elofsson (2007, 143) writes, "unilateral abatement is sometimes advocated in order to set a good example that will make other countries follow." The mechanism by which this is expected to work is often based on social norms and ethical considerations. Our analysis provides an alternative perspective. The optimal contract can be read as a menu of conditional pledges in which the principal offers emission reductions in return to emission reductions by the agent. Types with a higher WTP for abatement and/or lower abatement costs should be assigned higher emission reductions. However, for the agent it is individually rational to misrepresent his type so as to avoid the associated burden. Thus the principal offers to shoulder marginal abatement costs above those of the agent, if the latter accepts higher emission reductions in return (respectively, if the agent honestly reveals that he is of a type to which the contract assigns high emission reductions). Such conditional pledges have indeed been an element of recent climate policies, e.g. the European Union’s commitment to raise its emission reductions from 20% to 30% (of 1990 levels by 2020), provided that other major emitting countries follow suit (UNEP 2012). Also the recent Lima accords ask countries to state their Individually Nationally Determined Contributions (INDCs). Our analysis suggests that it might be a good idea to accompany these by conditional pledges as in the EU example given above.

The model in this paper has been motivated with the example of international climate policies. However, several other relationships exist that are characterized by private information, multilateral externalities, and type-dependent outside options, which have been at the core of the analysis. Examples are joint ventures and team production problems, for which it is often the case that one party contributes more than the other. It is an interesting topic of future research whether the mechanisms that have been analyzed in this paper can help to better understand such behavior.

References


\[^{19}\text{See Harstad (2015) on dynamic aspects of climate agreements.}\]
Journal of Public Economics 110(0), 101 – 111.


7 Appendix

A1: Proof of Lemma 1

We begin by comparing emissions in the first-best solution and in the Nash equilibrium. The former are given by (6), the latter follow from (4) and (5) as (after multiplying both sides of the first-order condition by \[\frac{1 - \alpha (1 - \theta_i)}{\theta_i}\] for easier comparison)

\[
\frac{1 - \alpha (1 - \theta_i)}{\theta_i} B_i'(x_i^0) - [1 - \alpha (1 - \theta_i)] D'(X^0) = 0, \quad i = 1, 2. \tag{30}
\]

Given the assumption of linear damage costs and noting that \[2 - \alpha (2 - \theta_1 - \theta_2) > 1 - \alpha (1 - \theta_i),\] comparison of the two conditions (6) and (30) yields \[B_1'(x_1^0(\theta_1)) < B_1'(x_1^1(\theta_1)) \iff x_1^0(\theta_1) > x_1^1(\theta_1)\] by concavity of the benefit function.

Turning to the optimal contract, at \[^2 = 2^2\] the first-order condition of the incentive constraint (10) can be written in terms of the agent’s payoff as

\[
\dot{U}(\theta_2) = -\left(\frac{1 - \alpha}{\theta_2^2}\right) B_2(x_2(\theta_2)) - \alpha D(X(\theta_2)) < 0. \tag{31}
\]

Similarly, applying the envelope theorem to agent’s payoff outside a contract, as given in (8), yields

\[
\dot{R}(\theta_2) = -\left(\frac{1 - \alpha}{\theta_2^2}\right) B_2(x_2^0(\theta_2)) - \alpha D(X^0(\theta_2)) < 0. \tag{32}
\]

Accordingly, \(\dot{U}(\theta_2)\) and \(\dot{R}(\theta_2)\) differ only by the emission level at which they are evaluated. It is well known that the incentive constraint (10) can be replaced by the local incentive constraint (31) and the monotonicity constraint \(\dot{x}_2(\theta) \leq 0\) (see, e.g., Bolton and Dewatripont 2005). Moreover, solving (7) at \[^2 = 2^2\] for \(s(\theta_2)\) and substitution into (9), the principal’s problem is to determine the emissions profile \(\{x_1(\theta_2), x_2(\theta_2), \theta_2 \in [\theta_2, \bar{\theta}_2]\}\) that maximizes her expected payoff,

\[
\int_{\theta_2}^{\bar{\theta}_2} \left[ \sum_{i=1}^{2} \frac{1 - \alpha (1 - \theta_i)}{\theta_i} B_i(x_i) - [2 - \alpha (2 - \theta_1 - \theta_2)] D(X) - U(\theta_2) \right] dF(\theta_2), \tag{33}
\]

subject to the ‘dynamic’ constraint (31), the ‘state’ constraint (11) and the monotonicity constraint, \(x_2(\theta_2) \leq 0\).

This optimal control problem can be solved using the Pontryagin principle, where \(x_i\) are the controls and \(\dot{U}\) is the state variable (see, e.g., Chiang 1992). Specifically, constraint (11) is a pure state constraint of the first order, because the controls appear after differentiating,

\[
\dot{U}(\theta_2) - \dot{R}(\theta_2) = -\frac{1 - \alpha}{\theta_2^2} B_2(x_2(\theta_2)) - \alpha D(X(\theta_2)) - \dot{R}(\theta_2). \tag{34}
\]
Hence one can use the indirect method and replace the state constraint (11) by

\[ \dot{U}(\theta_2) \geq \dot{R}(\theta_2) \text{ whenever } U(\theta_2) = R(\theta_2). \]  

(35)

This yields the Hamiltonian (\(\lambda(\theta_2)\) is the co-state of \(U(\theta_2)\) and the arguments are dropped from now on),

\[ H = \left[ \sum_{i=1,2} \frac{1 - \alpha (1 - \theta_i)}{\theta_i} B_i(x_i) - [2 - \alpha (2 - \theta_1 - \theta_2)] D(X) - U \right] f + \lambda \dot{U} \]  

(36)

and the Lagrangian (\(\mu(\theta_2)\) is the Lagrangian multiplier)

\[ \mathcal{L} = H + \mu \left( \dot{U} - \dot{R} \right). \]  

(37)

The conditions for the optimal contract are,

\[ \frac{\partial \mathcal{L}}{\partial x_1} = \left[ \frac{1 - \alpha (1 - \theta_1)}{\theta_1} B'_1 - [2 - \alpha (2 - \theta_1 - \theta_2)] D' \right] f - (\lambda + \mu) \alpha D' = 0, \]  

(38)

\[ \frac{\partial \mathcal{L}}{\partial x_2} = \left[ \frac{1 - \alpha (1 - \theta_2)}{\theta_2} B'_2 - [2 - \alpha (2 - \theta_1 - \theta_2)] D' \right] f - (\lambda + \mu) \left[ \frac{1 - \alpha}{\theta_2^2} B'_2 + \alpha D' \right] = 0, \]  

(39)

\[ \dot{\lambda} = - \frac{\partial \mathcal{L}}{\partial \dot{U}} = f, \]  

(40)

\[ \dot{U} = - \left( \frac{1 - \alpha}{\theta_2^2} \right) B_2 - \alpha D, \]  

(41)

\[ \frac{\partial \mathcal{L}}{\partial \mu} = \dot{U} - \dot{R} \geq 0, \quad \mu \geq 0, \quad \mu \left( \dot{U} - \dot{R} \right) = 0, \]  

(42)

\[ U(\theta_2) - R(\theta_2) \geq 0, \quad \mu [U(\theta_2) - R(\theta_2)] = 0, \]  

(43)

\[ \dot{\mu} \leq 0 \quad [= 0 \text{ when } U(\theta_2) > R(\theta_2)], \]  

(44)

\[ \lambda(\theta_2) \leq 0, \quad \lambda(\theta_2) [U(\theta_2) - R(\theta_2)] = 0, \]  

(45)

\[ \lambda(\theta_2) \geq 0, \quad \lambda(\theta_2) [U(\theta_2) - R(\theta_2)] = 0. \]  

(46)

Here, (40) and (41) are the standard differential equations for the co-state and state variable. The complementary slackness condition (43) assures that the constraint on the state variable (42) only applies when \(U(\theta_2) - R(\theta_2) = 0\). (44) restricts the dynamics of the Lagrangian multiplier if the state constraint binds. Finally, (45) and (46) are the transversality conditions which reflect that we have a truncated vertical initial and terminal line.

There are different combinations of binding and non-binding state constraints at the
lowest and highest type. First, suppose \( \lambda (\overline{\theta}_2) > 0 \). At \( \theta_2 = \overline{\theta}_2 \), (38) and (39) then imply

\[
\frac{1 - \alpha}{\theta_1} (1 - \theta_1) B_1' - [2 - \alpha (2 - \theta_1 - \overline{\theta}_2)] D' = \frac{\lambda (\overline{\theta}_2) + \mu (\overline{\theta}_2)}{f (\overline{\theta}_2)} \alpha D' > 0, \quad (47)
\]

\[
\frac{1 - \alpha}{\theta_2} B_2' - [2 - \alpha (2 - \theta_1 - \overline{\theta}_2)] D' = \frac{\lambda (\overline{\theta}_2) + \mu (\overline{\theta}_2)}{f (\overline{\theta}_2)} \left[ \frac{1 - \alpha}{\theta_2} B_2' + \alpha D' \right] > 0. \quad (48)
\]

At the first-best emission level, the l.h.s. is equal to 0 for both equations (see 6). Hence if \( \overline{\theta}_2 \) is a boundary type and using superscript \( b \) to indicate a boundary solution, one obtains \( x_i^b (\overline{\theta}_2) \leq x_i^1 (\overline{\theta}_2) < x_i^0 (\overline{\theta}_2) \), \( i = 1, 2 \). Here, the last inequality follows from the beginning of the proof, and the first inequality takes into account that for \( \alpha = 0 \) the principal’s contract emissions are first best so that \( x_i^1 (\overline{\theta}_2) = x_i^1 (\overline{\theta}_2) \). Observe that the inequalities imply \( X_i^b (\overline{\theta}_2) < X_i^0 (\overline{\theta}_2) \). Moreover, for a boundary type \( \mu (\theta_2) > 0 \) so that for these types from (42), (31) and (32),

\[
\bar{U} - \bar{R} = \frac{1 - \alpha}{\theta_2^2} \left[ B_2 (x_2^0) - B_2 (x_2^b) \right] + \alpha [D (X^0) - D (X^b)] = 0. \quad (49)
\]

Evaluating (49) at \( \overline{\theta}_2 \), there are three cases which all lead to a contradiction to either \( x_i^b (\overline{\theta}_2) < x_i^0 (\overline{\theta}_2) \) or \( X_i^b (\overline{\theta}_2) < X_i^0 (\overline{\theta}_2) \). Specifically, for \( \alpha = 0 \), it implies \( x_i^0 (\overline{\theta}_2) = x_i^0 (\overline{\theta}_2) \).

For \( \alpha \in (0, 1) \), it follows from \( X_i^0 (\overline{\theta}_2) > X_i^b (\overline{\theta}_2) \) that \( B_2 (x_2^0 (\overline{\theta}_2)) < B_2 (x_2^b (\overline{\theta}_2)) \). Finally, for \( \alpha = 1 \) condition (41) implies \( X_i^0 (\overline{\theta}_2) > X_i^b (\overline{\theta}_2) \). Hence we conclude that \( \lambda (\overline{\theta}_2) = 0 \).

Using \( \lambda (\overline{\theta}_2) = 0 \), integration of the co-state differential equation (40) over the interval \( [\theta_2, \overline{\theta}_2] \) leads to \( \lambda (\theta_2) = F(\theta_2) - 1 \). Substituting this into (38) and (39) yields (12) and (13), which proves statement (ii) of the lemma. Finally, \( \lambda (\theta_2) = 0 \) would imply \( \lambda (\theta_2) = F(\theta_2) \) after integration of (40) over the interval \( [\theta_2, \overline{\theta}_2] \). However, we have already shown that \( \lambda (\theta_2) = F(\theta_2) - 1 \) so that we have a contradiction. Therefore, \( \lambda (\theta_2) < 0 \) and the contract binds at the lowest type (statement i).

We now turn to statement (iii) of the lemma. The claim that the highest type, \( \overline{\theta}_2 \), implements the efficient emissions level follows straightforwardly from comparing the first-order conditions for the first-best solution (6) and for contract emissions (12 and 13), thereby noting that the contract does not bind at the highest type (see above) so that \( \mu (\overline{\theta}_2) = 0 \). For those \( \theta_2 < \overline{\theta}_2 \) that receive an interior solution of the contract, we have \( \mu (\theta_2) = 0 \) as well as \( 1 - F(\theta_2) > 0 \). Therefore, the right-hand side of (12) and (13) is negative. Given that the left-hand side equals zero at efficient emission levels (see 6) and \( B_i (x_i) \) is concave, it follows that contract emissions are above their efficient level.

By the same argument, this is also the case for those \( \theta_2 < \overline{\theta}_2 \) that receive a boundary solution if \( 1 - F(\theta_2) > \mu (\theta_2) \). By contradiction, assume that \( F(\theta_2) - 1 + \mu (\theta_2) \geq 0 \). Hence, the right-hand side of (13) is non-negative. Comparing this with condition (6) for first-best emissions, it follows that \( x_i^0 (\theta_2) \leq x_i^1 (\theta_2) < x_i^0 (\theta_2) \), where the second inequality follows from the beginning of this proof. For this constellation, condition (49) implies \( X_i^b (\theta_2) > X_i^0 (\theta_2) \). This in turn requires \( x_i^b (\theta_2) > x_i^0 > x_i^1 (\theta_2) \), where the second inequality
again follows from the beginning of this proof. However, for \( x_1^1 (\theta_2) > x_1^1 (\theta_2) \), the left-hand side of the condition for emissions in the optimal contract (12) must be negative because it is equal to zero at \( x_1^1 (\theta_2) \) from (6) and \( B_1(x_1) \) is concave. Hence also the right-hand side of (12) must be negative and, therefore, \( F(\theta_2) - 1 + \mu(\theta_2) < 0 \). Thus we have a contradiction and conclude that \( F(\theta_2) - 1 + \mu(\theta_2) < 0 \).

A2: Proof of Proposition 3

The problem is the same as in the proof of Lemma 1, except that the constraint \( x_1 (\theta_2) \leq x_1^0 \) is added. After accounting for the latter, Lagrangian (37) becomes,

\[
\tilde{L} = H + \mu \left( \bar{U} - \bar{R} \right) + \nu [x_1^0 - x_1 (\theta_2)],
\]

Hence the first-order condition with respect to \( x_1 \) (equation (38) in the original problem) is now

\[
\frac{\partial \tilde{L}}{\partial x_1} = \left[ 1 - \alpha \frac{1 - \theta_1}{\theta_1} B_1' - [2 - \alpha (2 - \theta_1 - \theta_2)] D' \right] f - (\lambda + \mu) \alpha D' - \nu = 0,
\]

due to the additional constraint on the principal’s emission,

\[
\nu [x_1^0 - x_1 (\theta_2)] = 0, \quad \nu \geq 0.
\]

All other conditions of the optimal contract – equations (39) to (46) in Appendix A1 – remain unchanged. Applying the same steps (starting in the paragraph before (47)) as in the proof of Lemma 1 yields \( \lambda (\theta_2) = F(\theta_2) - 1 \) and, therefore, (26) and (13) as the conditions that determine emissions.

For types \( \theta_2 \) that receive a boundary solutions \( (\mu(\theta_2) > 0) \) and for which \( x_1^1 (\theta_2) > x_1^0 \), it remains to show that \( X^b (\theta_2) = X^0 (\theta_2) \), where (in slight abuse of notation) superscript \( b \) now refers to the boundary solution of the amended contract (we used this in the paragraph before Proposition 3 for \( \alpha > 0 \)). For \( \alpha = 1 \), this follows immediately from (49) because the benefit term cancels. For \( \alpha \in (0, 1) \), remember that \( x_1^1 (\theta_2) = x_1^0 \) from (52). First, suppose that \( x_2^2 (\theta_2) < x_2^0 (\theta_2) \). From (49) it follows that \( X^b (\theta_2) > X^0 (\theta_2) \) so that \( x_1^1 (\theta_2) > x_1^0 \), a contradiction. Next, suppose that \( x_2^2 (\theta_2) > x_2^0 (\theta_2) \). From (49) it follows that \( X^b (\theta_2) < X^0 (\theta_2) \) so that \( x_1^0 (\theta_2) < x_1^1 \), which yields again a contradiction. The only remaining case is \( x_2^2 (\theta_2) = x_2^0 (\theta_2) \), for which \( X^b (\theta_2) = X^0 (\theta_2) \) and \( x_1^1 (\theta_2) = x_1^0 \) follow immediately from (49) because then benefit term cancels too.
Zuletzt erschienen /previous publications:

V-378-15 Carsten Helm, Franz Wirl, Climate policies with private information: The case for unilateral action
V-377-15 Klaus Eisenack, Institutional adaptation to cooling water scarcity in the electricity sector under global warming
V-376-15 Christoph Böhringer, Brita Bye, Taran Fæhn, and Knut Einar Rosendahl, Targeted carbon tariffs – Carbon leakage and welfare effects
V-375-15 Heinz Welsch, Philipp Biermann, Measuring Nuclear Power Plant Externalities Using Life Satisfaction Data: A Spatial Analysis for Switzerland
V-374-15 Erkan Gören, The Relationship Between Novelty-Seeking Traits And Comparative Economic Development
V-373-14 Charlotte von Möllendorff, Heinz Welsch
V-372-14 Heinz Welsch, Jan Kühling, Affective States and the Notion of Happiness: A Preliminary Analysis
V-371-14 Carsten Helm, Robert C. Schmidt, Climate cooperation with technology investments and border carbon adjustment
V-370-14 Christoph Böhringer, Nicholas Rivers, Hideyuki Yonezawa, Vertical fiscal externalities and the environment
V-368-14 Marius Paschen, Dynamic Analysis of the German Day-Ahead Electricity Spot Market
V-367-14 Heinz Welsch, Susana Ferreira, Environment, Well-Being, and Experienced Preference
V-366-14 Erkan Gören, The Biogeographic Origins of Novelty-Seeking Traits
V-365-14 Anna Pechan, Which Incentives Does Regulation Give to Adapt Network Infrastructure to Climate Change? - A German Case Study
V-364-14 Christoph Böhringer, André Müller, Jan Schneider, Carbon Tariffs Revisited
V-363-14 Christoph Böhringer, Alexander Cuntz, Diemtar Harhoff, Emmanuel A. Otoo, The Impacts of Feed-in Tariffs on Innovation: Empirical Evidence from Germany
V-362-14 Christoph Böhringer, Nicholas Rivers, Thomas Ruhterford, Randall Wigle, Sharing the burden for climate change mitigation in the Canadian federation
V-361-14 Christoph Böhringer, André Müller, Environmental Tax Reforms in Switzerland A Computable General Equilibrium Impact Analysis
V-360-14 Christoph Böhringer, Jared C. Carbone, Thomas F. Rutherford, The Strategic Value of Carbon Tariffs
V-359-13 Heinz Welsch, Philipp Biermann, Electricity Supply Preferences in Europe: Evidence from Subjective Well-Being Data
V-358-13 Heinz Welsch, Katrin Rehdanz, Daiju Narita, Toshihiro Okubo, Well-being effects of a major negative externality: The case of Fukushima
V-357-13 Anna Pechan, Klaus Eisenack, The impact of heat waves on electricity spot markets
V-355-13 Christoph Böhringer, Knut Einar Rosendahl, Jan Schneider, Unilateral Climate Policy: Can Opec Resolve the Leakage Problem?
V-354-13 Christoph Böhringer, Thomas F. Rutherford, Marco Springmann; Clean-Development Investments:An Incentive-Compatible CGE Modelling Framework
V-353-13 Erkan Gören, How Ethnic Diversity affects Economic Development?
V-352-13 Erkan Gören, Economic Effects of Domestic and Neighbouring Countries’ Cultural Diversity
V-351-13 Jürgen Bitzer, Erkan Gören, Measuring Capital Services by Energy Use: An Empirical Comparative Study
V-350-12 Heinz Welsch, Jan Kühling, Competitive Altruism and Endogenous Reference Group Selection in Private Provision of Environmental Public Goods
V-349-12 Heinz Welsch, Organic Food and Human Health: Instrumental Variables Evidence
V-348-12 Carsten Helm, Dominique Demougin, Incentive Contracts and Efficient Unemployment Benefits in a Globalized World
Christoph Böhringer, Andreas Lange, Thomas F. Rutherford, Optimal Emission Pricing in the Presence of International Spillovers: Decomposing Leakage and Terms-of-Trade Motives

Christoph Böhringer, Jared C. Carbone, Thomas F. Rutherford, Efficiency and Equity Implications of Alternative Instruments to Reduce Carbon Leakage

Christoph Böhringer, Brita Bye, Taran Fæhn, Knut Einar Rosendahl, Alternative Designs for Tariffs on Embodied Carbon: A Global Cost-Effectiveness Analysis

Klaus Eisenack und Leonhard Kähler, Unilateral emission reductions can lead to Pareto improvements when adaptation to damages is possible

Heinz Welsch and Jan Kühlung, Anti-Inflation Policy Benefits the Poor: Evidence from Subjective Well-Being Data

Heinz Welsch and Jan Kühlung, Comparative Economic Performance and Institutional Change in OECD Countries: Evidence from Subjective Well-Being Data

Carsten Helm and Stefan Pichler, Climate Policy with Technology Transfers and Permit Trading

Christoph Böhringer, Jared C. Carbone, Thomas F. Rutherford, Embodied Carbon Tariffs

Christoph Böhringer, Carolyn Fischer, and Knut Einar Rosendahl, Cost-Effective Unilateral Climate Policy Design: Size Matters

Christoph Böhringer and Victoria Alexeeva-Talebi, Unilateral climate policy and competitiveness: The implications of differential emission pricing

Christoph Böhringer, Bouwe Dijkstra, and Knut Einar Rosendahl, Sectoral and Regional Expansion of Emissions Trading

Carsten Helm and Franz Wirl, International Environmental Agreements: Incentive Contracts with Multilateral Externalities

Christoph Böhringer and Andreas Keller, Energy Security: An Impact Assessment of the EU Climate and Energy Package

Klaus Eisenack, Adaptation financing as part of a global climate agreement: is the adaptation levy appropriate?

Udo Ebert and Patrick Moyes, Inequality of Well-Being and Isoelastic Equivalence Scales

Udo Ebert and Heinz Welsch, Adaptation and Mitigation in Global Pollution Problems: Economic Impacts of Productivity, Sensitivity, and Adaptive Capacity

Udo Ebert, The redistribution of income when needs differ

Heinz Welsch and Jan Kühlung, How Has the Crisis of 2008-2009 Affected Subjective Well-Being?

Heinz Welsch, Stabilität, Wachstum und Well-Being: Wer sind die Champions der Makroökonomie?

Klaus Eisenack, The inefficiency of private adaptation to pollution in the presence of endogeneous market structure

Udo Ebert and Patrick Moyes, Talents, Preferences and Inequality of Well-Being

Christoph Böhringer and Knut Einar Rosendahl, Greening Electricity More Than Necessary: On the Excess Cost of Overlapping Regulation in EU Climate Policy

Udo Ebert, The decomposition of inequality reconsidered: Weakly decomposable measures

Udo Ebert, Inequality reducing taxation reconsidered

Heinz Welsch und Jan Kühlung, Nutzenmaxima, Routinen und Referenzpersonen beim nachhaltigen Konsum

Heinz Welsch, Jan Kühlung, Is Pro-Environmental Consumption Utility-Maximizing? Evidence from Subjective Well-Being Data

Jürgen Bitzer, Ingo Geishecker, and Philipp J.H. Schröder, Returns to Open Source Software Engagement: An Empirical Test of the Signaling Hypothesis

Edwin van der Werf, Unilateral climate policy, asymmetric backstop adoption, and carbon leakage in a two-region Hotelling model