Doctoral Thesis

Climate change, permit markets with refunding, and rules treaties

Author(s):
Oberpriller, Quirin

Publication Date:
2013

Permanent Link:
https://doi.org/10.3929/ethz-a-009995150

Rights / License:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection. For more information please consult the Terms of use.
Climate Change, Permit Markets with Refunding, and Rules Treaties

A dissertation submitted to
ETH ZURICH

for the degree of
DOCTOR OF SCIENCES

presented by
QUIRIN OBERPRILLER
Dipl. Ing., TU MUNICH
Master, ETH ZURICH
born on July 6, 1980
in Munich, Germany

accepted on the recommendation of
Prof. Dr. Hans Gersbach (ETH Zurich), examiner
Prof. Dr. Antoine Bommier (ETH Zurich), co-examiner

2013
Acknowledgment

I would like to express my gratitude to several people who have been a great help in devising and writing this thesis. It originated during my time as a research assistant at the Chair of Macroeconomics: Innovation and Policy at ETH Zurich and was supported by a grant from the Swiss National Science Foundation.

I am indebted to my supervisor Professor Hans Gersbach for the trust he put into me when hiring me and the patience he has shown, especially in the beginning of this project. In addition, I am grateful that he helped me improving my writing precision and skills as well as for providing the ideas and being an excellent scientific mentor of the Chapters 2 and 5.

Many thanks to Martin Scheffel, for being my coauthor of Chapter 5 and also for the recommendations he gave me in improving Chapters 4 and 6. He was always open for lively discussions. Many thanks also to Maik Schneider for the helpful comments on Chapter 4.

I am very grateful to Margrit Buser for the careful proofreading of the complete thesis. Not only did she correct numerous typos, but she also improved the readability and the comprehensibility of the thesis a lot.

My appreciations also go to my co-advisor Professor Antoine Bommier for taking on this burden, to Professors Reto Knutti and Niki Gruber for discussing with me several aspects of Chapter 6 and to Professor Clive Bell for helping me to improve my English writing skills.

Special thanks go to my colleagues Elias, Vitali, Marina, Philippe, Jürg, Oriana, Martin, Maïk, Uli, Volker, Oriol, Kamali, Noemi, Olivier, Oli, Christian, Markus and Stephan for having made these four years at the chair a cheerful endeavor with many ping-pong lessons, laughs as well as funny and inspiring moments.

Finally, thanks to Barbara who always supported me when writing the thesis, especially in the final moments and for enduring my endless talks about weird economic stuff. And of course to Lorenz for coming into my life.
Contents

Zusammenfassung 7

Thesis Summary 9

1 Introduction 11

2 Rules vs. Targets: Climate Treaties under Uncertainty 19
   2.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19
   2.2 The Model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26
   2.3 Global Social Optimum and No-Treaty Outcome . . . . . . . . . .. . . 28
   2.4 Target Treaty . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 36
   2.5 Rules Treaty . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 41
   2.6 Rules Treaty vs. Target Treaty . . . . . . . . . . . . . . . . . . . . . .. 51
   2.7 Treaty Formation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 53
   2.8 Extreme Events . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 54
   2.9 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 55
   2.10 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 58
   2.11 Table of variables . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 59
   2.12 Proofs . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 60

3 Uncertainties and Climate Change 65
   3.1 Uncertainties of Environmental Damages . . . . . . . . . . . . . . . . . 65
   3.2 Uncertainties of Abatement Cost . . . . . . . . . . . . . . . . . . . . . . 71

4 Small Coalitions and Climate Change 75
   4.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 75
   4.2 The Set-Up . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 85
   4.3 Benchmarks . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 87
   4.4 Simultaneous Moves and Linear Damages . . . . . . . . . . . . . . . . . 87
   4.5 Simultaneous Moves and Quadratic Damages . . . . . . . . . . . . . . . 90
Zusammenfassung


Summary

The reduction of greenhouse gas emissions has become a major concern of policymakers. Nevertheless, little has been achieved in terms of slowing down or even reversing the ever-increasing emissions. A major reason is that emission reductions are a global public good. A country that reduces emissions has to bear the associated costs, whereas all countries benefit from such abatement – and thus have strong incentives to free-ride on the emission reductions of others. As a consequence, an international agreement among the major emitters would be needed that obliges members to reduce their emissions significantly. Such an agreement, however, is still lacking, partly because there is no international authority that can enforce participation in such a treaty and compliance to its obligations.

In this thesis, we propose two newly-designed climate treaties that have the potential to circumvent the free-riding problem, thanks to the use of a refunding scheme. What is more, the ‘Rules Treaty’, proposed in Chapter 2, allows members countries to react to newly-arriving information optimally. This is a crucial property for a climate treaty, as uncertainties are pervasive in climate change policy – as shown in Chapter 3. The ‘Rules Treaty for Innovation and Abatement’ proposed in Chapter 5 tackles the free-riding issue in two areas, the innovation of new abatement technologies and the reduction of emissions. In Chapter 4, we examine in more general terms what kind of strategic effects a small group of countries forming a climate treaty should take into account with respect to the free-riding of the remaining outsiders. Finally in Chapter 6, we conjecture that common methods to represent the climate system in economic models lead to an underestimation of the irreversibility of climate change. We propose an alternative method, which features this irreversibility and, at the same time, is sufficiently simple to be useful for economic modeling.
1 Introduction

The level of greenhouse gases (GHGs) in the atmosphere is on the rise (Forster et al., 2007). The most important GHG is carbon dioxide (CO$_2$); its concentration rose from 280 ppm (parts per million) at the eve of the industrial revolution to 396 ppm in 2012. This is the highest level in 420,000 years (Petit et al., 1999) if not much longer (Hansen et al., 2008). In parallel, the average global temperature rose by approximately 0.76 °C, the sea level rose by approximately 0.17 m, the acidity of the oceans dropped by approximately 0.05 pH points, and the area covered by glaciers and sea-ice is shrinking (Solomon et al., 2007; Barnosky et al., 2012). If emissions continue to rise unabatedly, these changes will just be the beginning of a large-scale transformation of the earth system (Meehl et al., 2007), and various kinds of environmental damages are to be expected (Parry et al., 2007). The reduction of GHG emissions has thus become a major concern of policymakers.

Emission reductions are a global public good. A country that reduces emissions has to bear the associated costs, whereas all countries benefit from such abatement. Because there is no international authority to enforce participation and compliance, an effective international climate treaty is thus hard to come about. All countries prefer free-riding on other countries’ emission reduction (see Hoel (1992); Carraro and Siniscalco (1993); Barrett (1994) for early contributions). Without international coordination, any individual action of single countries is bound to remain without significant impact on long-term climate change.

This utmost importance of international coordination has been acknowledged for a long time. In 1992, the United Nations Framework Convention on Climate Change was established and subsequentially signed by all major countries. Under its umbrella, there have been almost continuous negotiations among its member states, with the aim to design and put into force an international climate treaty that achieves the ‘...stabilization of greenhouse gas concentrations in the atmosphere at a level that would

1 Source: [http://www.esrl.noaa.gov/gmd/ccgg/trends](http://www.esrl.noaa.gov/gmd/ccgg/trends) (retrieved on 1. April 2013)
1. Introduction

prevent dangerous anthropogenic interference with the climate system’.\textsuperscript{2} To make this rather vague objective more concrete, two thresholds have been gained significance. The increase of global average temperature as compared to preindustrial values should not exceed 2 °C, or alternatively, the CO\textsubscript{2} concentration should be stabilized at 450 ppm.\textsuperscript{3}

The first output of these negotiations was the Kyoto Protocol, devised in 1997 and put into force in 2005. By the end of 2012, and in accordance with the principle of ‘common but differentiated responsibilities’, the developed countries committed themselves to binding reduction targets by 5.2% on average, as compared to the 1990 level, while developing countries were put under no obligation: Emission reduction seemed to endanger their economic development and they contributed little to the cumulative global emissions, so far. It was understood that once they had caught up, in a follow up treaty they would have to reduce emissions as well.

This ‘differentiation’, as well as the Kyoto Protocol’s focus on binding reduction targets, was the one of the main reasons why the Kyoto Protocol has failed to bring the two major emitters on board. China, as a developing country, was not obliged to reduce its emissions under the Kyoto Protocol to begin with. As a response, the US – even though they signed the Kyoto Protocol – never ratified it, for the reasons that ‘it exempts 80% of the world, including major population centers such as China and India, from compliance, and would cause serious harm to the US economy’.\textsuperscript{4} Other countries like Japan and Canada ratified the Protocol, but did not stick to their promises. Russia ratified it in 2005. Compared to the 1990 reference level, Russia’s emissions had already plummeted due to the collapse of the Soviet Union. This allowed Russia to meet the Kyoto Protocol’s obligation effortlessly and even yielded prospects of selling the remaining emission rights. This left the EU, together with some smaller countries, the sole party in charge of serious emission reductions. Thus, it is no surprise that the current emission levels and their projected rise are clearly incompatible with the thresholds of 450 ppm or 2 °C (Höhne et al., 2012; Peters et al., 2012).

The ineffectiveness of the Kyoto Protocol is a vivid illustration of the free-rider problem’s severity. Because an international authority is lacking, voluntary promises to re-

\textsuperscript{2} United Nations (1992).

\textsuperscript{3} These thresholds are often used interchangeably. It is important to note, however, that the conversion factor between temperature and GHG concentrations, the so-called climate sensitivity, is highly uncertain and thus only probabilistic estimates can be made (Knutti and Hegerl, 2008; Roe and Baker, 2007). In addition, it is uncertain which emission trajectories would lead to these thresholds (Meinshausen et al., 2009).

duce emission have been so far the only means to tackle climate change. This approach seems to fail as the multitude of particular interests makes it very difficult to mutually agree upon significant reduction targets. Still, there have been desperate efforts to find a more effective successor treaty, which culminated with the 2009 breakdown in Copenhagen.

As a response to the failure of the UNFCCC process, alternative ways to achieve emission reductions have been put forward (Prins et al., 2010). Some renounce the idea of an all-encompassing climate treaty altogether (Barrett, 2008), whereas others devise international climate treaties that a better suited to account for the countries’ strong free-riding incentives – among other problems. In this thesis, we stick to the latter approach and propose an astutely-designed treaty.

Furthermore, climate policy is riddled with uncertainty, as neither abatement costs nor environmental damages can be determined with accuracy (Heal and Kriström, 2002). Any climate policy should take this fundamental problem into account. In particular, the actions to be taken should be both sensible from today’s point of view and also flexible with regard to new information. Or work follows the spirit of Knutti and Hegerl (2008), who propose that ‘as scientists continue to narrow the estimates of the climate sensitivity, and as the feasibility of emission reductions is explored, long-term emission targets can be adjusted on the basis of future insight.’

In Chapter 2, we propose a ‘Rules Treaty’. It comprises a set of rules instead of fixed abatement targets. With the help of a model, we show that these rules have the potential to solve the free-riding problem. Suppose an international permit market is already in place, where a country’s local planner can freely choose the permits he allocates to domestic firms. Under such a setting, free-riding would prevail. Overall, local planners would issue more permits than is globally optimal, as they do not take into account that their domestic emissions damage other countries. On top of this basic set-up, the Rules Treaty grants an international agency the right to generate an amount $\gamma \in (0, \infty)$ of additional permits for every permit issued by the local planner of a participating country. This international agency sells its permits to the participating countries’ firms, and completely redistributes the corresponding revenues to the participating countries, according to predetermined refunding shares. If all countries participate, local and global interest are fully aligned for the fraction $\frac{1}{1+\gamma}$ of each permit a local planner issues. For a very high amount $\gamma$, the sum of permits issued by the local

---

5 See Aldy et al. (2003); Bodansky et al. (2004); Kuik et al. (2008); Aldy and Stavins (2009).
6 This chapter is based on Gersbach and Oberpriller (2012).
1. Introduction

planners and the international agency thus approximates the globally optimal amount. If, in addition, we assume that every country is pivotal for the Rules Treaty to be put into force, the refunding shares can be set such that all countries participate voluntarily. The Rules Treaty thus prevents free-riding in both participation and compliance.

In addition, as the participants of a Rules Treaty are not obliged to adhere to a predetermined emission reduction target, they are able to react to newly-arriving information, about abatement costs and environmental damages for instance. This flexibility stems from the fact that the global emission level is determined concurrently to the actual information about abatement costs and damages, whereas only the rules were fixed when the negotiations took place. This is in stark contrast to the Kyoto Protocol, where emission targets were predetermined during the negotiations.

In Chapter 3, we provide a brief overview of the multitude of uncertainties associated with the damages and abatement costs of climate change.

In a more general context, in Chapter 4, we consider a world split into two when tackling climate change. There is a group of countries willing to join forces and act in the group’s best interest, whereas the remaining countries act in their own best interest. Crucially, in this chapter we assume that – due to a sense of responsibility, for instance – the group members do not shirk their duties, even though this might be to their individual advantage. We describe several effects this group should take into account when designing a climate treaty. The results are especially interesting if the group size is rather small, as direct benefits of emission reductions are of minor importance to such a group. Instead, its main concern is to decrease the remaining countries’ possibilities to free-ride. Such free-riding could be so strong that it overcompensates the gains of cooperation for the small group. As a result, members of the small group might be jointly better off if they do not form a climate treaty at all, and act individually.

Furthermore, under uncertainty, the members of a small group might be better off designing a treaty under which they commit to emission reduction instead of designing a treaty, such as the Rules Treaty, under which they are able to adjust their emissions to new information.

In Chapter 5, we return to treaty design. In addition to free-riding in abatement, we concern ourselves with free-riding in Research and Development (R&D) of new technologies that allow emission reductions at a lower cost. These new technologies are a crucial part of every policy to combat climate change. Clarke et al. (2006),

---

7 This chapter is based on joint work with Hans Gersbach and Martin Scheffel.
8 See Jaffe et al. (2003, 2005); de Coninck et al. (2008); Popp et al. (2009); Aldy et al. (2009);
for example, find that advanced technologies can reduce discounted global abatement costs considerably, especially for ambitious reduction targets. Stabilization at 450 ppm could be more than 50 percent less expensive (see their figures 5.12 and 5.13). Edmonds et al. (2007) find that the present value cost of achieving a 550 ppm target is $20 trillion smaller when new technologies are taken into account. As Aldy et al. (2009) point out, however, it is difficult to model these effects quantitatively. Barrett (2009) gives a comprehensive overview of the portfolio of technologies and their prospects to play a significant role in future emission reductions.

We will show that free-riding in abatement efforts are intertwined with free-riding in R&D efforts: we call this the ‘double free-riding problem’. There is underprovision of global R&D efforts, even if new technologies could be effectively protected by international patents and if licenses could be sold by the R&D country on an international license market. Buyer countries only take into account their local cost reduction potential and thus prices for new technologies are comparatively low. The R&D country’s revenues from selling the superior technology is thus not high enough to provide sufficient incentives for globally optimal R&D effort. Compared to Chapter 2, the inclusion of free-riding in R&D efforts widens the gap between the free-riding outcome and the globally optimal outcome.

To cope with the double free-riding problem, we augment the scope of the Rules Treaty with R&D efforts. This ‘Rules Treaty for Innovation and Abatement’ basically functions like the Rules Treaty, with the major difference that parts of the international agency’s revenues are given to the R&D country, if it licenses its technology free of charge to all participating countries. This reimbursement can be set such that it brings about the globally optimal R&D effort.

In Chapter 6, we leave the realm of climate treaties and address the question why disagreement exists – between climate scientists and economists as well as among economists themselves – on what would be a sensible emission reduction path. There are well-known reasons for this dissent. There is a divide regarding the value of natural systems, the choice of an appropriate discount rate or the treatment of inequality, to name just a few. We discuss a further point: Most researchers working on the economics of climate change oversimplify its nature.

In simple analytical economic models it us often assume that a fixed fraction of the atmospheric CO\textsubscript{2} above the preindustrial value ‘decays’ each year. This resembles the capital depreciation in models of economic growth, and allows to concentrate on the


1. Introduction

economic issues of climate change. However, this model is not in accordance with climate science, as it neglects the longevity of CO$_2$ in the atmosphere. Another point is that the warming response to a certain level of GHG has a time lag ranging from decades to centuries. This is sometimes neglected in economists’ models. Thus, the time needed to return to lower temperatures, if the GHG level were to decrease is underestimated.

The economists’ model of CO$_2$ ‘decay’ has been subject to criticism ever since it started being used, and there have been proposals for improvements. However, more accurate modeling approaches have found little resonance. We make a new attempt at solving this problem and propose to use a relatively new concept, the ‘climate carbon response’ (Matthews et al., 2009), which is remarkably easy and represents the complex climate system rather well. It says that for every unit of emitted CO$_2$, we commit the planet to a certain irreversible warming – at a time-scale of centuries. We show that this concept can be applied easily in economic models and might affect results, especially if one cares about longer time horizons.
2 Rules vs. Targets: Climate Treaties under Uncertainty*

2.1 Introduction

2.1.1 Motivation

There is now a broad consensus that the climate is changing and that this development is linked to the increasing stock of greenhouse gases (GHGs) resulting from human activity. Yet the international coordination necessary to realize substantial cuts in emissions is proving extremely difficult to achieve. This is due to a variety of reasons, but two particularly thorny problems stand out.

First, lower emissions are a global public good. The degree of climate change depends on the total stock of GHGs in the atmosphere; where the emissions take place is irrelevant, and the benefits of abatement accrue to all countries. Costs to reduce emissions, however, have to be borne by those individual countries that are undertaking abatement efforts. The voluntary abatement efforts of individual countries, if uncoordinated, remain below what is optimal from a global perspective. Furthermore, there is no international authority having the power to enforce desirable emission reductions. This results in the notorious free-riding problem: countries may not participate in a treaty, or if they do, they may not deliver on their promises.

Second, abatement costs and environmental damages are very difficult to predict. The cost-benefit analysis of any climate policy is therefore subject to much uncertainty. Finding the globally optimal level of emission reduction, for instance, involves much guesswork. Any binding target\(^1\), even an ex-ante optimal one, would turn out to be suboptimal upon the arrival of new information as to damages, for instance. Only

---

\(^*\) This chapter is based on Gersbach and Oberpriller (2012).

\(^1\) Suppose for the moment that we could prevent free-riding and enforce any binding emission reduction target.
continuous renegotiation could incorporate the steady trickle of new information, but this is exceedingly difficult at an international level. Thus, policy-makers hesitate to agree to binding reductions in the first place.\footnote{There are further reasons why international coordination is so difficult. Countries, e.g., differ with regard to technology levels, abatement costs and damage costs functions. Such heterogeneity will play an important role in Sections 2.5 and 2.6, where we discuss the incentives of countries to participate in climate treaties.}

The interplay of incentives to free-ride and the uncertainties about costs and benefits make it particularly difficult to design, evaluate and implement a suitable climate policy architecture. This paper develops a blueprint for a global climate treaty, using tradable emission permits as a policy instrument. This treaty has the potential to overcome these challenges. We introduce the principle that treaties should be confined to rules, so that we call it the ‘Rules Treaty’. Our main insight is that there exists a set of rules that will allow new information on abatement costs and environmental damages to be processed efficiently without requiring renegotiation.

In the remaining part of this introduction, we first discuss briefly the various uncertainties climate policy has to deal with. We proceed with a general description of our model, introduce the Rules Treaty and several benchmark scenarios. Then, we summarize our major results and review the relevant literature to put our results in a broader context. Finally, we outline the organization of the rest of the paper.

### 2.1.2 The Impact of Uncertainty

The current generation has to start planning – and paying – for emission abatement, but most of the potential damage caused by climate change will have to be borne by future generations. It is, however, difficult to determine what sacrifices should be made today to reduce future damage, since the valuation of the damage and the abatement costs are highly controversial issues.

Figure 2.1 displays the causal chain from GHG emissions to the economic valuation of global damage at each date. Each and every link involves uncertainty, and together, they add up to an immense uncertainty regarding damages. Links 1 to 3 involve uncertainties because the complexities of the earth system hinder predictions with small confidence bounds (Solomon et al., 2007). Link 4 is connected with the extent and effectiveness of changes in behavior and adaptation measures. Link 5 involves valuing these changes. Finally, aggregation across regions and generations is required, a step that involves value judgments about the distribution of real living standards at any point in time,
and the weights to be assigned to different generations. This is a matter on which there is much disagreement (see e.g. Stern (2007), Nordhaus (2007), Ackerman et al. (2009)).

Abatement costs are somewhat less uncertain. Estimates depend mainly on the predicted speed of technological change and its impact on costs, as well as the availability of technologies able to reduce either emissions or the stock of GHGs (Metz et al., 2007).

2.1.3 The Approach

We modify the model of Gersbach and Winkler (2011) in two respects: The first modification is to allow the international agency to rescale the total number of permits (see Section 2.1.4); the second is to introduce uncertainty, using a sequence of events as depicted in Figure 2.2.
uncertainty is resolved. Actions (distribution and trading of permits, emission choices, damage, etc.) take place under complete information (ex-post).

In this paper we have deliberately chosen a very simple set-up that allows to highlight how efficiently suitable Rules Treaties process information. In section 2.9 we outline several extensions in which Rules Treaties could be applied, including a multi-period framework with gradual revelation of uncertainty.

2.1.4 Benchmarks Scenarios and the Rules Treaty

Let us first define the ‘Global Social Optimum’ as the outcome which minimizes total global costs in each state of the world (ex-post optimality). Global costs are the sum of abatement costs and environmental damages over all countries.

Next, we will describe three different outcomes of international coordination (or non-coordination): the ‘No-Treaty Outcome’, the ‘Target Treaty’ and the Rules Treaty. All of them assume that an international permit market for GHG emissions is in place, and countries distribute a certain number of permits to domestic firms. The firms, in turn, may trade their permits. Our three approaches can be described as follows:

The No-Treaty Outcome is the non-cooperative solution in which countries freely choose the number of permits they will distribute to domestic firms. It also serves as the outside option for all parties during treaty negotiations.

Under the (ideal) Target Treaty, signatories cooperate and fix the total number of permits at an ex-ante optimal level. This total is then distributed among member countries and countries comply. The total number of permits and the distribution cannot be renegotiated ex-post. The latter property reflects that it is almost impossible for a large set of countries to embark on a series of successful renegotiations on targets when new information arrives. This has been abundantly demonstrated by the cumbersome, lengthy and often unsuccessful UNFCCC process to renegotiate the Kyoto Protocol.\(^3\) Assuming an ex-ante optimal number of permits under the Target Treaty establishes a benchmark, as it gives the Target Treaty the best chance to perform well.

The Rules Treaty works as follows: The permit market is administered by an international agency and obeys a set of rules that are negotiated ex-ante by the participating countries. These rules comprise a scaling factor and country-specific refunding shares. Under the Rules Treaty, participants can, ex-post, freely choose the number of emission

\(^3\) For a detailed description see Aldy and Stavins (2009).
permits they allocate to domestic firms. For every permit a country issues, the international agency is allowed to issue additional permits in accordance with the scaling factor. The agency auctions its permits to member countries’ firms and reimburses all its revenues to member countries according to the refunding shares.

2.1.5 Results

We will obtain two major insights regarding the Rules Treaty. First, it allows the processing of new information, as it fixes rules, but not the number of permits. Thus, the countries will be able, and willing, to adjust their permit issuance when they receive new information on damages or abatement costs. This is in stark contrast to treaties based on fixed targets, like the Target Treaty.

For large scaling factors, information processing is almost optimal under a Rules Treaty. The reason is that global costs decrease with the scaling factor, and approach the Global Social Optimum in each state of the world when scaling factors become large. This seems to be counterintuitive at first, and we will devote a large part of the paper to prove the result and to provide the intuition (see e.g. Section 2.5.2).

Second, the Rules Treaty can be designed to induce voluntary participation, i.e. all countries are better off ex-ante if they participate than if they choose the outside options – the No-Treaty Outcome in our case. This can be achieved by selecting a refunding rule that compensates those countries with expected abatement costs that are high and expected damages that are low for making higher abatement efforts under the Rules Treaty than they would make under the No-Treaty Outcome. If the No-Treaty Outcome is the outside option, all countries therefore participate voluntary.

The two major characteristics of the Rules Treaty described above imply that free-riding is prevented altogether if the scaling factor is sufficiently high. The first characteristic implies incentive compatibility and the second one that participation is individually rational. Beside the above advantages there is the drawback that arriving at a Rules Treaty might be difficult if countries are rather heterogeneous. Then, there is a trade-off between compliance under first-best rules and incentives to participate.

We further provide a detailed comparison between Target and Rules Treaties and show that the advantage of Rules Treaties persist when there are extreme events. Moreover, we show how Rules Treaties can be applied in a broader context. In particular,

\[4\] Information that reveals higher marginal damages or lower marginal abatement costs, for instance, will typically induce countries to tighten permit issuance.
uncertainty regarding business-as-usual emissions, more general functional forms and multi-period settings, as well as gradual revelation of information are suitable extensions that would preserve the benefits of Rules Treaties.

Finally, we discuss several critical features that need to be considered when Rules Treaties are used in practice.

### 2.1.6 Connection to the Literature

Our paper connects to the current research in several areas. Let us briefly highlight relevant findings and show how the Rules Treaty advances the literature.

**Climate Policy Architecture**

Many proposals for a suitable climate policy architecture have been put forward.\(^5\) The importance of climate treaties that can overcome free-riding in participation and compliance and that are flexible in the presence of new information has been stressed by Aldy et al. (2003). With the Rules Treaty developed in this paper, we introduce a treaty type that is able to address both issues at the same time.

**Free-riding**

Ever since Hoel (1992), Carraro and Siniscalco (1993) and Barrett (1994), it is well-known that free-riding is a major hurdle for effective international environmental agreements, as there exists no international authority enforcing participation and compliance. This paper tackles these two forms of free-riding but puts the main focus on compliance. Most of the literature – including the papers mentioned above – focus instead on participation. In the small literature that shares our focus on compliance, two general concepts have been put forward which complement our approach. One is, broadly speaking, to set up a fund which, once the money is available, supports abatement efforts.\(^6\) The other approach is ‘matching’ as proposed by Guttman (1978), Guttman and Schnytzer (1992) and put into the context of GHG abatement by Boadway et al. (2011). With symmetric countries, if a country chooses to abate one unit of GHGs domestically, all other countries would respond by reducing one unit as well. This yields globally optimal emission reductions. The matching approach relies on countries committing to their promised matching rates.\(^7\) In our model, once the Rules Treaty has been set up, no further commitment is necessary, as optimal abatement efforts are

---

\(^5\) For reviews, see Aldy et al. (2003), Bodansky et al. (2004), Kuik et al. (2008), and Aldy and Stavins (2009).

\(^6\) For different versions of this approach see e.g. Bradford (2004) and Gersbach et al. (2011).

\(^7\) For a detailed discussion see chapter 10.6 in Finus (2001).
a best response. Still, it is important that the rules of our treaty cannot be changed ex-post.

Uncertainty and Learning
Most of the literature on uncertainty and learning focuses on the trade-offs of intertemporal abatement decisions. Our model and results are complementary to this literature. We examine the effects of learning when countries commit to either targets or rules before uncertainty is resolved and abate under certainty.  

Prices vs. Quantities Under Uncertainty
This paper proposes a quantity regulation as the instrument to address climate change under uncertainty. Ever since Weitzman (1974), we know that permits are the preferred instrument if there is less uncertainty about expected marginal abatement costs than about marginal damages. Otherwise, taxes are preferable. Pizer (1997) uses a computable general equilibrium model to find that – in the context of climate change – expected welfare gains are five times higher under an optimal tax policy than under an optimal permit policy. This argument against permits, however, does not apply to the Rules Treaty suggested in our paper, as it allows countries to adjust their permits issuance to new information.

2.1.7 Organization of the Paper
The remainder of the paper is organized as follows: In Section 2.2 we describe the set-up of the model. In Section 2.3 we define and derive closed-form solutions for the Global Social Optimum and the No-Treaty Outcome. The same is done for the Target Treaty and the Rules Treaty, in Sections 2.4 and 2.5, followed by a comparison of both treaties in Sections 2.6. Section 2.7 examines one way to model the formation of the Rules Treaty. Section 2.8 discusses the parties’ behavior under the treaties when extreme events occur. Section 2.9 summarizes the results, suggests promising extensions, and draws more general conclusions.

Our approach to model the information structure is similar to the one used in Kolstad and Ulph (2008) and Kolstad and Ulph (2011). They find that learning may increase the expected global costs.
2. Rules vs. Targets: Climate Treaties under Uncertainty

2.2 The Model

2.2.1 Set-Up

We consider a model of a global economy with \( n \) politically autonomous countries indexed by \( i = 1, \ldots, n \) and run by a local planner. In each country \( i \), GHG emissions \( e_i > 0 \) arise from the activity of a representative firm, which faces a country-specific abatement cost \( C_i(e_i) \):

\[
C_i(e_i) = \frac{1}{2\phi_i^{\omega\phi}}(\bar{e}_i - e_i)^2, \quad e_i \in (0, \bar{e}_i), \quad \phi_i^{\omega\phi} > 0, \quad i = 1, \ldots, n. \tag{2.1}
\]

The abatement cost function describes the costs the firm incurs to reduce emissions. There are no abatement costs for \( e_i = \bar{e}_i \), which are business-as-usual emissions. \( \phi_i^{\omega\phi} \) is the abatement cost parameter of country \( i \). It depends on some aggregate event \( \omega \phi \), which can take on two values: \( \omega \phi \in \Omega_\phi := \{h_\phi, l_\phi\} \). The value \( h_\phi(l_\phi) \) indicates that abatement cost are high (low) in all countries, that is \( \phi_h^{\omega\phi} < \phi_l^{\omega\phi} \) for all \( i \). The aggregate events \( h_\phi \) and \( l_\phi \) occur with probability \( \pi \) and \( 1 - \pi \), respectively. Constant parameter ratios are a special case. There, for all \( i, \phi_i^{h\phi} = \lambda^\phi \phi_i^{l\phi} \) for some real number \( \lambda^\phi > 1 \).

The sum of all business-as-usual emissions across countries is denoted by \( \bar{E} \) with \( \bar{E} = \sum_{i=1}^{n} \bar{e}_i \), and the global emissions are \( E = \sum_{i=1}^{n} e_i \). Without loss of generality we normalize the initial global stock of GHGs to zero. Global emissions cause country-specific environmental damages, \( D_i(E) \), in each country, referred to as “damages”, so that

\[
D_i(E) = \frac{\beta_i^{\omega\beta}}{2}E^2, \quad \beta_i^{\omega\beta} > 0, \quad i = 1, \ldots, n. \tag{2.2}
\]

\( \beta_i^{\omega\beta} \) is the damage parameter of country \( i \) that depends on a second aggregate event \( \omega_\beta \), which can also take on two values: \( \omega_\beta \in \Omega_\beta := \{h_\beta, l_\beta\} \). Value \( h_\beta(l_\beta) \) indicates that damages in all countries are high (low), i.e. \( \beta_h^{\omega\beta} > \beta_l^{\omega\beta} \) for all \( i \). The aggregate events \( h_\beta \) and \( l_\beta \) occur with probability \( \sigma \) and \( 1 - \sigma \), respectively. Constant parameter ratios are again a special case. There, for all \( i, \beta_i^{h\beta} = \lambda^\beta \beta_i^{l\beta} \) for some real number \( \lambda^\beta > 1 \).

To describe the entire set of possible aggregate events, we define \( \Omega = \Omega_\phi \times \Omega_\beta = \{(l_\phi, l_\beta), (h_\phi, l_\beta), (l_\phi, h_\beta), (h_\phi, h_\beta)\} \). We use \( \omega = (\omega_\phi, \omega_\beta) \) with \( \omega \in \Omega \) to denote a state of the world. \( \text{Prob}(\omega) \) is the probability that \( \omega \) will occur (\( \sum_{\omega \in \Omega} \text{Prob}(\omega) = 1 \)). A special case is stochastically independent aggregate events \( \omega_\phi \) and \( \omega_\beta \). The probabilities are given in Table 2.1:

We assume that the probabilities and the values \( \beta_i^{\omega\beta} \) and \( \phi_i^{\omega\phi} \) for all states of the world
Table 2.1: Probabilities of states of the world if aggregate events are stochastically independent

<table>
<thead>
<tr>
<th></th>
<th>$h_\beta$</th>
<th>$l_\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_\phi$</td>
<td>$\pi \sigma$</td>
<td>$\pi (1 - \sigma)$</td>
</tr>
<tr>
<td>$l_\phi$</td>
<td>$(1 - \pi) \sigma$</td>
<td>$(1 - \pi)(1 - \sigma)$</td>
</tr>
</tbody>
</table>

and all countries are common knowledge. Furthermore, once the aggregate events have been realized, the state of the world is also common knowledge.

To shorten the notation we will henceforth write $\phi^{\omega}_i$ instead of $\phi_\omega^{\omega_i}$ and $\beta^{\omega}_i$ instead of $\beta_\omega^{\omega_i}$ if $\omega = (\omega_\phi, \omega_\beta)$ occurs. The use of the superscript $\omega$ will apply in the same way to other variables as well.

2.2.2 International Permit Market

As a building block for our analysis, we introduce an international permit market in the sense of Helm (2003), coupled with refunding in the sense of Gersbach and Winkler (2011), which operates after the aggregate events have occurred. In this and the next subsection, there is no administrative agency issuing additional permits, and we take the amount of emission permits each country has issued in state $\omega$, $\epsilon^{\omega}_i$, as given. Later, $\epsilon^{\omega}_i$ will be endogenized. The total number of permits $\mathcal{E}^{\omega} = \sum_{i=1}^{n} \epsilon_i^{\omega}$ is supplied in the market. For the functioning of the permit market, only $\mathcal{E}^{\omega}$ will matter. The possibility of trade guarantees that emission reductions will be cost effective.

Cost minimization behavior by the representative firms implies that in each country the marginal abatement costs will equal the permit price, $p^{\omega}$, in each state $\omega$:

$$p^{\omega} = -\frac{\partial C_i}{\partial \epsilon^{\omega}_i} = \frac{1}{\phi_i^{\omega}} (\bar{e}_i - \epsilon_i^{\omega}), \quad i = 1, \ldots, n. \quad (2.3)$$

Market clearing requires that the total number of permits equal global emissions:

$$\mathcal{E}^{\omega} = \sum_{i=1}^{n} \epsilon_i^{\omega} = \sum_{i=1}^{n} \epsilon_i^{\omega} = E^{\omega}. \quad (2.4)$$

To calculate the permit price in the permit market equilibrium for state $\omega$, we sum equations (2.3) over all countries and obtain

$$p^{\omega} = \frac{\sum_{i=1}^{n} \epsilon_i^{\omega} - \sum_{i=1}^{n} \epsilon_i^{\omega}}{\sum_{i=1}^{n} \phi_i^{\omega}} = \frac{\bar{E} - \mathcal{E}^{\omega}}{\Phi^{\omega}}, \quad (2.5)$$
where $\Phi^\omega := \sum_{i=1}^n \phi_i^\omega$ is referred to as the aggregate abatement cost parameter in state $\omega$.

### 2.2.3 Global costs

For later use we devise expressions for global costs when international permit markets are present. For any particular state $\omega$, we use cost minimization behavior by the firm, given by Equation 2.3, to rewrite optimal chosen abatement costs (Equation 2.1) in terms of the permit price:

$$C_i^\omega (p^\omega) = \frac{\phi_i^\omega}{2} (p^\omega)^2. \tag{2.6}$$

The sum of all abatement costs and damages across countries in a state $\omega$ are the global costs, denoted by $K^\omega$. Using Equations (2.2) and (2.6), we can express global costs in terms of the permit price and the total number of permits.

$$K^\omega = \sum_{i=1}^n \frac{\phi_i^\omega}{2} (p^\omega)^2 + \sum_{i=1}^n \frac{\beta_i^\omega}{2} (E^\omega)^2, \text{ and}$$

$$K^\omega = \Phi^\omega (p^\omega)^2 + B^\omega (E^\omega)^2, \tag{2.7a}$$

$$K^\omega = \sum_{i=1}^n \frac{\phi_i^\omega}{2} (p^\omega)^2 + \sum_{i=1}^n \frac{\beta_i^\omega}{2} (E^\omega)^2,$$

where $B^\omega := \sum_{i=1}^n \beta_i^\omega$ is referred to as the aggregate damage parameter.

Furthermore, the expected global costs, $\mathbb{E}[K]$, are given by

$$\mathbb{E}[K] = \sum_{\omega \in \Omega} \text{Prob}(\omega) \left[ \frac{\Phi^\omega}{2} (p^\omega)^2 + \frac{B^\omega}{2} (E^\omega)^2 \right]. \tag{2.8}$$

### 2.3 Global Social Optimum and No-Treaty Outcome

In this section we derive the Global Social Optimum (SO) and the No-Treaty Outcome (NT). They serve as benchmark scenarios for the treaties we will introduce in the next section. The No-Treaty Outcome occurs when no negotiation has taken place or negotiations have failed, as countries can still operate an international permit market in such cases. It serves as the outside option, i.e. it establishes the level of costs countries have to undercut in order to be willing to participate in a treaty extending beyond the mere presence of an international permit market. Note that both the global social planner and participants in the permit market act after the state of the world has been realized.
Throughout the paper we focus on the interior solutions, in the sense that no country will choose $e_i = 0$ and reduce all emissions. Corner solutions of the type $e_i = \bar{e}_i$ cannot occur, as in such a case marginal costs of abatement are zero, while marginal damages are positive.

### 2.3.1 Global Social Optimum

The sequence of events in the Global Social Optimum is shown in Figure 2.3.

In every state $\omega$, a social planner ex-post minimizes global costs, $K^\omega$, with respect to the emissions of countries, $e_i^\omega$. Using Equations (2.1) and (2.2), the problem is given by

$$\min_{\{e_i^\omega\}_{i=1}^n} \sum_{i=1}^n \left[ \frac{1}{2\phi_i^\omega} (\bar{e}_i - e_i^\omega)^2 + \frac{\beta_i^\omega}{2} (E^\omega)^2 \right],$$

subject to $\sum_{i=1}^n e_i^\omega = E^\omega$.

The necessary first-order conditions are

$$\frac{1}{\phi_i^\omega} (\bar{e}_i - e_i^\omega) = B^\omega E^\omega \quad i = 1, \ldots, n. \quad (2.10)$$

The first-order conditions say that in the Global Social Optimum, the cost of abating an additional unit of emissions in a country $i$ (LHS) has to equal the global damages caused by that additional unit (RHS).

**Proposition 1 (Global Social Optimum)**

*In the Global Social Optimum, in every state of the world $\omega$, the uniquely determined*
values of global emissions, $E^{SO,\omega}$, and local emissions, $e_i^{SO,\omega}$, are

$$E^{SO,\omega} = \frac{\bar{E}}{1 + B^{\omega}\Phi^{\omega}}, \quad \text{and}$$

$$e_i^{SO,\omega} = \bar{e}_i - \phi_i^{\omega} \frac{B^{\omega}\bar{E}}{1 + B^{\omega}\Phi^{\omega}}, \quad i = 1, \ldots, n.$$  

(2.11a) (2.11b)

The proof is given in Section 2.12.

We note that the Global Social Optimum can also be implemented by distributing any vector of emission permits $\{e_i^{\omega}\}_{i=1}^n$ to countries such that $\sum_{i=1}^n e_i^{\omega} = E^{SO,\omega}$ and by letting countries trade in the international permit market. In this case, the equilibrium price amounts to

$$p^{SO,\omega} = \frac{B^{\omega}\bar{E}}{1 + B^{\omega}\Phi^{\omega}}.$$  

For the remainder of the paper we assume that emissions as determined by Equation (2.11b) are non-negative.\(^9\)

**Assumption 1**

*For all countries $i = 1, \ldots, n$ and all states $\omega$ the following condition holds:*

$$\bar{e}_i \geq \phi_i^{\omega} \frac{B^{\omega}\bar{E}}{1 + B^{\omega}\Phi^{\omega}}.$$  

**2.3.2 No-Treaty Outcome**

**Description**

The sequence of events in the No-Treaty Outcome is shown in Figure 2.4:

After the state of the world has been realized, the local planner in each country is free to issue any amount of emission permits, $e_i^{\omega}$, which he distributes to the representative firm free of charge.\(^{10}\) Given the total supply of permits, the equilibrium price is obtained via

\(^9\) Depending on the target temperature it might be necessary that several countries store more carbon than they emit in the second half of the century (Clarke et al., 2009). This could be accomplished, e.g., using carbon capture and storage in power plants fueled with biomass. Our results can be extended to such scenarios with negative emissions if we add an initial stock of greenhouse gases inherited from previous periods.

\(^{10}\) The permits could also be auctioned by the local planner. It is well known that auctioning does not affect the cost-minimizing emission choice of firms and the local costs of a country if markets for permits are competitive. It only affects the internal distribution of costs and benefits within each
2. Rules vs. Targets: Climate Treaties under Uncertainty

Figure 2.4: Sequence of events in the No-Treaty Outcome

Equation (2.5), and the firms choose their emissions, \( e_i^\omega \), according to Equation (2.3). Firms buy or sell permits on the permit market such that the number of permits they hold matches their emissions.

Equilibrium Notation

At this stage, it is useful to differentiate the equilibrium notions used in the paper. In each state of the world, an equilibrium in the international permit market is characterized by a price of permits at which supply equals demand. An overall equilibrium for the No-Treaty case (henceforth equilibrium) is characterized by

- the Nash equilibrium of the local planners’ permit choices,
- the abatement decisions by firms (taking permit prices as given), and
- market clearing.

Equilibrium Characterization

Each country \( i \) chooses a number of permits, \( e_i^\omega \), in such a way as to minimize local costs. This is the sum of the firm’s abatement costs, local damage, and the cost arising from permit trading. It takes the permit choices of the other countries, \( e_j^\omega, j \neq i \), as given.

\[
\min_{\epsilon_i^\omega} \left[ \frac{\phi_i^\omega}{2} (p^\omega)^2 + \frac{\beta_i^\omega}{2} (E_i^\omega)^2 + p^\omega (e_i^\omega - \epsilon_i^\omega) \right], \quad i = 1, \ldots, n, \tag{2.12}
\]

country between the public and firms.
subject to (2.3), (2.4), and (2.5). We use the following derivatives:

\[
\frac{dp^\omega}{d\epsilon_i^\omega} = \frac{\partial p^\omega}{\partial E^\omega} \frac{dE^\omega}{d\epsilon_i^\omega} = -\frac{1}{\Phi^\omega}(1), \quad \text{and} \quad \frac{de_i^\omega}{d\epsilon_i^\omega} = \frac{\partial e_i^\omega}{\partial p^\omega} \frac{dp^\omega}{d\epsilon_i^\omega} = -\phi_i^\omega \left( -\frac{1}{\Phi^\omega} \right) = \frac{\phi_i^\omega}{\Phi^\omega},
\]

and obtain the necessary first-order conditions:

\[
-\phi_i^\omega \Phi^\omega p^\omega + \beta_i^\omega E^\omega - \frac{1}{\Phi^\omega}(e_i^\omega - \epsilon_i^\omega) - p^\omega (1 - \phi_i^\omega \Phi^\omega) = 0, \quad i = 1, \ldots, n. \tag{2.13}
\]

The four terms represent various effects on the costs of a country \(i\) if it issues one more permit.\(^\text{11}\)

1. \(-\phi_i^\omega \Phi^\omega p^\omega\): decrease of the firms’ abatement costs. As there is one more permit in the market, the firm will increase its emissions by \(\frac{de_i^\omega}{d\epsilon_i^\omega} = \phi_i^\omega \Phi^\omega\). Its abatement costs decrease by the marginal abatement costs multiplied by the increase in emissions.

2. \(+\beta_i^\omega E^\omega\): increase in local damages. One more permit causes an equivalent increase of global emissions.

3. \(-\frac{1}{\Phi^\omega}(e_i^\omega - \epsilon_i^\omega)\): change in trade revenues. The permit price decreases by \(\frac{1}{\Phi^\omega}\). The firm has to pay less for the permits it needs to cover its emissions, \(e_i^\omega\), but it also receives less for the permits it sells, \(\epsilon_i^\omega\). Depending on whether the firm is a net seller (\(e_i^\omega < \epsilon_i^\omega\)) or buyer (\(e_i^\omega > \epsilon_i^\omega\)) of permits, the revenues will decrease or increase, respectively.

4. \(-p^\omega (1 - \phi_i^\omega \Phi^\omega)\): value of the remainder of the additional permit. From the one additional permits it receives, the firm needs the fraction \(\phi_i^\omega \Phi^\omega\) to cover its additional emissions, and it sells the remaining fraction \((1 - \phi_i^\omega \Phi^\omega)\) at price \(p^\omega\).

Note that the first and fourth effect sum up to \(p^\omega\) and the net benefit for country \(i\) from the first and the fourth effect is simply the value of the additional permit. Hence, we can rewrite the first-order conditions (2.13) as

\[
p^\omega - \frac{1}{\Phi^\omega}(e_i^\omega - \epsilon_i^\omega) = \beta_i^\omega E^\omega, \quad i = 1, \ldots, n. \tag{2.14}
\]

The marginal benefits of issuing a permit (LHS) equal the local marginal damages (RHS).

\(^\text{11}^\)To simplify presentation, we represent the marginal effects by the issuance of one more permit. We are aware that this is not exactly a marginal change.
Proposition 2 (No-Treaty Outcome)

We obtain a unique equilibrium of the No-Treaty Outcome in each state $\omega$ as a vector of permit issuance $\{\epsilon_{i}^{NT,\omega}\}_{i=1}^{n}$ and the associated total number of permits, $E^{NT,\omega}$, the permit price, $p^{NT,\omega}$, a vector of country specific emissions, $\{e_{i}^{NT,\omega}\}_{i=1}^{n}$ and costs, $\{K_{i}^{NT,\omega}\}_{i=1}^{n}$, given by

$$E^{NT,\omega} = \frac{\bar{E}}{1 + \beta^{\omega} \Phi^{\omega}}, \quad (2.15a)$$

$$p^{NT,\omega} = \frac{B^{\omega} \bar{E}}{n + B^{\omega} \Phi^{\omega}}, \quad (2.15b)$$

$$e_{i}^{NT,\omega} = \bar{e}_{i} - p^{NT,\omega} \phi_{i}^{\omega}, \quad i = 1, \ldots, n, \quad (2.15c)$$

$$\epsilon_{i}^{NT,\omega} = \bar{e}_{i} - p^{NT,\omega} \phi_{i}^{\omega} \left(\frac{\beta_{i}^{\omega}}{\phi_{i}^{\omega}} + \frac{\beta_{i}^{\omega} - \bar{\beta}^{\omega}}{\beta^{\omega}}\right), \quad i = 1, \ldots, n, \quad (2.15d)$$

$$K_{i}^{NT,\omega} = \frac{1}{2} \left(\frac{p^{NT,\omega}}{\phi_{i}^{\omega}}\right)^{2} - \frac{1}{2} \left(E^{NT,\omega}\right)^{2} + p^{NT,\omega} e_{i}^{NT,\omega} \left[\epsilon_{i}^{NT,\omega} - e_{i}^{NT,\omega}\right], \quad i = 1, \ldots, n. \quad (2.15e)$$

\(\beta^{\omega} := \frac{1}{n} \sum_{i=1}^{n} \beta_{i}^{\omega}\) is the average damage parameter. The proof is given in Section 2.12.

The difference between country-specific emissions and permit issuance in equilibrium is

$$e_{i}^{NT,\omega} - \epsilon_{i}^{NT,\omega} = p^{NT,\omega} \phi_{i}^{\omega} \left(\frac{\beta_{i}^{\omega}}{\phi_{i}^{\omega}} - \frac{\bar{\beta}^{\omega}}{\beta^{\omega}}\right). \quad (2.16)$$

Analogous to the findings in Helm (2003), countries with damage parameters lower than the average are net sellers and vice versa.

With regard to permit issuance, two polar cases could happen in theory. First $\epsilon_{i}^{NT,\omega} > \bar{e}_{i}$, so that a country issues more permits than its business-as-usual emissions. Second, $\epsilon_{i}^{NT,\omega} < 0$ which means that a country destroys permits in order to reduce the global supply of permits. We allow for both polar cases.

Comparing Equations (2.15a) and (2.11a), we observe the standard result that the total emissions chosen by countries are higher than in the Global Social Optimum. The No-Treaty Outcome is inefficient, as the local planner of a country does not take into account other country’s damages when choosing the number of permits.

**Free-Rider Problem**

To illustrate the free-rider problem further, let us take $k$ replica of the economy with $n$ countries each. In this case, the total number of permits in the No-Treaty Outcome
2. Rules vs. Targets: Climate Treaties under Uncertainty

<table>
<thead>
<tr>
<th>Country</th>
<th>Business-as-usual</th>
<th>Abatement</th>
<th>Damage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \bar{e}_i )</td>
<td>( \phi_i^{h} )</td>
<td>( \rho_i^{h} )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \phi_i^{l} )</td>
<td>( \rho_i^{l} )</td>
</tr>
<tr>
<td>1</td>
<td>60</td>
<td>0.2</td>
<td>0.6</td>
</tr>
<tr>
<td>2</td>
<td>70</td>
<td>0.4</td>
<td>0.7</td>
</tr>
<tr>
<td>3</td>
<td>80</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>Probabilities</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
</tr>
</tbody>
</table>

**Table 2.2:** Parameters of our example

and the total emissions in the social optimum are given by

\[
E_{NT, \omega} = \frac{\bar{E}}{1 + \beta^\omega \Phi^\omega}, \text{ and}
\]

\[
E_{SO, \omega} = \frac{\bar{E}}{1 + kB^\omega \Phi^\omega}.
\]

In the limit, as \( k \) approaches infinity, the socially optimal emissions are zero, whereas global emissions approach a constant greater than zero in the No-Treaty Outcome.

### 2.3.3 Participation Constraint

In this paper we discuss two treaties. A country signs a treaty voluntarily if the expected cost of participating is lower than in the ex-ante outside option, which we assume to be the No-Treaty Outcome. Essentially, every country is pivotal on whether a treaty is formed or not because both treaties require unanimous agreement to be put into force. This represents the most favorable scenario for achieving consensus on forming a treaty. If this participation constraint is fulfilled for all countries, we call a treaty implementable:

**Definition 1 (Implementable Treaty)**

* A treaty is implementable if, for each country \( i = 1, \ldots, n \), the expected cost of participating in the treaty is no higher than in the No-Treaty Outcome.

### 2.3.4 Example

Throughout the paper, we will illustrate the outcomes with an example. We consider three countries. Table 2.2 summarizes their parameters.
The aggregate events are assumed to be stochastically independent and occur with probability $\frac{1}{2}$. Hence, each event $\omega = (\omega_\phi, \omega_\beta)$ occurs with probability $\frac{1}{4}$.

**Figure 2.5:** No-Treaty Outcome in state of the world $\omega = (h_\phi, h_\beta)$. The black curves display the local cost for a country $i$ (see cost function (2.12)) in the No-Treaty Outcome, given that the other countries choose permits at the equilibrium levels. The minima of the local cost curves represent the equilibrium. Black vertical lines are the equilibrium permit issuance and red vertical lines are the equilibrium emissions. Blue vertical lines are the emissions in the Global Social Optimum. The ends of the abscissas correspond to business-as-usual emissions.

Figure 2.5 shows that in the No-Treaty Outcome all firms emit more GHGs than in the Global Social Optimum. However, they issue fewer permits than their business-as-usual emissions. Country 1’s damage parameter is below average and, according to Equation 2.16, its firm is a net seller of permits. Abatement, $\bar{e}_i - e_i^{NT}\omega$, by the firm in Country 1 is half of Country 2’s firm and a third of Country 3’s firm, which reflects the difference in the abatement cost parameters (Equation 2.15c).
2.4 Target Treaty

Fixed targets are the approach taken by the currently most important attempt to slow down climate change: the UNFCCC process, including the Kyoto Protocol. In this section, we present our model of such a type of treaty, the Target Treaty (TT). We assume that the Target Treaty minimizes expected global costs. Correspondingly, it is an ‘optimal’ Target Treaty. This assumption is not made to reflect the state of current target treaties such as the Kyoto Protocol. Instead, the optimal Target Treaty serves as a benchmark, as it represents the best solution that can be achieved if countries commit to emission reduction targets.

2.4.1 Treaty Description

The sequence of events and functions is shown in Figure 2.6. Target Treaty negotiations take place before the state of the world has been realized. This reflects the fact that once countries have agreed on targets, continuous renegotiation after the arrival of new information is impossible. During the negotiation, stage an ex-ante optimal total number of permits, $E^{TT}$, is determined. Countries receive permits according to an allocation, $\{\epsilon_i^{TT}\}_{i=1}^n$, such that $\sum_{i=1}^n \epsilon_i^{TT} = E^{TT}$. Formally we define the Target Treaty as follows:

**Definition 2 (Target Treaty)**

A Target Treaty (TT) is an aggregate number of permits $E^{TT}$ and an allocation of emission permits $\{\epsilon_i^{TT}\}_{i=1}^n$, with $\sum_{i=1}^n \epsilon_i^{TT} = E^{TT}$. $E^{TT}$ minimizes expected global costs.

For the purpose of this paper, we assume that the Target Treaty is binding.

**Assumption 2**

The allocation of emission permits $\{\epsilon_i^{TT}\}_{i=1}^n$ is binding for all countries and in all states.
This assumption implies that in any state of the world, the number of permits countries give to its firms cannot exceed or fall below $\epsilon_T^{TT}$. The former implies complete enforceability of the permit allocation. The latter could in principle happen if, e.g., a country in a particular state is highly affected by climate change and therefore wants to lower the aggregate number of permits by not using part of its allocation. We rule out this possibility, as it only occurs in extreme cases or with few and heterogeneous countries participating in the treaty.\textsuperscript{12} In Section 2.8 we will discuss cases in which targets are not exhausted by countries when an extreme event happens.

In the following, we first illustrate the ex-ante design of the Target Treaty and later derive solutions for the ex-post stage.

### 2.4.2 Ex-ante: Design of the Target Treaty

In this section we determine the ex-ante globally optimal number of permits, $E^{TT}$, and discuss implementable permit allocations.

#### Total Number of Permits

Suppose there is a designer seeking to minimize the expected global cost $E[K^{TT}]$ (see Equation (2.8)) with respect to $E$, the total number of permits:

$$
\min_{E} \sum_{\omega} \left\{ \text{Prob}(\omega) \left[ \frac{\Phi^\omega}{2} \left( p^\omega \right)^2 + \frac{B^\omega}{2} E^2 \right] \right\},
$$

subject to Equation (2.5).

The necessary condition is

$$
\sum_{\omega} \{ \text{Prob}(\omega) \left[ -p^\omega + B^\omega E \right] \} = 0.
$$

\textsuperscript{12} The derivation of the ex-ante optimal target becomes much more cumbersome if ex-ante the potential destruction of permits in some extreme event has to be taken into account. If countries are homogeneous, a sufficient condition that targets are binding in all events is

$$
\pi \lambda^\phi + 1 - \pi \leq n \left( \sigma + \frac{1 - \sigma}{\lambda^\beta} \right).
$$

This condition is always satisfied if the number of countries is sufficiently large.
As the objective function (2.17) is strictly convex, the necessary condition is also sufficient. We use Equation (2.5) and the assumption that the aggregate events are independent to determine the unique $E_{TT}$:

$$E_{TT} = \frac{\bar{E}A}{D + A},$$

(2.18)

where $D := \sigma B^h + (1 - \sigma)B^l$ and $A := \left(\frac{\pi}{\Phi^h} + \frac{1 - \pi}{\Phi^l}\right)$.\(^{13}\)

Using Equations (2.5) and (2.8), the expected global cost in the Target Treaty, $E[K^{TT}]$, amounts to

$$E[K^{TT}] = \frac{A}{2} \left(\frac{\bar{E}D}{D + A}\right)^2 + \frac{D}{2} \left(\frac{\bar{E}A}{D + A}\right)^2.$$  

**Allocation of Permits**

Apart from the feasibility constraint $\sum_{i=1}^{n} \epsilon_{i}^{TT} = E_{TT}$, the determination of $E_{TT}$ does not pin down the country-specific allocation of permits, $\epsilon_{i}^{TT}$. An example of a feasible allocation is a Target Treaty denoted by $\overline{TT}$, such that for each country the permit allocation equals the expected emissions in the Global Social Optimum:

$$\epsilon_{i}^{TT} = E[\epsilon_{i}^{SO}] = E[\bar{\epsilon}_{i} - \frac{\bar{E}D}{(D + A)\Phi} \phi_{i}] = \bar{\epsilon}_{i} - \frac{\bar{E}D}{D + A} \left(\frac{\sigma \phi^{h} + (1 - \sigma) \phi^{l}}{\Phi^{h} + (1 - \sigma) \Phi^{l}}\right), \quad i = 1, \ldots, n.$$  

For the purpose of this paper we do not need to model the detailed process of how permits are allocated across countries. It is nevertheless useful to show the set of implementable Target Treaties. It is characterized by the following condition:

$$E[K_{i}^{TT}(\epsilon_{i}^{TT})] \leq E[K_{i}^{NT}], \quad \forall i = 1, \ldots, n,$$

(2.19)

where $E[K_{i}^{TT}(\epsilon_{i}^{TT})]$ and $E[K_{i}^{NT}]$ are the expected cost for country $i$ in a Target Treaty with permit allocation $\epsilon_{i}^{TT}$ and in the No-Treaty Outcome, respectively. We obtain the following lemma:

**Lemma 1 (Implementable Target Treaty)**

There exists an implementable Target Treaty if and only if the expected global cost of the Target Treaty is no higher than in the No-Treaty Outcome, i.e. if and only if $E[K^{TT}] \leq E[K^{NT}]$.

\(^{13}\) If there is no uncertainty (e.g. $\pi = \sigma = 0$), Equation (2.18) collapses to Equation (2.11a).
The proof of Lemma 1 is standard and given in Section 2.12. The intuition is straightforward. Once a treaty is globally less costly than the outside option and the redistribution of permits (i.e., costs) across countries does not generate frictions, the treaty is ex-ante implementable.\footnote{We note that $TT$ might not be implementable.}

### 2.4.3 Ex-post Outcome

In what follows, we consider the solution under a Target Treaty once the state of the world has been realized. The overall equilibrium corresponds to the equilibrium in the international permit market, as the amount of permit the countries’ local planner distribute to their firms is fixed according to the allocation $\{e^T_T\}_i^n$. The allocation is independent of the state of the world. The equilibrium permit price, however, depends on this state.\footnote{The price only depends on the aggregate event $\omega$.} Combining Equations (2.5) and (2.18), we obtain

$$p^{TT,\omega} = \frac{\bar{E}D}{(D + A)\Phi^\omega}.$$  

(2.20)

Given the price, firms choose the equilibrium level of emissions, $e^{TT,\omega}_i$, in accordance with Equation (2.3):

$$e^{TT,\omega}_i = \bar{e}_i - \frac{\bar{E}D}{(D + A)\Phi^\omega}\phi^\omega_i, \quad i = 1, \ldots, n.$$  

(2.21)

Inspection of Equation (2.21) shows that emissions are independent of the permit allocation.

Assumption 1 does not suffice to guarantee that emissions as determined by Equation (2.21) do not fall below zero. To ensure that $e^{TT,\omega}_i$ is non-negative, we assume for the remainder of the paper:

**Assumption 3**

For all countries $i = 1, \ldots, n$ and all states $\omega$, the following condition holds:

$$\bar{e}_i \geq \frac{\bar{E}D}{(D + A)\Phi^\omega}\phi^\omega_i.$$
Proposition 3 (Target Treaty)
In a Target Treaty, the total number of permits, $E^{TT}$, is given by Equation (2.18). There exists a set of implementable permit allocations $\{\epsilon^{TT}_i\}_{i=1}^n$ if $E[K^{TT}] \leq E[K^{NT}]$. In equilibrium, the permit price $p^{TT,\omega}$ in each state $\omega$ is given by (2.20), the vector of country-specific emissions $\{e^{TT,\omega}_i\}_{i=1}^n$ is determined according to (2.21) and costs $\{K^{TT,\omega}_i\}_{i=1}^n$ are given by

$$K^{TT,\omega}_i = \frac{\phi^2}{2} (p^{TT,\omega})^2 + \frac{\beta^2}{2} (E^{TT})^2 + p^{TT,\omega} (e^{TT,\omega}_i - \epsilon^{TT}_i), \quad i = 1, \ldots, n.$$  

Finally, we observe that in general $E^{TT}$ is not ex-post optimal.

Lemma 2 (Target Treaty does not Implement Global Social Optimum)
If there is uncertainty about the damage and abatement cost parameters, there exists no Target Treaty that can implement the Global Social Optimum.

The proof of Lemma 2 is given in Section 2.12. The intuition is as follows: under a Target Treaty emission reductions are fixed before the state of the world is known. Hence when new information arrives, countries are not able to react but have to stick to the old, now suboptimal, reduction targets. This is a severe disadvantage of all treaties involving fixed targets.

2.4.4 Example

We illustrate the properties and potential problems of the Target Treaty using the numerical example introduced in Section 2.3.4. Figure 2.7 shows the ex-ante globally optimal number of permits, $E^{TT}$, and the ex-post globally optimal level of emissions in each state of the world, $E^{SO,\omega}$. We observe that the Target Treaty is not ex-post optimal. In the state $\omega = (h_\phi, l_\beta)$ abatement costs are high and damages low. Hence, the optimal global emissions, $E^{SO,(h_\phi,l_\beta)}$, are rather high, and $E^{TT}$ is too low. By contrast in state $\omega = (l_\phi, h_\beta)$ abatement costs are low and damages high. $E^{SO,(l_\phi,h_\beta)}$ is rather low, and $E^{TT}$ is too high. In the other two states, these two effects partly cancel out and $E^{TT}$ is quite close to the Global Social Optimum.

16 The parameters fulfill Assumptions 2 and 3.
2. Rules vs. Targets: Climate Treaties under Uncertainty

Figure 2.7: The black curves depict the global costs in the four different states. The vertical lines show emissions in the Global Social Optimum (black) and the Target Treaty (green).

2.5 Rules Treaty

In this section, we discuss our proposal for a climate treaty, the Rules Treaty. We start by describing how it functions, then we present the equilibrium in the ex-post stage and highlight the advantages of the Rules Treaty. Finally, we discuss treaty formation in the ex-ante stage.

2.5.1 Treaty Description

Figure 2.8 shows the functioning and sequence of events in the Rules Treaty. It is similar to the No-Treaty Outcome, as countries can freely select the number of permits, $e^\omega_i$, once the state of the world is realized. However, there are two important differences.
First, there exists an international agency that is entitled to issue an additional number of permits $E_A^\omega = \gamma \sum_{i=1}^n \epsilon_i^\omega = \gamma E_C^\omega$, where $\gamma \in [0, \infty)$ is called the scaling factor. $E_C^\omega$ denotes the aggregate number of permits issued by the countries. Hence, the total number of permits in the market is $E_T^\omega = E_A^\omega + E_C^\omega = (1 + \gamma)E_C^\omega$.

Second, the agency sells its permits to the firms. The revenues generated are fully refunded to the countries according to a refunding rule $\{\rho_i\}_{i=1}^n$, with $\sum_{i=1}^n \rho_i = 1$. $\rho_i$ is the share of the agency’s revenues refunded to country $i$. The refunding rule does not depend on the abatement efforts of countries and is independent of the state of the world.

The crucial difference between the Rules Treaty and the Target Treaty is the following: During negotiations, in the Rules Treaty countries only fix the scaling factor and the refunding rule. In the Target Treaty, by contrast, emission targets are fixed.

Formally, we define the Rules Treaty as follows:

**Definition 3 (Rules Treaty)**

A Rules Treaty (RT) consists of the following set of parameters: $\{\gamma, \{\rho_i\}_{i=1}^n\}$, $\gamma \in [0, \infty)$ is the scaling factor, and $\{\rho_i\}_{i=1}^n$, with $\sum_{i=1}^n \rho_i = 1$, is the refunding rule.

In the following, we explore the functioning of the Rules Treaty and work backwards to solve for an implementable and optimal Rules Treaty.

1. Ex-post: Permit issuance: Upon realization of the state of the world $\omega$, each country’s local planner chooses a number of permits, given $\gamma$ and $\{\rho_i\}_{i=1}^n$.

2. Ex-ante: Design of the Rules Treaty: During negotiations, countries set up a Rules
Treaty \( \{ \gamma, \{ \rho_i \}_{i=1}^n \} \) that has to be implementable and minimizes the expected global cost.

### 2.5.2 Ex-post Outcome

In this section, we describe the overall equilibrium (henceforth ‘equilibrium’) in each state of the world for the Rules Treaty. It is defined by the same three components (Nash equilibrium of the local planners’ permit choices, abatement decisions of firms, market clearing) as those introduced in Section 2.3.2 for the No-Treaty Outcome.

#### Equilibrium Characterization

In every state \( \omega \), each country \( i \) receives a refund, \( r^\omega_i \), according to its share, \( \rho_i \), of the revenues the agency generates by selling permits. It is given by

\[
r^\omega_i = \rho_i p^\omega \gamma E^\omega C.
\]

Each country \( i \) chooses a number of permits, \( \epsilon^\omega_i \), to minimize its local costs, taking the permit choices of the other countries, \( \epsilon^\omega_j \neq i \), as given.

\[
\min_{\epsilon^\omega_i} \left[ \frac{\phi^\omega_i}{2} (p^\omega)^2 + \frac{\beta^\omega_i}{2} (E^\omega_i)^2 + p^\omega (e^\omega_i - \epsilon^\omega_i) - \rho_i \gamma p^\omega E^\omega C \right], \quad i = 1, \ldots, n,
\]

subject to Equation (2.3), \( E^\omega = E^\omega_T = (1 + \gamma) \sum_{i=1}^n \epsilon^\omega_i \) and \( p^\omega = \frac{E - E^\omega_T}{\Phi^\omega} \). We use the following derivatives:

\[
\frac{dp^\omega}{d\epsilon^\omega_i} = \frac{\partial p^\omega}{\partial E^\omega_T} \frac{dE^\omega_T}{d\epsilon^\omega_i} = -\frac{1}{\Phi^\omega} (1 + \gamma), \quad \text{and}
\]

\[
\frac{de^\omega_i}{dp^\omega} = \frac{\partial e^\omega_i}{\partial p^\omega} \frac{dp^\omega}{d\epsilon^\omega_i} = -\phi^\omega_i \left( -\frac{1 + \gamma}{\Phi^\omega} \right) = (1 + \gamma) \frac{\phi^\omega_i}{\Phi^\omega}
\]

and obtain the necessary first-order conditions for a cost minimum:

\[
-(1 + \gamma) \phi^\omega_i p^\omega + (1 + \gamma) \beta^\omega_i E^\omega_T - \frac{(1 + \gamma)}{\Phi^\omega} (e^\omega_i - \epsilon^\omega_i) - p^\omega \left( 1 - (1 + \gamma) \frac{\phi^\omega_i}{\Phi^\omega} \right)
\]

\[
-\rho_i \gamma p^\omega + \rho_i \frac{1 + \gamma}{\Phi^\omega} \gamma E^\omega_C = 0, \quad i = 1, \ldots, n.
\]
The six terms represent various effects on the costs of a country $i$ if it issues one more permit. Effects 1-4 can be interpreted in the same way as in the No-Treaty Outcome, but they are amplified by the scaling factor. The new effects 5 and 6 are due to refunding.

1. $-(1 + \gamma) \frac{\partial e^\omega}{\partial \omega} p^\omega$: the decrease of the firms’ abatement costs. As there are $(1 + \gamma)$ more permits in the market, the firm increases its emissions by $\frac{\partial e^\omega}{\partial \omega} = (1 + \gamma) \frac{\partial e}{\partial \omega}$. Its abatement costs decrease by the marginal abatement costs times the increase in emissions.

2. $+(1 + \gamma) \beta^\omega_i E_T^\omega$: the increase in local damages. $(1 + \gamma)$ more permits cause an equivalent increase of global emissions.

3. $-(1 + \gamma) \frac{\partial e^\omega}{\partial \omega}(e^\omega_i - \epsilon^\omega_i)$: the change in trade revenues. The permit price decreases by $\frac{1 + \gamma}{\Phi}$. The firm has to pay less for the permits it needs to cover its emissions, $e^\omega_i$, but also receives less for the permits it sells, $\epsilon^\omega_i$. Depending on whether the firm is a net seller ($e^\omega_i < \epsilon^\omega_i$) or buyer ($e^\omega_i > \epsilon^\omega_i$) of permits, the revenues will decrease or increase, respectively.

4. $-p^\omega \left(1 - \frac{\partial e^\omega}{\partial \omega} (1 + \gamma)\right)$: the value of the remainder of the additional permit. From the one additional permits it receives, the firm needs the fraction $\frac{\partial e^\omega}{\partial \omega} (1 + \gamma)$ to cover its additional emissions, and it sells the remaining fraction $(1 - \frac{\partial e^\omega}{\partial \omega} (1 + \gamma))$ at price $p^\omega$ (if $\gamma$ is sufficiently high, it has to buy the remaining fraction).

5. $-\rho_i \gamma p^\omega$: the additional refund the country receives. The agency auctions $\gamma$ additional permits at price $p^\omega$.

6. $+\rho_i \frac{1 + \gamma}{\Phi} \gamma E_C^\omega$: the value of the original refund decreases. As the price of the permits drops, the agency makes less revenue by selling the original permits, $\gamma E_C$.

Again the first and fourth effect sum up to $p^\omega$. Hence we rewrite the first-order conditions (2.23) as

$$p^\omega (1 + \rho_i \gamma) - \frac{1 + \gamma}{\Phi^\omega} [e^\omega_i + \rho_i \gamma E_C^\omega - e^\omega_i] = \beta^\omega_i E_T^\omega (1 + \gamma).$$

(2.24)

The marginal benefits of issuing a permit (LHS) equal the local marginal damages (RHS).

In the next lemma, we provide the necessary and sufficient conditions to ensure that for a country $i$ the cost function (2.22) is convex.

---

17 For ease of presentation and comparison to the No-Treaty Outcome, we represent the marginal effects by the issuance of one more permit. We are aware that this is not exactly a marginal change.
Lemma 3 (Convexity of Cost Function)
The local cost function (2.22) is strictly convex in $\epsilon^\omega_i$ in a state $\omega$ and for a country $i$ if and only if

$$2\rho_i\gamma \geq \frac{\phi^\omega_i}{\Phi^\omega}(1 + \gamma) - 2 - \beta^\omega_i \Phi^\omega(1 + \gamma). \tag{2.25}$$

The proof of Lemma 3 is omitted as it is straightforward. Proposition 4 summarizes the properties of the equilibrium in the ex-post stage.

Proposition 4 (Rules Treaty)
Suppose we set up a Rules Treaty such that Condition 2.25 holds in every state $\omega$ and for all countries $i$. Then there exists a unique equilibrium under a Rules Treaty with the vector of permit issuance, $\{\epsilon^\omega_iRT\}_{i=1}^n$, the corresponding total number of permits issued from the countries and the agency, $E^\omega_{RT}$, the permit price, $p^{RT,\omega}$, the vector of country specific emissions, $\{e_i^{RT,\omega}\}_{i=1}^n$, and costs, $\{K^\omega_i^{RT}\}_{i=1}^n$:

$$E^\omega_{RT} = \frac{\tilde{E}}{1 + \frac{\beta^\omega}{\gamma}} \tag{2.26a}$$

$$p^{RT,\omega} = \frac{\Phi^\omega \tilde{E}}{1 + \frac{\beta^\omega}{\Phi^\omega} + \frac{n-1}{1+\gamma}} \tag{2.26b}$$

$$e_i^{RT,\omega} = \tilde{e}_i - p^{RT,\omega}\phi^\omega_i \tag{2.26c}$$

$$K^\omega_i^{RT} = \frac{\phi^\omega_i}{2}(p^{RT,\omega})^2 + \frac{\beta^\omega_i}{2}(E^\omega_{RT})^2 \tag{2.26d}$$

$$+ p^{RT,\omega}(e_i^{RT,\omega} - e_i^{RT,\omega}) - \rho_i p^{RT,\omega} \gamma e^C_{RT,\omega} \tag{2.26e}$$

The proof of Proposition 4 is given in Section 2.12. The next corollary highlights the fact that the (expected) global costs will decrease in $\gamma$. Furthermore, it characterizes the polar cases:

Corollary 1 (Properties of the Rules Treaty)

(i) In all states $\omega$,

$$\frac{dK^{RT,\omega}(\gamma)}{d\gamma} < 0 \text{ for } \gamma \in [0, \infty). \tag{2.26f}$$

(ii) If $\gamma = 0$, the outcome of the Rules Treaty is the same as in the No-Treaty
2. Rules vs. Targets: Climate Treaties under Uncertainty

Outcome. Thus, $E_{T}^{RT,\omega}(\gamma = 0) = E^{NT,\omega}$ and $K_{T}^{RT,\omega}(\gamma = 0) = K^{NT,\omega}$ for all $\omega$.

(iii) If $\gamma \to \infty$, the outcome of the Rules Treaty approximates the Global Social Optimum. Thus, $\lim_{\gamma \to \infty} E_{T}^{RT,\omega} \to E^{SO,\omega}$ and $\lim_{\gamma \to \infty} K_{T}^{RT,\omega} \to K^{SO,\omega}$.

The proof of point (i) of Corollary 1 is given in Section 2.12. Points (ii) and (iii) follow directly from Equations (2.26)a-(2.26)e by setting $\gamma = 0$ and by taking the limit $\gamma \to \infty$. 18

The intuition of Corollary 1 is as follows: If a country issues an additional permit, changes in costs — at a global level and for the issuing country — can be separated into two parts. First, the benefits from using or selling the additional permit fully accrue to the issuing country, and thus local costs decline. The additional costs, however, have to be borne by all countries. Hence the issuing country has incentives to free-ride and overissue. 19 Second, $\gamma$ additional permits are auctioned by the international agency, thus affecting global costs. As refunding shares are fixed, each country’s costs move in the same direction, albeit with different weights. 20 As a consequence, local and global interests are fully aligned. The higher $\gamma$ is, the more important is the latter part. This holds irrespective of how high the refunding share of the issuing country is.

To further fix ideas, suppose a Rules Treaty with a very high scaling factor is in place and the aggregate number of permits is as in the Global Social Optimum. If a country $i$ issues an additional permit, its local costs will decrease as described above. However, there are $1 + \gamma$ additional permits in the market. Therefore, global costs, which, per definition, have been minimal at the Global Social Optimum, will increase steeply. As each country bears a fraction of the increase of global cost, the local costs of all countries including country $i$ will increase. As $\gamma$ is very high, the latter change completely dominates the former. If a country reduces permit issuance by one unit, local costs will increase both because the country can use or sell one permit less and because global costs increase steeply. Hence, neither increasing nor decreasing permit issuance is a profitable deviation for country $i$.

18 A remark is in order here. $\lim_{\gamma \to \infty} E_{T}^{RT,\omega} \to 0$ but the individual choices of $\epsilon_{i,\omega}^{RT}$ will in general not be zero. It is, however, possible to set the refunding shares $\{\rho_{i}\}_{i=1}^{n}$ such that $\lim_{\gamma \to \infty} \epsilon_{i,\omega}^{RT} \to 0$ for all countries $i$ in some state $\omega$. Inspection of (2.26d) shows that we have to set $\rho_{i}$ in such a way that the second term vanishes in state $\omega$:

$$
\rho_{i,\omega}^{SO} = \frac{\bar{e}_{i} - \Phi^{\omega} p_{SO,\omega} (\frac{\Phi}{\Phi} + \frac{\beta}{\beta})}{E - 2\Phi^{\omega} p_{SO,\omega}}. 
$$

(2.27)

Note that it is not possible to set $\rho_{i}$ such that (2.27) is satisfied in all states $\omega$.

19 This is exactly what happens in the No-Treaty Outcome.

20 The Countries’ costs move in the same direction as long as the refunding share is positive.
Figure 2.9: Equilibrium permit choices in the Rules Treaty as a function of $\gamma$ for two different refunding rules (upper and lower part). In the upper part, refunding shares are given by Equation (2.27). The state of the world is $\omega = (h_{\phi}, h_{\beta})$. Blue solid lines: $\epsilon_{i}^{RT,\omega}$. Blue dashed lines: $(1 + \gamma)\epsilon_{i}^{RT,\omega}$. Black dashed horizontal lines are $\epsilon_i = 0$ and $\epsilon_i = \bar{\epsilon_i}$.

Example

In the following, we use the example introduced in subsection 2.3.4 to show ex-post outcomes of the Rules Treaty. In accordance with Equation (2.26d), Figure 2.9 shows that every country will decrease its equilibrium issuance of permits, $\epsilon_{i}^{RT,\omega}$, when $\gamma$ is increased. In the upper part, the refunding share for all $i$ is $\rho_i = \rho_i^{0,\omega}$, therefore $\epsilon_{i}^{RT,\omega}$ approaches zero when the scaling factor becomes very large. For any other choice of the refunding rule, $\epsilon_{i}^{RT,\omega}$ remains bounded away from zero, as exemplified in the lower part of Figure 2.9.

Figure 2.10 depicts the difference of the equilibrium in a Rules Treaty with $\{\gamma = 1, \{\rho_i^{0,\omega}\}_{i=1}^n\}$ and the No-Treaty Outcome. It is apparent that in the Rules Treaty
2. Rules vs. Targets: Climate Treaties under Uncertainty

Figure 2.10: Upper part: No-Treaty Outcome. Lower part: Rules Treaty \( \{ \gamma = 1, \{ \rho_i^{0, \omega} \}_{i=1}^n \} \). State of the world \( \omega = (h_\phi, h_\beta) \). The black curves show the cost for a country \( i \) given that the other countries choose permits at equilibrium levels. The minima of the local cost curves represent the equilibrium. Black vertical lines are equilibrium permit issuance, and red vertical lines are equilibrium emissions. Dashed black vertical lines are \( (1 + \gamma) \epsilon_i^{RT, \omega} \). Ends of the abscissas correspond to business-as-usual emissions (except for Country 1).

the emissions of countries are smaller than in the No-Treaty Outcome. As \( \gamma = 1 \), the agency will issue the same number of permits as the countries. But \( \epsilon_i^{RT, \omega} \) is still smaller than in the No-Treaty Outcome. As with the No-Treaty Outcome, countries with damage parameters lower than the average — here Country 1 — will issue more permits (directly and indirectly via the agency) than their firms emit GHGs (compare dashed line with red line).
2.5.3 Ex-ante

In the following, we consider the formation of the Rules Treaty and illustrate it with an example.

**Design of the Rules Treaty**

Suppose there is a designer aiming at minimizing the expected global cost, \( \mathbb{E}[K^{RT}] \) under the constraint that the Rules Treaty is implementable.\(^{21}\) Corollary 1 reduces the designer’s problem to finding the largest \( \gamma \in [0, \infty) \) for which an implementable refunding rule exists. For any \( \gamma \), a refunding rule \( \{\rho_i\}_{i=1}^n \) is implementable if and only if

\[
\mathbb{E}[K_i^{RT}(\gamma, \rho_i)] \leq \mathbb{E}[K_i^{NT}] \quad \forall \ i = 1, \ldots, n,
\]

where \( \mathbb{E}[K_i^{RT}(\gamma, \rho_i)] \) are the expected costs of country \( i \) in the Rules Treaty. We show in Proposition 5 that an implementable refunding rule exists for all \( \gamma \in [0, \infty) \). As a prerequisite, we calculate the minimal refunding share, \( \rho_i^{\text{min}}(\gamma) \), that guarantees the participation of a country \( i \) in the Rules Treaty for a given \( \gamma \). It is defined by \( \mathbb{E}[K_i^{RT}(\rho_i^{\text{min}}(\gamma), \gamma)] = \mathbb{E}[K_i^{NT}] \). This yields

\[
\mathbb{E}\left[\frac{\phi_i}{2}(p_i^{RT})^2 + \frac{\beta_i}{2}(e_i^{RT})^2 + p_i^{RT}(e_i^{RT} - \epsilon_i^{RT}(\rho_i^{\text{min}})) - p_i^{RT}\gamma \mathcal{E}_C^{RT}\right] = \mathbb{E}[K_i^{NT}].
\]

We solve for \( \rho_i^{\text{min}}(\gamma) \), using \( \epsilon_i^{RT}(\rho_i^{\text{min}}) \) given by Equation (2.26d). The unique solution is:

\[
\rho_i^{\text{min}}(\gamma) = \frac{\mathbb{E}\left[\frac{\phi_i}{2}(p_i^{RT})^2 - (p_i^{NT})^2 + \frac{\beta_i}{2}(e_i^{RT})^2 - (\mathcal{E}_i^{RT})^2 - \frac{n\beta_i}{B}[(p_i^{NT})^2 - \gamma + \frac{n(1+\gamma)(p_i^{RT})^2}{n(1+\gamma)(p_i^{RT})^2} + (p_i^{NT})^2] - \frac{1}{1+\gamma}(p_i^{RT})^2\Phi\right]}{\mathbb{E}[(p_i^{RT})^2]}. \quad (2.28)
\]

**Proposition 5 (Implementable Rules Treaty)**

For any \( \gamma \in [0, \infty) \), suppose that Condition 2.25 holds for all \( \rho_i \geq \rho_i^{\text{min}}(\gamma) \). Then there exists at least one implementable refunding rule, \( \{\rho_i\}_{i=1}^n \) for any \( \gamma \in [0, \infty) \). Hence the corresponding Rules Treaty is implementable.

\(^{21}\) Apart from implementability, we do not need for the purposes of this paper to model the detailed process of how refunding shares are allocated across countries. However, for completeness, we show the Nash Bargaining Solution in Section 2.7.
The proof of Proposition 5 is given in Section 2.12. Proposition 5 implies, notably, that a very large value of $\gamma$ is implementable. Hence an implementable Rules Treaty is able to approximate the Global Social Optimum. We summarize this important result in the following corollary.

**Corollary 2 (Global Social Optimum Approximately Implementable)**

For very large $\gamma$, suppose that Condition 2.25 holds for all $\rho_i \geq \rho_i^{\text{min}}(\gamma)$. Then there exists an implementable Rules Treaty that approximately establishes the Global Social Optimum.

Again the intuition for Proposition 5 and Corollary 2 is straightforward. If the scaling factor is greater than zero, global costs in the Rules Treaty will be smaller than in the No-Treaty Outcome. As we have not put any constraints on the way the refunds can be distributed, it is always possible to choose a set of refunding parameters ensuring that all countries are better off. In particular, this is true if the scaling factor is very high.

**Example**

In this section we fine-tune the intuition of Proposition 5 with the example. The upper part of Figure 2.11 shows the maximal $\gamma$ for which the refunding rule $\{\rho_1 = 0.2, \rho_2 = 0.2, \rho_3 = 0.6\}$ is implementable in the three-country example. It demonstrates that Country 1 and 2 are only better off with the Rules Treaty if approximately $\gamma < 2$. Therefore, $\gamma > 2$ cannot be implemented. However, we can increase the refund share of countries 1 and 2 at the expense of Country 3, e.g. by setting $\{\rho_1 = 0.3, \rho_2 = 0.3, \rho_3 = 0.4\}$. This is shown in the lower part of Figure 2.11. In this case, every value $\gamma \in [0, \infty)$ is implementable. Figure 2.12 depicts $\rho_i^{\text{min}}(\gamma)$ as well as the Nash-bargaining (NB) solution for equal bargaining power. Note that for the refunding rule $\{\rho_1 = 0.3, \rho_2 = 0.3, \rho_3 = 0.4\}$ used in the lower part of Figure 2.11, $\rho_i > \rho_i^{\text{min}}(\gamma)$ holds for all $\gamma$ and $i$.  

22 For each $\gamma \in [0, \infty)$, the Nash-bargaining solution, $\{\rho_1^{NB}, \rho_2^{NB}, \rho_3^{NB}\}$ is obtained by

$$\{\rho_1^{NB}, \rho_2^{NB}\} \in \arg\max_{\rho_1, \rho_2} \left\{ (K_1^{NT} - K_1^{RT}(\gamma, \rho_1))^{\frac{1}{3}} (K_2^{NT} - K_2^{RT}(\gamma, \rho_2))^{\frac{1}{3}} (K_3^{NT} - K_3^{RT}(\gamma, 1 - \rho_1 - \rho_2))^{\frac{1}{3}} \right\}$$

and $\rho_3^{NB} = 1 - (\rho_1^{NB} + \rho_2^{NB})$.

The refunding rule in that case is such that

$$K_1^{NT} - K_1^{RT}(\gamma, \rho_1^{NB}) = K_2^{NT} - K_2^{RT}(\gamma, \rho_2^{NB}) = K_3^{NT} - K_3^{RT}(\gamma, \rho_3^{NB}).$$
2.6 Rules Treaty vs. Target Treaty

2.6.1 The General Case

In this section we compare the Rules Treaty with the Target Treaty. As we have shown in Lemma 2 and Corollary 2, the Target Treaty is not ex-post optimal whereas the Rules Treaty is ex-post optimal when the scaling factor approaches infinity. However, politicians might hesitate to agree to a very high scaling factor, one possible reason being that it would deprive them of room to maneuver during an economic downturn. For such cases, the comparison we state in Corollary 3 becomes relevant.
Corollary 3 (Comparison in the General Case)

(I) If the expected global costs of the Target Treaty are no higher than in the No-Treaty Outcome, the Rules Treaty is ex-ante superior to the Target Treaty if and only if $\gamma > \gamma^{\text{crit}}$. $\gamma^{\text{crit}} \in (0, \infty)$ is the unique critical value where the expected global costs of both treaties are the same.

(II) If the expected global costs of the Target Treaty are higher than in the No-Treaty Outcome, the Rules Treaty is ex-ante superior to the Target Treaty for any $\gamma \in (0, \infty)$.

The existence of $\gamma^{\text{crit}}$ in the first part of the corollary is a consequence of two results in Corollary 2. First, for $\gamma$ very high, the expected global costs are globally optimal, i.e. minimal. Second, the global costs in the Rules Treaty increase when $\gamma$ is decreased.
2.6.2 The Homogeneous Case

Next, we compare the treaties when countries are completely homogeneous. Consider the following implementable and symmetric treaties: a Rules Treaty $RT_{\text{Symm}} = \{ \{1/n\}_{i=1}^n, \gamma \text{ very large} \}$ and a Target Treaty $TT_{\text{Symm}} = \{ \{\frac{\gamma TT}{n}\}_{i=1}^n \}$. We find the following corollary:

Corollary 4 (Comparison in the Homogeneous Case)

If countries are completely homogeneous, $RT_{\text{Symm}}$ is ex-ante Pareto superior to $TT_{\text{Symm}}$. The reason is that in both treaties the global costs, which are lower for $RT_{\text{Symm}}$ than for $TT_{\text{Symm}}$, are shared equally among countries.

2.6.3 Alternative Assumptions on Target Treaties

Finally, let us discuss how the comparison – and thus the value of $\gamma^{\text{crit}}$ – changes when we alter the two crucial assumptions of the Target Treaty, namely that it is ex-ante optimal and non-renegotiable. Suppose we relax the former assumption and consider circumstances when countries agree on an ex-ante sub-optimal target treaty. Then, ex-ante global costs would increase, causing $\gamma^{\text{crit}}$ to decrease. Thus the Rules Treaty would be even more favorable. Suppose the countries manage to renegotiate the targets from time to time. This would improve the outcome of a Target Treaty. In essence, this is the idea of the UNFCCC process. The Kyoto protocol fixed the targets for the period 2008-2012 and a follow-up treaty was supposed to fix targets for later periods. This approach turned out to be extremely cumbersome, as repeated and frequent renegotiations among a large set of countries pose an almost insurmountable challenge. Under the Rules Treaty, renegotiations are not necessary, as adjustments occur automatically.

2.7 Treaty Formation

Up to this point, the only constraint we have imposed on the Rules Treaty during treaty formation is that it has to be implementable, i.e. every country must be ex-ante better
off than in the No-Treaty Outcome. This leads to the condition that $\rho_i \geq \rho_i^{\min}$ must hold for all $i$. In the proof of Proposition 5, we showed that $\sum_{i=1}^{n} \rho_i^{\min} \leq 1$. So far we have not modeled how the remaining refunding shares, $1 - \sum_{i=1}^{n} \rho_i^{\min}$, are distributed. In the following, we remedy this omission. We assume that a Nash-bargaining (NB) game takes place and countries have bargaining power $\pi_i$. For each $\gamma \in [0, \infty)$, the Nash Bargaining game solves

$$\min_{\{\rho_i\}_{i=1}^{n}} \left\{ \prod_{i=1}^{n-1} \left[ (\mathbb{E}[K_{i}^{NT}] - \mathbb{E}[K_{i}^{RT}(\gamma, \rho_i)])^{\pi_i} \left( \mathbb{E}[K_{n}^{NT}] - \mathbb{E}[K_{n}^{RT}(\gamma, 1 - \sum_{i=1}^{n-1} \rho_i)] \right) \right]^{\pi_n} \right\}.$$ 

The resulting refunding rule, $\{\rho_i^{NB}\}_{i=1}^{n}$, is such that for all $i$ and $\gamma \in [0, \infty)$,

$$\mathbb{E}[K_{i}^{NT}] - \mathbb{E}[K_{i}^{RT}(\gamma, \rho_i^{NB})] = \pi_i \left( \mathbb{E}[K_{n}^{NT}] - \mathbb{E}[K^{RT}(\gamma)] \right).$$

As a first example, consider the extreme case where a country $j$ has all the bargaining power, i.e. $\pi_j = 1$. Country $j$’s refunding share is such that all the global cost saving from the Rules Treaty, as compared to the No-Treaty Outcome, will accrue to it. All other countries $i \neq j$ receive the refunding share, $\rho_i^{\min}$, that merely guarantees their participation. Next, consider the case where all countries are symmetric and have equal bargaining power. The resulting refunding shares after bargaining will be $\rho_i = \frac{1}{n}$ for all countries $i$. This corresponds to $RT^{Symm}$ as discussed in Chapter 2.6, if the scaling factor is very large.

### 2.8 Extreme Events

We next explore the consequences of extreme events. Suppose, there is a small, but not negligible, possibility that climate change will result in catastrophic damages. In such a situation, both the Target Treaty and the Rules Treaty will allow countries to react. However, information processing is superior in the Rules Treaty.

In a Target Treaty, the optimal ex-ante number of permits will turn out much too high, and countries may benefit from unilateral destroying permits.\(^{26}\) Two cases might occur. First, all countries have more permits than they would choose in the No-Treaty Outcome. In such a case, the outcome is the same as in the No-Treaty Outcome, as all countries profit from destroying some of their permits up to the levels of the No-Treaty Outcome. Second, some countries have more and others have fewer permits.

\(^{26}\) In the formal model, Assumption 2 excludes such cases.
than in the No-Treaty Outcome. Those who have more might choose to destroy some of their permits. At all events if information arrives indicating that marginal damages will be very high, some permits may not be used in equilibrium. In this sense, there is some reaction to new information in the Target Treaty in the case of a extreme event. However, it tends only to approximate the No-Treaty Outcome in this respect.

Countries participating in the Rules Treaty are fully able to react to information that damages are very high and correspondingly choose low numbers of permits. Because the Rules Treaty induces optimal adjustments to new information independently of its type, this holds true no matter how high damages are and how unlikely they were ex-ante.

2.9 Discussion

2.9.1 Scope of Rules Treaties

We have addressed two deficiencies in current attempts to set up international climate treaties on the basis of reduction targets: the underprovision of climate protection and the difficulties inflexibility towards new information. In a simple model we have shown that the Rules Treaty has the potential to overcome both problems. First, it can be designed in such a way that the countries have incentives to abate emissions up to a level that is optimal from a global perspective. Second, by construction, it induces countries to react optimally to new information.

Our simple model is sufficient for showing these advantages. Several extensions could provide further useful insights. Let us discuss some of them and point out those features that would be critical if the Rules Treaties were to be used in practice.

2.9.2 More General Functional Forms

The shapes of the damage and abatement cost functions are important, as marginal costs and benefits are largely determined by them. Our assumption that both functions are quadratic enhances technical convenience. The exact choice of the shape does not influence our qualitative results as long as the functional form is similar with respect to smoothness and as long as marginal abatement and damage costs are increasing.

\[^{27}\text{Smoothness means continuously differentiable abatement and damage costs.}\]
However, completely different functional forms, such as a damage function which includes a singularity – a GHG stock above which damages are extremely high –, might affect the result of the Target Treaty radically. In the Target Treaty higher abatement could be justified as an insurance against such catastrophic climate change. In our one-period model, the Rules Treaty remains unaffected by these problems. However, this may no longer be true in a multi-period model.

2.9.3 Uncertain Business-as-usual Emissions

Many countries – especially developing ones – fear that binding emission targets might threaten their long-term development objectives. Under a treaty that sets targets to counteract this problem one can set these developing countries’ targets in relation to the projected business-as-usual emissions. Such projections, however, are notoriously difficult to make. Pizer (1997) pointed out that such uncertainty is one of the major reasons why targets are not suitable. The outcome under the Rules Treaty, however, is not affected by this uncertainty of business-as-usual emissions projections, as ex-post outcomes remain optimal.

2.9.4 Multi-Period Framework

A multi-period framework would allow to address more general and gradual forms of information revelation, and to take into account the accumulation of GHGs across time. In particular, information about damages may arrive over many periods due to the inertia of the climate system. Information about abatement costs, however, could be revealed in those periods where abatement takes place. Moreover, the instantaneous damage function could be modified in such a way that the damage not only depends on the current stock of GHGs, but also on the rate of GHG accumulation in previous periods. Such extensions to multi-periods do not invalidate the use of Rules Treaties. Once the rules are in place, we will observe continuous adjustments over several periods in such a setting, to the extent that information about abatement technologies and damages are revealed and abatement efforts are distributed optimally across time.

There are, however, extensions to multi-period settings that raise new issues with regard to climate treaties, including Rules Treaties – which should be addressed in future research. As an example, let us consider technological constraints on abatement levels, due to lock-in effects. Lock-in effects are relevant in the energy system, for instance, as power plants typically have long economic lifetimes. Such constraints might imply that
emission levels cannot decrease below a certain fraction of the level reached in previous period(s).

2.9.5 Negotiation Process

There are various ways to model treaty formation. We have chosen a simple Nash bargaining process. More complex bargaining protocols could be considered. For instance, there are various types of sequential bargaining protocols that could be used to determine the refund shares in the Rules Treaty and the permit allocation in the Target Treaty.

The Rules Treaty allows to separate efficiency from equity during negotiations.\(^{28}\) The efficiency is solely embodied in the scaling factor. The distribution of global costs is independent thereof, as it is determined by the refunding rule. A change in the refunding share of one country does not affect the efficiency of the Rules Treaty; for this change has to be compensated by an adjustment of other countries’ share – as the sum of the shares adds up to one. Equity and efficiency tend to be more intertwined in negotiations about Target Treaties. A change in the target for one country simultaneously affects distribution and efficiency, unless other countries agree to offset these changes by adjusting their own targets. In this sense, the separation of efficiency and equity is easier to achieve under a Rules Treaty than under a Target Treaty.

2.9.6 Vulnerabilities of the Rules Treaty

In practical applications, several issues have to be considered. First, countries are likely to misrepresent their environmental damages and abatement costs in order to improve their bargaining position and receive a higher refund. This will hamper negotiations about the parameters in the Rules Treaty. Such attempts at deception, however, are common in any climate treaty negotiation.

Second, there are even countries that may have incentives to prevent other countries from reducing emissions. A country like Saudi Arabia, for instance, which heavily depends on the export of fossil fuels, or like Russia who, in addition, might gain from moderate climate change. Inducing participation of such countries would require large refunding shares and may force other countries to accept negative shares. There are several ways to adjust Rules Treaties in such circumstances. First, as a radical solution,

\(^{28}\) This, in turn, may increase the likelihood of success in multilateral negotiations in the first place.
such countries could be excluded from the treaty. To avoid significant inefficiencies, however, only a small fraction of countries – i.e. responsible for a small share of global emissions – could be excluded. Second, one could establish a hierarchical scheme. In a first step, countries form clusters\textsuperscript{29} and those clusters negotiate on the refunding shares and the scaling factor. In a second step, the countries within the cluster negotiate on the number of permits they want to issue as a cluster and on the allocation to individual countries. This impairs the negative influence of individual countries that are not interested in reducing emissions as long as these countries form a minority within the cluster.

\section*{2.10 Conclusion}

Rules Treaties offer an alternative to Target Treaties. While no blueprint for a treaty is a panacea, our main idea to focus on rules rather than targets could be a useful principle in further attempts to solve a global public-good problem as pressing as the mitigation of climate change.

\textsuperscript{29} These could be similar to the groups of countries that formed during the negotiations under the United Nation Framework of Climate Change: Europe, US, G77 and China, Small Island Nation States, for instance.
## 2. Rules vs. Targets: Climate Treaties under Uncertainty

### 2.11 Table of variables

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i$</td>
<td>country</td>
</tr>
<tr>
<td>$n$</td>
<td>number of countries</td>
</tr>
<tr>
<td>$SO$</td>
<td>Global Social Optimum</td>
</tr>
<tr>
<td>$NT$</td>
<td>No-Treaty Outcome</td>
</tr>
<tr>
<td>$TT$</td>
<td>Target Treaty</td>
</tr>
<tr>
<td>$RT$</td>
<td>Rules Treaty</td>
</tr>
<tr>
<td>$\bar{e}_i$</td>
<td>business-as-usual emissions of country $i$</td>
</tr>
<tr>
<td>$\bar{E}$</td>
<td>global business-as-usual emissions</td>
</tr>
<tr>
<td>$e_i$</td>
<td>GHG emissions of country $i$</td>
</tr>
<tr>
<td>$\epsilon_i$</td>
<td>emission permits of country $i$</td>
</tr>
<tr>
<td>$\mathcal{E}$</td>
<td>global emission permits, $\sum_{i=1}^{n} \epsilon_i$</td>
</tr>
<tr>
<td>$\mathcal{E}_T$</td>
<td>In RT: total amount of emission permits, $\sum_{i=1}^{n} \epsilon_i(1 + \gamma)$</td>
</tr>
<tr>
<td>$\mathcal{E}_C$</td>
<td>In RT: emission permits issued by countries, $\sum_{i=1}^{n} \epsilon_i$</td>
</tr>
<tr>
<td>$C_i$</td>
<td>abatement costs of country $i$</td>
</tr>
<tr>
<td>$\phi_i$</td>
<td>abatement cost parameter of country $i$</td>
</tr>
<tr>
<td>$\Phi$</td>
<td>$\sum_{i=1}^{n} \phi_i$</td>
</tr>
<tr>
<td>$p$</td>
<td>price of emission permit</td>
</tr>
<tr>
<td>$E$</td>
<td>global emission of GHG</td>
</tr>
<tr>
<td>$\beta_i$</td>
<td>damage parameter of country $i$</td>
</tr>
<tr>
<td>$B$</td>
<td>$\sum_{i=1}^{n} \beta_i$</td>
</tr>
<tr>
<td>$\bar{\beta}$</td>
<td>average damage parameter</td>
</tr>
<tr>
<td>$D_i$</td>
<td>(environmental) damage of country $i$</td>
</tr>
<tr>
<td>$K_i$</td>
<td>local costs of country $i$</td>
</tr>
<tr>
<td>$K$</td>
<td>global costs, i.e. $\sum_{i=1}^{n} K_i$</td>
</tr>
<tr>
<td>$\mathbb{E}[K_i]$</td>
<td>expected local costs</td>
</tr>
<tr>
<td>$\mathbb{E}[K]$</td>
<td>expected global costs</td>
</tr>
<tr>
<td>$\rho_i$</td>
<td>refund share of country $i$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>scaling factor</td>
</tr>
<tr>
<td>$r_i$</td>
<td>refund for country $i$</td>
</tr>
<tr>
<td>$\omega_\phi$</td>
<td>aggregate event (abatement costs); $\omega_\phi \in {h_\phi, l_\phi}$</td>
</tr>
<tr>
<td>$\omega_\beta$</td>
<td>aggregate event (environmental damages); $\omega_\beta \in {h_\beta, l_\beta}$</td>
</tr>
<tr>
<td>$\omega$</td>
<td>state of the world $\omega = \omega_\phi \times \omega_\beta$</td>
</tr>
<tr>
<td>$\Omega$</td>
<td>set of states of the world: $\omega \in \Omega$</td>
</tr>
<tr>
<td>$\pi$</td>
<td>probability of $h_\phi$</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>probability of $h_\beta$</td>
</tr>
<tr>
<td>$D$</td>
<td>$\pi^h B^h + \pi^l B^l$</td>
</tr>
<tr>
<td>$A$</td>
<td>$(\sigma^h + \sigma^l)$</td>
</tr>
<tr>
<td>$\rho_{\text{min}}(\gamma)$</td>
<td>given $\gamma$, refunding share where $\mathbb{E}[K^{RT}_i] = \mathbb{E}[K^{NT}_i]$</td>
</tr>
<tr>
<td>$\gamma_{\text{crit}}$</td>
<td>scaling factor where $\mathbb{E}[K^{RT}_i] = \mathbb{E}[K^{TT}_i]$</td>
</tr>
</tbody>
</table>
2. Rules vs. Targets: Climate Treaties under Uncertainty

2.12 Proofs

Proof of Proposition 1

Summing up the first-order conditions (2.10) yields

$$\tilde{E} - E^\omega = B^\omega E^\omega \Phi^\omega. \quad (2.29)$$

This can be rearranged to yield Equation (2.11a).

Next, we solve Equation (2.10) for $e_i$:

$$e^\omega_i = \bar{e}_i - B^\omega E^\omega \phi^\omega_i. \quad (2.30)$$

Plugging in Equation (2.11a) yields Equation (2.11b) □

As Equation (2.9) is strictly convex, the solution is unique. Hence the necessary conditions are also sufficient.

Proof of Proposition 2

Summing up the necessary first-order conditions (2.14) across countries, one obtains

$$p^\omega n = B^\omega E^\omega. \quad (2.31)$$

From Equation (2.5) we conclude

$$\frac{\tilde{E} - E^\omega}{\Phi^\omega} n = B^\omega E^\omega. \quad (2.32)$$

Equation (2.32) can be rewritten to yield Equation (2.15a). Plugging Equation (2.15a) into Equation (2.31) gives Equation (2.15b). To derive Equation (2.15d), we rewrite Equation (2.14) as

$$e^\omega_i = \Phi^\omega [p^\omega - \beta_i^\omega E^\omega] + e^\omega_i,$$

and plug in $E^\omega = \frac{p^\omega}{B} n$ and Equation (2.3). Thus we obtain

$$e^\omega_i = \Phi^\omega \left[ p^\omega - \beta_i^\omega \frac{p^\omega}{B} n \right] + \bar{e}_i - p^\omega \phi^\omega_i$$

$$\iff e^\omega_i = \bar{e}_i - p^\omega \Phi^\omega \left( \frac{\phi^\omega_i}{\Phi^\omega} + \frac{\beta_i^\omega - \bar{\beta}^\omega}{\beta} \right).$$

As the cost function (2.12) is strictly convex, the solution is unique. Hence the necessary
2. Rules vs. Targets: Climate Treaties under Uncertainty

conditions are also sufficient.

\[\square\]

Proof of Lemma 1

Suppose all countries \(j \neq i\) obtain permits \(\epsilon_{TT}^j\) such that \(E[K_{TT}^j] = E[K_{NT}^j]\). If \(E[K_{TT}] > E[K_{NT}]\), it must be the case that \(E[K_{TT}^i] > E[K_{NT}^i]\) for the remaining country \(i\). Hence there cannot exist an implementable Target Treaty.

If \(E[K_{TT}] \leq E[K_{NT}]\), it must be the case that \(E[K_{TT}^i] \leq E[K_{NT}^i]\) for the remaining country \(i\). Hence the Target Treaty is implementable. \(\square\)

Proof of Lemma 2

In order to be socially optimal for all realizations \(\omega\), \(E^{TT} = E^{SO,\omega}\) must hold for all \(\omega\). According to Equation (2.11a) and (2.18), this is true if \(\frac{A}{B+\Phi} = \frac{1}{1+\Phi_{\omega}}\) for all \(\omega\). This is not possible, since \(\frac{1}{1+\Phi_{\omega}}\) differs in at least two states of the world. \(\square\)

Proof of Proposition 4

Summing up the necessary first-order conditions (2.24) across countries, one obtains

\[p^\omega(n + \gamma) - \frac{1 + \gamma}{\Phi_{\omega}}[\mathcal{E}_C + \gamma \mathcal{E}_C - \mathcal{E}_T] = \mathcal{B}^\omega \mathcal{E}_{TT}^\omega(1 + \gamma).\]

As \(\mathcal{E}_{TT}^\omega = \mathcal{E}_{TT}^\omega(1 + \gamma)\), this is

\[p^\omega(n + \gamma) = \mathcal{B}^\omega \mathcal{E}_{TT}^\omega(1 + \gamma). \quad (2.33)\]

From Equation (2.5) we conclude

\[\frac{\bar{E} - \mathcal{E}_{TT}^\omega}{\Phi_{\omega}}(n + \gamma) = \mathcal{B}^\omega \mathcal{E}_{TT}^\omega(1 + \gamma). \quad (2.34)\]

Equation (2.34) can be rewritten to yield Equation (2.26a). Plugging Equation (2.26a) into Equation (2.33) gives Equation (2.26b). To derive Equation (2.26d), we rewrite Equation (2.24)

\[\mathcal{E}_{TT}^\omega = \Phi_{\omega} \frac{1}{1 + \gamma} \left[p^\omega(1 + \rho_i \gamma) - \frac{\rho_i}{\mathcal{B}^\omega(1+\gamma)} \mathcal{E}_{TT}^\omega(1 + \gamma)\right] - \rho_i \gamma \mathcal{E}_T^\omega + \mathcal{E}_{TT}^\omega + \epsilon_i^\omega,\]

and plug in \(\mathcal{E}_{TT}^\omega = \frac{p^\omega \omega n + \gamma}{\mathcal{B}^\omega},\) and Equations (2.3) and (2.5). Thus we obtain

\[\mathcal{E}_{TT}^\omega = \mathcal{B}^\omega \frac{1}{1 + \gamma} \left[p^\omega(1 + \rho_i \gamma) - \frac{p^\omega}{\mathcal{B}^\omega(1+\gamma)}(1 + \gamma) \beta_i^\omega\right] - \rho_i \gamma \left(\bar{E} - p^\omega \Phi_{\omega}\right) + \epsilon_i - p^\omega \phi_i^\omega\]

61
\[ \varepsilon_i^\omega = \frac{\Phi^\omega}{1 + \gamma} p^\omega [1 + \mu_i] - \frac{\rho^\omega n + \gamma}{1 + \gamma} \Phi^\omega \beta_i^\omega - \frac{\gamma}{1 + \gamma} \rho_i (\bar{E} - p^\omega \Phi^\omega) + \bar{e}_i - p^\omega \phi_i^\omega \]

\[ \varepsilon_i^\omega = \frac{1}{1 + \gamma} \frac{\rho^\omega \Phi^\omega}{1 + \gamma} + \rho_i \frac{\gamma}{1 + \gamma} \Phi^\omega - \frac{\gamma}{1 + \gamma} \frac{\beta_i^\omega}{B^\omega} p^\omega \Phi^\omega - \frac{n}{1 + \gamma} \frac{\beta_i^\omega}{B^\omega} p^\omega \Phi^\omega \\
- \frac{\gamma}{1 + \gamma} \rho_i (\bar{E} - p^\omega \Phi^\omega) + \bar{e}_i \left( \frac{1}{1 + \gamma} + \frac{\gamma}{1 + \gamma} \right) - p^\omega \phi_i^\omega \left( \frac{1}{1 + \gamma} + \frac{\gamma}{1 + \gamma} \right) \]

\[ \varepsilon_i^\omega = \frac{\bar{e}_i}{1 + \gamma} + \frac{\gamma}{1 + \gamma} \left( \rho_i p^\omega \Phi^\omega + \bar{e}_i - p^\omega \phi_i^\omega - \rho_i \bar{E} + \rho_i p^\omega \Phi^\omega - \frac{\beta_i^\omega}{B^\omega} p^\omega \Phi^\omega \right) \\
+ \frac{1}{1 + \gamma} (p^\omega \Phi^\omega - p^\omega \phi_i^\omega - n \frac{\beta_i^\omega}{B^\omega} p^\omega \Phi^\omega) \]

\[ \varepsilon_i^\omega = \frac{\bar{e}_i}{1 + \gamma} + \frac{\gamma}{1 + \gamma} \left\{ -\rho_i (\bar{E} - 2p^\omega \Phi^\omega) + \bar{e}_i - p^\omega \Phi^\omega \left( \frac{\phi_i^\omega}{\Phi^\omega} - \frac{\beta_i^\omega}{B^\omega} \right) \right\} \\
+ \frac{1}{1 + \gamma} p^\omega \Phi^\omega \left( -\frac{\phi_i^\omega}{\Phi^\omega} + \frac{\beta_i^\omega}{\beta_i^\omega} \right) \]

\[ \varepsilon_i^\omega = \frac{\bar{e}_i}{1 + \gamma} - \frac{\gamma}{1 + \gamma} \left\{ \rho_i (\bar{E} - 2p^\omega \Phi^\omega) - \left[ \bar{e}_i - p^\omega \Phi^\omega \left( \frac{\phi_i^\omega}{\Phi^\omega} - \frac{\beta_i^\omega}{B^\omega} \right) \right] \right\} \\
- \frac{1}{1 + \gamma} p^\omega \Phi^\omega \left( \frac{\phi_i^\omega}{\Phi^\omega} + \frac{\beta_i^\omega}{\beta_i^\omega} \right). \]

\[ \square \]

**Proof of Corollary 1(i)**

We use Equations (2.4) and (2.5) and rewrite Equation (2.7b) as

\[ K^{RT,\omega}(E^{RT,\omega}(\gamma)) = \frac{\Phi^\omega}{2} \left( \frac{\bar{E} - E^{RT,\omega}(\gamma)}{\Phi^\omega} \right)^2 + \frac{B^\omega}{2} \left( E^{RT,\omega}(\gamma) \right)^2. \]

Hence,

\[ \frac{d K^{RT,\omega}}{d \gamma} = \frac{\partial K^{RT,\omega}}{\partial E^{RT,\omega}} \frac{d E^{RT,\omega}}{d \gamma}. \]
For $\gamma \in [0, \infty)$, $E^{RT,\omega} \in (E^{SO,\omega}, E^{NT,\omega}]$. $K^{RT,\omega}(E^{RT,\omega})$ is by construction at its minimum for $E^{RT,\omega} = E^{SO,\omega}$ and convex in $E^{RT,\omega}$. Hence, $\frac{\partial K^{RT,\omega}}{\partial E^{RT,\omega}} > 0$. Furthermore, from Equation (2.26a) we obtain $\frac{dE^{RT,\omega}}{d\gamma} < 0$. □

Proof of Proposition 5
Summing up $\rho_i^{min}(\gamma)$ as given by Equation (2.28) leads to

$$\sum_{i=1}^{n} \rho_i^{min}(\gamma) = \frac{\mathbb{E}[K^{RT}(\gamma)] - \mathbb{E}[K^{NT}]}{\mathbb{E}[(p^{RT})^2 \frac{\gamma}{1+\gamma} \Phi]} + 1. \quad (2.35)$$

This sum is no bigger than 1 for any $\gamma \in [0, \infty)$, as the expected global costs under the Rules Treaty are no higher than in the No-Treaty Outcome and the denominator is positive.

Hence, for any $\gamma \in [0, \infty)$, as $\sum_{i=1}^{n} \rho_i^{min}(\gamma) \leq 1$, there exists a Rules Treaty $\{\{\rho_i\}_{i=1}^{n}, \gamma\}$ with $\rho_i \geq \rho_i^{min}(\gamma) \forall i$ which, by definition of $\rho_i^{min}(\gamma)$, is implementable. □

Proof of Lemma 4
$RT^{Symm}$ approximately implements the Global Social Optimum in all states $\omega$, and in equilibrium each country $i$ issues the same number of permits $\epsilon_i^{RT,\omega} \approx \frac{E^{SO,\omega}}{n}$. Hence, country-specific costs are distributed symmetrically in both treaties.$^{30}$ Furthermore, $\mathbb{E}[K^{RT^{Symm}}] < \mathbb{E}[K^{TT^{Symm}}]$. □

$^{30}$ If emission permits and the refunding shares are not symmetrical but chosen arbitrarily, one might imagine a case where the Rules Treaty is not a Pareto Improvement compared to the Target Treaty even though countries are symmetrical. Suppose, e.g., that $\mathbb{E}[K^{TT}] - \frac{n-1}{n} \mathbb{E}[K^{NT}] \leq \frac{E^{SO,\omega}}{n}$ and $\mathbb{E}[K^{TT}] \leq \mathbb{E}[K^{NT}]$. Then there exists an implementable Target Treaty $TT^{Asymm}$ where $n - 1$ countries $j \neq i$ obtain permits such that $\mathbb{E}[K_j^{TT}] = \mathbb{E}[K_j^{NT}]$, and country $i$ receives the remaining permits. As a consequence, country $i$ is ex-ante better off in this Target Treaty than in the Rules Treaty $RT^{Symm}$. Hence, $RT^{Symm}$ is not Pareto superior to $TT^{Asymm}$. 

63
3 Uncertainties and Climate Change

This chapter provides a brief overview of the uncertainties associated with the economics of climate change. We will first discuss issues related to environmental damages and then turn to abatement costs. We will argue that uncertainty is a key factor when discussing the consequences of climate change as well as the possibilities to mitigate them.

3.1 Uncertainties of Environmental Damages

Let us examine these uncertainties step by step from the emission of greenhouse gases (GHGs) to the final evaluation of damages as shown in Figure 3.1. They can be grouped into three main categories: uncertainties in the physical description of the earth system (steps 1-3), in the adaptive capacities of societies (step 4) and finally in the evaluation of damages (steps 5 - 6).

![Figure 3.1: Multiple uncertainties in the economic evaluation of damages caused by GHGs emissions: every link in the causal chain is associated with uncertainties.](image)

Note that we will not discuss uncertainties with regard to the emissions path humans will choose in the future. As explained in Knutti et al. (2008), for instance, this un-

---

1 Parts of Chapter 6 are complementary to this chapter.
3. Uncertainties and Climate Change

certainty is the main reason why predictions of the temperature increases until 2100 differ.

3.1.1 Physical Description of the Earth System

Step 1: Emissions of GHGs → Atmospheric stock of GHGs
Humans add a variety of GHGs to the atmosphere, the most important of them is CO₂. Once it has entered the atmosphere, it is partitioned between the main CO₂ reservoirs: the shallow and deep ocean, the atmosphere and the biosphere. The short-term partitioning is rather well constrained (Broecker et al., 1979; Sabine et al., 2004; Canadell et al., 2007; Raupach et al., 2009). Predictions of the medium and long-term behavior, however, are much less certain, because the carbon cycle is altered through climate change and there are many, yet little-understood, feedback mechanisms (Falkowski et al., 2000; Sarmiento and Gruber, 2002). That is why the future fraction of anthropogenic CO₂ that will remain in the atmosphere is difficult to assess. Consequently, the stock of GHGs for a given path of emissions is uncertain as well.

Step 2: Stock of GHGs → Global temperature change

By themselves, GHGs are not harmful. It is the climate change they cause that raises concern. A measure of climate change for any given stock of additional GHGs in the atmosphere is the resulting long-run increase in the global average temperature $\Delta T_{global}$. It is common to represent this function by a number that is easy to grasp, the climate sensitivity. The climate sensitivity is $\Delta T_{global}$ if the CO₂ concentration is kept constant at twice the pre-industrial level. In 1979, climate sensitivity was estimated to be within a range of 1.5 – 4.5 °C, with a best guess of 3 °C (Charney et al., 1979). There exists a multitude of feedbacks in the climate system whose behavior is still not well understood (Bony et al., 2006), and whose influence on the climate sensitivity parameter are notoriously difficult to predict (Roe and Baker, 2007). Thus, despite thorough research it was not possible to constrain this range any further. During the last years, on the contrary, scientists became aware of some feedbacks that are strongly non-linear. Although these remain highly speculative, there is a fear that values of climate sensitivity that are well above 4.5 °C cannot be ruled out in the long run and there is no upper bound agreed upon (Knutti and Hegerl, 2008).

Climate sensitivity predicts the temperature in the long-run equilibrium. Because the oceans warm slowly it might take centuries to fully reach that equilibrium (Roe and Bauman, 2012). Thus another important concept is the transient climate response.
3. Uncertainties and Climate Change

It is the temperature increase at the point in time where the CO$_2$ concentration is twice the preindustrial value (Frame et al., 2006). There is much more consensus about that value, as it is related to observable warming more directly, and the ill-constrained long-term feedbacks are of minor importance.$^2$

**Step 3: Global temperature change → Local changes**

While the global average temperature change $\Delta T^{\text{global}}$ is a simple way to put a lot of information into a single number, it can also be misleading. The actual damages in a specific country $i$ depend on local changes in a variety of climatic factors, $\Delta X_i$, including temperature, precipitation (Trenberth et al., 2003) – and sea level change, for instance. The further we down the prediction scale to local level, the higher the uncertainties (Christensen et al., 2007). As to temperature, the local changes tend to be more pronounced over land and less pronounced over oceans than the global ones, as the oceans have a larger heat capacity. Thus, the countries’ temperature changes are higher than the global number might on average suggest. The most extreme example is the arctic region, whose temperature increase will be much more extreme than the global one because reflective snow and ice will be replaced by open ocean and land (Meehl et al., 2007).

In a warmer world, the sea level, $\Delta SL$, will rise because of two effects. First, water expands when becoming warmer. Second, the run-off from melting glaciers and ice sheets increases the water volume in the ocean. Predictions of the latter are extremely ill-constrained and are subject to much uncertainty (Rahmstorf, 2010). Although there are slight local differences, sea level rise has little regional components.

Smooth projections of climate change may give a wrong impression of the real danger. First, the prediction of average changes masks the fact that the transition into a warmer world will be accompanied by more extreme events (Katz and Brown, 1992; Meehl et al., 2007). Second, there are several tipping points in the climate system, none of which can be predicted precisely (Lenton et al., 2008). A tipping point occurs when a certain property of the climate system reaches a point where strongly nonlinear effects start to take place, which dominate the external forcing. Then a small perturbation suffices to cause an abrupt change into a different stable state. An example is the Greenland ice sheet. Warming at the periphery lowers the altitude of the ice sheet as a whole. More melting takes place because of the higher temperatures at lower altitudes. This

---

$^2$ This is why Allen and Frame (2007) criticize the standard approach, which concentrates on abatement targets of GHGs, based on a best guess of climate sensitivity, that is afterwards obeyed strictly. Instead, they recommend to concentrate on a temperature target and adopt emissions accordingly. This suggestion is in line with our proposal made in Chapter 2.
causes a positive feedback loop which, at a certain point, cannot be reversed, even if global temperatures went back to preindustrial levels. Only colder temperatures such as the ones of the last ice age, could reverse this trend. Most tipping points are potentially harmful to humans, although there are also examples of beneficial ones, such as the strengthening of the West-African monsoon, leading to more rainfall in the Sahel/Sahara zone. Tipping points can also be societal, such as higher food prices caused by droughts, which can contribute to revolutions, such as the ones of the Arab Spring.

### 3.1.2 Adaptive Capacities of Societies

**Step 4: Local changes → Local impacts**

The importance of local impacts depends not only on the local characteristics of climate change, but also on the vulnerability and the socio-economic development of the countries concerned. Damages will rise steeply if the adaptation capabilities of the affected societies are exceeded. Consider the example of the Netherlands and Bangladesh. Both will be highly affected by the rising sea level, but the Netherlands are more able to handle such consequences, as the country is rich and is able to build dikes at a sophisticated level, as well as to implement other adaptation measures.

### 3.1.3 Evaluation of Damages

**Step 5: Local impacts → Economic valuation of local impacts**

The economic valuation of local impacts is a highly controversial issue. Even if we could perfectly predict how the climate changes at regional level and what the extent of its impacts will be, how to measure damages remains a controversial issue.

It is easy to determine direct damages of a certain weather extreme, for instance. A causal relationship between a specific event and climate change is more difficult to make and is only possible through a long-run mean – if at all. Damages due to more gradually changes such as the see-level rise have to be valued as well. In addition, there are impacts on the environment that cannot be measured directly in monetary terms. Nevertheless to include these effects into a cost-benefit framework, one has to give them some monetary value. There exist several methods to do so, e.g. the willingness to pay for these changes not to occur or willingness to accept compensation for these changes to occur, but none of them is undisputed in the literature (Fankhauser et al., 1997). Non-economic impacts might play an even bigger role when the environmental degradation
3. Uncertainties and Climate Change

grows, because the relative importance of environmental goods will rise (Sterner and Persson, 2008). Our current spending for food, for instance, greatly understates our willingness-to-pay for food.

**Step 6: Local damages across time → Net present value of global damage**

Suppose we could agree on how to measure local damages, and we could predict these perfectly. The question remains how to aggregate these local damages across countries and time. Such aggregation is necessary if we want to compare the impact of climate change with the cost of moderation. As this is an ethical question, it cannot be based on natural science.

There is also the issue to add up damages across countries. Assuming that the current global wealth distribution is unfair, the monetary valuations of local environmental damages in different countries have to be adjusted to these asymmetries. A loss of, say, 1$ in a poor country is more serious than a loss of 1$ in a rich country. To account for this problem, equity weighting, $w_i$, was proposed (Fankhauser et al., 1997). The poorer a country, the higher the weight of damages in that country. The use of equity weights has met criticism on the ground that equity considerations should not be linked to climate change issues.

One also has to determine the weight of future generations, which boils down to the appropriate choice of the consumption discount rate $\rho$. As benefits are will be reaped in the future, whereas cost occur earlier, the choice of $\rho$ is very important. The lower it is, the higher the weight of future generations and the more abatement must be done by the present generation.

It is common to highlight two influences on $\rho$. The first is the pure rate of time preference, $\delta$, which reflects impatience. The second is the parameter $\eta$ which reflects the degree of inequality aversion and risk aversion at the same time. The effect of $\delta$ is unambiguous: the higher it is, the higher is $\rho$. The effect of an increase in $\eta$ depends on the framework chosen and is ambiguous. One the one hand, a higher $\eta$ results in more inequality aversion. As the future is assumed to be richer than the present, $\rho$ increases. On the other hand, considering the current and future inequality across countries, $\rho$ decreases when $\eta$ increases (Heal, 2009). In addition, a higher $\eta$ also means more risk aversion. As we are uncertain about the consequences of climate change, a risk averse

---

3 It does not satisfies the equity concerns of society.

4 One could go one step further and aggregate utilities at the level of individuals, but the underlying problems remain the same.

5 There is an large literature on this topic. A good starting point is Dasgupta (2008).

6 Most of the literature assumes exogenous economic growth.
society rather errs on the side of caution, so that \( \rho \) decreases (Gollier, 2002). Which of these effects dominates depends on the specifications of the model that is used.

The role of uncertainty has attracted much attention lately. The first attempts to model the economics of climate change mostly disregarded uncertainty and used a deterministic approach (Nordhaus, 1980). Weitzman (2009) – among others – has shown that these deterministic models might be flawed, as uncertainty, in combination with risk aversion, can easily outweigh all other influences. The fundamental reason for this is that when dealing with climate change, one does not know the probability distribution of the relevant parameters. The most famous example is the distribution of the climate sensitivity parameter which – as already discussed – has no upper bound agreed upon.

In the model of Weitzman (2009) risk averse agents are theoretically willing to spend an infinite amount to prevent climate change from happening because of these fat-tailed distributions. This illustrates the importance of uncertainty. It also shows that expected utility theory reaches its limits when dealing with this kind of structural uncertainty, so that other approaches might be more useful. One such approach is the smooth ambiguity model (Klibanoff et al., 2005, 2009) which has been applied in the context of climate change in Heal (2009), for instance.

### 3.1.4 Summary

It is common to calculate the expected net present value of the global damages, \( D \), due to a certain path of global emissions, \( E = \{E_t\}_{t=0}^\infty \), with a formula in the form of Equation (3.1), which integrates many of the issues discussed above.

\[
E[D(E)] = E \left[ \sum_t \left( \frac{1}{1 + \rho} \right)^t \sum_i w_i^t D_i^t \left( \Delta T_i^t(E), \Delta Precip_i^t(E), \Delta SL_i^t(E), ... \right) \right]. \tag{3.1}
\]

The social costs of carbon is a related concept. It is the damage caused by the emission of an additional ton of carbon at a given time, and it is calculated by summing up the discounted total damages caused by this ton as long as it remains in the atmosphere. Most estimates use intertemporal optimization in a cost-benefit framework where the SCC are the shadow price of carbon along the optimal emission path.

In the literature, estimates of the social costs of carbon spread over a huge span and there is no consensus as to the bounds of this span – a further illustration of the huge uncertainties connected to climate change. In Tol (2005)’s review of 103 SSC estimates,
the mean is 93, and a median is 14 US$/tC. The lowest estimates are slightly negative, while on the other side there are estimates well above 1000 US$/tC.

To find out where such a wide range comes from, Hope (2007) determined the influence of a range of parameters on the social costs of carbon in the PAGE2002 model. He found that six scientific and seven economic parameters have a major influences:

1. Climate sensitivity
2. Pure rate of time preference
3. Non-economic impact parameter
4. Equity weight parameter
5. Half life of global warming
6. Economic impact parameter
7. Proportion of CO$_2$ emitted to air

All the listed parameters have either a substantial degree of uncertainty or are disputed, which explains the huge range of SCC and the overall uncertainty associated with environmental damages.

### 3.2 Uncertainties of Abatement Cost

In this section we want to discuss the uncertainties associated with the estimation of abatement costs. They are not as high as those of environmental damages, but are substantial nevertheless. There are several reasons why estimates differ.\(^7\)

In different sectors, emission reductions may be done at different costs, the most important sectors being: energy supply, transport, buildings, industry, agriculture, forestry, and waste management. It is difficult to tell where – and how much – reduction will be done in which sectors, and correspondingly, how high the aggregate costs are.

There are two major methods to model abatement costs. The bottom-up approach estimates the capabilities of different sectors to reduce emissions and sums up the resulting costs. The top-down approach works at the macroeconomic level, uses aggregated parameters and usually predicts the cost as a percentage of GDP. Because the approaches are so different, results potentially differ as well.

\(^7\) Metz et al. (2007) provide a thorough overview over many of the aspects we touch upon.
The cost estimates of economic models differ because every model has its own way to incorporate the multitude of factors that influence future abatement costs. Examples are the choice of the discount factor, the effect of efficiency measures, the predicted costs of fossil fuels and the existence of no-regret options. Yet, one may assume that the most important factor is technological change.

There already exist many ways to decarbonize the energy system, such as solar or wind power. However, it is still uncertain whether and at which costs renewable energy systems can replace the systems based on fossil fuels. From the medium to long run, an additional element of uncertainty is the discovery of yet-unknown or hypothesized ways to generate energy, such as fusion power. In addition, some models start form the assumption that the technological change is completely exogenous and correspondingly recommend to wait until abatement becomes less expensive due to the availability of such improved technology. Others take the effects of learning-by-doing into account, and recommend to start significant abatement efforts now without waiting for the ‘deus ex machina’ of exogenous technological change (Ackerman et al., 2009).

Instead of focusing on the monetary cost of reducing GHGs alone, one could also define costs in more general terms. One component of the abatement costs could be renouncing the consumption of polluting goods. In such a framework, parts of the costs are measured by how a lower level of consumption affects utility. For example, a less carbon intensive way of living comes at a great cost for people who are used to driving fuel-consuming cars or to flying a lot. A price increase of the kind of mobility that relies on fossil fuels would be costly for them. To the extent that they get used to a different life-style their loss of utility – and thus the abatement costs – would decrease. Such a projection of abatement efforts on utility levels is an additional element of uncertainty.

Taking co-benefits of mitigation into account also reduces abatement cost for society. Reduced car traffic, less local pollution, less noise, less NO\textsubscript{x}-emissions are common examples. The abating firms often do not take into account such positive externalities. Of course, there might also be negative effects of emission reductions: Substituting fossil fuels with biological ones, for instance, reduces food production.

Finally, while many studies focus on CO\textsubscript{2}, other GHGs also contribute to global warming significantly, such as N\textsubscript{2}O, CH\textsubscript{4} and FCCs. Including them into abatement efforts allows to reach a certain reduction target at a lower cost, as marginal abatement efforts are usually increasing for each gas and as reduction might be less costly for non-CO\textsubscript{2} GHGs.
4 Small Coalitions and Climate Change

4.1 Introduction

4.1.1 Motivation

Reducing the emissions of greenhouse gases (GHGs) is a global public good. Yet no international authority can force countries to reduce their emissions above and beyond what is in their national interest. A possible solution to this well-known free-riding problem is to unite the countries’ particular interests under a climate treaty. With a treaty, however, the free-riding problem resurfaces—in the form of the participation and compliance problem. Climate treaties, that is, have to be designed in such a way that countries participate voluntarily and stick to their promises. So far, the literature primarily dealt with the question which coalition structure arises under such circumstances. A strand of research focuses on whether a climate treaty comprising all countries can be sustained (Chander and Tulkens, 1995; Tulkens, 1997). The answer is affirmative if one assumes that if a single country or a group of countries leaves the coalition, the remaining countries will stop cooperation and continue as singletons. This is a punishment strong enough to deter deviators. It has been questioned, however, whether this kind of punishment is credible (Barrett, 2005). Once a country has deviated, it might be in the best interest of the remaining countries to continue cooperation. This idea is captured in the concept of internal and external stability (D’Aspremont et al., 1983). It assumes that if countries enter or leave a treaty, the other countries do not punish actively. They merely adjust emissions to reflect the changed coalition size. Thus, in this strand of the literature, leaving a coalition has no severe consequences, and countries will rather choose not to participate in a treaty while free-riding on the efforts of the other countries. As a result, stable coalitions sizes are very small (Carraro and Sinis-
A third strand of literature uses the equilibrium concept of ‘Farsighted Stability’. It assumes that a country leaves a coalition if it is worse off in the current situation than in the new equilibrium coalition structure that obtains upon deviation (Chwe, 1994; Ray and Vohra, 2001; De Zeeuw, 2008). This makes free-riding more costly than if the remaining countries merely readjust emissions, but less costly than if all countries leave the coalition. As a result, stable coalitions of various sizes emerge.

These lines of research are important to understand why climate coalitions do not form easily. They remain silent, however, on outcomes for the case a coalition of a certain size has already formed. To the best of my knowledge, this has not been discussed in the literature yet. In this paper, therefore, we do not model the participation game, but take its outcome as given. We acknowledge that small coalitions ought to be considered more stable and put a special focus on them. However, considering the various stability concepts explained above and taking into account that there might be additional factors that potentially stabilize a coalition, we take it that in principle a stable coalition of any size could arise.

The starting point of our model is thus an already existing group of countries forming a climate treaty. This group could be of any size and its members will not leave the climate treaty. The remaining countries, the outsiders, act as singletons. We assume, that is, an exogenously given and stable coalition structure.

As in the literature cited above, we assume that members set their emissions taking into account merely the impact on other members, but do not consider outsiders’ damages. A possible extension to this paper would thus be to assume that coalition members are altruistic and take damages inflicted upon outsiders into account as well.

A detailed analysis of the interaction between members and outsiders, and of its variations with the group size allows us to pinpoint several novel effects. These findings might be relevant for such a group when contemplating to set up a climate treaty.

We put a special focus on uncertainty with respect to abatement costs and climate damages, as such uncertainty is pervasive in the context of climate change. The possibility to react to new information seems to be a crucial advantage when designing a climate treaty (Gersbach and Oberpriller, 2012). Commitment to an emission-reduction target, on the other hand, might give members a strategic advantage. As we will show, which

---

1 For a detailed discussion of issues related to stability of international agreements, see also Finus (2003, 2008).
2 See Section 4.1.2.
effect dominates depends both on the coalition size and on the countries’ vulnerability to climate change.

4.1.2 The Stability of Climate Coalitions

Suppose there is a climate coalition of a size that would not be considered stable in the sense of Barrett (1994). Considering costs of reducing emissions and local environmental damages, the story goes, at least one country would have incentives to leave that coalition and other countries cannot stop it, as their only means of punishment is to adjust their emission levels. Here we identify some reasons why such a coalition might be stable nevertheless. These considerations will not be included into the formal model.

The membership in a climate treaty might be linked to topics where countries benefit from the integration into a global framework. As an example, consider a treaty that links climate and free-trade. Leaving such a treaty, a country on the one hand benefits from free-riding with respect to emission abatement but, on the other hand, deprives itself of the benefits from free-trade. There might still be countries not willing to participate in such a treaty, but it is not unreasonable to assume that a stable coalition size might comprise a significant part of the world.

The climate treaty could be subject to a minimum participation clause (Black et al., 1993). This means that a treaty enters into force only if a predefined number of participants is reached. This, in turn, raises the question how this number came about. Still, a minimum participation clause is relevant in practice. Most notably, it was used for the Kyoto Protocol.4

Finally, negotiations are not always controlled by perfectly self-interested countries. Out of a sense of responsibility, several countries could be determined to act. The EU, for example, announced to reduce its GHGs emission until 2050 compared to 1990 levels by 80–95%.5 Such an unilateral and stark promise is not perfectly consistent with our assumption that coalition members merely consider their own environmental damages, but certainly includes elements of altruism towards other countries. It it shows in any

---

3 For the sake of argument we only consider external stability.
4 The Kyoto Protocol was meant to come into force only if 55 parties were to participate. In addition the participants belonging to the Annex 1 Group were to account for at least 55% of the emissions in 1990 in the Annex 1 Group.
4. Small Coalitions and Climate Change

case that countries are not merely driven by economic interests.

4.1.3 The Set-Up

We use a simple model where the world is divided into several countries which experience local costs for two reasons: they suffer from (environmental) damages from the emission of GHGs and they invest into costly abatement efforts. Countries are homogeneous with respect to damages and abatement costs. There is a fixed coalition structure, consisting of one group of member countries that cooperate and set up a climate treaty, whereas the remaining outsiders act as singletons. Obviously, for this separation to arise in the first place, in the – not modeled – preceding participation game, countries have to be heterogeneous in some respect.

As depicted in Figure 4.1, we examine several scenarios, which have a varying degree of complexity and divide the game into three stages. In the third stage, damages and abatement takes place. We distinguish between scenarios of constant and increasing marginal damages (not shown in Figure 4.1) and abatement decisions are taken either simultaneously or sequentially. In the latter case, members move first in stage 3.1 and outsiders move second in stage 3.2. We refer to the former as the simultaneous-move game (in abatement decisions) and to the latter as the sequential-move game (in abatement decisions). Countries always act under certainty in Stage 3.

The set-up of the climate treaty occurs under certainty or uncertainty, respectively. Under certainty, the state of the world realizes in Stage 1 and the treaty is set up in Stage 2. Under uncertainty, the climate treaty is set up in Stage 1, at which time countries know the probability distribution of the environmental damages and abatement costs. In Stage 2 uncertainty is resolved. This separation allows us to model the effects of uncertainty and learning.

Members set up a treaty under which they fully cooperate and comply with their promises. Under certainty, there is only one such treaty, to which we will simply refer to as the ‘coalition’ (C). Under uncertainty, however, treaties can differ with respect

---

6 It would be straightforward to relax this assumption and allow for heterogeneous countries with respect to damages and abatement costs. However, it would complicate the analysis and contribute little to our main agenda.

7 Even though it might be demanding for the reader to keep track of all of them, this approach is useful. The simple scenarios allow us to spotlight several effects. The more complex ones allow us to see how these effects interact, and they give rise to some new effects.

8 The game as a whole is always sequential, as there are three stages; by ‘simultaneous’ and ‘sequential’ we only refer to the abatement decision in the third stage.

9 This approach is similar to Kolstad and Ulph (2008, 2011) and Gersbach and Oberpriller (2012).
4. Small Coalitions and Climate Change

Figure 4.1: Sequence of events and description of scenarios.
1) Certainty/Uncertainty,
2) Simultaneous/Sequential-move game,
3) Linear/Quadratic damages.

To how much flexibility they grant to countries – or, as other side of the medal – to which degree countries commit. To highlight the difference, we compare two extreme scenarios. Under the ‘Flexibility Treaty’ (FT), member countries react optimally to newly-arriving information. Under a FT, that is, a member promises to do whatever is optimal for all members whenever new information arrives.\footnote{In Gersbach and Oberpriller (2012), we propose a set of rules that establishes this property.} Under a ‘Commitment Treaty’ (CT) member countries commit to abatement targets when the treaty is put into force. The members’ targets must not be changed, even if new information would demand it. This separation allows us to compare the importance of learning, on the one hand, and the advantages of commitment, on the other hand. In the simultaneous-move game, we assume that the outsiders of a CT commit to abatement efforts in Stage 1 as well. In the sequential move scenario, we assume – more realistically – that outsiders cannot commit, and choose their abatement efforts under certainty in Stage 3 instead. In Table 4.1 we summarize, for the scenario under uncertainty, the different stages at which members and outsiders decide about their abatement, i.e. they either commit to abatement or directly abate.
4. Small Coalitions and Climate Change

<table>
<thead>
<tr>
<th>Scenarios:</th>
<th>Simultaneous</th>
<th>Sequential</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treaties:</td>
<td>FT</td>
<td>CT</td>
</tr>
<tr>
<td>Members</td>
<td>Abate in Stage 3</td>
<td>Commit in Stage 1</td>
</tr>
<tr>
<td>Outsiders</td>
<td>Abate in Stage 3</td>
<td>Commit in Stage 1</td>
</tr>
</tbody>
</table>

Table 4.1: Stages where abatement decisions take place under uncertainty.

4.1.4 Results

Let us show the most important results of the various scenarios step by step. As results of simple scenarios often carry over to more complex ones, we focus on newly-arising effects of each step.

Simultaneous-Move Game

In the simultaneous-move game, the outsiders are always better off than the members – a manifestation of the free-rider problem.

Linear Damages Under Certainty

For all countries – members and outsiders –, local costs decrease when the coalition size rises, because more cooperation is better for all; we call this the ‘Cooperation Effect’.

Linear Damages Under Uncertainty

For any given coalition size, the expected local costs are lower for all countries when a FT is in place than under a CT. This is because countries can exploit the fact that they act under certainty; we call this the ‘Learning Effect’. Furthermore, suppose that a CT is in place. If the uncertainty is high, or the coalition size is small, members would jointly prefer not to be part of a CT at all, as acting as singletons allows them to act under certainty instead of having to commit to an abatement effort ex-ante. In such a situation, the Learning Effect is more relevant than the Cooperation Effect.

Quadratic Damages Under Certainty

If damages are quadratic, the countries’ abatement efforts are not independent anymore; for the marginal damages within a country increase when other countries decrease their abatement efforts. In this setting, the members’ local abatement efforts first increase and then decrease with coalition size.\(^\text{11}\) To see this, consider an increase in the coalition size by one member. If the abatement efforts of the remaining outsiders would stay the same, the grown coalition size would result in an increase of members’ local

\(^{11}\) This is true unless the vulnerability to climate change is unreasonably small, in which case members’ local abatement efforts monotonically increase with the coalition size.
abatement efforts in accordance to the Cooperation Effect. The remaining outsiders, however, reduce their abatement efforts as, roughly speaking, there is now one country more to free-ride upon. To compensate for the leakage, members increases local abatement levels for all coalition sizes compared to a situations where no leakage would take place. We call this the ‘Leakage Effect’. The smaller the coalition size, the stronger this effect. Thus members of a small coalition have high local abatement levels. If a certain coalition size is reached, the number of outsiders is small enough, so that leakage does not play such a dominant role any more. Thus, being already at a high level, the members’ local abatement levels, in fact, decrease with the coalition size. The easing pressure to compensate for leakage decreases the members’ local abatement efforts – more than the Cooperation Effect increases them.

Because of this particular behavior of the members’ local abatement effort, members’ local costs also first increase and then decrease with the coalition size. As a corollary, for a small enough coalition size, members are worse off than if there would be no climate treaty at all.

A further result of this setting is that, if the coalition size is small and countries are quite vulnerable to climate change, the members’ local costs might decrease with the countries’ vulnerability. The reason is that the abatement efforts of the many free-riders are higher if their vulnerability is higher. As a result, global emissions decrease. This effect is big enough to compensate the members’ increased vulnerability and rising abatement costs.

*Quadratic Damages Under Uncertainty*

Under a CT, all countries’ abatement commitments are higher than their expected abatement effects under a FT, as they act more cautiously in the CT. Thus, by choosing a CT, members are able to decrease the outsiders’ free-riding to a certain degree. We call this the ‘Positive Uncertainty Effect’. If the coalition size is small and, at the same time, countries are quite vulnerable to climate change, the coalition members might rather opt for a CT than for a FT, even though this deprives them of the possibility to use new information. The Positive Uncertainty Effect outweighs the Learning Effect, as outsiders are the majority and damages have a relatively high importance as compared to the abatement costs.

---

12 Remember that countries are homogeneous. Thus, we refer to the increase of vulnerability in all countries. It would, of course, never be beneficial for a country if only its own vulnerability were to increase.
4. Small Coalitions and Climate Change

Sequential Move Game

Linear Damages Under Certainty
This setting corresponds exactly to the simultaneous-move game.

Quadratic Damages Under Certainty
With increasing marginal damages, members benefit from being able to move first, as the outsiders will partly compensate any decrease in the members’ abatement effort. Compared to the simultaneous-move game, global abatement efforts decrease and correspondingly, damages increase in all countries. Nevertheless, members are better off, as the decrease in abatement costs compensates the increase in damages. This is the ‘General First-Mover Advantage’. Outsiders, however, are worse off, as both damages and abatement costs increase. If the coalition size is small, the General First-Mover Advantage is strong enough to make members better off than outsiders.

This result indicates that in the real world, it might be beneficial for single countries or small groups of countries to be the first ones that commit to a low abatement level, by building coal power plants, for instance.

Linear Damages Under Uncertainty
As marginal damages are constant, the abatement efforts of members and outsiders are independent of each other and there is no General First-Mover Advantage. Still, if a CT is in place, members benefit from the fact that outsiders act under certainty and are thus better off than in the simultaneous-move game. We call this the ‘Certainty First-Mover Advantage’. If the CT has few members and uncertainty is high, members, in expectation, might even be better off than outsiders. For members of a FT, there is no Certainty First-Mover Advantage.

Quadratic Damages Under Uncertainty
In this scenario, the analysis becomes too involved to give closed-form solutions. Numerical results indicate that there is an additional advantage for members of a CT. Knowing that outsiders act under certainty not only makes members better off for any given abatement effort, it even allows them to commit to less abatement effort. We call this the ‘Abatement First-Mover Advantage’. The more outsiders there are, the stronger this effect is. The Certainty First-Mover Advantage and the Abatement First-Mover Advantage together might outweigh the Learning Effect, especially for small coalition sizes, in which case the CT is more attractive for members than the FT.

Outsiders are always worse off under a sequential-move game, as compared to a simultaneous one. In the real world, they will thus try to act simultaneously. This is, however,
4. Small Coalitions and Climate Change

particularly unlikely in the case of the CT, as outsiders cannot credibly commit – for the very reason that they are not part of any treaty. For the FT, on the other hand, simultaneous action from outsiders is more likely, as this merely means that they have to reduce emissions at the same time as the members.

4.1.5 Uncertainty and Learning in the Literature

Our paper connects to current research from several areas with regard to uncertainty and climate change. Let us briefly highlight relevant findings and show how our proposal complements the existing literature.

It is well-known that under a single global planner, the value of additional information is positive. The reason is that the global planner could simply choose to ignore new information (Lange and Treich, 2008; Kolstad and Ulph, 2008). Put into our context, a global planner would rather act ex-post, when learning has taken place, than ex-ante. For the same reason, the FT is superior to the CT if the climate treaty includes all countries. In such a case, the CT corresponds to the ex-ante global planner solution, whereas the FT corresponds to the ex-post global planner solution. We extend this literature for cases where only a subset of countries cooperates, and show that, for members, the value of information might be negative if the coalition size is small and vulnerability to climate change is high. Globally, the value of information remains positive.

The information structure we use to determine the effects of uncertainty and learning is very similar to the one in Kolstad and Ulph (2008, 2011). The treaty setup (Stage 1) is separated from abatement (Stage 3) by the resolution of uncertainty (Stage 2).\(^\text{13}\)

Kolstad and Ulph (2008) compare three scenarios: complete learning, partial learning and no-learning. Under complete learning, coalition formation and abatement decisions occur under certainty. Under partial learning, coalition formation takes place under uncertainty and abatement decisions under certainty. Under no-learning, both are under uncertainty. The case of partial learning corresponds to the FT, whereas no-learning corresponds to the CT.\(^\text{14}\) Kolstad and Ulph (2008) find that expected global costs under

\(^{13}\) We do not intend to advance the literature on the question whether, under the prospect of future learning, there should be more or less abatement in the present (Lange and Treich, 2008; Ulph and Ulph, 1997). Thus there is no abatement under uncertainty in our model.

\(^{14}\) Under both treaties, the countries have the same amount of information when they abate, while under the CT, the countries have credibly committed to targets that have been set before learning took place, so that those countries cannot use this information. In this sense, the CT is equivalent to the no-learning scenario.
a scenario where there is (partial or complete) learning might, in fact, be higher than under a no-learning scenario – which renders the value of information negative. This is complementary to our results as, apart for the information structure, their model set-up is quite different from ours.

In Kolstad and Ulph (2008), the value of information might be negative if instead of a single global planner, there were many local planners which interact non-cooperatively in both the coalition formation game and the abatement decisions.15 The equilibrium coalition size decreases when (expected) environmental damages are higher,16 which, in turn, increases the expected global costs under learning as compared to no-learning.

This is because in the states where environmental damages turn out to be high, there is, in addition, less cooperation. Thus, global costs are a convex function of the damage realization. In the model of Kolstad and Ulph (2008), abatement choices are discrete (full/no abatement), and it is assumed that for each country, full abatement would be the globally optimal choice in all states of the world. This allows the authors to concentrate on the effect described above. Yet, it masks the beneficial effect of learning we describe in our model, namely the state-dependent adjustment of the optimal abatement effort. In addition, we assume a fixed coalition size in both treaties, whereas the result of Kolstad and Ulph (2008) crucially depend on the fact that the sizes may vary.

Dellink et al. (2007) compare Kolstad and Ulph (2008)’s ‘complete and no-learning’ scenario, using a more realistic numerical model with continuous abatement possibilities and heterogeneous countries. They come to the conclusion that results are sensitive to the underlying assumptions, and thus the possibility of learning does not necessarily yield a negative global impact. In addition, the differences are rather small in absolute terms, as the concept of external/internal stability does not allow big coalition sizes, no matter whether learning takes place or not. To the best of our knowledge, there is no study that models partial learning versus no-learning under more realistic assumptions,17 but we expected that this conclusion would not be different.

15 In Kolstad and Ulph (2008), the number of signatories in the coalition is determined according to the internal and external stability criteria.
16 Also see Barrett (1994).
17 This is the more relevant comparison: we cannot think of a real-world situation corresponding to the complete learning scenario in cases where uncertainty is relevant nevertheless.
4. Small Coalitions and Climate Change

4.1.6 Organization of the Paper

The remainder of the paper is organized as follows: In Section 4.2 we describe the set-up of the model. In Section 4.3 we define and derive closed-form solutions for the Global Social Optimum and the No-Treaty Outcome. In Sections 4.4 and 4.5, we examine the simultaneous-move game for linear and quadratic damages, under certainty and uncertainty. The same is done for the sequential-move game in Section 4.6. Section 4.7 concludes.

4.2 The Set-Up

There are $N$ countries that are homogeneous with respect to the (environmental) damages and abatement costs that we consider in the model. The local costs of a country are thus

$$k = \frac{\beta}{L}(\bar{E} - A)^L + \frac{1}{2}a^2,$$

where $\bar{E}$ are global business-as-usual emissions, $A$ is the global abatement effort and $a$ the local abatement effort. $\beta > 0$ is the environmental damages parameter, and depicts the countries’ vulnerability to climate change. Under certainty, $\beta$ is known, under uncertainty it is a random variable. $L \in \{1, 2\}$ is a parameter that determines whether damages are linear or quadratic. Abatement costs are quadratic. An exogenously-given number of countries $s$ ($1 < s \leq N$) forms a climate treaty.

Under certainty, coalition $C$ forms. Under uncertainty, countries know the probability distribution of the damage parameter when setting up the treaty. We consider two types to climate treaties: a Flexibility Treaty (FT) and a Commitment Treaty (CT).

---

18 We do not consider which instrument a country uses to abate. A permit market, a carbon tax or command-and-control could be imagined.

19 We could have written the local costs as $\frac{\tilde{\beta}}{\phi}(\tilde{E} - A)^L + \frac{\tilde{\phi}}{2}a^2$, with $\tilde{\beta}$ being the damage parameter and $\phi$ the abatement cost parameter. Dividing by $\phi$, rescales the local costs with $\beta = \frac{\tilde{\beta}}{\phi}$ being the ratio of damage to abatement cost parameter. Without loss of generality, we set $\phi = 1$, such that $\beta = \tilde{\beta}$. Thus we refer to $\beta$ simply as the environmental damage parameter.

20 Note that the value of $\beta$ is crucial in determining how severe the free-riding problem is. If $\beta \gg 1$, either vulnerability is very high or it is comparably cheap to abate. Thus, it is in the countries’ local interest to abate nearly all emissions. Similarly, if $\beta \ll 1$, vulnerability is low, or abatement is prohibitively expensive, so that it is globally efficient to abate very little locally. Thus, under both scenarios, free-riding is a small concern. For intermediate values, however, free-riding becomes an issue.
4. Small Coalitions and Climate Change

In addition, there is a ‘No-Treaty Outcome’ (NTO).\(^{21}\)

**Definition 4 (Flexibility Treaty (FT))**

*In scenarios under uncertainty, under a FT, members set their local abatement effort ex-post (Stage 3) such that, in each state of the world, joint costs of the coalition are minimized.*

**Definition 5 (Commitment Treaty (CT))**

*In scenarios under uncertainty, under a CT, members commit to a local abatement effort ex-ante (Stage 1) such that, in expectation, joint costs of the coalition are minimized.*

**Definition 6 (No-Treaty Outcome (NTO))**

*If the NTO prevails, each country sets the local abatement effort ex-post such that in each state of the world, its local costs are minimized.*

Under the FT, the members make a common and credible declaration of intent that once learning has taken place and the damage parameter is certain (Stage 3), they will set the abatement level at the optimal level for the members. This allows to benefit from learning. Under the CT, the members credibly commit ex-ante (Stage 1) to an abatement level. Thus, they are not able to benefit from learning.

Members fully cooperate, but play a non-cooperative game against non-members – the outsiders – to determine abatement efforts. Outsiders play a non-cooperative game against members and other outsiders.\(^{22}\)

Two scenarios are relevant with respect to outsiders abatement decisions. In the simultaneous-move game, all countries decide simultaneously. In the sequential-move game, the members decide first and the outsiders second. The latter scenario is more realistic when considering the CT, because members can commit to an abatement decision more credibly than outsiders.

We use the following notation throughout the paper. Members are denoted by subscript \(M\) and outsiders by subscript \(O\). Superscripts denote the equilibrium outcomes for the

---

\(^{21}\) The NTO corresponds to a coalition size of one \((s = 1)\) with simultaneous abatement. In scenarios with uncertainty, it corresponds to the FT with \(s = 1\), as abatements occurs under certainty in the NTO.

\(^{22}\) We are aware that it is slightly inconsistent to assume that on the one side, countries do not behave according to some stability concept when deciding to participate, but on the other side, abatement efforts are determined with the Nash equilibrium concept. However, the same ambiguity – albeit, arguably, to a lesser degree – is present in all the literature on external/internal stability mentioned. Participation decisions of countries are a Nash equilibrium, while members’ compliance is simply assumed.
4. Small Coalitions and Climate Change

corresponding treaty. \(k_{CT}^M\), for instance, denotes the local costs for a member of the Commitment Treaty. Superscript \(\omega\) denotes a state of the world outcome or realization. To prevent superfluous notation, we will refrain from indexing if the context provides enough information.

4.3 Benchmarks

The Global Social Optimum (GSO) is defined as the globally efficient outcome under certainty. Local abatement efforts for all countries are\(^{23}\)

\[
a^{GSO} = N \beta \left( \frac{\tilde{E}}{\beta N^2 + 1} \right)^{L-1}.
\]

The No-Treaty Outcome obtains when no country participates in a treaty, and all free-ride instead. According to the set-up in this paper, in such a case, all countries act under certainty. Local abatement efforts for all countries are\(^{24}\)

\[
a^{NTO} = \beta \left( \frac{\tilde{E}}{\beta N + 1} \right)^{L-1}.
\]

4.4 Simultaneous Moves and Linear Damages

In this section, we set \(L = 1\). Thus, local costs are

\[
\beta(\tilde{E} - A) + \frac{1}{2}a^2.
\]

We first consider the situation under certainty and then under uncertainty.

\(^{23}\)To see this, set \(s = N\) for Equations (4.2) and (4.7).

\(^{24}\)To see this set, \(s = 1\) for Equations (4.2) and (4.7).
4. Small Coalitions and Climate Change

4.4.1 Certainty

Given a coalition \( C \) with coalition size \( s \), members and outsiders solve

\[
\begin{align*}
\min_{a_M} s\beta(\bar{E} - A) + \frac{1}{2}a_M^2,
\min_{a_O} \beta(\bar{E} - A) + \frac{1}{2}a_O^2.
\end{align*}
\]

Using \( \frac{dA}{da_M} = s \) and \( \frac{dA}{da_O} = 1 \), the local abatement effort of members and outsiders are

\[
\begin{align*}
a_C^M(s) &= \beta s, \quad (4.2a) \\
a_C^O(s) &= \beta. \quad (4.2b)
\end{align*}
\]

As damages are linear, the members’ abatement and the outsiders’ abatement are independent from each other. Global abatement reads

\[
A^C(s) = sa_C^M(s) + (N - s)a_C^O(s) = \beta(s^2 - s + N).
\]

Local costs for members and outsiders as well as in the NTO are

\[
\begin{align*}
k_C^M(s) &= \beta \bar{E} - \beta^2(\frac{1}{2}s^2 - s + N), \quad (4.3a) \\
k_C^O(s) &= \beta \bar{E} - \beta^2(s^2 - s + N - \frac{1}{2}), \quad (4.3b) \\
k^{NTO} &= \beta \bar{E} - \beta^2(N - \frac{1}{2}). \quad (4.3c)
\end{align*}
\]

Comparing these local costs, we find

Lemma 4 (Simultaneous Moves – Linear Damages – Certainty)

For all coalition sizes \( 1 < s \leq N \),

1. local costs are lower for outsiders than for members,
2. local costs decrease in the coalition size for members and outsiders,
3. as a corollary of 2., the NTO is the worst case outcome for members and outsiders.

An increase in coalition size decreases the local costs of all countries, due to the higher degree of concerted action. We call this the ‘Cooperation Effect’. Outsiders have lower local costs than members, as they benefit from the members’ high abatement efforts, whereas their own abatement efforts are low.
4. Small Coalitions and Climate Change

4.4.2 Uncertainty

In this subsection, the damage parameter $\beta$ is uncertain at the time the treaty is set up. We consider first the FT, then the CT, and then compare the outcomes.

**Flexibility Treaty**

Under a FT, all countries act under certainty. Thus, at the time the treaty is set up the expected local costs for members and outsiders – and if the NTO obtains – are the expectations of the local costs under certainty (see Equations (4.3)):

$$E[k_{FT}^M(s)] = E[\beta] \tilde{E} - E[\beta^2](\frac{1}{2}s^2 - s + N), \quad (4.4a)$$

$$E[k_{FT}^O(s)] = E[\beta] \tilde{E} - E[\beta^2](s^2 - s + N - \frac{1}{2}). \quad (4.4b)$$

$$E[k^{NTO}] = E[\beta] \tilde{E} - E[\beta^2](N - \frac{1}{2}). \quad (4.4c)$$

**Commitment Treaty**

Under a CT, the members commit to local abatement efforts and, in this section, commit simultaneously with the outsiders. The results are thus analogous to Equations (4.3), except that $\beta$ is being replaced with its expected value:

$$E[k_{CT}^M(s)] = E[\beta] \tilde{E} - E[\beta^2](\frac{1}{2}s^2 - s + N), \quad (4.5a)$$

$$E[k_{CT}^O(s)] = E[\beta] \tilde{E} - E[\beta^2](s^2 - s + N - \frac{1}{2}). \quad (4.5b)$$

**Comparison**

Inspection of Equations (4.4) and (4.5) shows

**Lemma 5 (Simultaneous Moves – Linear Damages – Uncertainty)**

For any coalition size $1 < s \leq N$,

1. under a FT, expected local costs for members are lower than if the NTO prevails,

2. expected local costs for members and outsiders are lower under a FT than under a CT.
The second point follows from $E[\beta^2] > E^2[\beta].$ Intuitively, the countries’ ability to make good use of information makes the FT less costly than the CT. We call this the ‘Learning Effect’.

Next, let us define $s_{1}^{crit}$ as the number of participants in a CT such that $E[k_{CT}^{CT}(s_{1}^{crit})] > E[k^{NTO}]$. This yields

$$s_{1}^{crit} = 1 + \sqrt{\frac{E[\beta^2]}{E^2[\beta]} - 1}(2N - 1) > 1.$$ 

The inequality follows from $E[\beta^2] > E^2[\beta]$ and implies

**Proposition 6 (Simultaneous Moves – Linear Damages – Uncertainty)**

If $s < s_{1}^{crit}$, the expected local costs for members of a CT are higher than in case the NTO prevails.

If $s < s_{1}^{crit}$, the Learning Effect dominates the Cooperation Effect. The CT deprives members – and outsiders – of the ability to process new information, whereas they act under certainty in the NTO. In such a situation, members would jointly prefer not to form a coalition at all. The larger the uncertainty – i.e. the higher the variance of $\beta$ –, the bigger is $s_{1}^{crit}$. A higher uncertainty increases the Learning Effect, but has no influence on the Cooperation Effect. There exists the possibility that $s_{1}^{crit} > N$, i.e. the NTO is jointly favored, even if all countries participate in a CT.

Proposition 6 shows a possible reason why countries have been rather unwilling to participate in a climate treaty where they have to commit to abatement targets: the ‘critical mass’ might not be reached yet to overcome the disadvantages of fixed targets with respect to uncertainty. The linear model, therefore, already provides a good intuition of how valuable it is to be able to react to newly-arriving information. We proceed with a slightly more sophisticated model which will yield further insights.

### 4.5 Simultaneous Moves and Quadratic Damages

In this section, we set $L = 2$. Thus, local costs are

$$\frac{\beta}{2}(\bar{E} - A)^2 + \frac{1}{2}a^2.$$ 

\(^{25}\) The definition of the variance is $Var(X) = E[X^2] - E^2[X]$. With $Var(X) > 0$, this claim follows.
Marginal damages are increasing and thus countries react to the abatement decisions of other countries. We first consider the situation under certainty and then under uncertainty.

4.5.1 Certainty

Equilibrium

Members and outsiders of the coalition $C$ solve

$$\min_{a_M} \beta \frac{(\bar{E} - A)^2}{2} + \frac{1}{2} a_M^2,$$

$$\max_{a_O} \beta \frac{(\bar{E} - A)^2}{2} + \frac{1}{2} a_O^2.$$

It is straightforward to derive local and global abatement efforts and the total abatement efforts of members, $A_M^C$, and outsiders, $A_O^C$, i.e. the sum of all members’ and of all outsiders’ local abatement efforts, respectively. In addition, we calculate the first derivatives with respect to the coalition size $s$.

$$a_M^C(s) = s \beta \frac{\bar{E}}{\beta(N + s^2 - s) + 1},$$

$$A_M^C(s) = s^2 \beta \frac{\bar{E}}{\beta(N + s^2 - s) + 1},$$

$$a_O^C(s) = \beta \bar{E} \frac{1}{\beta(N + s^2 - s) + 1},$$

$$A_O^C(s) = (N - s) \beta \bar{E} \frac{1}{\beta(N + s^2 - s) + 1},$$

$$A^C(s) = \beta \bar{E} \frac{(N + s^2 - s)}{\beta(N + s^2 - s) + 1},$$

$$a_M'(s) = \beta \bar{E} \frac{\beta(N - s^2) + 1}{(\beta(N + s^2 - s) + 1)^2},$$

$$A_M'(s) = \beta \bar{E} \frac{\beta s(2N - s) + 2s}{(\beta(N + s^2 - s) + 1)^2},$$

$$a_O'(s) = -\beta \bar{E} \frac{\beta(2s - 1)}{(\beta(N + s^2 - s) + 1)^2},$$

$$A_O'(s) = -\beta \bar{E} \frac{\beta s(2N - s) + 1}{(\beta(N + s^2 - s) + 1)^2},$$

$$A'(s) = \beta \bar{E} \frac{2s - 1}{(\beta(N + s^2 - s) + 1)^2}.$$

Comparative Statics of Abatement Levels

We find that the bigger the coalition size, the higher are the global abatement efforts ($A^C'(s) > 0$ for all $1 < s \leq N$). Yet, as the coalition size increases, the local abatement of the remaining outsiders decreases ($a_O'(s) < 0$ for all $1 < s \leq N$). Specifically, for every unit of increase in global abatement, the outsiders reduce their local efforts by
4. Small Coalitions and Climate Change

The members’ increase in abatement efforts for an increase in the coalition size is extenuated by more free-riding of the remaining outsiders. If coalition size increases, the total effort of outsiders decreases thus for two reasons ($A'_O(s) < 0$ for all $1 < s \leq N$): the number of outsiders shrinks and each outsider abates less. Correspondingly, the members’ total abatement effort increases more rapidly with $s$ than the global abatement efforts. We call this additional effort of the members the ‘Leakage Effect’.

To examine the behavior of the members’ local abatement effort more clearly, we define $s_a$ as the coalition size where these efforts are maximal: $A'_O(s_a) \equiv 0$. It is easy to see that $s_a = \sqrt{N + \frac{1}{\beta}} > 1$. Thus,

**Proposition 7 (Simultaneous Moves – Quadratic Damages – Certainty)**

*If the coalition size increases, for $s < s_a$, the members’ local abatement efforts increase, while they decrease for $s > s_a$.*

With increasing marginal damages, it is thus not true any more that a greater coalition size increases the member’s local abatement efforts for all coalition sizes. Thus, although members abate more in total when the coalition size increases, beyond $s_a$, this is only because the coalition size increases – not because individual members abate more.

To obtain the intuition for this non-monotonic behavior of the members’ local abatement decisions, consider an increase of the coalition size by one member. The members of the grown coalition face two opposing forces with regard to changing their local abatement decision. First, because the coalition size increased, there is one more country to take into account. This strengthened Cooperation Effect implies more abatement by each member. Second, because the new member contributes to the joint abatement efforts instead of free-riding, this relieves the other members’ burden to counteract free-riding. The weakened Leakage Effect implies less local abatement by each member. The Cooperation Effect dominates at smaller coalition sizes, and the Leakage Effect at larger ones. To see this more clearly, consider Figure 4.2. For the moment, only consider the solid curves which represent the simultaneous-move game. For the step from $s = 5$

---

26 $\frac{A'_O(s)}{A'_O(s)} = -\beta$ for all $1 < s \leq N$.

27 If $\beta < \frac{1}{N-1}$, it follows that $s_a > N$ and the members’ local abatement efforts increase monotonically for all feasible coalition sizes $s \leq N$. In this explanation, we assume, on the contrary, that $\beta > \frac{1}{N-1}$. This seems reasonable, given footnote 20.

28 Throughout the paper, we will use examples to illustrate the results. Unless stated otherwise, we use the following parameters: $N = 6$ and $E = 100$. Under certainty, $\beta = 0.2$; under uncertainty, there are two equally likely states of the world, with $\beta = 0.1$ and $\beta = 0.3$, respectively. These numbers are for illustration only and are not meant to represent a real-world counterpart.
4. Small Coalitions and Climate Change

to $s = 6$, the Leakage Effect dominates, and each member abates less. Prior to joining, the single outsider contributed little and members (partly) compensated. After joining, the ex-outsider increases its local abatement efforts, which allows the other members to decrease theirs. The increased Cooperation Effect plays only a minor role. For the step from $s = 1$ to $s = 2$, on the other hand, members increase their abatement effort. The Cooperation Effect dominates because the coalition size has doubled, whereas, as the coalition size remains small, the outsiders’ free-riding remains high.

Figure 4.2: Local (upper left) and total (upper right) abatement efforts and local costs of members (blue curves), outsiders (red curves) and average (or global, respectively) values (green curves) as a function of $s$. Solid curves are for the simultaneous-move game, dashed ones for the sequential-move game.
Comparative Statics of Costs

Local costs of members and outsiders, and the respective first derivative with respect to the number of participants $s$ are:

$$k_C^M(s) = \beta E^2 \frac{\beta s^2 + 1}{2(\beta(N + s^2 - s) + 1)^2}, \quad k_C^O(s) = \beta E^2 \frac{\beta + 1}{2(\beta(N + s^2 - s) + 1)^2},$$

(4.8a)

$$k_C^M'(s) = \beta^2 E^2 \frac{-\beta s^3 + s(\beta N - 1) + 1}{(\beta(N + s^2 - s) + 1)^3}, \quad k_C^O'(s) = -\beta^2 E^2 \frac{(1 + \beta)(2s - 1)}{(\beta(N + s^2 - s) + 1)^3}.$$  

(4.8b)

The Number of Participants

It holds that $k_O'(s) < 0$ for all $1 < s \leq N$. This is a consequence of the previous results, i.e. that global abatement increases with coalition size, whereas the outsiders’ abatement decreases.

To analyze the members’ costs, define $s_{\text{cost}}$ by $k_M'(s_{\text{cost}}) = 0$. It can be shown that $1 < s_{\text{cost}} < N$. For a coalition size below the threshold $s_{\text{costs}}$, the members’ local costs increase in $s$, and they decrease beyond it. The reason for the former is that the sharp increase in abatement costs outweighs the decreased damages. This could make it rather difficult for small-size coalitions to increase slightly in size. Only if the coalition grows beyond the threshold $s_{\text{costs}}$ do the beneficial effects of cooperation outcompete the detrimental leakage effects.

Next, define $s_{\text{crit}}^2$ as $k_M''(s_{\text{crit}}^2) = k_{\text{NTO}}$. As the members’ local costs initially increase and then decrease, it must be that $s_{\text{crit}}^2 > 1$. If coalition size is small, the Leakage Effect and the resulting high abatement costs for the members are, therefore, another reason why a coalition of countries would rather not put a climate treaty into force.

These results are illustrated in Figure 4.2 and summarized in the following proposition

**Proposition 8 (Simultaneous Moves – Quadratic Damages – Certainty)**

1. If the coalition size is increased, local costs of outsider decrease for all $1 < s \leq N$ and local costs of members increase for $s < s_{\text{costs}}$ and decrease for $s > s_{\text{costs}}$.

2. If the coalition size is $s < s_{\text{crit}}^2$, then local costs of members are higher than if the NTO prevailed.

The Damage Parameter

$^{29}$ From Equation (4.8), $s_{\text{cost}}$ is given by the implicit cubic equation $-\beta s_{\text{cost}}^3 + s_{\text{cost}}(\beta N - 1) + 1 = 0$. First, $k_M''(1) > 0$, as $-\beta + \beta N - 1 + 1 > 0$. Second, $k_M''(N) < 0$, as $-\beta N^3 + N(\beta N - 1) + 1 < 0$. 

94
4. Small Coalitions and Climate Change

In the remainder of Section 4.5, we concentrate on the members and refrain from showing results for outsiders, as these do not yield any new insights.

We define \( \beta^T(s) := \frac{1}{N+s^2-s} \) and find\(^{30}\)

**Lemma 6 (Simultaneous Moves – Quadratic Damages – Certainty)**

*For all coalition sizes \( 1 < s \leq N \), if \( \beta \) increases the members’*

1. abatement costs increase and
2. damages increase if \( \beta < \beta^T(s) \) and decrease if \( \beta > \beta^T(s) \).

The first point is obvious. The higher the vulnerability to climate change, the more members abate, and the higher the abatement costs. The reason for the second point is that for \( \beta \) sufficiently high, the increase in global abatement levels overcompensates the increase in \( \beta \).

Next, we examine the derivative of members’ local costs with respect to \( \beta \).

Define \( s^R(N) := \sqrt{1+4N-1} \). It is easy to see that \( 1 < s^R(N) < N \) for all \( N \). Furthermore, define \( \beta^R(N, s) := \frac{1}{N-s^2-s} \). We find that\(^{31}\)

**Proposition 9 (Simultaneous Moves – Quadratic Damages – Certainty)**

*If and only if \( s < s^R(N) \) and \( \beta > \beta^R(N, s) \), then the members’ local costs decrease in the damage parameter.*

If the coalition size is below threshold \( s^R(N) \), and at the same time, the vulnerability to climate change is above threshold \( \beta^R(N, s) \), the members’ damages decrease faster than abatement costs increase when \( s \) increases. Intuitively, if \( s < s^R(N) \), there are many free-riders, which – despite their comparatively low abatement efforts – have a big influence on the global abatement level. Increasing the damage parameter induces the free-riders to abate more. This decreases the damages of all countries and, in particular, overcompensates the increasing abatement costs of members. These trade-offs are illustrated in Figure 4.3.

\(^{30}\) This follows straightforward by the inspection of the first derivatives of costs. See Section 4.10, Equations (4.11b) and (4.11e).

\(^{31}\) Inspection of (4.11h) in Section 4.10 shows that if \( s^2 + s - N > 0 \), then \( k_M^C(\beta) > 0 \) for all \( \beta \). If \( s^2 + s - N < 0 \) and \( \beta < \frac{1}{N-s^2-s} \), then \( k_M^C(\beta) > 0 \). If \( s^2 + s - N < 0 \) and \( \beta > \frac{1}{N-s^2-s} \), then \( k_M^C(\beta) < 0 \).
4. Small Coalitions and Climate Change

Figure 4.3: Left: Local abatement efforts of members (blue curve) and outsiders (red curve) as a function of $\beta$ with $s = 2$ and $N = 18$ ($s^R(18) = 3.772$). Right: Members’ local costs (blue curve), abatement costs (cyan curve) and damages (black curve).

4.5.2 Uncertainty

We now examine the case where damages are quadratic and there is uncertainty.

Abatement Efforts

In the CT, the abatement decisions of both members and outsiders are determined ex-ante. Thus, the results are the same as under certainty (see Equations (4.7)), except that $\beta$ is replaced with its expected value $E[\beta]$. Hence,

\[
a_{CT}^M = s \frac{E}{(N + s^2 - s) + \frac{1}{E[\beta]}}, \quad a_{CT}^O = \frac{E}{(N + s^2 - s) + \frac{1}{E[\beta]}}, \quad A_{CT} = \frac{E(N + s^2 - s)}{(N + s^2 - s) + \frac{1}{E[\beta]}}.
\]
In the FT, abatement decisions are taken ex-post and the expectations of Equations (4.7) yield the expected abatement

\[
E[a^\text{FT}_M] = E[s \frac{\bar{E}}{(N + s^2 - s) + \frac{1}{\beta}}],
E[a^\text{FT}_O] = E[s \frac{\bar{E}}{(N + s^2 - s) + \frac{1}{\beta}}],
E[A^\text{FT}] = E[s \frac{\bar{E}(N + s^2 - s)}{(N + s^2 - s) + \frac{1}{\beta}}].
\]

Comparing the abatement levels of the CT and the FT, it is easy to see that

\textbf{Lemma 7 (Simultaneous Moves – Quadratic Damages – Uncertainty)}

For all coalition sizes \(1 < s \leq N\) and all \(\beta\), it holds that abatement commitments under a CT are higher than the expected abatement efforts under a FT for outsiders, members and thus also globally.

Intuitively, under a CT, countries choose abatement levels to ‘cover’ all states of the world, and thus act more cautiously.

\textbf{Expected Costs}

In the following analysis, we will find that if all possible realizations of \(\beta\) are above or below a certain threshold, this is a sufficient condition to determine under which treaty the members have lower expected local costs, lower expected damages and lower expected abatement costs. In such a case, the function is convex or concave, respectively, at the support of \(\beta\). If, however, the support of \(\beta\) is on both sides of the threshold, we would need additional information about the distribution of \(\beta\) to be able to make a comparison. We refrain from doing so because our purpose is merely to highlight which possibilities exist.

We find

\textbf{Lemma 8 (Simultaneous Moves – Quadratic Damages – Uncertainty)}

For all coalition sizes \(1 < s \leq N\),

1. If \(\beta > (\leq) 2 \beta^\text{FT}(s)\) for all \(\beta\), then the members’ expected local damages are higher (smaller) under a FT than under a CT,
2. If \(\beta < (>) \frac{1}{2} \beta^\text{FT}(s)\) for all \(\beta\), then the members’ expected local abatement costs are higher (smaller) under a FT than under a CT.

\textsuperscript{32} This follows from \(\frac{1}{1+x^n} = -\frac{2}{(1+z)^2}\).

\textsuperscript{33} This follows straightforward from Equations (4.11c) and (4.11f), given in Section 4.10. The costs under a FT are smaller than under a CT if a cost function is concave (second derivative negative).
If the vulnerability exceeds the threshold \(2\beta^T(s)\) in all realizations, the expected damages are higher under the FT than under the CT. Expected abatement costs, however, are lower. And the reverse is true for the threshold \(1/2\beta^T(s)\). This shows that vulnerability enters the members’ expected local costs through two oppositional channels.

Next, as regards expected local costs of a member under a CT and a FT, the following holds\(^{34}\)

**Proposition 10 (Simultaneous Moves – Quadratic Damages – Uncertainty)**

*If and only if* \(s < s^R(N)\) *and* \(\beta > 2\beta^R(N, s)\) *for all* \(\beta\), *then the members’ expected local costs under the FT are higher than under the CT.*

If the coalition size exceeds the threshold \(s^R(N)\), members are, in expectation, better off under the FT than under the CT, regardless of their vulnerability to climate change. This threshold is always exceeded for the grand coalition – or equivalently for a single global planner.

If, however, the coalition size is below the threshold \(s^R(N)\) and at the same time all realizations of \(\beta\) are above the threshold \(2\beta^R(N, s)\), the members are, in expectation, worse off under the FT than under the CT. An intuitive explanation runs along the following lines: As the threshold \(s^R(N)\) is always below \(N\), if a coalition of less than size \(s^R(N)\) forms, global abatement, in expectation, is inefficiently low under both treaty types. Inspection of Lemma 7 shows that, in expectation, outsiders abate more under a CT than under a FT. Thus, under the CT, the uncertainty about the vulnerability under a CT somewhat curbs the detrimental effects of free-riding. We call this the ‘Positive Uncertainty Effect’. This effect is especially relevant for smaller coalition sizes, as there are more free-riders. It also is more relevant when the vulnerability to climate change is high, as high abatement effort to mitigate the climate damages are more important, relative to the possibility that abatement costs turn out to be higher than necessary. As a consequence, if the coalition size is small and vulnerability is high, the CT’s benefits for members from the Positive Uncertainty Effect outweigh the FT’s benefits from the Learning Effect.

### 4.5.3 Summary

The results of Lemmas 6 and 8, and Propositions 9 and 10 are summarized in Table 4.2 and illustrated in Figure 4.4. Figure 4.5 illustrates the expected local abatement efforts

\(^{34}\) See Section 4.11.
4. Small Coalitions and Climate Change

and the expected local costs for members and outsiders, where, for now, only the Sim cases are relevant.

<table>
<thead>
<tr>
<th>If $\beta$ increases</th>
<th>abatement costs decrease</th>
<th>damages decrease</th>
<th>local costs decrease</th>
</tr>
</thead>
<tbody>
<tr>
<td>never</td>
<td>if $\beta &gt; \beta^T(s) := \frac{1}{N+s^2-s}$</td>
<td>if $\beta &gt; \beta^R(N, s) := \frac{1}{N-s^2-s}$ and $s &lt; s^R(N) := \sqrt{1+4N-1}$</td>
<td>(Blue dashed curve)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Green dashed curve)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Expected costs under a FT are higher than under a CT if for all $\beta$</th>
<th>abatement costs</th>
<th>damages</th>
<th>local costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta &lt; \frac{1}{2}\beta^T(s)$ (Red solid curve)</td>
<td>$\beta &gt; 2\beta^T(s)$ (Blue solid curve)</td>
<td>$\beta &gt; 2\beta^R(N, s)$ and $s &lt; s^R(N)$ (Green solid curve)</td>
<td></td>
</tr>
</tbody>
</table>

**Table 4.2:** Summary of the main results of Section 4.5.

**Figure 4.4:** For the explanation of curves see Table 4.2. The black solid horizontal line is $s^R(6) = 2$.  

99
4.6 Sequential Move Game

In this section, we once again compare members and outsiders. Members move first, outsiders move second.

Under certainty, only the scenario of quadratic damages is relevant. If damages are linear, abatement decisions are independent and thus, there is no difference between the simultaneous and the sequential case.

Under uncertainty, for the CT, the members commit to their abatement decision ex-ante, whereas outsiders decide ex-post. This seems more realistic than assuming – as we did in Section 4.5 – that outsiders commit ex-ante, simultaneously with the members. Outsiders may find it hard to credibly commit to an abatement decision because of the very fact that they do not belong to a treaty. Under the FT, both the members and the outsiders decide ex-post, but sequentially.

For the sequential-move game, we change the notation to highlight the difference between the sequential and the simultaneous case. Subscript 1 refers to the first mover, i.e. the member, and subscript 2 refers to the second mover, i.e. the outsider. If appropriate, equilibria are denoted by the superscript $\text{Seq}$ in the sequential case and by $\text{Sim}$ in the simultaneous case.

4.6.1 Certainty and Quadratic Damages

To solve this scenario, we use backward induction. First, we examine Stage 3.2, where outsiders choose abatement efforts, given the members’ efforts. Second, we examine Stage 3.1, where members choose efforts anticipating the outsiders’ efforts.

**Stage 3.2: Outsiders**

An outsider solves

$$\min_{a_2} \beta \frac{1}{2} (\bar{E} - A)^2 + \frac{1}{2} a_2^2.$$ 

As $a_1^C$ is already fixed and the abatement decisions of the other outsiders are taken as given, it holds that $\frac{\partial \delta}{\partial a_2} = 1$. We find

$$a_2^C = \frac{\beta (\bar{E} - sa_1^C)}{\beta (N - s) + 1}.$$
4. Small Coalitions and Climate Change

Stage 3.1: Members

A member solves
\[ \min_{a_1} s \frac{\beta}{2} (\bar{E} - A)^2 + \frac{1}{2} a_1^2. \]

With \( A = sa_1 + (N - s)a_2^C = sa_1 + (N - s)\frac{\beta(E - sa_1)}{\beta(N - s) + 1} \), we find \( \frac{dA}{da_1} = 1 - s - (N - s)\frac{\beta s}{\beta(N - s) + 1} \) and thus
\[ a_1^C = \frac{s\bar{E}\beta}{s^2\beta + (\beta(N - s) + 1)^2}. \]

Summary and Comparison

Comparing the abatement level and costs of the sequential scenario with the simultaneous scenario it is easy to see that\(^{35}\)

**Lemma 9 (Sequential/Simultaneous Moves – Quadratic Damages – Certainty)**

1. For all coalition sizes \( 1 < s \leq N \), \( a_1^C < a_M^C, a_2^C > a_O^C \) and, \( A^{Seq} < A^{Sim} \).
2. For all coalition sizes \( 1 < s \leq N \), \( k_1^{Seq} < k_M^{Sim} \) and \( k_2^{Seq} > k_O^{Sim} \).

For outsiders, the local costs are higher in the sequential case for two reasons: they abate more, i.e. abatement costs are higher, and global abatement efforts are lower, i.e. damages are higher. For the members, the local costs are lower in the sequential case, as the lower abatement costs outweigh the higher damages. Thus, they have a ‘General First-Mover Advantage’.\(^{36}\) Define \( s^P := \frac{\beta N + 1}{\beta + 1} \). If the coalition size is smaller than this threshold, the General First-Mover Advantage is strong enough to make members better off than outsiders.\(^{37}\)

**Proposition 11 (Sequential Moves – Quadratic Damages – Certainty)**

If \( s < s^P \), then members’ local costs are smaller than the outsiders’ local costs.

This result is illustrated in Figure 4.2. It shows the advantage members have in the sequential-move game, especially for small coalition sizes. It also shows that the out-

---

\(^{35}\) See Equations (4.12).

\(^{36}\) This result is in accordance with the common notion of the first-mover advantage. Our setting differs from the standard one – price competition of firms – because damages increase if countries move sequentially. Still, the first movers could choose the abatement level of the simultaneous-move game, inducing the second movers to also choose the simultaneous-move game outcome. Thus, it has to be that first movers are (weakly) better off in the sequential-move game than under the simultaneous-move game.

\(^{37}\) Compare Equations (4.12d) and (4.12e).
siders’ local costs increase more than the members’ local costs decrease. Furthermore, Figure 4.2 illustrates that in a sequential-move game, the members are better off as compared to the NTO for all $1 < s \leq N$. This holds true as members can always replicate the NTO if they move first.

These results could imply that in the real world, small groups of countries compete to be the first ones to benefit from commitment to a low abatement level, knowing that the others have to get the coals out of the fire.

4.6.2 Uncertainty

The sequential-move game under uncertainty is straightforward to solve for the FT. For the linear case, they do not differ at all from the simultaneous case, and for the quadratic case one simply has to take the expectations of the results of Section 4.6.1. We thus focus on the CT, where interesting differences to the simultaneous-move game arise.

**Linear Damages**

We start with the easier case of linear damages, where the countries’ abatement decisions are independent. Thus, there is no General First-Mover Advantage for members. Under the CT, the members commit to abatement efforts ex-ante, whereas the outsiders make their decisions ex-post, i.e. under certainty. Thus,

$$a_{2}^{CT,\omega}(s) = \beta^{\omega}, \quad a_{1}^{CT,\omega}(s) = E[\beta]s.$$  

Thus, the expected local costs for the members and the outsiders are

$$E[k_{1}^{CT}] = E[\beta]\bar{E} - \frac{1}{2}s^{2}E^{2}[\beta] + E[\beta^{2}](s - N), \quad (4.9a)$$

$$E[k_{2}^{CT}] = E[\beta]\bar{E} - s^{2}E^{2}[\beta] + E[\beta^{2}](s - N + \frac{1}{2}). \quad (4.9b)$$

Comparing Equations (4.4a), (4.5a), and (4.9), we find\(^{38}\)

**Proposition 12 (Sequential/Simultaneous Moves – Linear Damages – Uncertainty)**

1. $E[k_{M}^{FT, Sim}] < E[k_{1}^{CT, Seq}] < E[k_{M}^{CT, Sim}]$ for all $1 < s \leq N$,

\(^{38}\) Remember that with linear damages, there is not difference in the FT between the sequential- and the simultaneous-move game.
4. Small Coalitions and Climate Change

2. \( E[k_1^{CT, Seq}] \geq E[k_2^{CT, Seq}] \) if \( s^2 \geq \frac{E[\beta^2]}{E[\beta]} \)

For members, the FT remains the best treaty, as all countries act under certainty. The difference to the CT shrinks, however, if moves are sequential, as outsiders decide under certainty. We call this the ‘Certainty First-Mover Advantage’ for members. The smaller the coalition size or the higher the uncertainty, the bigger is the Certainty First-Mover Advantage. It might even be big enough to make the members better off than the outsiders.\(^{39}\)

**Quadratic Damages**

We solve this scenario using backward induction. First, we examine the ex-post stage 3.2 where the outsiders choose their abatement efforts given the members’ commitment. Second, members choose their abatement commitment ex-ante in Stage 1, anticipating the outsiders’ decision ex-post. As this scenario solves analogously to the certain case – with the slight complication that we have to allow for expected values – we refrain from a detailed derivation and immediately show the results. The members’ commitment, the outsiders’ expected abatement efforts and the global expected abatement efforts are

\[
\begin{align*}
\hat{a}_1^{CT} &= \frac{\bar{E}}{s + E[\beta(N-s)+1]} \\
E[\hat{a}_2^{CT}] &= \frac{\beta \bar{E}}{(\beta(N-s)+1)(\beta(N-s)+1)} \\
E[A^{Seq}] &= \frac{\bar{E}}{\beta(N-s)+1} \left( \frac{1}{\beta(N-s)+1} + \frac{1}{s^2E[(\beta(N-s)+1)^2]} \right)
\end{align*}
\]

These results are rather complex, and analytical comparisons are much more involved and much less meaningful than before. Instead, we use previous results in combination with Figure 4.5 to show the potential consequences of this setting.

Clearly, for all coalition sizes, members of the FT and the CT, respectively, have lower expected costs in the sequential-move game as compared to the simultaneous one. An important cause for this is the General First-Mover Advantage. For members of the CT, there is also the contribution of the Certainty First-Mover Advantage as well as the ‘Abatement First-Mover Advantage’.

To illustrate this novel effect, we include a scenario, \( \check{C}T^{Seq} \) in Figure 4.5, where both

\(^{39}\) Compare this to Proposition 5 which states that, in the simultaneous-move game, the members’ expected costs under the CT are higher than the outsiders’ for all coalition sizes and all \( \beta \).
4. Small Coalitions and Climate Change

Figure 4.5: 1) $FT_{Sim}^{T}$ (blue); members and outsiders abate simultaneously ex-post, 2) $FT_{Seq}^{T}$ (cyan); members and outsiders abate sequentially ex-post, and 3) $CT_{Sim}^{T}$ (red); members and outsiders commit simultaneously ex-ante, 4) $CT_{Seq}^{T}$ (orange); members commit ex-ante and outsiders abate ex-post, 5) $\tilde{CT}_{Seq}^{T}$ (thin orange); members and outsiders commit sequentially ex-ante.

The solid curves are for members, the dashed ones for outsiders. On the left side, there are expected (or committed) abatement efforts and on the right side, the expected local costs.

members and outsiders commit ex-ante, but sequentially. Correspondingly, the members still have the General First-Mover Advantage, but do not benefit any more from the outsiders acting under certainty. To offset, the members commit to more abatement under $\tilde{CT}_{Seq}^{T}$ than under $CT_{Seq}^{T}$ (compare the thin orange curve and the thick orange curve). This indicates that under $CT_{Seq}^{T}$, the members, exploiting the fact that the outsiders act under certainty, decrease their abatement efforts.

This cuts both ways. If outsiders were able to commit ex-ante, this would induce the
members to increase their abatement efforts. If there are few outsiders, this effect would even have the potential to make outsiders, in expectation, better off if they committing ex-ante instead of deciding on abatement ex-post. This is true even tough they remain followers and deprive themselves of the possibility to use new information.

### 4.6.3 Summary

Let us summarize the results of this section and point out further implications. If the members manage to move first – i.e. under a CT, commit to abatement efforts before outsiders do or, under a FT, abate first – their expected local costs drop for both treaty types, because of the General First-Mover Advantage. Under the CT, there are two additional first-mover advantages in their favor. Members benefit because outsiders abate under certainty – the Certainty First-Mover Advantage. And members commit to less abatement, as they know that the outsiders will act under certainty – the Abatement First-Mover Advantage.

When deliberating which treaty to form, the latter two effects strengthen the case for a CT, especially if the coalition size is small. In addition, in the real world, it seems to be more likely that members of a CT manage to move first, as compared to members of a FT.

### 4.7 Conclusion

We started from the assumption that there are several countries willing to cooperate when tackling climate change, leaving aside the remaining countries. We found several effects that such a group should take into account when considering to set up a climate treaty. We focused on the effects of uncertainty and on the influence of group size. If the group is rather small, commitment might be more important than the ability to react to new information. In addition, it might be better for a small group not to cooperate.

---

40 We stress that the Certainty First-Mover Advantage and the Abatement First-Mover Advantage are distinct, even though they both depend on the outsiders acting under certainty. The former decreases members’ expected damages for any abatement commitment, whereas because of the latter members reduce their abatement commitment.

41 Members of a CT can commit to abatement efforts more easily than outsiders. For a FT, on the other hand, both members and outsiders abate under certainty, and thus will rather act simultaneously.
4.8 Table of variables

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSO</td>
<td>Global Social Optimum</td>
</tr>
<tr>
<td>NTO</td>
<td>No-Treaty Outcome</td>
</tr>
<tr>
<td>CT</td>
<td>Commitment Treaty</td>
</tr>
<tr>
<td>FT</td>
<td>Flexibility Treaty</td>
</tr>
<tr>
<td>N</td>
<td>number of countries</td>
</tr>
<tr>
<td>$\bar{E}$</td>
<td>global business-as-usual emissions</td>
</tr>
<tr>
<td>$\beta$</td>
<td>damage parameter</td>
</tr>
<tr>
<td>a</td>
<td>local abatement efforts</td>
</tr>
<tr>
<td>A</td>
<td>global abatement efforts</td>
</tr>
<tr>
<td>a_c</td>
<td>local abatement costs</td>
</tr>
<tr>
<td>d</td>
<td>local damages costs</td>
</tr>
<tr>
<td>k</td>
<td>local costs (sum of local abatement efforts and local damages)</td>
</tr>
<tr>
<td>L</td>
<td>$L = 1$ in the linear damage scenario; $L = 2$ in the quadratic damage scenario</td>
</tr>
<tr>
<td>Sim</td>
<td>Simultaneous-move game with respect to abatement</td>
</tr>
<tr>
<td>Seq</td>
<td>Sequential-move game with respect to abatement</td>
</tr>
</tbody>
</table>

For Simultaneous-move game with respect to abatement:

- $s_1^{crit} = E[k_M^{C}(s_1^{crit})] = E[k^{NTO}]$ for Sim in the linear damage scenario under uncertainty.
- $s_2^{crit} = k_M^{C}(s_2^{crit}) \equiv k^{NTO}$; for Sim in the quadratic damage scenario under certainty.

For Sequential-move game with respect to abatement:

- $s_a^{crit} = 0$; for Sim in the quadratic damage scenario under certainty.
- $s_{cost}^{crit} = 0$ for SIM in the quadratic damage scenario under certainty.
- $s^R(N) = \sqrt{\frac{1 + 4N - 1}{2}}$
- $\beta^R(N, s) = \frac{1}{N - s^2 - s}$
- $\beta^T(s) = \frac{1}{N + s^2 - s}$
4.9 Effects that Influence the Members’ Local Costs

In the following, we list all effects that influence the local costs of members. Table 4.3 shows in which scenarios the respective effects play a role.

1. Cooperation Effect: For all countries, local costs decrease as the coalition size increases.

2. Learning Effect: Under uncertainty, countries are better off if they – or others – can decide on their abatement decisions in Stage 3 under certainty.

3. Leakage Effect: If damages are quadratic, members partly compensate for the free-riding of outsiders, thereby incurring higher abatement costs. The smaller the coalition size, the stronger is this effect.

4. Positive Ignorance Effect: Under uncertainty and quadratic benefits, all countries commit to higher abatement levels in Stage 1 under a CT than their expected abatement decision would be under a FT. Choosing a CT in the simultaneous move game, members can force outsiders to commit and thus decrease their free-riding behavior to a certain extent.

5. General First-Mover Advantage: In the sequential-move game and with quadratic damages, members – which move first – can decrease their abatement efforts in anticipation of the outsiders’ reaction.

6. Certainty First-Mover Advantage: Under uncertainty, in the sequential-move game, members of a CT benefit from the fact that outsiders abate under certainty.

7. Abatement First-Mover Advantage: Under uncertainty, in the sequential-move game and with quadratic damages, members of a CT know that outsiders abate under certainty, which allows them to decrease their abatement commitment more than due to the General First-Mover Advantage alone.

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Simultaneous</th>
<th>Sequential</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treaties</td>
<td>Certainty</td>
<td>Uncertainty</td>
</tr>
<tr>
<td>Linear</td>
<td>1</td>
<td>1,2</td>
</tr>
<tr>
<td>Quadratic</td>
<td>1,3</td>
<td>1,2,3,4</td>
</tr>
</tbody>
</table>

Table 4.3: Scenarios where effects as listed above play a role.
4. Small Coalitions and Climate Change

4.10 Functions

The following are the local abatement costs \( ac \), local damages \( d \), and local costs \( k \) for the simultaneous moves game with quadratic damages under certainty.

\[
\begin{align*}
ac_M^C(\beta) &= \bar{E}^2 \frac{\beta^2 s^2}{2(\beta(N + s^2 - s) + 1)^2},
\quad (4.11a) \\
ac_M'(\beta) &= \bar{E}^2 \frac{\beta s^2}{(\beta(N + s^2 - s) + 1)^3},
\quad (4.11b) \\
ac_M''(\beta) &= \bar{E}^2 \frac{s^2 (1 - 2\beta(N + s^2 - s))}{(\beta(N + s^2 - s) + 1)^4},
\quad (4.11c) \\
d_M^C(\beta) &= \bar{E}^2 \frac{\beta}{2(\beta(N + s^2 - s) + 1)^2},
\quad (4.11d) \\
d_M'(\beta) &= \bar{E}^2 \frac{1 - \beta(N + s^2 - s)}{2(\beta(N + s^2 - s) + 1)^3},
\quad (4.11e) \\
d_M''(\beta) &= \bar{E}^2 \frac{(N + s^2 - s)(\beta(N + s^2 - s) - 2)}{(\beta(N + s^2 - s) + 1)^4},
\quad (4.11f) \\
k_M^C(\beta) &= \bar{E}^2 \frac{\beta^2 s^2 + \beta}{2(\beta(N + s^2 - s) + 1)^2},
\quad (4.11g) \\
k_M'(\beta) &= \bar{E}^2 \frac{\beta(s^2 + s - N) + 1}{2(\beta(N + s^2 - s) + 1)^3},
\quad (4.11h) \\
k_M''(\beta) &= \bar{E}^2 \frac{\beta N^2 - 2N(\beta s + 1) - s(\beta s^2 - \beta s + s - 2)}{(\beta(N + s^2 - s) + 1)^4},
\quad (4.11i)
\end{align*}
\]

In the sequential-move game with quadratic damages under certainty, abatement efforts and local costs are

\[
\begin{align*}
a_1^C &= \frac{s \bar{E} \beta}{s^2 \beta + (\beta(N - s) + 1)^2},
\quad (4.12a) \\
a_2^C &= \frac{\beta \bar{E}(\beta(N - s) + 1)}{s^2 \beta + (\beta(N - s) + 1)^2},
\quad (4.12b) \\
A_{seq} &= \frac{\beta \bar{E} \left(s^2 + (N - s)(\beta(N - s) + 1)\right)}{s^2 \beta + (\beta(N - s) + 1)^2},
\quad (4.12c) \\
k_1^C &= \frac{b \bar{E}^2}{2 (b^2(N - s)^2 + b(2N + s^2 - 2s) + 1)},
\quad (4.12d) \\
k_2^C &= \frac{b(b + 1) \bar{E}^2(b(N - s) + 1)^2}{2 (b^2(N - s)^2 + b(2N + s^2 - 2s) + 1)^2},
\quad (4.12e)
\end{align*}
\]
4. Small Coalitions and Climate Change

4.11 Proofs

Proof of Proposition 10
Inspection of the numerator of (4.11i) shows that $k_M^{C''}(\beta) > 0$ iff

$$\beta N^2 - 2N(\beta s + 1) - s(\beta s^3 - \beta s + s - 2) > 0$$

$$\Leftrightarrow 2(s - N - \frac{1}{2}s^2) + \beta \left( N - s^2 - s \right) \left( N + s^2 - s \right) > 0$$

$N + s^2 - s > 0$ for all $1 < s \leq N$. $s - N - \frac{1}{2}s^2 < 0$ for all $1 < s \leq N$. Finally, $N - s^2 - s < (>)0$ if $s > (<) s^R(N) = \sqrt{1 + 4N + 1}$. Thus, for $s > s^R(N)$, the inequality is false. If on the other hand, $s < s^R(N)$, the inequality is true (false) if $\beta > (<) \beta^R(N, s) = \frac{2}{N-s^2-s}$. If the inequality is true (false), Function (4.11i) is convex (concave) in $\beta$ and thus the members’ expected local costs are higher (lower) under the FT than under the CT. □
5 Climate Treaties for Innovation and Abatement*

5.1 Introduction

5.1.1 Motivation

Mitigation of climate change has become a major concern of policy-makers, which has called for a series of international negotiations aiming to reduce emissions of greenhouse gases (GHGs). So far, however, these negotiations did not lead to reductions sufficient to slow down global warming. Moreover, there is a growing consensus that the development of new abatement technologies that reduce abatement costs substantially is indispensable to achieve the current emission reduction objectives.\footnote{Barrett (2009) provides a comprehensive overview of the technology portfolio which he expects to play a significant role in future emission reductions. Examples include wind and solar power or carbon capture and storage. However, as pointed out by Aldy et al. (2009), quantifying the benefits of new technologies is associated with large uncertainties.}

This paper focuses on two interrelated reasons for the failure of international climate negotiations: free-riding in abatement efforts and free-riding in innovation efforts. We call this the \textit{double free-riding problem}. Free-riding in abatement efforts originates from the fact that local emissions have a negative impact on other countries. Hence, avoiding local emissions is a global public good, and the standard free-riding argument for the underprovision of abatement efforts applies.\footnote{For reviews see Aldy et al. (2003), Bodansky et al. (2004), Kuik et al. (2008), and Aldy and Stavins (2009).}

Free-riding in innovation efforts occurs since countries have a low willingness to pay for licenses of superior technologies. The reason is that buyers of licenses do not take into account that the increase in abatement due to a superior technology decreases damages

---

* This chapter is based on joint work with Hans Gersbach and Martin Scheffel.
in other countries. As a consequence, countries have low incentives to innovate.\(^3\) Such free-riding in innovation is aggravated when the development of new abatement technologies requires scientific advancements through basic research.\(^4\) As basic research is a public good in this context, it is subject to the standard underprovision problem. A similar underinvestment problem would occur if such scientific advancements can be achieved by privately-funded applied research. To sum up, incentives of countries and firms to invest into new abatement technologies fall below the socially optimal levels.

This *double free-riding problem* is a severe obstacle for international climate policy, as there is no international authority that is able to enforce contractual obligations between the countries. Efficient climate treaties should thus satisfy two criteria. First, they have to be self-enforcing in the sense that countries participate voluntarily and comply with the intended targets. Second, climate treaties have to deal with both, innovation and abatement.\(^5\)

The purpose of this paper is, first, to develop a suitable economic framework to analyze the double free-riding problem and, second, to propose a blueprint for a climate treaty that addresses the double free-riding problem. Our proposal consists of a set of rules and does not involve specific emission targets. So far, such rules treaties have focused exclusively on free-riding in abatement (see Gersbach and Winkler (2011) and Gersbach and Oberpriller (2012)). Developing rules treaties for the double free-riding problem regarding abatement and innovation can contribute to the ongoing debate on the optimal design of environmental agreements regarding emission reduction and new abatement technologies.\(^6\)

### 5.1.2 Model and Result

There are \(n\) homogeneous countries with a representative polluting firm in each of them. Local emissions cause global environmental damage. Countries issue emission permits to their domestic firm and invest into the innovation of new abatement technologies.

\(^3\) For an alternative and interesting formulation of low incentives to invest in new abatement technologies, see Goeschl and Perino (2013).

\(^4\) In general, the output of basic research projects have no direct commercial use, which ultimately calls for public funding of basic research. E.g. see Nelson (1959) or more recently Cozzi and Galli (2009) and Gersbach et al. (2013).

\(^5\) See Jaffe et al. (2003), Jaffe et al. (2005), de Coninck et al. (2008), Popp et al. (2009), Aldy et al. (2009), and Acemoglu et al. (2012), for instance.

\(^6\) For a summary, see de Coninck et al. (2008) who identify four types of treaties that promote new technologies: (1) knowledge sharing and coordination; (2) research, deployment and demonstration (RD&D); (3) technology transfer; and (4) technology deployment mandates, standards, and incentives.
which are fully protected by international patents. We refer to the country with the most advanced technology as the *technology leader*. There is an international market on which licenses for these patents are traded between countries. Firms participate in an international market for emission permits, and exert abatement effort, using the country’s abatement technology.

The first part of this paper is devoted to solving for the subgame perfect equilibrium of the laissez-faire outcome to which we refer as the *no-treaty outcome* (NTO). A comparison with the *global social optimum* (GSO), in which a global social planner directly chooses abatement and innovation efforts for all countries, allows us to assess and decompose the strong impact of the *double free-riding problem* on the NTO: abatement efforts and innovation efforts are substantially below their globally optimal levels.

As we assume that intellectual property rights are fully enforceable and that a market for licensing abatement technologies exists, it is surprising at first sight that this does not solve the free-riding problem in innovation. The reason is as follows: The higher the technology level of a country, the lower the marginal abatement costs and, as a consequence, the higher the domestic firm’s abatement efforts. Countries value the license for a more advanced technology according to the net benefit of changes in abatement costs and damages. Specifically, they do not take into account that a superior technology level and the ensuing increase in abatement of the domestic firm, also decreases damages in other countries. Hence, the revenues the technology leader receives when selling its licenses to all other countries falls short of the amount that would induce the technology leader to the globally optimal innovation efforts.

In the second part of this paper, we propose a blueprint of an international climate treaty based on rules – the *rules treaty for innovation and abatement* (RTIA) – to cope with the inefficiencies of the double free-riding problem. The RTIA consists of three elements.

First, a country is only allowed to allocate a fraction $\mu$ of the permits it issues to its domestic firms. The remainder is given to an international agency that sells them. Second, the revenues of the agency are fully redistributed to the participating countries. If the technology leader licenses its technology for free to other countries instead of selling it, it qualifies for an share $\rho^{CT}$ of the revenues. The remainder of the revenues is equally distributed to all countries, each receiving a share $\rho = (1 - \rho^{CT})/n$. Third, the refund to the technology leader is reduced by a lump-sum $F$ which is equally redistributed to all other countries. The set of rules is thus parameterized by $\{\mu, \rho, \rho^{CT}, F\}$, where $\mu$ denotes the allocation parameter and $\{\rho, \rho^{CT}, F\}$ denote the refunding parameters.
We show that the RTIA solves the double free-riding problem if the parameters are chosen appropriately. First, local and global interests are aligned for the fraction of permits that have to be given to the international agency. A low allocation parameter $\mu$ thus prevents free-riding in abatement efforts. Second, if sufficiently high clean-tech refunds are given to the technology leader for free-licensing its technology, the RTIA is able to yield globally optimal innovation efforts. Thus, choosing $\rho^{CT}$ appropriately prevents free-riding in innovation efforts. Third, for any allocation and refunding parameters, there exists a lump-sum reduction $F$ of the clean-tech refund that guarantees the implementability of the globally optimal innovation efforts for any given abatement effort.

The underlying mechanism that induces the technology leader to exert more innovation efforts under the RTIA is not obvious. Without a clean-tech refund, the refund each country receives exactly compensates the domestic firms’ costs to buy permits from the international agency. If clean-tech refunding is introduced, the total refund the technology leader receives exceeds its domestic firm’s permit buying costs. This net benefit rises with the technology level which is an incentive for the technology leader to increase innovation efforts. The reason for this is subtle. With a clean-tech refund in place, refunding alone provides a disincentive for the technology leader to innovate. If the technology leader increases innovation efforts, global emissions decrease. Because the permit price is unaffected by changes in the technology level in our model, technological advancements reduce the international agency’s revenues and thus refunds. As the technology leader receives a higher share of the agency’s revenues, it is especially affected by this decrease in revenues. To counteract this effect, the technology leader issues emission permits more aggressively. The ensuing decrease of the domestic firm’s permit-buying costs always exceeds the loss from lower refunds. Thus, the introduction of the clean-tech refund gives the technology leader an incentive to invest into new abatement technologies. In addition, this incentive increases in $\rho^{CT}$, which allows to target the optimal innovation efforts by choosing $\rho^{CT}$ appropriately.

5.1.3 Organization of the Paper

The rest of the paper is organized as follows. Section 5.2 outlines the model. Section 5.3 computes the GSO, which serves as the benchmark for the subsequent analysis. In Section 5.4, we analyze the double free-riding problem. In Section 5.5, we propose the RTIA, characterize the induced equilibrium and show that the RTIA can implement the GSO. In Section 5.6, we provide a variety of extensions of the model, and in Section
5.7, we explore critical aspects for the practical implementation of climate treaties for abatement and innovation. Finally, Section 5.8 concludes the paper.

5.2 The Model

There are n homogeneous countries indexed by i and j with a representative GHG emitting firm in each of them. Each country suffers from environmental damage caused by global emissions. Emissions in one country affect all other countries to the same extent.

Countries issue emission permits to their domestic firm and invest into the innovation of new abatement technologies. Technologies are fully protected by international patents, and there exists an international market for abatement technologies. Marginal abatement costs are decreasing in the abatement technology level such that the local planner can lower the firm’s marginal abatement cost either by exerting innovation efforts or by buying a superior abatement technology on the international license market.

Firms participate in an international market for emission permits and exert abatement efforts using the country’s abatement technology. Contractual obligations resulting from the market for emission permits are completely enforceable. Thus, firms have to hold sufficient emission permits to cover their actual emissions.

In the following, we provide a detailed description of the local (country-specific) costs and the international markets for emission permits and abatement technologies.

5.2.1 The Local Costs

Let $\bar{\tau}$ and $\overline{\chi}$ denote a country’s emission and technology level in the absence of abatement and innovation, that is the business-as-usual levels. Local costs consist of 5 components:

First, abatement costs are $\frac{1}{2}(\bar{\tau} - \tau_i)^2$. They are quadratic in abatement efforts $\bar{\tau} - \tau_i$, where $\tau_i$ denotes firm $i$’s emissions, and linear in the technology parameter $\chi_i$. The lower $\chi_i$, the more efficient is the abatement technology. Thus, referring to $\chi$ as the technology level, a lower value indicates a more superior abatement technology.

Second, developing a superior technology level, $\chi_i (0 < \chi_i \leq \overline{\chi})$, entails costs given by
\( \alpha((\sum \chi_i)^2 - 1) \), where \( \alpha \) denotes a common innovation cost parameter.\(^7\) For simplicity, we assume that innovation is a deterministic process. Hence, choosing innovation efforts is equivalent to choosing the technology level.

Third, countries buy and sell licenses of superior abatement technologies on an international market at the price \( \pi \). Hence, technology leaders reduce their local costs by selling the superior abatement technology, while local costs for buyer countries increase.

Fourth, there are the firm’s costs of trading permits, \( p(e_i - \epsilon_i) \), where \( p \) is the permit price and \( \epsilon_i \) is the allocation of permits to the countries’ firm. If allocated permits do not suffice to meet the actual emissions, the firm has to buy additional permits at the market price \( p \). Otherwise, the firm can sell the permits.

Fifth, the emitted GHGs by the firms cause environmental damage, \( \beta E \), where \( E = \sum e_i \) are global emissions and \( \beta \) is the marginal damage.

### 5.2.2 International Market for Emission Permits and Firms’ Abatement Choice

There is a competitive international market for emission permits on which all contractual obligations are enforceable. In this section, we describe how this market operates and derive the firms’ abatement choice. The resulting formulas will be applicable throughout the paper.

Initially, each firm holds a number of permits, \( \epsilon_i \), which have been allocated by the local planner and which we take as given for the moment.\(^8\) The firms have to trade permits such that their chosen levels of emissions, \( e_i \), are covered by permits. Thus, in equilibrium, the total number of emission permits, \( \mathcal{E} \), are equal to global emissions, such that \( \mathcal{E} = E \).

Firms are cost minimizing and take the permit price as given. Hence, firms choose emissions such that their marginal abatement costs equal the permit price

\[
p = \chi_i(\bar{e} - e_i) \quad \Leftrightarrow \quad e_i = \bar{e} - \frac{p}{\chi_i}.
\]  

\(^{(5.1)}\)

Note that, in principle, innovation costs are the sum of investments by the public and the firm. From the country’s perspective only total outlay for innovation matters. Whether and how much firms contribute to developing new technologies does not play a role as we focus on domestic welfare, and thus, transfers and cost sharing within a country do not matter.

\(^{8}\) The number of permits the firms receive depends on whether there is a treaty or not and on the parameters of the treaty. These specifications will be analyzed later in the paper.
For convenience and future reference, we use Equation (5.1) to rewrite abatement costs as

$$\frac{\chi_i}{2}(\bar{e} - e_i)^2 = \frac{p^2}{2\chi_i}. \tag{5.2}$$

Local emissions decrease with an increase in the technology level (i.e. $\chi_i$ declines) and an increase in the price. The price is determined by market clearing on the emission permit market

$$p = \frac{\bar{E} - \mathcal{E}}{\sum_{i=1}^{n} \frac{1}{\chi_i}}, \tag{5.3}$$

where $\bar{E} = \sum_i \bar{e}$ denotes the global business-as-usual emission level.

### 5.2.3 International Market for Abatement Technology

We assume that the superior technology can be protected by patents recognized by all countries. Hence, countries at the technological frontier can protect their knowledge and sell licenses to other countries. If two countries have the same technology level, Bertrand competition drives the price at which countries are able to sell their patents down to zero. If country $i$ is the global technology leader, i.e. $\chi_i < \min_{j \neq i} \chi_j$, it is a monopolist for the technology advantage, $\min_{j \neq i} \chi_j - \chi_i$, and can sell it to all countries at a positive price.\(^9\)

### 5.3 Global Social Optimum

In this section we characterize the global social optimum (GSO). Suppose there exists a global social planner who directly chooses innovation efforts and emissions in each country, thereby dealing with both free-rider problems efficiently. Because duplication of innovation efforts is costly and innovation is deterministic, only one country conducts innovation and the technology developed in this country is distributed to all other countries.

\(^9\) Suppose, for instance, that Germany is the technology leader in building coal power plants that can capture and store carbon (CCS). Many countries, on the other hand, know how to make standard solar panels. Hence, countries would have to pay Germany for the blueprints to build a CCS coal power plant, but the technology to produce standard solar panels would be available at zero price for countries that do not yet possess this technology.
Since all countries have the same technology levels, $\chi$, the convexity of the abatement cost function implies that in a first best solution all countries choose the same emission levels which we denote by $e$. Hence, the social planner’s optimization problem reads

$$\min_{e, \chi} \left\{ n \left( \beta E + \frac{\chi}{2} (e - \bar{e})^2 \right) + \alpha \left( \frac{(X)^2}{\chi} - 1 \right) \right\}$$

subject to

$$\chi - \bar{\chi} \leq 0.$$  \hspace{1cm} (5.5)

The GSO is characterized as follows:\footnote{The proof is deferred to Section 5.12.1.}

**Proposition 13 (Global Social Optimum)**

The globally optimal innovation efforts and emissions are given by

$$\chi^{GSO} = \min \left\{ \frac{4\alpha \bar{\chi}}{n^3 \beta^2}, 1 \right\} \bar{\chi}, \hspace{1cm} (5.6a)$$

$$E^{GSO} = \bar{E} - \frac{n^2 \beta}{\chi^{GSO}}, \hspace{1cm} (5.6b)$$

$$e^{GSO} = \frac{E^{GSO}}{n}. \hspace{1cm} (5.6c)$$

For later comparison, it is useful to consider the hypothetical permit price that would result if $E = E^{GSO}$. Using Equations (5.3) and (5.6b) yields

$$p^{GSO} = n \beta. \hspace{1cm} (5.7)$$

If we assume $\chi^{GSO} < \bar{\chi}$, global emissions are $E^{GSO} = \bar{E} - \frac{n^2 \beta}{\chi^{GSO}}$ and thus highly nonlinear and decreasing in the number of countries.\footnote{In order to guarantee $E > 0$, we assume for the rest of the paper $\bar{E} > \frac{n^2 \beta^2}{4\alpha \bar{\chi}}$.} The strong sensitivity of $E^{GSO}$ with respect to $n$ stems from the fact that innovation efforts and abatement efforts are mutually reinforcing. If $n$ is high, it is optimal to innovate heavily, as all countries benefit. The strong improvement in technology makes abatement cheap, which finally leads to substantial reductions in global emissions.
5. No-Treaty Outcome

This section analyzes the subgame perfect equilibrium of a laissez-faire economy with operating international markets for emission permits and abatement technologies. We refer to this equilibrium as the no-treaty outcome (NTO). The sequential structure of the game is as follows:

**Stage 1** Each local planner $i$ chooses innovation efforts, which determines the pre-trade technology level denoted by $\chi'_i$, i.e. $\chi^\text{innovation} \rightarrow \chi'_i$.

**Stage 2** Technologies are traded on an international market, resulting in the post-trade technology level, $\chi_i$, i.e. $\chi'_i \rightarrow \chi_i$.

**Stage 3** Each local planner allocates a number of emission permits to the domestic firm. The firm makes its abatement choice using $\chi_i$ and trades emission permits on the international permit market.

We solve for the subgame perfect equilibrium using backward induction. Afterwards, we contrast the subgame perfect equilibrium of the NTO to that of the GSO, in order to decompose the impact of the double free-riding problem on abatement and innovation.

### 5.4.1 Stage 3: Permit Choice, Permit Trade and Abatement

In the final stage, expenses for innovation efforts and technology trade are sunk. Taking the technology level as given, the cost minimization problem of an arbitrary local planner at this stage is

$$\min_{\epsilon_i} \left\{ \beta E + \frac{\chi_i}{2} (\bar{e} - e_i)^2 + p (e_i - \epsilon_i) \right\},$$

where $p = p(\mathcal{E})$ and $e_i = e_i(p(\mathcal{E}))$ for all $i$ are determined by the equilibrium in the permit market and optimal abatement decisions of firms – Equations (5.1) and (5.3).

The first-order condition reads

$$\beta - \chi_i(\bar{e} - e_i) + \frac{1}{\chi_i \sum_{j=1}^n \chi_j} \left[ \frac{1}{\chi_i \sum_{j=1}^n \chi_j} - 1 \right] - \frac{1}{\sum_{j=1}^n \chi_j} (e_i - \epsilon_i) = 0 \Leftrightarrow \beta + \frac{e_i - \bar{e}}{\sum_{j=1}^n \chi_j} + p \left( \frac{1}{\chi_i \sum_{j=1}^n \chi_j} - 1 \right) = 0.$$

---

12 Whether this occurs through grandfathering or auctioning does not matter for our purpose; for this choice merely affects distribution of welfare within the country.
Using the firm’s optimal abatement choice, Equation (5.1), delivers

\[ \epsilon_i = e_i + \frac{1}{\sum_{j=1}^{n} x_j} (p - \beta). \]  

(5.8)

Summing over all countries, the price for an emission permit in the NTO computes as

\[ p^{NTO} = \beta. \]  

(5.9)

Thus, the price for an emission permit in the NTO equals the local marginal damage, whereas the hypothetical price in the GSO equals the global marginal damage. The difference is due to the fact that a local planner does not take into account that the emissions of the domestic firm inflict damages on other countries. Specifically, \( np^{NTO} = p^{GSO} \) reflects free-riding in emission abatement, as the lower price in the NTO translates into lower abatement efforts.

Plugging Equations (5.1) and (5.9) into Equation (5.8) yields

\[ \epsilon_i = e - \beta \frac{1}{x_i}. \]  

(5.10)

Emission permits issued by a country are independent of the emission permits issued by the other countries. This result is a direct consequence of the assumption that marginal damages are constant and simplifies our analysis substantially.

Furthermore, comparing Equations (5.1) and (5.10), local planners allocate exactly the number of emission permits firms need to cover their emission level. In other words, there is no trade on the international permit market. This comes at no surprise since we assumed that countries are homogeneous with respect to abatement costs and marginal damages. While the market for emission permits plays no role in the NTO,\(^{13}\) it will be central for the analysis of the RTIA.

### 5.4.2 Stage 2: Technology Trade

Because innovation efforts have been carried out in Stage 1, the pre-trade technology levels \( x_i' \) are given for all countries. Without loss of generality, countries can be ranked and labeled according to their pre-trade technology: \( x_1' \leq x_2' \leq x_3' \leq \cdots \leq x_n' \leq \overline{x} \).

\(^{13}\) If there is no international market for emission permits, the local planner chooses emissions (e.g. via taxes or national permits) such that marginal damages equal marginal costs, that is \( \beta = \chi (\overline{\tau} - e_i) \).
Clearly, the post-trade technology level for the technology leader satisfies $\chi_1 = \chi'_1$.

Selling the own technology to other countries is a profitable strategy as it, first, generates profits and, second, increases abatement efforts by other countries which leads to a reduction of damages. For technology level $\chi'_2$, there is Bertrand competition between Country 1 and 2 that drives the equilibrium price for technology level $\chi'_2$ down to zero and leads to complete diffusion of technology level $\chi'_2$ to countries 3, $\ldots$, $n$. Country 1, in contrast, is a monopolist for abatement technology level $\chi \in [\chi_1, \chi'_2)$. To maximize revenues, Country 1 always sets the price schedule such that other countries buy the frontier technology $\chi_1$. The technology price thus depends on $\chi_1$ and $\chi'_2$ and is denoted by $\pi(\chi_1, \chi'_2)$.

Country 1 sets $\pi(\chi_1, \chi'_2) \geq 0$ such that Country 2 weakly prefers buying a license to use technology level $\chi_1$ instead of using technology level $\chi'_2$. Hence, given that the costs of innovation efforts are already sunk at the second stage and no permits are traded, $\pi(\chi_1, \chi'_2)$ satisfies

$$\beta \left( E - p^{NTO} \left( \frac{1}{\chi_1} + \frac{1}{\chi'_2} + \sum_{i=3}^{n} \frac{1}{\chi_i} \right) \right) + \frac{p^{NTO^2}}{2\chi_1} + \pi(\chi_1, \chi'_2) \leq \beta \left( E - p^{NTO} \left( \frac{1}{\chi_1} + \frac{1}{\chi'_2} + \sum_{i=3}^{n} \frac{1}{\chi_i} \right) \right) + \frac{p^{NTO^2}}{2\chi'_2}, \quad (5.11)$$

where on the left-hand side there are the local costs of Country 2 in case it buys technology level $\chi_1$ at price $\pi(\chi_1, \chi'_2)$ and on the right-hand side there are the local costs in case it does not buy and uses $\chi'_2$ instead. Observe that this comparison is independent of the buying decisions made by others.

Solving for the technology price and applying condition (5.9), we obtain

$$\pi^{NTO}(\chi_1, \chi'_2) = \frac{\beta^2}{2} \left( \frac{1}{\chi_1} - \frac{1}{\chi'_2} \right). \quad (5.12)$$

The equality follows from assuming that the seller can extract the buyer’s total willingness to pay.$^{14}$

By an analogous argument, other Countries $j = 3, \ldots, n$ buy a license at price $\pi(\chi_1, \chi'_2)$ as well. As a consequence, at the end of Stage 2 when all technology trade has taken place, all countries employ the frontier technology level $\chi_1$ and each country 2, 3, $\ldots$, $n$ pays $\pi^{NTO}(\chi_1, \chi'_2)$ to Country 1.

$^{14}$ For further discussion of alternative pricing mechanisms, see Section 5.6.3.
5. Climate Treaties for Innovation and Abatement

5.4.3 Stage 1: Innovation Efforts Choice

Countries are ex-ante homogeneous and initially endowed with the business-as-usual abatement technology level $\bar{\chi}$. The question which country becomes country 1, 2, ..., $n$ (ranked and labeled according to the post-innovation technology level) is the outcome of the coordination game in Stage 1. Because of Bertrand competition in Stage 2, in any subgame perfect equilibrium, there are at most two countries committed to innovation – Country 1 and 2.

Country 1 exerts innovation efforts to lower the sum of local damages and abatement costs and to increase the revenues from selling the technology, $(n-1)\pi(\chi_1, \chi'_2)$. Because Country 2 buys and uses $\chi_1$, it solely exerts innovation efforts in order to lower the price Country 1 can charge for a license.

For given $\chi'_2$, the cost minimization problem for Country 1 with respect to innovation efforts reads

$$\min_{\chi_1} \left\{ \beta \left( \frac{\bar{E} - n\beta}{\chi_1} \right) + \frac{\beta^2}{2\chi_1} + \alpha \left( \left( \frac{\bar{X}}{\chi_1} \right)^2 - 1 \right) - (n - 1)\pi(\chi_1, \chi'_2) \right\}$$

subject to

$$\chi_1 - \bar{\chi} \leq 0.$$

In a similar vein, for given $\chi_1$, the cost minimization problem of Country 2 reads

$$\min_{\chi'_2} \left\{ \beta \left( \frac{\bar{E} - n\beta}{\chi_1} \right) + \frac{\beta^2}{2\chi_1} + \alpha \left( \left( \frac{\bar{X}}{\chi'_2} \right)^2 - 1 \right) + \pi(\chi_1, \chi'_2) \right\}$$

subject to

$$\chi'_2 - \bar{\chi} \leq 0.$$

Solving both optimization problems gives

$$\chi_1 = \min \left\{ \frac{4\alpha\bar{X}}{(3n - 2)\beta^2}, 1 \right\} \bar{\chi}, \quad (5.13a)$$

$$\chi'_2 = \min \left\{ \frac{4\alpha\bar{X}}{\beta^2}, 1 \right\} \bar{\chi}. \quad (5.13b)$$

\[15\] See Section 5.12.2.
5. Climate Treaties for Innovation and Abatement

5.4.4 Subgame Perfect Equilibrium

The following proposition summarizes the results for the NTO as derived by backward induction. Section 5.10 provides a graphical illustration.

**Proposition 14 (No-Treaty Outcome)**

*Up to relabeling of countries, there exists a unique subgame perfect equilibrium characterized as follows:

**Stage 1:** In equilibrium, countries choose innovation efforts (pre-trade technology level)

\[
\chi_{1}^{NTO} = \min \left\{ \frac{4a \chi}{(3n - 2) \beta^2}, 1 \right\}, \chi,
\]

\[
\chi_{2}^{NTO} = \min \left\{ \frac{4a \chi}{\beta^2}, 1 \right\}, \chi,
\]

\[
\chi_{j}^{NTO} = \chi, \quad \forall j = 3, \ldots, n.
\]

**Stage 2:** There is complete technology diffusion: First, the license to use technology level \(\chi_{2}^{NTO}\) is traded at price zero. Second, all but Country 1 buy a license to use the technology frontier, \(\chi_{1}^{NTO}\), at technology price

\[
\pi(\chi_{1}^{NTO}, \chi_{2}^{NTO}) = \frac{\beta^2}{2} \left( \frac{1}{\chi_{1}^{NTO}} - \frac{1}{\chi_{2}^{NTO}} \right).
\]

**Stage 3:** Each local planner issues \(\epsilon_{i}^{NTO} = \bar{e} - \frac{\beta}{\chi_{1}^{NTO}}\) emission permits to its domestic firm. The domestic firm chooses an emission level \(\epsilon_{i}^{NTO} = \epsilon_{i}^{NTO}\) and the equilibrium permit price is \(p^{NTO} = \beta\).

Let \(K_{i}\) denote the local costs of country \(i\) in equilibrium. In addition, define \(K^{NoR&D}\) as the local costs when countries do not innovate. In this case, the game collapses to Stage 3 only. With respect to local costs, we derive the following result:16

**Corollary 5**

The distribution of costs across countries in the subgame perfect equilibrium characterized in Proposition 14 satisfies

\[
K_{i}^{NTO} \leq K_{2}^{NTO} \leq K_{1}^{NTO} \leq K^{NoR&D}, \quad \forall i = 3, \ldots, n.
\]

---

16 The proof is deferred to Section 5.12.3.
All countries benefit from innovation. Countries \( j = 3, \ldots, n \) are (weakly) better off than countries 1 and 2, as the latter are not fully compensated for their innovation efforts.

### 5.4.5 Decomposition of Free-Riding Effects

We will show that compared to the single free-riding problem with respect to abatement efforts alone – as discussed in Gersbach and Oberpriller (2012), for instance – the impact of the double free-riding problem is especially severe. The reason is that introducing innovation in abatement technology leads to an additional free-riding problem that is strengthened by free-riding in abatement.

To see this more clearly, it is useful to define the following variants of the game: First, under autarky (AUT), there is no technology trade, i.e. Stage 2 would be absent from the game. Second, under INN, countries act cooperatively with respect to innovation but non-cooperatively with respect to abatement. Thus, countries cooperate to minimize global costs in Stage 1 and 2, while anticipating a non-cooperative behavior in Stage 3 of the game. Third, under ABA, countries act non-cooperatively with respect to innovation but cooperatively with respect to abatement. Thus, in Stage 3 countries cooperate to minimize global costs. For ease of exposition, we only consider interior solutions for the abatement technology and additionally assume that \( \chi_2 = \bar{\chi} \). The results are given in Table 5.1.\(^{17}\)

<table>
<thead>
<tr>
<th></th>
<th>( \chi )</th>
<th>( E )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global social optimum (GSO)</td>
<td>( \frac{1}{n^3} \frac{4\alpha \bar{\chi}^2}{\beta^2} )</td>
<td>( -n^4 \frac{n^3 \beta^3}{4\alpha \bar{\chi}^2} )</td>
</tr>
<tr>
<td>Cooperation w.r.t abatement (ABA)</td>
<td>( \frac{1}{n^2} \frac{4\alpha \bar{\chi}^2}{\beta^2} )</td>
<td>( -n^3 \frac{n^3 \beta^3}{4\alpha \bar{\chi}^2} )</td>
</tr>
<tr>
<td>Cooperation w.r.t innovation (INN)</td>
<td>( \frac{1}{n(2n - 1)} \frac{4\alpha \bar{\chi}^2}{\beta^2} )</td>
<td>( -n(2n - 1) \frac{n^3 \beta^3}{4\alpha \bar{\chi}^2} )</td>
</tr>
<tr>
<td>No-treaty outcome (NTO)</td>
<td>( \frac{1}{3n - 2} \frac{4\alpha \bar{\chi}^2}{\beta^2} )</td>
<td>( -(3n - 2) \frac{n^3 \beta^3}{4\alpha \bar{\chi}^2} )</td>
</tr>
<tr>
<td>Autarky (AUT)</td>
<td>( \frac{1}{\beta^2} )</td>
<td>( -\frac{n^3 \beta^3}{4\alpha \bar{\chi}^2} )</td>
</tr>
</tbody>
</table>

**Table 5.1:** Level of abatement technology and global emissions for various outcomes.

Regarding the technology level, the step from autarky to the NTO induces the technology leader to exert more innovation efforts, as the frontier technology can now be

---

\(^{17}\) For a detailed derivation of the results, see Section 5.12.4.
sold to all other countries. Thus, $\chi$ decreases by a factor of $n$. Buyer countries only pay for their local costs reductions, such that the technology price is still too low to induce globally optimal innovation.

Next, there is a wedge in the technology level of order $n^2$ between $\chi^{NTO}$ and $\chi^{GSO}$. The intuition behind this result is as follows: Enforcing cooperation with respect to innovation (compare $\chi^{NTO}$ and $\chi^{INN}$) internalizes the positive externality of using a better abatement technology. This delivers a decrease of $\chi$ of order $n$. Enforcing cooperation with respect to abatement (compare $\chi^{NTO}$ and $\chi^{ABA}$) internalizes the positive externality of exerting more abatement efforts. It is an important insight to see that this also delivers a decrease of $\chi$ of order $n$. The reason is that countries anticipate that they will increase their abatement efforts in Stage 3. They are thus willing to pay more for a better abatement technology in Stage 2, which, in turn, incentivize the technology leader to increase innovation efforts. Moving from the NTO to the GSO, with cooperation in both innovation and abatement efforts (compare $\chi^{NTO}$ and $\chi^{GSO}$), the technology parameter decreases by order $n^2$ – one $n$ for cooperation in innovation and abatement, each.

For global emissions, the wedge between $E^{NTO}$ and $E^{GSO}$ is of order $n^3$. In a setting where the technology level is fixed, this wedge would be only of order $n$. Thus, introducing innovation increases the free-riding problem by order $n^2$. This is the reason why the double free-riding problem is especially severe. Specifically, enforcing cooperation in innovation increases abatement by a factor of $n$ (compare $E^{NTO}$ and $E^{INN}$), as $\chi$ decreases by a factor of $n$. In contrast, enforcing cooperation in abatement increases abatement by a factor of $n^2$ (compare $E^{NTO}$ and $E^{ABA}$). The reason is that in addition to the factor of $n$ that stems solely from cooperation in abatement, $\chi$ decreases by a factor of $n$.

### 5.5 Rules Treaty for Innovation and Abatement

This section introduces and models the outcome of an international climate agreement that has the potential to attenuate the negative impact of the double free-riding problem. We call it the *rules treaty for innovation and abatement* (RTIA). We show that for any RTIA that aims at the optimal innovation efforts for given abatement efforts there exists a subgame perfect equilibrium. In addition, in Section 5.5.6 we show that all countries participate in the RTIA voluntarily in case the treaty fails if a single country leaves.
5. Climate Treaties for Innovation and Abatement

5.5.1 Rules

The rules of the RTIA are characterized by the set of parameters \( \{\mu, \rho, \rho^{CT}, F\} \) that add to the previously discussed stages of the laissez-faire game in several ways. While there is no change of Stage 1, the technology leader has two choices in Stage 2:\(^{18}\) either it sells the technology as in the NTO or it shares the frontier technology with all participating countries. In the Stage 3, of every permit the local planner issues, only the fraction \( \mu \in (0, 1) \) is allocated to the domestic firm. The remainder goes to an international agency, which, in turn, sells these permits to firms at the prevailing market price. The agency completely redistributes its revenues according to the refunding parameters \( \{\rho, \rho^{CT}\} \), where \( \rho \) is the fraction of the revenues that each country receives and \( \rho^{CT} = 1 - n\rho \) is the clean-tech refund that the technology leader receives on top if it licenses its technology for free to other countries. The technology leader’s refund is reduced by a lump-sum \( F \), while the other countries receive \( F/(n - 1) \) on top.

We show that in Stage 2, the technology leader always prefers to license its superior technology for free instead of selling it to the other countries. We use backward induction to solve for the subgame perfect equilibrium under the RTIA and derive the refunding and lump-sum parameters that implement the globally optimal innovation efforts for any given allocation parameter.

5.5.2 Stage 3: Permit Choice, Permit Trade and Abatement

In Stage 3, expenses for innovation efforts and technology trade are already sunk and the technology level is given. The agency’s revenues are given by \((1 - \mu)pE\) and the firms’ emissions are given by Equation (5.1). The number of permits issued under the RTIA, however, differs between Country 1 and Countries \( j = 2, \ldots, n \); for Country 1 in addition receives the clean-tech refund and thus faces different incentives to issue permits. The local planners in Country 1 and 2 face the following optimization problem

\(^{18}\) In principle, several countries could be technology leaders, in which case the clean-tech refund would have to be shared equally. However, because countries could claim the complete refund by slightly increasing their innovation efforts at little additional costs, there will be only one technology leader in equilibrium.
with respect to the permit issuance

$$
\min_{\epsilon_1} \left\{ \beta E + \frac{p^2}{2\chi} + p(e - \mu \epsilon_1) - (\rho + \rho^{CT})(1 - \mu)pE + F \right\},
$$

$$
\min_{\epsilon_2} \left\{ \beta E + \frac{p^2}{2\chi} + p(e - \mu \epsilon_j) - \rho(1 - \mu)pE - \frac{F}{(n-1)} \right\},
$$

respectively. Solving the optimization problems and imposing market clearing on the market for emission permits yields\(^{19}\)

$$
p^{RTIA} = \frac{n\beta}{(n-1)\mu + 1}.
$$

The permit price under the RTIA is a function of the allocation parameter \(\mu\), but it is independent of the refunding parameters \(\{\rho, \rho^{CT}, F\}\). If \(\mu = 1\), the permit price is equal to the one obtained in the NTO. In contrast, if \(\mu = 0\), the permit price corresponds to the hypothetical price in the GSO. This shows that the RTIA has the potential to overcome the free-riding problem with respect to abatement. The reason is that for the fraction of permits the local planner gives to the agency, local and global interests are fully aligned.\(^{20}\)

Using Equation (5.15), the first-order conditions of the cost minimization problems yield

$$
\epsilon_{1}^{RTIA} = e + (n - 1)\frac{1 - \mu}{\mu} \rho^{CT}(\bar{\sigma} - 2e),
$$

$$
\epsilon_{j}^{RTIA} = e - \frac{1 - \mu}{\mu} \rho^{CT}(\bar{\sigma} - 2e),
$$

$$
E^{RTIA} = \epsilon_{1}^{RTIA} + \sum_{j \neq 1} \epsilon_{j}^{RTIA} = \bar{E} - \frac{n^2 \beta}{((n-1)\mu + 1)\chi}.
$$

### 5.5.3 Stage 2: Technology Exchange

As will be shown in Stage 1, the refunding parameters \(\{\rho, \rho^{CT}, F\}\) can be set such that Country 1 always prefers sharing the frontier technology over selling the technology and renouncing the clean-tech refund. This conjecture already completes the analysis of Stage 2.

---

\(^{19}\) See Section 5.12.5.

\(^{20}\) This is reminiscent of the analysis in Gersbach and Winkler (2011) and Gersbach and Oberpriller (2012), who focus on pure abatement treaties.
5.5.4 Stage 1: Innovation Efforts Choice and the Refunding Scheme

Before solving Stage 1, as a preliminary step, we determine the globally optimal technology level for a given allocation parameter $\mu$. We refer to this as the constrained globally optimal technology level $\chi^{*}(\mu)$. Solving the optimization problem for a given $\mu$ yields

$$\chi^{*}(\mu) = \min \left\{ \frac{4\alpha \chi}{np_{RTIA}(2n\beta - p_{RTIA})}, 1 \right\} \chi$$

$$\Leftrightarrow \chi^{*}(\mu) = \min \left\{ \frac{4\alpha \chi((n - 1)\mu + 1)^2}{n^3 \beta^2(2(n - 1)\mu + 1)^2}, 1 \right\} \chi$$

Note that for $\mu \to 1$, the technology level remains superior to that in the NTO. The lower $\mu$, the higher is the constrained optimal technology level. For $\mu \to 0$, the technology level approaches the one in the GSO.

We now show that the RTIA can induce $\chi^{*}(\mu)$ as a subgame perfect equilibrium by setting $\rho^{CT}$ and $F$ appropriately. Country 1 solves

$$\min_{\chi_1} \left\{ \beta E + \frac{p^2}{2\chi_1} + p(e - \mu e_1) - (\rho + \rho^{CT})(1 - \mu)pE + F \right\}.$$ 

Solving this optimization problem and aligning the resulting technology level with $\chi^{*}(\mu)$, we obtain

**Proposition 15 (Optimal Refunding Rule and Local Costs Comparison)**

*For any allocation parameter $\mu \in (0, 1)$, there exist $\{\rho^{*}(\mu), \rho^{CT*}(\mu)\}$ that induce Country 1 to exert the constrained optimal innovation efforts $\chi^{*}(\mu)$. Furthermore, there exists $F^{*}(\mu)$ that equalizes local costs across countries and guarantees the existence of*

---

21 See Section 5.12.6.
22 The proof is deferred to Section 5.12.7.
a subgame perfect equilibrium. Specifically,

\[
\rho_{CT^*}(\mu) = \frac{\alpha \left( \left( \frac{\chi}{\chi^* (\mu)} \right)^2 - 1 \right) \chi^* (\mu) \left( (n - 1) \mu + 1 \right)^2}{(1 - \mu) n^3 \beta^2},
\]

\[
\rho^* (\mu) = \frac{1 - \rho_{CT^*} (\mu)}{n}, \text{ and}
\]

\[
F^* (\mu) = \frac{n(n - 1)p(\mu)^2 (2n \beta - p(\mu))^2}{16 \alpha \chi^2} + \frac{\alpha (n - 1)}{n}.
\]

There are three remarks on Proposition 15: First, there is only one country exerting the constrained globally optimal innovation efforts. Otherwise, countries would have to split the clean-tech refund. Exerting slightly more innovation efforts at little costs and thereby claiming the complete refund would be a profitable strategy in such a case.

Second, \( \rho_{CT} \) and \( F \) fulfill distinct roles. Increasing \( \rho_{CT} \) shifts the minimum of the local costs of the technology leader with respect to the innovation efforts towards higher innovation efforts. Specifically, \( \rho_{CT^*} (\mu) \) is such that the minimum is at \( \chi^* (\mu) \). For \( F = 0 \), the local costs of the technology leader are below the local costs of the other countries. However, this cannot be an equilibrium, as it would be a profitable strategy of other countries to increase innovation efforts slightly above the efforts of the current technology leader in order to claim the total clean-tech refund. This bidding-up process continues to such a high level of innovation efforts that the local costs of the current technology leader equals that of the other countries. However, this cannot be an equilibrium either, as it is not the minimum of the current technology leader’s local costs with respect to innovation efforts. Thus, the current technology leader would profit when decreasing innovation efforts. As a result, no equilibrium exists for \( F = 0 \).

To ensure the existence of an equilibrium, the lump-sum parameter \( F = F^* (\mu) \) is set such that at the point where the technology leader’s local costs with respect to innovation efforts are at a minimum, the local costs of the technology leader are equal to the remaining countries’ local costs.

Third, lump-sum parameters satisfying \( F > F^* (\mu) \) also implement the same subgame-perfect equilibrium. However, if \( F \) is too high, this violates the participation constraint of the technology leader (see Section 5.5.5). Therefore, we restrict our proposal, without loss of generality, to \( F = F^* (\mu) \).
5.5.5 Subgame Perfect Equilibrium

For notational convenience, define $RTIA^*(\mu) := \{\mu, \rho^*(\mu), \rho^{CT^*}(\mu), F^*(\mu)\}$ as the RTIA that induces the constrained globally optimal innovation efforts for any given allocation parameter $\mu$.

The following proposition summarizes the subgame perfect equilibrium under the RTIA. A graphical illustration of the RTIA and a comparison to the NTO is provided in Section 5.10.

**Proposition 16 (RTIA)**

For any allocation parameter $\mu \in (0, 1)$, the associated constrained globally optimal refunding and lump-sum parameters are determined in Proposition 15. The equilibrium abatement technology, the number of emission permits issued and the equilibrium price of emission permits is given by Equations (5.17), (5.16), and (5.15).

The next corollary compares local costs among countries under the RTIA and in the NTO.

**Corollary 6 (Local Costs under the RTIA)**

For any allocation parameter $\mu \in (0, 1)$,

1. The local costs under the $RTIA^*(\mu)$ are the same for all countries and,
2. The local costs for all countries under the $RTIA^*(\mu)$ are strictly lower than in the NTO.

**Proof of Corollary 6**

The cost equivalence under the $RTIA^*(\mu)$ was already part of Proposition 15. The second part follows from earlier results that under the $RTIA^*(\mu)$ both the abatement level as well as the technology level are (weakly) higher than in the NTO.

Comparing Corollary 5 and Corollary 6 shows one crucial difference between the NTO and the RTIA. In the NTO, Country 1 can only charge the amount others are willing to pay voluntarily. As a consequence, Country 1 has higher costs than the other countries. In contrast, under the RTIA, Country 1 gets fully compensated via the clean-tech refund (net of the lump-sum payment) and thus has the same costs as the other countries. This implies that, in Stage 2, Country 1 is better off free-licensing the frontier technology and receiving the clean-tech refund than selling the technology.

The next corollary shows that the non-cooperative outcome under the RTIA approximates the GSO if $\mu \to 0$ and the refunding and lump-sum parameter are set according
to Proposition 15.

**Corollary 7 (RTIA Approximates GSO)**

Suppose \( \mu \to 0 \), then the non-cooperative outcome under the RTIA* approximates the GSO.

**Proof of Corollary 7**

The proof consists of two steps. First, for \( \mu \to 0 \) we have \( p^{RTIA} \to p^{GSO} \). Thus, the RTIA deals efficiently with the free-rider problem in abatement. Second, for \( \mu \to 0 \) and \( \rho^{CT} = \rho^{CT*}(\mu) \), the technology level under the RTIA approaches the globally optimal one, i.e. \( \chi^*(\mu) \to \chi^{GSO} \). Thus, the RTIA deals efficiently with the free-rider problem in innovation efforts.

\( \square \)

### 5.5.6 Participation

So far, we have focused on globally efficient innovation efforts and abatement decisions of countries once the RTIA has been formed (compliance). In this section, we investigate whether it is individually rational for countries to join the RTIA (participation). For this purpose, we first make the strong assumption that the RTIA becomes effective only if all countries participate. In this case, a country’s local planner knows that if he does not join, the NTO obtains. Specifically, we find:\(^{23}\)

**Proposition 17 (Participation)**

Suppose the RTIA demands unanimous agreement. Participating in the RTIA*(\( \mu \)) is individually rational for all countries for any \( \mu \in (0, 1) \).

This result is in line with the result that the \( \gamma \)-core is non-empty if side-payments are feasible (Chander and Tulkens (1995), Hoel (1992)). The notion of the \( \gamma \)-core assumes that once a single country or a group of countries leave an agreement the remaining countries stop cooperation and continue as singletons.\(^{24}\) The strong punishment for deviators is the main driver of the positive result of Proposition 17. Whether this punishment is credible is subject to enduring discussions, and a variety of alternative stability concepts have been proposed.\(^{25}\)

---

\(^{23}\) The proof is deferred to Section 5.12.8.

\(^{24}\) Our assumption of unanimous agreement is a special case, as it only allows single countries to deviate but not the deviation of a group of countries.

5.6 Extensions

This section discusses refinements of the model with respect to the innovation process, technology diffusion and trading of the abatement technology under the NTO. Furthermore, we show that under a refunding treaty that only covers abatement, countries are worse off than under the RTIA.

5.6.1 Stochastic vs. Deterministic innovation

In the model discussed so far, innovation deterministically improves the abatement technology. This assumption has been made for reasons of tractability, but can also be justified on the grounds that improvements in the overall abatement technology involve a large number of – potentially stochastic – innovations. Invoking a law of large numbers, the overall technology level evolves deterministically, although the elementary innovations may be stochastic.

If innovation is modeled stochastically, the RTIA would still have the form developed in the previous section. However, rewarding innovation efforts has to be adjusted. For instance, if the individual success probability is low, it is possible that in the GSO more than one country innovate. This occurs as long as the expected gain from increasing the overall success probability outweighs the expected loss of costly duplications. In such a situation, the clean-tech refund has to be adjusted such that more countries have incentives to invest.

5.6.2 Technology Diffusion

So far, we have assumed that patent laws are fully enforceable, guaranteeing that technologies cannot be used by other countries unless they acquire a license. However, including technology diffusion by imitation or theft is straightforward. A first approach follows Golombek and Hoel (2005) and includes diffusion through effective investment, i.e. own investment plus gain from technology diffusion. A second approach allows for different spill-over rates between signatories and non-signatories. Clearly, with technology diffusion, the NTO worsens as there will be less innovation efforts; for the technology leader can extract less revenue from selling its technology. If all countries participate, the setup of the RTIA remains valid even when there are technological spill-overs, and its capacity to improve upon the NTO becomes even larger.
5.6.3 Alternative Pricing Mechanisms for Abatement Technology

In the NTO, when technology trading takes place in Stage 2, we implicitly assume that the seller has all bargaining power, and trading of the abatement technology is bilateral. In the following we discuss the impact of these assumptions in more detail.

Suppose the seller country keeps all bargaining power, but the buyer countries form a coalition. Trading the superior technology with this coalition instead of bilateral trading allows the seller country to increase the technology price per country by a factor of \((2n - 3)\) as the coalition members internalize the positive externality on other member countries when using a superior abatement technology. Therefore, the willingness to pay increases. Correspondingly, the seller exerts more innovation efforts and, compared to the original NTO, local costs for all countries would be lower in such a laissez-fair outcome. Yet, for each country, local costs under the latter are still strictly larger than under the \(RTIA^*(\mu)\) for all \(\mu \in (0,1)\). In any case, it seems that our original assumption of bilateral trading is more realistic. The formation of a buyer coalition is plagued by a free-riding problem of its own, as buyers have strong incentives to leave such a coalition and benefit from the higher technology level of the remaining members of the buyer coalition.

Alternatively, suppose there is bilateral bargaining, but bargaining power is distributed more equally among sellers and buyers. Clearly, this reduces the technology price, leads to less innovation efforts and increases the local costs of all countries in the NTO. The higher the bargaining power of the buyers, the stronger is this effect and the higher are the benefits of introducing the RTIA.

5.6.4 Rules Treaty for Abatement Only

We now establish the importance of supplementing climate treaties with instruments targeting at innovation of new abatement technologies. Specifically, we show that although treaties focusing solely on abatement create incentives to exert additional innovation efforts, the efforts fall significantly short of the optimal level.

Consider a variant of the rules treaty in which there is no clean-tech refund, i.e. \(\rho^{CT} = 0\). We refer to this specific treaty as a rules treaty for abatement (RTA). If only the
technology leader exerts innovation efforts, the technology price in the RTA is\textsuperscript{26}

\[
\pi^{RTA}(\chi^{RTA}, \chi) = \left( p^{RTA} \beta - \frac{p^{RTA} \beta^2}{2} \right) \left( \frac{1}{\chi^{RTA}} - \frac{1}{\chi} \right).
\] (5.18)

Since the price for emission permits is independent of the technology level, we have \( p^{RTA} = p^{RTIA} \). As the technology price cannot be negative, we set \( \pi^{RTA} = 0 \) if \( p^{RTA} > 2 \beta \). In such circumstances, the technology leader licenses its technology for free, even without receiving the clean-tech refund.\textsuperscript{27} The abatement technology in the RTA is

\[
\chi^{RTA} = \begin{cases} 
\min \left\{ \frac{4 \alpha \chi}{2(2n-1)p^{RTIA} \beta - np^{RTIA} \beta^2}, 1 \right\} \chi, & \text{if } p^{RTA} \leq 2 \beta, \\
\min \left\{ \frac{4 \alpha \chi}{2np^{RTIA} \beta - p^{RTIA} \beta^2}, 1 \right\} \chi, & \text{if } p^{RTA} > 2 \beta.
\end{cases}
\]

Comparing this to the outcome of the RTIA yields

\textbf{Proposition 18}

1. The technology level in the RTA is always inferior to that in the RTIA, i.e. for all \( \mu \in (0, 1) \), \( \chi^{RTA} > \chi^{RTIA} \).

2. Global emissions in the RTA are higher than under the RTIA for all \( \mu \in (0, 1) \).

Thus, abatement treaties alone cannot solve the double free-riding problem. This problem becomes more severe if cooperation with respect to abatement is high, as a higher willingness to abate worsens the negative effects of free-riding in innovation. A graphical illustration of the RTA and a comparison to the RTIA is deferred to Section 5.10.

\textbf{5.7 Practical Considerations}

Having discussed refinements and extensions of the analytical economic model in the previous section, we now focus on more practical issues concerning the implementation of treaties for innovation and abatement.

\textsuperscript{26} See Section 5.12.9.

\textsuperscript{27} The detailed argument why buyer countries might not be willing to pay a positive price for frontier technologies is provided in Section 5.11.
5.7.1 Assigning the Clean-Tech Refund

In our model, the entire country innovates and receives the clean-tech refund. In reality, there is much more disaggregation within countries, as many entities have to cooperate to develop new technologies. Broadly speaking, there is a hierarchy of basic and applied research. Basic research produces ideas and theories as well as prototypes, and has to be publicly funded, as it is not patentable. Applied research delivers marketable applications of the output of basic research, and is typically performed by private companies which acquire enforceable patents for their investment.

Clearly, the clean-tech refund has to cover both types of investment, and there are several ways to organize the funding of these different investments. First, the state obtains the clean-tech refund and does both – financing basic research and subsidizing private firms for applied research – requiring that firms license their marketable technologies for free.

Second, the clean-tech refund is split into two parts. One part is channeled directly to governments for funding basic research. The other part is given to private firms that have licensed new abatement technologies costlessly to other firms. The refund serves as a substitute for foregone licensing revenues.

However, one has to be aware of several possible inefficiencies arising when the firms’ innovation efforts are subsidized by the public, in particular since subsidies tend to crowd out private investment. How refunding for technology improvement should be organized and how competition of private firms can be put to use requires further research and lies beyond the scope of this paper.

5.7.2 Verifying Innovation and Abatement Efforts

Verification matters for both parts of the treaty – abatement and innovation. Local abatement efforts are prone to manipulation. China refused to allow outside monitoring, insisting on its sovereignty and non-interference in internal affairs (Harris, 2010). Many sources of GHGs, e.g. land-use-change, still lack appropriate and commonly-accepted accounting techniques (Nilsson et al., 2001). However, our proposal is not affected by the problem of verifying additionality, which is a major concern in credit systems like

28 There might be capacity constraints on innovation efforts, as there might not be enough scientists to scale up the innovation efforts, at least in the short run. The costs of redirecting innovation efforts to improve abatement technologies are part of the costs of innovation, as expressed by $\alpha((\frac{\Sigma}{\lambda})^2 - 1)$. 

135
the Clean Development Mechanism.\textsuperscript{29}

The verification of innovation success and the ensuing reduction of abatement costs is difficult to tackle as well. It requires reliable documentation of already-accomplished reductions, and objective measuring tools to predict future reductions. Thus, verification is a delicate matter and will entail a bureaucratic apparatus with its inefficiencies and flaws.

\textbf{5.7.3 Coping with Heterogeneity}

Heterogeneities across countries play an important role in the practical implementation of the RTIA. Only the developed countries have the capacity and knowledge base to operate at the frontier of abatement technology. In principle, this is no concern for the RTIA, as it suffices that two countries are able to innovate, which would even facilitate coordination on the investing country.

On the other side, developing countries tend to lack the capacity to absorb new abatement technologies. Thus, benefits from technology improvements may be smaller, due to insufficient adaptation. To some degree, this also defuses the perception that clean-tech refunding is aggravating inequalities between developing and industrialized countries. Fairness concerns could be taken into account by using different allocation parameters $\mu_D$ and $\mu_I$ for developing and industrialized countries, respectively. When $\mu_D > \mu_I$, firms in developing countries comparatively need to buy less permits than their counterparts in industrialized countries.

The abatement possibilities consist of numerous different technologies. Thus, if countries are heterogeneous, two further issues arise. First, each country’s optimal technology portfolio may differ from the others’ portfolio.\textsuperscript{30} Second, it is unlikely that one country leads in all technologies. Several countries might lead, each of them in one or several technologies. In such circumstances, several countries will receive refunds to the extent to which their developed technology reduces global emissions.

\textsuperscript{29} The problem of additionality occurs as the emission reduction achieved with a particular project have to be verified. This requires to estimate the emissions that would have occurred otherwise.

\textsuperscript{30} Whether a technology is suited for a country depends on its geographical conditions and its level of development. For example, wind power replacing coal in electricity production is less useful in a country where there is little or erratic wind. A developing country will not benefit from a more efficient gas power plant design if it lacks the capacity to build such plants in the first place.
5.7.4 Embedding the RTIA in Broader Treaties

Embedding incentives to develop new technology into climate treaties leads to a larger menu of policies. There is an extensive discussion in the literature whether widening the strategy space eases the participation problem. The main idea is to link emission reductions with topics where international cooperation is advantageous and at the same time, not hampered by free-riding problems. Our proposal can be seen along these lines, as it gives the participants access to advanced technologies.

5.8 Conclusion

In this paper, we discussed how the combination of technological and environmental policies can slow down climate change. To address the double free-riding problem, the development of new technologies is an indispensable ingredient of such policies. New technologies allow the reduction of the same amount of emissions at lower costs and thus, *ceteris paribus*, lead to lower emissions. We showed that leaving innovation to market forces alone leads to globally inefficient technology levels, especially if free-riding in abatement efforts takes place. In particular, we argued that buyer of licenses of superior technologies have a particularly low willingness to pay in such circumstances which stifles innovation.

We outlined a rules treaty for innovation and abatement that has the potential to solve this double free-riding problem with an allocation rule combined with a refunding rule. With the help of a simple model, we have shown that setting these rules implements the global social optimum. Together with a discussion of extensions and practical issues, the current proposal may help to identify alternative paths for future climate negotiations.

---

31 E.g. see Barrett (2005), Chapter 16.
5. Climate Treaties for Innovation and Abatement

5.9 Table of Variables

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>number of countries</td>
</tr>
<tr>
<td>$i, j$</td>
<td>country indices</td>
</tr>
<tr>
<td>$e$</td>
<td>business-as-usual emissions</td>
</tr>
<tr>
<td>$e_i$</td>
<td>GHG emissions of country $i$</td>
</tr>
<tr>
<td>$E$</td>
<td>global business-as-usual emissions</td>
</tr>
<tr>
<td>$E$</td>
<td>global GHG emissions</td>
</tr>
<tr>
<td>$e_i$</td>
<td>emission permits issued by local planner to domestic firm</td>
</tr>
<tr>
<td>$E$</td>
<td>global emission permits issued</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>innovation cost parameter</td>
</tr>
<tr>
<td>$\beta$</td>
<td>(environmental) damage parameter</td>
</tr>
<tr>
<td>$\chi$</td>
<td>business-as-usual abatement technology</td>
</tr>
<tr>
<td>$\chi_i'$</td>
<td>pre-trade abatement technology of country $i$</td>
</tr>
<tr>
<td>$\chi_i$</td>
<td>post-trade abatement technology of country $i$</td>
</tr>
<tr>
<td>$p$</td>
<td>price of emission permit</td>
</tr>
<tr>
<td>$\pi$</td>
<td>price of abatement technology</td>
</tr>
<tr>
<td>$K_i$</td>
<td>local costs of country $i$</td>
</tr>
<tr>
<td>$\mu$</td>
<td>allocation parameter</td>
</tr>
<tr>
<td>$\rho$</td>
<td>general refunding parameter</td>
</tr>
<tr>
<td>$\rho^{CT}$</td>
<td>clean-tech refunding parameter</td>
</tr>
<tr>
<td>$GSO$</td>
<td>global social optimum</td>
</tr>
<tr>
<td>$NTO$</td>
<td>no-treaty outcome</td>
</tr>
<tr>
<td>$RTIA$</td>
<td>rules treaty for innovation and abatement</td>
</tr>
<tr>
<td>$RTA$</td>
<td>refunding treaty for abatement only</td>
</tr>
<tr>
<td>$INN$</td>
<td>NTO with cooperation in innovation only</td>
</tr>
<tr>
<td>$ABA$</td>
<td>NTO with cooperation in abatement only</td>
</tr>
</tbody>
</table>

5.10 Graphical Illustrations

The following figures provide an example of some results we derived in the main text. The parameters are: $n = 4$, $\beta = 15$, $\overline{\chi} = 10$, $\overline{e} = 1300$ and $\alpha = 35$.\footnote{The values are such that in the NTO: $\chi_2' = \overline{\chi}$. They are merely for illustrative purpose and are not calibrated.}

Figure 5.1 compares the local costs in the NTO with those under a $RTIA^*(\mu = 0.001)$. The latter approximates the GSO. For the NTO starting at $\chi_1 = \overline{\chi}$, local costs for Country 2 decrease when Country 1 increases its innovation efforts, i.e. when $\chi_1$ decreases. Under the RTIA local costs for Country 2 remain constant. For the NTO and the RTIA, local costs of Country 1 decrease until the equilibrium minimum point is
reached. Increasing innovation efforts even further, local costs of Country 1 rise steeply, as innovation costs start dominating. In the NTO, Country 1’s local costs are higher than Country 2’s, which is in accordance with Corollary 5. It holds that $\chi_{GSO} < \chi_{NTO}$, which shows that the non-cooperative outcome with respect to innovation is globally inefficient. In the RTIA, at $\chi_1 = \chi$ average local costs are lower than in the NTO, as there is, in addition, cooperation in abatement. Furthermore, average local costs are minimal at Country 1’s local costs minimum, which shows that the RTIA can implement the constraint globally optimal outcome.

Figure 5.2 compares the local costs of the scenarios we considered. For all countries,
local costs under the RTIA*(μ) for any μ ∈ (0, 1) are strictly below the local costs in the NTO. This holds true even if μ → 1. The reason is that albeit the agencies revenues approach zero, ρCT approaches infinity and the RTIA thus, in the limit, still induces technology levels superior to those found in the NTO. If μ → 0, the local costs under the RTIA approach the GSO.

Figure 5.3 illustrates the outcome of the RTA and compares it to the RTIA. At μ = 1, χRTIA < χRTA = χNTO. When μ decreases, χRTA declines but πRTA stays rather constant. Hence Country 1’s main incentive to do further innovation is to lower its damages via improving the technology that all countries use. At the point where πRTA declines rather abruptly, χRTA remains constant as the damage effect and the decrease in revenues approximately compensate each other. For πRTA = 0, χRTA declines again. χRTIA, on the other hand, is a smooth function of μ.

33 Remember that all countries have the same local costs in the equilibrium under the RTIA.
5. Climate Treaties for Innovation and Abatement

Figure 5.3: The blue curves are the outcomes of the RTA, the red ones of the RTIA. Note that permit prices are the same in the RTIA and RTA and that there is no pricing of technology in the RTIA.

5.11 Negative Willingness to Pay in the RTA

In this section we discuss the rather surprising result that in the RTA as $\mu$ decreases, the buyer country’s willingness to pay for the superior technology decreases and finally even turns negative.

Consider Equation (5.18). The first term in the first bracket is positive and linear in the permit price, as it stems from the local damages reduction. The second term in the brackets is negative and quadratic in the permit price, as it stems from the cost increase in the abatement costs. Surprisingly, although the superior technology decreases abatement cost for a given amount of abatement, the total abatement costs increase. The reason is that better technology induces higher abatement levels, which overcompensates the cost decrease for a certain level. As a consequence, for $p_{RTA}$ sufficiently high, the quadratic term dominates and buying the technology makes the country worse.
off. This result clearly depends on how the damage and abatement costs are modeled and, in our case, is driven by the assumption of linear damages on the one side and quadratic abatement costs on the other side.

To get further intuition, consider a setting where a firm faces the decision of either buying (or not selling) permits or reducing emissions. Suppose that permit prices are fixed and the value of the permits the firm is being grandfathered is sunk. Hence, a firm minimizes \( pe + \frac{1}{2}(\pi - e)^2 \), and with \( e = \pi - \frac{p}{\chi} \) we get \( p\pi - \frac{p^2}{2\chi} \). The firms’ costs unambiguously decrease in the technology level. The firm level, however, is not the case we are considering. Instead, our focus is on the national level where the local damages, \( \beta E \), are also accounted for. In addition, in our setting there are no costs associated with buying emission permits as the local planner creates exactly the amount needed by the firm. Hence, the term \( pe \) is dropped and other terms related to permit trading or refunding do not enter either.\(^{34}\) Knowing that he can create the amount of permits needed, the local planner only cares about the balance between abatement costs on the one side and environmental damages on the other side.

The result hinges on the assumption that the firms always use the best available technology. To see this, consider the case where the willingness to pay is negative \( (p^{RTA} > 2\beta) \). If the local planner could influence the firms’ decision, he would induce the firm to use \( \overline{\chi} \) instead of \( \chi^{RTA} \), even though the technology leader does not charge anything for it. This influence could be direct via a command and control approach. A more indirect way would be to promise the firms that, for whatever they emit, they receive both the fraction of permits the local planner is allowed to keep and, in addition, the revenues of the refunding. Thus, the only costs the firm would be left with are the abatement costs and would thus voluntarily use \( \overline{\chi} \). The RTA has to be set up in a way that prevents such a counterproductive behavior of countries.

\(^{34}\) Note that parts of the permits are channeled through the refunding scheme. Due to symmetry, the price firms pay the agency to buy the permits exactly equals the refunds the country receives.
5.12 Proofs

5.12.1 Proof of Proposition 13

Let $\lambda$ denote the Lagrange multiplier associated with the constraint. The Karush-Kuhn-Tucker conditions read

$$\frac{\partial}{\partial e} = -\chi (\bar{e} - e) + n\beta = 0,$$

$$\frac{\partial}{\partial \chi} = \frac{n}{2} (\bar{e} - e)^2 - 2\alpha \frac{\chi^2}{\chi^3} + \lambda = 0,$$

$$\chi - \bar{\chi} \leq 0, \quad \lambda \geq 0, \quad \lambda (\chi - \bar{\chi}) = 0.$$

Summing up over all countries and rearranging terms yields

$$E = \bar{E} - \frac{n^2 \beta}{\chi},$$

$$\chi^3 = \frac{4\alpha \bar{\chi}^2}{n(\bar{e} - e)^2 + 2\lambda}.$$

Suppose the constraint is non-binding. In this case, the complementarity slackness condition yields $\lambda = 0$, and since $\bar{e} - e = \frac{n\beta}{\chi}$, we get $\chi = \frac{4\alpha \bar{\chi}^2}{n^2 \beta^2}$. This a solution to the optimization problem if and only if it satisfies the constraint $\chi - \bar{\chi} \leq 0$, which is true as long as $4\alpha \bar{\chi} \leq n^2 \beta^2$. Otherwise, the globally optimal technology level is $\chi = \bar{\chi}$. □

5.12.2 Proof of Equations (5.13a) and (5.13b)

Let $\lambda_1$ denote the Lagrange multiplier associated with the constraint. The Karush-Kuhn-Tucker conditions are given by

$$\frac{n\beta^2}{\lambda_1^2} - \frac{\beta^2}{2\lambda_1^2} - 2\alpha \frac{\bar{\chi}^2}{\lambda_1} + (n - 1) \frac{\beta^2}{2\lambda_1} + \lambda_1 = 0,$$

$$\chi_1 - \bar{\chi} \leq 0, \quad \lambda_1 \geq 0, \quad \lambda_1 (\chi_1 - \bar{\chi}) = 0.$$

Suppose the constraint is non-binding. Then, $\lambda_1 = 0$. Solving for $\chi_1$ yields

$$\chi_1 = \frac{4\alpha \bar{\chi}}{(3n - 2)\beta^2} \bar{\chi}.$$
5. Climate Treaties for Innovation and Abatement

This is a solution as long as $\chi_1 \leq \chi$, which is true if $4\alpha\chi \leq (3n - 2)\beta^2$. Otherwise, $\chi_1 = \chi$. Hence, $\chi_1 = \min \{ \frac{4\alpha\chi}{(3n - 2)\beta^2}, 1 \} \chi$.

Let $\lambda_2$ denote the Lagrange multiplier associated with the constraint. The Karush-Kuhn-Tucker conditions read

$$-2\alpha\frac{\chi^2}{\chi_2^2} + \frac{\beta^2}{2\chi_2^2} + \lambda_2 = 0$$

$$\chi_2' - \chi \leq 0, \lambda_2 \geq 0, \lambda_2 (\chi_2' - \chi) = 0$$

Suppose the constraint is non-binding. Then, $\lambda_2 = 0$. Solving for $\chi_2'$ yields

$$\chi_2' = \frac{4\alpha\chi^2}{\beta^2},$$

which is a solution as long as $\chi_2' \leq \chi$ or, equivalently, $4\alpha\chi \leq \beta^2$. Otherwise, $\chi_2' = \chi$. Hence, $\chi_2' = \{ \frac{4\alpha\chi^2}{\beta^2}, 1 \} \chi$. □

5.12.3 Proof of Corollary 5

The cost functions are

$$K_1^{NTO} = \beta \left( E - \frac{n\beta}{\chi_1} \right) + \frac{\beta^2}{2\chi_1} + \alpha \left( \frac{\chi}{\chi_1} \right)^2 - 1 \pi(\chi_1, \chi_2),$$

$$K_2^{NTO} = \beta \left( E - \frac{n\beta}{\chi_1} \right) + \frac{\beta^2}{2\chi_1} + \alpha \left( \frac{\chi}{\chi_2} \right)^2 - 1 + \pi(\chi_1, \chi_2),$$

$$K_3^{NTO} = \beta \left( E - \frac{n\beta}{\chi_1} \right) + \frac{\beta^2}{2\chi_1} + \pi(\chi_1, \chi_2),$$

$$K^{NoR&D} = \beta \left( E - \frac{n\beta}{\chi} \right) + \frac{\beta^2}{2\chi},$$

For convenience, define $\gamma = \frac{4\alpha\chi}{(3n - 2)\beta^2}$. There are three possible cases to consider:

**Case 1:** Suppose $4\alpha\chi \geq (3n - 2)\beta^2$. This implies $\chi_1 = \chi_2', \chi_2 = \chi$, $\forall j = 3, \ldots, n$. Because there is no innovation efforts, this directly leads to $K_j^{NTO} = K_2^{NTO} = K_1^{NTO} = \chi$, $\forall j = 3, \ldots, n$.

**Case 2:** Suppose $\beta^2 \leq 4\alpha\chi < (3n - 2)\beta^2$. This implies $\chi_1' < \chi_2' = \chi_2 = \chi$, $\forall j = 3, \ldots, n$. Because Country 2 exerts no innovation efforts, $K_j^{NTO} = K_2^{NTO}$, $\forall j = 3, \ldots, n$.

144
3, ..., n. First, the cost difference between Country 2 and Country 1 is given by

\[ K_2^{NTO} - K_1^{NTO} = \alpha \left( \left( \frac{\Sigma}{\chi_2} \right)^2 - \left( \frac{\Sigma}{\chi_1} \right)^2 \right) + n \pi(\chi_1, \chi_2) \]

\[ = \frac{\beta^2}{4\gamma \chi} \left( 1 - \gamma \right) \left( 2n - (1 + \gamma)(3n - 2) \right) \]

Since \( \gamma \in (0, 1] \), the first two factors are unambiguously positive. The third factor attains its maximum as \( \gamma \to 0 \) and amounts to \( 2 - n \), which is non-positive as long as \( n \geq 2 \). This establishes \( K_2^{NTO} < K_1^{NTO} \). Second, the cost difference between Country 1 and the benchmark without innovation reads

\[ K_1^{NTO} - \bar{K} = \left( \frac{\beta^2}{2} - n\beta^2 \right) \left( \frac{1}{\chi_1} - \frac{1}{\chi} \right) + \alpha \left( \left( \frac{\Sigma}{\chi_1} \right)^2 - 1 \right) - (n - 1) \pi(\chi_1, \chi_2) \]

\[ = -\frac{(\gamma - 1)^2(3n - 2)\beta^2}{4\gamma \chi} < 0, \]

which establishes \( K_1^{NTO} < \bar{K} \).

**case 3:** Suppose \( \beta^2 > 4\alpha \chi \). This implies \( \chi_1 < \chi_2' < \chi_2 = \chi \), \( \forall j = 3, ..., n \). First, for an arbitrary country \( j = 3, ..., n \), the cost difference to Country 2 is given by

\[ K_j^{NTO} - K_2^{NTO} = -\alpha \left( \left( \frac{\Sigma}{\chi_2} \right)^2 - 1 \right) < 0 \]

which establishes \( K_j^{NTO} < K_2^{NTO} \), \( \forall j = 3, ..., n \). Second, the cost difference between Country 2 and Country 1 is given by

\[ K_2^{NTO} - K_1^{NTO} = \alpha \left( \left( \frac{\Sigma}{\chi_2} \right)^2 - \left( \frac{\Sigma}{\chi_1} \right)^2 \right) + n \pi(\chi_1, \chi_2') \]

\[ = -\frac{3(n - 1)^2\beta^2}{4\gamma(3n - 2)\chi} < 0, \]

which establishes \( K_2^{NTO} < K_1^{NTO} \). Third, the cost difference between Country 1 and the benchmark without innovation reads

\[ K_1^{NTO} - K_{NoR&D} = \left( \frac{\beta^2}{2} - n\beta^2 \right) \left( \frac{1}{\chi_1} - \frac{1}{\chi} \right) + \alpha \left( \left( \frac{\Sigma}{\chi_1} \right)^2 - 1 \right) - (n - 1) \pi(\chi_1, \chi_2') \]

\[ = \frac{\beta^2}{4\gamma \chi} \left( -(3n - 2)\gamma^2 - 2(1 - 2n)\gamma - n - \frac{(n - 1)^2}{3n - 2} \right) \]

The term in brackets is a quadratic Equation in \( \gamma \) that attains its maximum at
5. Climate Treaties for Innovation and Abatement

\[ \gamma = \frac{1-2n}{3n-2} < 0 \] and is monotonically decreasing for \( \gamma > \frac{1-2n}{3n-2} \). Given the support \( \gamma = (0, 1] \), evaluating the term in brackets at \( \gamma \to 0 \) yields \( -\left(n + \frac{(n-1)^2}{3n-2}\right) \), which is always negative. Hence, \( K_1^{NTO} < K^{NoR&D} \).

Summing up, \( K_j^{NTO} \leq K_2^{NTO} \leq K_1^{NTO} \leq K^{NoR&D} \), \( \forall j = 3, \ldots, n \). \( \square \)

5.12.4 Proof of Table 5.1

In general the local costs function for Country 1 is:

\[ K_1 = \beta E + \frac{p^2}{2\chi_1} + \alpha \left( \left( \frac{\chi}{\chi_1} \right)^2 - 1 \right) - (n-1) \pi \]

**Case AUT:** For autarky, \( p = \beta, \ E = E - \frac{\mu}{\chi_1} \) and \( \pi = 0 \). The cost minimization problem with respect to the abatement technology reads

\[ \min_{\chi_1} \left\{ \beta \left( E - \frac{\beta}{\chi_1} \right) + \frac{\beta^2}{2\chi_1} + \alpha \left( \left( \frac{\chi}{\chi_1} \right)^2 - 1 \right) \right\} \]

Solving the first order condition

\[ \frac{\beta^2}{\chi_1^2} - \frac{\beta^2}{2\chi_1^3} - \frac{2\alpha \chi^2}{\chi_1^3} = 0 \]

for \( \chi_1 \) yields the autarky technology level given in the table.\(^{35}\)

**Case INN:** For cooperation with respect to innovation only, \( p = \beta, \ E = E - n \frac{\mu}{\chi_1} \) and \( \pi = 0 \). Furthermore we take into account that Country 1 considers global damages and global abatement costs. The cost minimization problem with respect to the abatement technology reads

\[ \min_{\chi} \left\{ n \left( \beta \left( E - n \frac{\beta}{\chi} \right) + \frac{\beta^2}{2\chi} \right) + \alpha \left( \left( \frac{\chi}{\chi} \right)^2 - 1 \right) \right\} \]

Solving the first order condition

\[ n^2 \frac{\beta^2}{\chi^2} - n \frac{\beta^2}{2\chi^3} - 2\alpha \frac{\chi^2}{\chi_1^3} = 0 \]

for \( \chi_1 \) yields the cooperation with respect to innovation-only technology level given

\(^{35}\) An alternative way to compute the autarky technology level is based on the solution of the RTIA. Setting \( \mu \to 1 \) gives the desired result.
in the table.

**Case ABA:** For cooperation with respect to abatement only, \( p = n\beta, E = \bar{E} - \frac{p}{\chi^2} \) and \( \pi = (p\beta - E\beta)(\frac{1}{\chi^2} - \frac{1}{\chi}) \). The cost minimization problem with respect to the abatement technology reads

\[
\min_{\chi_1} \left\{ \beta \left( E - \frac{n\beta}{\chi_1} \right) + \frac{n^2\beta^2}{2\chi_1} + \alpha \left( \frac{\chi_1}{\chi^2} - 1 \right) - (n-1)(n\beta^2 - \frac{n^2\beta^2}{2}) \left( \frac{1}{\chi_1} - \frac{1}{\chi} \right) \right\}
\]

Solving the first order condition

\[
\frac{n^2\beta^2}{\chi^2_1} - \frac{n^2\beta^2}{2\chi_1^2} - 2\alpha \frac{\chi_1^2}{\chi^2} + (n-1)(n\beta^2 - \frac{n^2\beta^2}{2}) \frac{1}{\chi_1} = 0
\]

for \( \chi_1 \) yields the technology level for cooperation with respect to abatement only as given in the table.

The global emission follow using \( E = \bar{E} - \frac{p}{\chi} \). \( \square \)

### 5.12.5 Proof of Equation (5.15)

The respective first-order conditions read

\[
\frac{d K_1^{RTIA}}{d \epsilon_1} = -\frac{\chi}{n}(\bar{\epsilon} - \epsilon) + \beta + p\left( \frac{1}{n} - \mu \right) - \frac{\chi}{n}(e - \mu\epsilon_1) - (\rho + \rho^{CT})(1 - \mu) \left( p - \frac{\chi}{n} \epsilon \right) = 0
\]

\[
\frac{d K_2^{RTIA}}{d \epsilon_2} = -\frac{\chi}{n}(\bar{\epsilon} - \epsilon) + \beta + p\left( \frac{1}{n} - \mu \right) - \frac{\chi}{n}(e - \mu\epsilon_2) - \rho(1 - \mu) \left( p - \frac{\chi}{n} \epsilon \right) = 0
\]

Summing over all countries yields

\[
\sum_{i=1}^{n} \left( -\frac{\chi}{n}(\bar{\epsilon} - \epsilon) + \beta + p\left( \frac{1}{n} - \mu \right) - \frac{\chi}{n}(e - \mu\epsilon_i) \right) - (n\rho + \rho^{CT})(1 - \mu) \left( p - \frac{\chi}{n} \epsilon \right) = 0
\]

\[\leftrightarrow \quad -p + n\beta + p(1 - \mu n) - \frac{\chi}{n}(E - \mu\epsilon) - (n\rho + \rho^{CT})(1 - \mu) \left( p - \frac{\chi}{n} \epsilon \right) = 0\]

\[\leftrightarrow \quad p = \frac{n\beta}{(n-1)\mu + 1},\]

where the last line follows from the market clearing condition \( \epsilon = E \) and \( n\rho + \rho^{CT} = 1 \). \( \square \)
5.12.6 Proof of Equation (5.17)

The respective cost minimization problem reads

\[
\min_{\chi} \left\{ n \left( \frac{p_{RTIA}^2}{2\chi} + \beta \left( \frac{E}{\chi} - n \frac{p_{RTIA}}{\chi} \right) \right) + \alpha \left( \left( \frac{\chi}{\chi} \right)^2 - 1 \right) \right\}
\]

subject to

\[
\chi - \overline{\chi} \leq 0.
\]

Let \( \lambda \) denote the Lagrange multiplier associated with the constraint. The Karush-Kuhn-Tucker conditions read

\[
- \frac{np_{RTIA}^2}{2\chi^2} + \beta \frac{n^2 p_{RTIA}}{\chi^2} - 2\alpha \frac{\chi^2}{\chi^3} + \lambda = 0, \\
\chi - \overline{\chi} \leq 0, \lambda \geq 0, \lambda (\chi - \overline{\chi}) = 0.
\]

Suppose the constraint is non-binding. In this case, the complementarity slackness condition yields \( \lambda = 0 \) and we get

\[
\chi^* = \frac{4\alpha \overline{\chi}^2}{np_{RTIA}(2n\beta - p_{RTIA})}
\]

This a solution to the optimization problem if and only if it satisfies the constraint \( \chi - \overline{\chi} \leq 0 \), which is true as long as \( 4\alpha \overline{\chi} \leq n^3 \beta^2 (1 + \frac{(n-1)\mu}{(n-1)\mu + 1}) \). Otherwise, the globally optimal technology level is \( \chi^* = \overline{\chi} \).

5.12.7 Proof of Proposition 15

First, we determine the first-order condition of local costs of Country 1 (5.14a) with respect to the technology level under a RTIA \( \{\mu, \rho, \rho^{CT}, F\} \). Using Equations (5.1), \( E = \overline{E} - \frac{np}{\chi^1} \) and (5.16a) together with \( \frac{\partial E}{\partial x} = \frac{np}{\chi^1}, \frac{\partial e}{\partial x} = \frac{p}{\chi^1} \) and \( \frac{\partial \epsilon}{\partial x} = \frac{p}{\chi^1} - 2(n-1) \frac{1-\mu}{\mu} \rho^{CT} \frac{p}{\chi^1} \), the first-order condition reads

\[
\frac{n\beta p}{\chi^1} - \frac{p^2}{2\chi^1} - 2\alpha \frac{\chi^2}{\chi^1} + p \left( \frac{p}{\chi^2} - \mu \left( \frac{p}{\chi^2} - 2(n-1) \frac{1-\mu}{\mu} \rho^{CT} \frac{p}{\chi^2} \right) \right) - (\rho + \rho^{CT})(1 - \mu) \frac{np}{\chi^1} = 0
\]
Substituting for the permit price (5.15) yields

\[ \rho^{CT} = \frac{4\alpha \chi_2 - p(2n\beta - p)}{2p^2(1 - \mu)(n - 1)}. \]  

(5.19)

Suppose there is an interior solution of (5.17). Setting \( \chi_1 = \chi^*(\mu) = \frac{4\alpha \chi_2}{np(2n\beta - p)} \) yields the clean-tech parameter that ensures that \( \chi^*(\mu) \) is the minimum of Country 1’s local costs

\[ \rho^{CT^*} = \frac{(n - 1)\mu + 0.5}{1 - \mu}. \]  

(5.20)

Plugging \( \rho^{CT^*} \) and \( \chi_1 = \chi^*(\mu) = \frac{4\alpha \chi_2}{np(2n\beta - p)} \) into the local costs functions (5.14a) and (5.14b) and using the constraint optimal amount of emission permits (5.16a) and (5.16b) yields

\[ K_{RTIA}^1(\chi_1 = \chi^*(\mu)) = -\frac{n^2p^2(2n\beta - p)^2}{16\alpha \chi^2} - \alpha + \beta E + F, \]

\[ K_{RTIA}^2(\chi_1 = \chi^*(\mu)) = \beta E - \frac{F}{n - 1}. \]

Next, given \( \mu \), we define \( \overline{F}(\mu) \) as the lump-sum parameter that equalizes local costs in countries 1 and 2: \( K_{1}^{RTIA}(\chi_1 = \chi^*(\mu)) = K_{2}^{RTIA}(\chi_1 = \chi^*(\mu)) \).

\[ \overline{F}(\mu) = \frac{n(n - 1)p(\mu)^2(2n\beta - p(\mu))^2}{16\alpha \chi^2} + \frac{\alpha(n - 1)}{n}. \]  

(5.21)

Consider a RTIA \( \{\mu, \rho, \rho^{CT}, F\} \). If \( F < \overline{F}(\mu) \) an equilibrium that implements \( \chi^*(\mu) \) does not exist. The proof is as follows: Suppose there are two countries A and B, and country A is the technology leader with \( \chi_A = \chi^*(\mu) \). Since \( F < \overline{F}(\mu) \), local costs in country B are higher than local costs in country A. By exerting slightly more innovation efforts than country A, country B would steal the complete clean-tech refund and lower its local costs. Clearly, country A responds to increase its innovation efforts as well. This process continues until costs are equalized between country A and B. As a consequence, one country, say A, would exert innovation efforts while the other, say B, would not innovate. However, this cannot be an equilibrium because given that country B does not exert innovation efforts, country A chooses \( \chi_A = \chi^*(\mu) \). Thus, for an equilibrium to exist, \( F \geq \overline{F}(\mu) \).

Furthermore, if \( F \) is sufficiently large, the costs for Country 1 can exceed the costs for
5. Climate Treaties for Innovation and Abatement

either leaving the treaty or the costs for selling the technology in Stage 2. Thus, in our treaty proposal, we restrict to \( F = F(\mu) \) since this always guarantees the existence of a constrained optimal subgame perfect equilibrium by construction. □

5.12.8 Proof of Proposition 17

Because \( K_{i}^{NTO} \leq K_{2}^{NTO} \leq K_{1}^{NTO} \leq K^{NoRkd} \), \( \forall i = 3, \ldots, n \) (see Corollary 5) and \( K_{j}^{RTIA} = K^{RTIA} \), \( \forall j = 1, \ldots, n \) (see Proposition 15), it suffices to show that \( K_{i}^{NTO} \geq K^{RTIA} \).

First, suppose \( \chi^{NTO} = \overline{\chi} \) and \( \chi^{RTIA} = \underline{\chi} \). This directly yields \( K_{i}^{NTO} \geq K^{RTIA} \). Next, suppose \( \chi^{NTO} = \overline{\chi} \) and \( \chi^{RTIA} < \underline{\chi} \). Due to optimization, we immediately know that \( K^{RTIA}(\chi^{RTIA}) < K^{RTIA}(\overline{\chi}) = K_{i}^{NTO} \). Hence, it remains to show that for interior solutions of the technology level in the NTO and the RTIA \( K_{i}^{NTO} \geq K^{RTIA} \).

The cost under RTIA is given by

\[
K^{RTIA} = \frac{\chi}{2} (\overline{\gamma} - e)^2 + \beta E + p^{RTIA} (e - \mu \epsilon_j) - \rho (1 - \mu) p^{RTIA} E
\]

\[
= \frac{p^{RTIA2} \chi}{2\chi} + \beta \left( \frac{E - np^{RTIA}}{\chi} \right) + (1 - \mu) \rho^{CT} \frac{p^{RTIA2} \chi}{\chi}
\]

\[
= \beta \frac{\overline{E} - \alpha}{n} - \frac{np^{RTIA2} (2n\beta - p^{RTIA})^2}{16 \alpha \chi^2},
\]

where the second line follows from the equilibrium conditions on the emission market, zero-profits for the agency controlling the refunding scheme and the optimal amount of emission permits issued by Countries \( j = 2, \ldots, n \). The last line additionally uses the optimal technology level and the optimal clean-tech refund for given \( \mu \). Since \( p^{RTIA} \) is decreasing in \( \mu \), the cost function \( K^{RTIA} \) is increasing in \( \mu \). Hence, it suffices to show that \( K^{RTIA|\mu=1} \leq K_{i}^{NTO} \). Evaluating the cost at \( \mu = 1 \) yields

\[
K^{RTIA|\mu=1} = \beta \frac{\overline{E} - \alpha}{n} - \frac{n(2n - 1)^2 \beta^4}{16 \alpha \chi^2}.
\]
The cost function under NTO is given by

\[ K_{3\text{NTO}} = \frac{\chi}{2} (\bar{e} - e)^2 + \beta E + p_{\text{NTO}} (e - \varepsilon) + \pi(\chi_{\text{NTO}}, \chi_{2\text{NTO}}) \]

\[ = \beta E - \beta^2 \frac{(n - 1)}{\chi_{\text{NTO}}} - \frac{\beta^2}{2} \frac{1}{\chi_{2\text{NTO}}} \]

\[ = \beta E - \frac{4 (n - 1) (3n - 2) \beta^2}{16 \alpha \chi^2} - \frac{\beta^2}{2} \frac{1}{\chi_{2\text{NTO}}}, \]

where the second line follows from the equilibrium conditions on the emission market, the optimal amount of emission permits issued by countries \( i = 3, \ldots, n \) and the equilibrium price for the frontier technology. The last line additionally uses the optimal technology level.

**Case 1:** Suppose \( 4\alpha \bar{\chi} < \beta^2 \). Then, \( \chi_{2\text{NTO}} = \frac{4 \alpha \bar{\chi}^2}{\beta^2} \) and

\[ K_{3\text{NTO}} - K_{\text{RTIA}}|_{\mu=1} = \frac{\alpha}{n} - \frac{\beta^4}{16 \alpha \bar{\chi}} \left[ 10 - 21 n + 16 n^2 - 4 n^3 \right]. \]

The term in square brackets is decreasing in \( n \) and equals to zero for \( n = 2 \). Hence, the cost difference is strictly positive.

**Case 2:** Suppose \( 4\alpha \bar{\chi} \geq \beta^2 \). Then, \( \chi_{2\text{NTO}} = \bar{\chi} \) and

\[ K_{3\text{NTO}} - K_{\text{RTIA}}|_{\mu=1} = \frac{\alpha}{n} - \frac{\beta^4}{16 \alpha \bar{\chi}} \left[ 8 - 21 n + 16 n^2 - 4 n^3 \right] - \frac{\beta^2}{2} \frac{1}{\bar{\chi}} \]

The term in square brackets is unambiguously negative and since \( 4\alpha \bar{\chi} \geq \beta^2 \), we get

\[ K_{3\text{NTO}} - K_{\text{RTIA}}|_{\mu=1} \geq \frac{\alpha}{n} - \alpha \left[ 8 - 21 n + 16 n^2 - 4 n^3 \right] - \frac{\beta^2}{2} \frac{1}{\bar{\chi}} \]

\[ \geq \frac{\alpha}{n} - \alpha \left[ 8 - 21 n + 16 n^2 - 4 n^3 \right] - \frac{\beta^2}{2} \frac{1}{\bar{\chi}} \]

\[ \geq \frac{\alpha}{n} - \alpha \left[ 8 - 21 n + 16 n^2 - 4 n^3 \right] - 2 \alpha \]

\[ \geq \frac{\alpha}{n} - \alpha \left[ 10 - 21 n + 16 n^2 - 4 n^3 \right]. \]

The expression in square brackets is decreasing in \( n \) and equals to zero for \( n = 2 \). Hence, the cost difference is strictly positive.

Combining case 1 and case 2 completes our proof.
5. Climate Treaties for Innovation and Abatement

5.12.9 Proof of Equation (5.18)

To derive Equation (5.18), we consider the case in which only one country innovates, i.e. \( \chi'_2 = \bar{\chi} \). Countries are symmetric and \( \rho^{CT} = 0 \). Thus \( \rho = \frac{1}{n} \) and \( e^{RTA} = e^{RTA} \).

For this reason the local costs of trading, \( p^{RTA}(e^{RTA} - \mu e^{RTA}) \) and the refund revenue, \( \rho(1-\mu)p^{RTA}E^{RTA} \), add up to zero irrespective of whether a country buys the technology or not. As a consequence, we we can directly apply Equation (5.11) to calculate the technology price \( p = p^{RTA} \). \( \square \)
6 A Climate Change Economist’s Guide to Modeling the Earth System

6.1 Introduction

6.1.1 Motivation

There is considerable disagreement in the literature about the adequate response to climate change. Roughly, a dividing line can be drawn between the ‘anxious’ persons, which call for urgent and drastic reductions of greenhouse gas (GHG) emissions and the ‘calm’ ones, which recommend minor reductions in the short run, to be increased smoothly with time.\(^1\) One may also divide the professionals dealing with this question into those with a background in climate science and those with a background in economics.\(^2\) It came to my attention that a majority of climate scientists belongs to the ‘anxious’ group, whereas economist are split between the ‘anxious’ and the ‘calm’.

Many answers have been offered for this conundrum. In this paper, we would like to highlight a point which, in my opinion, has gotten less attention than it deserves. Before doing so, let us set the stage and briefly discuss the different approaches used by climate scientists and economists to support their recommendations. This already leads us to some common answers to the said conundrum.

Climate scientist are primarily concerned with leaving the earth in a condition similar to the one we inherited, as captured in the concept of sustainability, for instance (Brundtland and World Commission on Environment and Development, 1987). They

\(^1\) We note that there are also the ‘sceptics’, which do not believe in man-made climate change and thus see no point in reducing emissions.

\(^2\) Of course, the picture is more nuanced than these divisions suggest. But for the sake of the argument, let us take this simplified view of the world.
Also put high emphasis on the risks associated with climate change. This mindset is the reason why most climate scientists subscribe to the idea that certain thresholds should not be crossed, in order to ensure that the earth system remains close to the present state. When selecting these thresholds climate scientists place little attention to the costs of emission reduction. The most widely-accepted thresholds not to be exceeded are a temperature increase of $2\, ^\circ C$, or a GHG level in the atmosphere of 450 ppm. Both can only be achieved through significant emission cuts (Höhne et al., 2012; Peters et al., 2012) so that climate scientists generally belong to the ‘anxious’ group.

The economists’ method for tackling this issue is an intertemporal cost-benefit analysis, in which they weigh the damages from emissions against the costs of emission reduction, and thereby derive the ‘optimal’ amount of emissions. Such outcomes mainly depend on the way damages and abatement costs are modeled, together with the choice of the parameters that enter into such models. To the extent that there is disagreement among economists on both issues, their recommendations may be either ‘anxious’ or ‘calm’.

A very important source of disagreement is the weight future generations should be given when deciding about current actions. Economists answer this question by discounting the future’s consumption (Dasgupta, 2008). The higher the discount rate, the higher the value of current consumption as compared to future consumption. In a world without uncertainty, it is customary to separate the discount rate into two components. According to the first component, the pure time discount rate, the future’s consumption has less value, simply because it lies in the future. There is no objectively correct way to set the pure time discount rate. Some scientists argue that based on the observed behavior of people in the market place, it should be significantly above zero (Nordhaus, 2007). Other scientists, mostly listing normative arguments, state that the pure time discount rate should be very small, or even zero (Stern, 2007; Ramsey, 1928). The second component is the growth discount rate. Its idea is that if future generations will be richer than the present one, we ought to care more about the present than about the future. This is a reason for discounting under the assumption that there will be economic growth. If an economist uses a high discount rate for any of these two reasons, it is likely that his model’s ‘optimal’ emissions belong to the ‘calm’ philosophy. Lately, economists account for inequality not only between, but also within

---

3 The economists’ cost-benefit analysis might seem ‘superior’ to the climate scientists’ one-sided approach. Yet, see Ackerman et al. (2009) for weaknesses of the cost-benefit analysis in the context of climate change.

4 This question is crucial because the environmental damages of climate change will mostly affect future generations, whereas the costs of emission reduction would occur now.

generations (Heal, 2009), for relative prices (Sterner and Persson, 2008) as well as for uncertainty (Weitzman, 2009; Gollier, 2002). Explicitly or implicitly, including these issues into a model decreases the discount rate and thus has the potential to make the model’s ‘optimal’ emissions ‘anxious’ ones.

In this paper, we highlight a less appreciated but potentially quite relevant reason why economists’ recommendations are of the ‘calm’ sort: We conjecture that those recommendations might be based on inappropriate models of the climate system. The example on which this paper puts a major focus is the commonly-used ‘CO$_2$ Decay Model’, which assumes that a fixed fraction of the atmospheric CO$_2$ above the preindustrial value ‘decays’ each year. This resembles capital depreciation in models of economic growth. While it allows to concentrate on the economic issues of climate change, it neglects the longevity of CO$_2$ in the atmosphere, among other problems, which renders it difficult to align with climate science.

The CO$_2$ Decay Model was proposed by William Nordhaus in Nordhaus (1980), which was among the first papers that put climate change into an intertemporal cost-benefit framework. It maps the complex climate system into simple formulas stating that the

“[t]he purpose of this paper is not realism – rather it is to simplify the system so that it can be readily and intuitively understood, and so that we can see which variables should be focused on in research and policy.”

This point was extended in Nordhaus and Boyer (2000):

“... it is desirable to have parsimonious representations, to have models that are structural (in the sense of reflecting solid scientific or economic underpinning), and to rely on models whose essential findings are robust to changes in the specifications.”

In general this is true about all economics models. They should be parsimonious and, at the same time, accurate. Finding the right balance between these, often contradictory, conditions is up to the economist’s assessment. Because the social systems to be represented by simple economic models are often very complex – and not governed by some law unambiguously agreed-upon, it is seldom straightforward to assess the accuracy of economic models in general. Instead, such accuracy often lies in the eye of the beholder.\(^5\)

Earth system models seem to fare better. Many concept have a solid foundation in

\(^5\) Consider the divide of macroeconomic models into neoclassical and Keynesian economics, for instance.
physical and chemical laws, about which there is little disagreement.\(^6\) It is straightforward to test whether an economist’s model reflects these well-known properties. If it does not, there should be a discussion about the implications of such mismatch, within the specific context of the model. Otherwise, the outcomes and implications of such a model have to be examined with prudence.

Along these lines, we will show that the CO\(_2\) Decay Model is neither structural nor robust. For instance, a paper’s main concern could be the near-term future where emissions increase exponentially. Alternatively, a paper’s main concern could be how high permissible emissions are that would stabilize the atmospheric CO\(_2\) concentration at a certain level. As we will see, in the former case, the CO\(_2\) Decay Model might be a reasonable representation, whereas in the latter it is not.

Shortly after Nordhaus published his widely-cited book Nordhaus (1994), on which he promoted the use of the CO\(_2\) Decay Model, some of the criticism put forward here has already been made (Price, 1995; Schultz and Kasting, 1997). Along with the criticism, alternatives were proposed, mostly dealing with a more accurate depiction of the longevity of CO\(_2\) in the atmosphere. However, these alternatives lack the simplicity of the CO\(_2\) Decay Model. Judged by the CO\(_2\) Decay Model’s still widespread use, the advantages of its simplicity often seem to have outweighed the possible lack of accuracy.

As the earth system is very complex, it is not straightforward to judge whether an economist’s model is an appropriate representation in the context considered in the paper. We thus provide in this paper a tailor-made overview of those properties of the earth system that are relevant for economic modeling as a useful tool for further research.

To understand why it is important to represent both the economics and the earth system accurately one can imagine the intertemporal cost-benefit analysis of climate change like a seesaw. The costs of emission reduction are on the one side, and the benefits of prevented damages on the other. The economic optimum would correspond to the seesaw being in the balance, for which both the ‘weight’ and the ‘distance’ to the junction point matter. The distance mainly depends on how far away in the future damages or costs occur, and by which rate they are discounted. The further away in the future and the higher the discount rate, the smaller the distance. As to the weights, they are first an economic question. Emission reduction costs are in the realm of economics per definition, and the valuation and aggregation of impacts for a

\(^6\) This is true notwithstanding the many aspects of the earth system that are little understood.
certain level of climate change are deeply rooted in economics as well. Climate science affects the weights as well. It investigates to what extent the climate changes for a given concentration of CO$_2$ in the atmosphere. In addition, climate science deals with the issue over which time horizon the impacts of the emitted CO$_2$ affect the earth system. As these impact will stay for a very long time, this finding implies that the damage weights have to be put on the seesaw very often. Using the seesaw metaphor, it is easy to see that the longevity issue is especially relevant if the discount rate is low. Otherwise, even if damages are put on the seesaw very often, this would be at close distance to the junction point, and thus would have little effect on the overall balance.

Finally, let us follow Nordhaus (2008)’s invitation

“Apologies are extended to those who feel that their discipline has been grossly oversimplified. Along with the apologies go invitations to help improve our understanding by providing better parsimonious representations of the crucial geophysical or economic processes. In modeling, small is genuinely beautiful”;

and suggest to use a relatively new concept, the ‘climate carbon response’ (Matthews et al., 2009). It states that for every unit of CO$_2$ that is emitted into the atmosphere, the earth temperature irreversibly increases by a certain amount. This linear relationship holds for a wide range of emission scenarios. The climate carbon response represents the complex climate system rather well. In addition, it is remarkably simple and thus suited for economic modeling. It implies that sooner or later net emissions of carbon into the atmosphere should decrease to zero. The increase of cumulative emissions over time increases the marginal damages up to the point where they equal the marginal abatement costs of zero emissions.

### 6.1.2 Organization of the Paper

The paper continues as follows: In Section 6.2, we describe those parts of the earth system that are relevant for economic modeling and the concept of the climate carbon response. In Section 6.3, we depict several earth system models used in the economic literature and assess their potentials and limitations. In Sections 6.4, we describe three common pitfalls that might occur if erroneous earth system models are used in an economists’ model without testing their appropriateness. In Section 6.5, we illustrate the points discussed in the previous sections with numerical examples and show that the concept of climate carbon response can be easily applied in economic modeling.
6.2 The Earth System

In this section, we convey an overview of that properties of the earth system that are important to assess the accuracy of economics models of climate change. Explicitly or implicitly these models comprise two components of the earth system. The first is a carbon cycle model, which investigates the fate of anthropogenic CO\textsubscript{2} after it has been emitted into the atmosphere. The second is a warming model, which investigates the temperature increases due to the elevated concentration of CO\textsubscript{2} in the atmosphere.

We will summarize the current knowledge about these parts of the earth system, giving special attention to uncertainties. Whether these properties should be included in an economic model or not is at the model-makers’ discretion, of course. Yet, to make this choice he has to be aware of those properties. We simplify as far as this is possible without compromising correctness. For more details on the data and on the physical, biological and chemical processes involved, the interested reader is referred to the literature cited.

Even though CO\textsubscript{2} is not the only GHG, we stay with the literature and concentrate on CO\textsubscript{2} for three reasons. First, other GHGs such as Methane and Nitrous Oxide, currently only contribute about half as much to the warming effect as CO\textsubscript{2} and coincidentally their effect might be balanced by the cooling effect of aerosols (Forster et al., 2007). Second, other GHGs’ relative role is expected to decline in the future, as their influence will remain constant, whereas CO\textsubscript{2}’s concentration is expected to increase (National Research Council, 2011). Third, other GHGs and aerosols are, by far, not as long-lived as CO\textsubscript{2}.

6.2.1 The Natural Carbon Cycle

The main reservoirs of CO\textsubscript{2} – or more precisely the carbon embodied in it – are, in decreasing size, the solid earth, the deep ocean (≈ 37,100 GtC), the surface ocean (≈ 900 GtC), the biosphere (≈ 2,300 GtC) and finally, the atmosphere (597 GtC) (Denman et al., 2007). Due to physical, chemical or biological processes, carbon interchanges between these reservoirs. This ‘natural’ carbon cycle is very complex and works on different timescales. The solid earth – mainly rocks – interacts so slowly – on timescales of 10,000 years and more – that it does not play a role on human timescales, although it is by far the biggest reservoir. It may thus be neglected. The ocean is generally

\footnote{Halocarbons are an exception, but they play a minor role.}
divided into a well-mixed surface zone (50-100m), with the deep ocean beneath it. The deep ocean contains vast amounts of carbon, but is rather inert. An exchange of carbon with other reservoirs occurs through mixing with the surface ocean, with a turnover time of several centuries. As we will see, the deep ocean plays a decisive role in the carbon cycle in the long run. The surface ocean, the biosphere and the atmosphere are in direct contact with each other and exchange carbon at timescales of years. We will henceforth call them the ‘fast-mixing reservoirs’. The surface ocean exchanges carbon with the atmosphere whenever there is a concentration gradient.\textsuperscript{8} The biosphere exchanges carbon with the atmosphere through respiration/decay and photosynthesis.

The atmosphere is at the center of the natural carbon cycle and it has the smallest reservoir size. As a consequence, small changes in the natural carbon cycle have the potential to affect the atmosphere’s CO\textsubscript{2} concentration disproportionately.\textsuperscript{9} With the climate changing, it is not self-evident that the atmosphere’s carbon concentration remains as stable as it was since the last ice age, where carbon sinks and carbon sources have been roughly in balance on timescales larger than a decade.

\textbf{6.2.2 The Human Perturbation of the Carbon Cycle}

Through the burning of fossil fuels and land-use change, humans emit large amounts of CO\textsubscript{2} into the atmosphere and disturb the natural carbon cycle. One may imagine this anthropogenic carbon as a perturbation superimposed on the natural carbon cycle, which enters the atmosphere from the solid earth (fossil fuel burning) and the biosphere (land-use change). Note that land-use change corresponds to a rearrangement within the fast-mixing reservoirs. Fossil fuel emissions, on the other hand, enter the fast-mixing reservoirs from outside, and remain for a very long time (Archer et al., 2009). Once anthropogenic carbon enters the atmosphere, some parts are quickly taken up by the other fast-mixing reservoirs, which can thus be called the short-term sinks. When the short-term sinks are saturated, the deep ocean starts to play a major role, which can thus be called the long-term sink. As we will see, a basic understanding of this distinction between short-term and long-term sinks and the way they take place, is a prerequisite to test the accuracy of economic models of the climate system.

\textsuperscript{8} Such gradients are mainly due to biological and physical processes in the ocean (Sarmiento and Gruber, 2006). Since humans started emitting CO\textsubscript{2} into the atmosphere, this concentration increase in the atmosphere led to a net flow of carbon from the atmosphere to the ocean.

\textsuperscript{9} During the ice ages, for example, the atmosphere contained about one-third less carbon, triggered by small changes of the earth’s orbit around the sun.
To illustrate this complex response pattern of the carbon cycle to human perturbations, let us, step by step, follow the response to an instantaneous emission of large amounts of carbon into the atmosphere. Albeit unrealistic, this is a well-known exercise. It is an intuitive and standardized way to depict the fraction of carbon that stays in the atmosphere after it has been emitted as a function of time. It is called the ‘unit decay function’. This illustration goes a long way in understanding the response of the carbon cycle to more realistic scenarios of anthropogenic emissions.

**Fast-Mixing Reservoirs – The Short-Term Sinks**

A few years after the anthropogenic CO$_2$ perturbation has entered the atmosphere, approximately half of it remains. The other half has mixed into the surface ocean and has been taken up by the biosphere, both of which currently playing an equally important role in this process.$^{10}$ The biosphere takes up parts of the anthropogenic carbon, as the increased level of atmospheric CO$_2$ improves the efficiency of photosynthesis and thus increases the level of carbon stored in the vegetation. This is the so-called ‘CO$_2$ fertilization’ effect (Norby et al., 2005). In addition, vegetation might grow again in areas where there had been forest clearing in the past. The strength of the biosphere sink cannot be constrained directly, it is rather determined as the remainder of the changes in the other, better constrained, reservoirs.

The taking-up by the surface ocean can, to some extent, be explained through simple physics. CO$_2$ is readily soluble in water, and as the surface ocean and the atmosphere are in direct contact to each other, the anthropogenic carbon perturbation in the atmosphere enters the surface ocean, until the two are in equilibrium. However, this picture becomes more complex if the perturbation is big enough to alter the chemistry of the ocean. Water turns acid when it takes up carbon, and the higher the acidity, the lower the ocean’s chemical capacity to take up CO$_2$.\textsuperscript{11} Thus, the equilibrium fraction of anthropogenic carbon in the surface ocean decreases when cumulative emissions increase. This decrease of the ocean’s chemical capacity is known for a long time (Bolin and Eriksson, 1958) and can be quantified quite precisely.

\textsuperscript{10} This short-term partitioning is reasonably well constrained (Broecker et al., 1979; Canadell et al., 2007; Sabine et al., 2004; Raupach et al., 2009).

\textsuperscript{11} Currently, out of every 20 CO$_2$ molecules that enter the ocean, 19 are transformed chemically, and the atmosphere is in physical equilibrium with the remaining one molecule. As a result, the chemical capacity of the ocean is much larger than implied by pure physical solubility alone. The side effect of this chemical transformation is that the oceans become more acid. This impairs the chemical transformation, such that more CO$_2$ remains unaltered.
The fraction of atmospheric CO$_2$ that has not been taken up by short-term sinks causes the climate to change. This, in turn, alters the short-term sink strengths through many different channels. Current predictions see an overall decrease of the short-term sink strength with climate change. That is, there is a carbon-climate feedback which seems to be positive. There is, however, little agreement on the exact magnitude of this feedback.$^{12}$

The major effects of climate change on the surface ocean’s sink strength are rather well-known. Climate change warms the ocean, lowering its physical solubility of CO$_2$. Thus, the sink strength of the surface ocean not only decreases chemically, but also physically. In addition, warming primarily affects the uppermost layer of the surface ocean and thus increases the stability of the surface ocean’s vertical structure. This stratification hinders mixing and thus reduces the surface ocean’s sink strength. The main reason why it is so difficult to quantify the carbon-climate feedback lies in the unpredictability of the biosphere’s sink.$^{13}$ To quantify it, predictions of local climate change as well as of the biosphere’s response to this altered conditions are needed. This involves many, as yet, little-understood processes.

The possible impacts of ‘known unknowns’, such as the potential for sudden releases of large amount of methane stored in the melting permafrost as well as ‘unknown unknowns’ give reason for further concern. A reason why the existence of unknown unknowns has been suggested is the discovery that there was a major warming event 56 million years ago, the so-called PETM (Cohen et al., 2007). It was triggered by a massive increase of GHGs of unknown source, possibly methane.

Following Friedlingstein et al. (2006) and Roe (2009), let us define the carbon-climate feedbacks in a more formal way.$^{14}$ To be able to quantify feedbacks, a reference system has to be defined. In this case, it is the increase in atmospheric CO$_2$ concentration by 2100, as compared to a scenario where carbon-climate feedbacks are neglected (Friedlingstein et al., 2006).

The increase in atmospheric

$^{12}$ Predictions range from less than 50 ppm to more than 200 ppm for the additional increase of the atmospheric CO$_2$ concentration by 2100, as compared to a scenario where carbon-climate feedbacks are neglected (Friedlingstein et al., 2006).

$^{13}$ Temperature or humidity changes might constrain the efficiency of photosynthesis to such an extent that a further increase in CO$_2$ would cease to fertilize the vegetation ever more. Similarly, the CO$_2$ fertilization might be hampered by the availability of nitrogen. (The availability of nitrogen does not change with climate change. Hence this is no feedback process. Notwithstanding, it impairs CO$_2$ fertilization.) Vegetation patterns will change, thus altering the natural carbon cycle and affecting the sink strength in ways that are little known yet. Global warming increases the decay processes in soils which are very temperature-sensitive, such that overall, the biosphere might well turn into a net source of CO$_2$ (Davidson and Janssens, 2006).

$^{14}$ The concept of feedbacks is also discussed in Section 6.2.3.

$^{15}$ ‘U’ stands for uncoupled.
CO₂ causes the climate to warm, which, in turn, causes the atmospheric CO₂ to change to $S_C^{\text{C}}$. This is the carbon-climate feedback. It is positive if $S_C > S_U$ and negative otherwise. The coupled and the uncoupled atmospheric CO₂ are related according to

$$S_C = 1/(1 - g)S_U.$$  

The feedback strength $-\infty < g < 1$ can be approximated as follows: Without feedbacks, the changes of CO₂ in the reservoirs can be described as

$$\Delta C^{\text{U}}_{\text{Bio}} = \beta_{\text{Bio}}S_U,$$

$$\Delta C^{\text{U}}_{\text{SO}} = \beta_{\text{SO}}S_U,$$

where $\Delta C_{\text{Bio}}$ and $\Delta C_{\text{SO}}$ are the increase of the carbon inventory in the biosphere and the surface ocean, respectively. The parameters $\beta_{\text{Bio}}$ and $\beta_{\text{SO}}$ describe the partitioning between the reservoirs. Including feedbacks, this reads

$$\Delta C^{\text{C}}_{\text{Bio}} = \beta_{\text{Bio}}S_C - \kappa_{\text{Bio}}\Delta T,$$

$$\Delta C^{\text{C}}_{\text{SO}} = \beta_{\text{SO}}S_C - \kappa_{\text{SO}}\Delta T,$$

where the $\kappa$s are parameters that describe the feedbacks. Inspection of the latter equation shows that the feedback is positive if $\kappa > 0$. The increase in temperature is

$$\Delta T = \alpha S_C,$$

where $\alpha$ is a parameter that describes how responsive the planet is to an increase in atmospheric CO₂. Combining these equations yields

$$g = \alpha \frac{\kappa_{\text{Bio}} + \kappa_{\text{SO}}}{1 + \beta_{\text{Bio}} + \beta_{\text{SO}}}.$$  

It shows that the feedback strength is the higher, the larger the feedback parameters are. Furthermore, the feedback strength increases in the planet’s responsivity to an increase in atmospheric CO₂. This nicely illustrates the reinforcing effects of a – say – positive feedback. The higher the temperature increase triggered by an initial increase in atmospheric CO₂, the stronger the resulting additional increase of atmospheric CO₂, which then triggers an even higher temperature increase. Note that this process converges as long as $g < 1$.  

---

16 “C” stands for coupled.
The Deep Ocean – The Long-Term Sink

After some years, within the fast-mixing reservoirs the separation of the anthropogenic CO$_2$ perturbation that entered the atmosphere is completed. Atmospheric carbon would henceforth remain constant, if it were not for the deep ocean. It mixes with the surface ocean and takes up parts of the anthropogenic perturbation stored in the surface ocean. This, in turn, enables the surface ocean to take up some of the remaining anthropogenic carbon from the atmosphere. After several centuries, this process reaches its final equilibrium. As the deep ocean’s reservoir size is by far the biggest, one might think that the deep ocean has taken up almost all of the anthropogenic carbon perturbation by then. The fraction of the perturbation that each reservoirs holds in the end, so the reasoning might be, has to correspond to the relative sizes of the reservoirs to begin with. However, this intuition is wrong. Instead, a substantial fraction of the emitted CO$_2$ will remain in the atmosphere forever at the timescales of 1000 years we consider here. How high this permanent fraction is, depends on the size of the perturbation. The reason is the same as explained above for the surface ocean: The deep oceans’ chemical capacity decreases, as it acidifies.\footnote{At timescales of thousands of years, parts of the ocean floor start to dissolve, lowering the acidity of the deep ocean and thus increase the ocean’s chemical capacity again. As a consequence, after several tens of thousands of years, the deep ocean takes up a substantial fraction of the initial perturbation. This extremely slow process is the reason why the deep ocean currently holds vast amounts of carbon, but does not take up the corresponding fraction of the perturbation at the timescales we consider.}

Because of the long time it takes to reach the final equilibrium, we are not only interested in the final fraction of the perturbation that remains in the atmosphere – as for the short-term sink –, but we should also address the question how fast the deep ocean takes up carbon in the meanwhile. This depends on the deep ocean’s turn-over time, which is reasonably well constrained.

Numerical Example

Figure 6.1 depicts the unit decay functions for several scenarios stemming from three different papers, Maier-Reimer and Hasselmann (1987) [MRH], Sarmiento et al. (1992) [SAR], Caldeira and Kasting (1993) [CAK]. MRH and SAR simulate the response to an instantaneous doubling (2\times PID) and quadrupling (4\times PID) of preindustrial values. Note that the remaining airborne fraction is higher in the 4\times PID scenarios of MRH and SAR. CAK’s method is slightly different. They simulate the remaining fraction of a small pulse of emissions on top of certain baseline scenarios. Baselines are, first, that
the stock remains at preindustrial value (280 ppm), second, that the stock increases to 750 and is stabilized there (750 ppm) and, third, 5000 GtC – assumed to be the total stock of fossil fuels – are being emitted in total, using a logistic growth equation such that CO$_2$ stock peaks at close to 2000 ppm and decreases thereafter (5000 GtC). Especially in the 5000 GtC scenario, the remaining fraction is very high, as due to high baseline emissions, sinks are saturated. Note that these models merely simulate the ocean sinks. None of them includes the biosphere’s sink or carbon-climate feedbacks.

![Unit Decay Function](image)

**Figure 6.1:** Unit decay functions of several scenarios in Maier-Reimer and Hasselmann (1987) [MRH], Sarmiento et al. (1992) [SAR], Caldeira and Kasting (1993) [CAK].

### 6.2.3 Global Warming

**An Overview**

GHGs alone are not harmful.\(^\text{18}\) It is the climate change they cause that raises concerns. Therefore, the second issue to be discussed here is how a certain time trajectory of GHGs in the atmosphere translates into a time trajectory of climate change. It is customary to divide the response of the earth system to a change in GHGs conceptually into two steps. First, GHGs cause a radiative forcing almost instantaneously. Second,

\(^{18}\) In this section, we use GHGs instead of CO$_2$ as the ensuing explanations hold true for all GHGs.
due to this radiative forcing, the earth’s temperature slowly changes. Before explaining each step in turn, it might be useful to illustrate the concepts with an example.

Suppose you have a huge pot of water that has room temperature, put it on a small gas cooker and turn on the cooker. It will take some time until the water in the pot has warmed to such a level that the energy lost by the pot to the room equals the energy input of the gas cooker. This is analogous to the earth system. The radiative forcing of the GHGs corresponds to the gas cooker whose flame reacts instantaneously to a level change, and the slow response of the earth system corresponds to the huge water pot.

Since the last ice-age, there was an approximate balance between the input of the short-wave radiation from the sun and the long-wave radiation the earth system emitted into space, such that the temperature remained approximately constant. The increase of GHGs disturbed this balance as the GHGs impair the earth’s ability to emit long-wave radiation into space, whereas they do not affect the short-wave radiation received from the sun. Radiative forcing, $\Delta RF$, is a measure to express the extent of this radiative disturbance caused by GHGs (the level change of the gas cooker). This disturbance occurs almost instantaneously and stays as long as GHG levels are elevated. There is little controversy on the magnitude of this radiative forcing caused by an increase of GHGs.\footnote{Assume that the pot is big enough such that the water never reaches the boiling point.}

As a reaction to this disturbance, the planet’s radiative budget is in a temporary planetary imbalance $\Delta PI(t)$. Following Roe (2009), this can be expressed as

$$\Delta PI(t) = \Delta RF + \frac{1}{\lambda} \Delta T(t),$$

(6.1)

where $\lambda$ is the climate sensitivity parameter and $\Delta T(t)$ is the temperature increase. Define $t^{eq}$, as the time when the planet has equilibrated in response to the radiative forcing and the planetary imbalance is zero, $\Delta PI(t^{eq}) = 0$. Thus,

$$\Delta T(t^{eq}) = \lambda \Delta RF.$$

The climate sensitivity parameter plays a crucial role in climate science as it is a measure by how much the planet warms for a given radiative forcing. Despite intensive research on this subject, the climate sensitivity parameter’s value is still quite uncertain.\footnote{The radiative forcing of aerosols, on the other hand, is highly uncertain. They tend to cool the planet as they reflect incoming radiation from the sun, thus counteracting the warming of the GHGs. This phenomena is called global dimming.}
(Knutti and Hegerl, 2008), as there are many feedbacks in the climate system that remain incompletely understood (Bony et al., 2006). Furthermore, the climate sensitivity parameter is a nonlinear function of the feedback, such that relatively small uncertainties in the feedback strength have the potential to cause much larger uncertainties in the climate sensitivity parameter (Roe and Baker, 2007).

To see this, let us define the climate feedbacks in a more formal way. The planetary imbalance caused by a positive radiative forcing causes the planet to warm. To be able to quantify the feedback strengths, as mentioned above, a reference system is needed, on which the feedbacks act upon. In this case, it is common to use the fact that outgoing long-wave radiation increases to the fourth power in temperature. Due to this ‘Plank Response’, an increase in the temperature of the earth system restores the planetary balance. The Planck response is used as a reference system, as it is the dominating effect and it is estimated to yield $\lambda_P = 0.31 \, \text{°C}^2 \text{W}^{-1}$. Suppose radiative forcing corresponds to a doubling of preindustrial CO$_2$ levels, $\Delta RF(2\times) \approx 4 \, \text{W/m}^2$. The equilibrium temperature due to the Plank response, without any feedbacks, would thus be $\sim 1.2 \, \text{°C}$.

The earth system, however, has feedbacks galore. As the earth warms because of the Plank Response other climate variables $x$ are affected which, in turn, cause additional radiative forcing. Define this additional radiative forcing per temperature increase due to a certain climate variable as $c_x = \frac{\partial P_I}{\partial x} \frac{\partial x}{\partial T}$. As an example take the Water Vapor Feedback. It occurs because warming increases the amount of water vapor in the atmosphere, which is a potent GHG and thus increases the radiative forcing beyond the original forcing. This feedback is positive as, per definition, positive feedbacks increases the temperature response as compared to the Planck Response, whereas negative ones dampen it. At short to medium timescales the most important climate feedbacks are the positive Water Vapor Feedback, the probably positive Cloud Feedback, the positive Surface Albedo Feedback, and the probably negative Lapse Rate Feedback.$^{21}$

The magnitude of these feedbacks is highly uncertain – especially for the Cloud Feedback. For $\Delta RF(2\times)$, the equilibrium temperature response increases from $\sim 1.2 \, \text{°C}$ for the pure Plank Response to a best-guess range of $\sim 2 - 4.5 \, \text{°C}$, with no upper

---

$^{21}$ The Cloud Feedback occurs as climate change alters cloud patterns, which both influence incoming solar radiation and outgoing long-wave radiation. The Surface Albedo Feedback occurs as climate change melts snow and ice, which reflects incoming solar radiation back to space. Snow and ice are replaced by land or the ocean, that absorb incoming solar radiation and thus warm. The Lapse Rate Feedback occurs as global warming tends to be stronger in the upper parts of the atmosphere. As a consequence, the temperature decrease with height (the lapse rate) decrease, which, in turn, weakens the greenhouse effect. For a more detailed discussion, see Bony et al. (2006).
bound agreed upon (Knutti and Hegerl, 2008). This temperature increase is commonly referred to as the ‘climate sensitivity’. As early as 1979, the climate sensitivity has been estimated to be in the range of 1.5 – 4.5 °C, with a best guess of 3 °C (Charney et al., 1979)

To illustrate why this range remained so large, we define the climate sensitivity parameter in terms of the Plank Response and the dimensionless feedback strengths, \( f_x = \lambda_P c_x \)

\[
\lambda := \frac{\lambda_P}{1 - \sum_x f_x}.
\]  
(6.2)

This shows that the climate sensitivity parameter is highly non-linear in the overall feedback strength, \( \sum_x f_x \). If the overall feedback strength is normally-distributed around a moderate strength of, say, \( \sum_x f_x = 0.65 \), this translates into a distribution of \( \lambda \) that is highly skewed towards higher values, without upper bound (Roe and Baker, 2007).

So far, we have discussed the equilibrium response. However, it takes a long time to reach this equilibrium\(^{23} \) and the higher the overall feedback strength the longer it takes (Roe and Bauman, 2012). To illustrate this, it is useful to assume that the planetary imbalance causes the planet to warm according to \( \Delta P I(t) = C \frac{d \Delta T(t)}{dt} \), where \( C \) is the thermal inertia of the earth system. This allows us to rewrite Equation (6.1) and derive Roe (2009)’s illustration of the warming behavior of the earth system

\[
\frac{d \Delta T(t)}{dt} + \frac{\Delta T(t)}{\tau} = \frac{\Delta RF}{C},
\]  
(6.3)

where the response time of the earth system, \( \tau \), is

\[
\tau = \frac{C \lambda_P}{1 - \sum_x f_x}.
\]  
(6.4)

It is easy to see that the response time – and thus also \( t^{eq} \) – increases in the magnitude of the Plank Response, increase in the terminal inertia, as well as in the overall feedback strength.

Several further points have to be mentioned in this context. First, several highly speculative but potential strong feedbacks are usually not included in the estimates of the climate sensitivity parameter and could start playing a role in the very long run.\(^{24} \)

\(^{22} \)To exclude non-physical results, it has to hold that \( \sum_x f_x < 1. \)

\(^{23} \)Hope (2007), based on Cubasch et al. (2001), estimates the half-life of the global response to an increase in GHGs in a range of 25-75 years, for instance.

\(^{24} \)The ice-sheets react on timescales of millennia, for instance.
note these feedback strength as \( f_g \) and define \( \hat{\lambda} := \lambda_0/(1 - \sum_x f_x - \sum_y f_y) \) as the ‘earth system sensitivity’. The equilibrium time, \( \hat{t}^{eq} > t^{eq} \) is the timer where \( \Delta T(\hat{t}^{eq}) \approx \hat{\lambda} \Delta RF \).

If the radiative forcing were to be maintained over several millenia, the earth system sensitivity would be the relevant measure and could potentially be twice the climate sensitivity (Hansen et al., 2008).

Second, the linear relationships of Equation (6.3) are useful for illustrative purposes, especially if the radiative forcing is not very high and not maintained over a long time period. It might be grossly misleading otherwise, as the strength of many feedbacks depends on the state of the climate system. If the disturbance is high, the earth might be a highly non-linear system.

Third, it is common practice to use the temperature increase as a proxy for climate change, but this glosses over several important issues. It abstracts from local changes and other climate variables such as precipitation and extreme events. It omits the fact that the sea-level would be rising, long after temperatures would have been stabilized (Meehl et al., 2005). It neglects the possibly of tipping points, defined as a temperature threshold, which, if crossed long enough, causes parts of the earth system to switch from one equilibrium state to another (Lenton et al., 2008).

**Beyond Climate Sensitivity**

The idea of stabilizing the level of GHG, and thus of ‘stabilizing’ the climate, was brought to prominence in the UNFCCC declaration in 1992 which demanded

“...stabilization of greenhouse gas concentrations in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system.”

Climate sensitivity is a way of estimating how high the interference with the climate system would be if GHG concentrations were to be stabilized. In addition, it is a useful measure to compare the temperature responses of different full-scale climate models.

For practical purposes, climate sensitivity is of limited use for two reasons. The one is that the requirement to maintain GHG long enough to reach equilibrium seems to be an unrealistic and not very meaningful goal. Humans might react to the climate change they experience instead of following stubbornly a stabilization goal set a long time ago. As Matthews et al. (2012) explain, this “would imply continued emissions

---

25 This is because the deep ocean warms slowly and it increases its volume while warming. In addition, melting of ice-sheets is a slow process.

at a changing level consistent with the level of natural sinks that evolve over time in a manner difficult to quantify.” The other reason is that the stabilization of CO$_2$ does not correspond to the stabilization of temperature as long as the temperature increase has not reached its equilibrium value. To the extent that we care about the non-equilibrium temperature response as well, the use of climate sensitivity will be misleading.

A more useful concept for this purpose is the transient climate response. It is the temperature increase at the point in time where the CO$_2$ concentration has reached twice the preindustrial value, with the concentration increasing at a rate of 1% starting from the preindustrial level (Frame et al., 2006). There is much more consensus about its value, as it is related more closely to observable warming and the ill-constraint long-term feedbacks are of minor importance (Meinshausen et al., 2009).

Consider our water-pot example once again. Starting at room temperature, you turn on the gas stove and turn the nob slowly to a certain level and leave it there. Transient climate response corresponds to the water’s temperature at the time you reached this level. Climate sensitivity corresponds to the water’s temperature at a much later moment, when equilibrium is reached.

### 6.2.4 Carbon Climate Response

More recently, several other concepts emerged that describe the impact of global warming, which are suited for our purpose (Allen et al., 2009; Matthews et al., 2009). They start from the observation that there is less uncertainty if one directly links the total amount of CO$_2$ emitted up to date to the current temperature increase – without the detour via the carbon cycle.

Let us focus on Matthews et al. (2009)’s proposal. They suggest that there is a linear relationship between irreversible global warming – at the timescales of centuries –, and the total amount of CO$_2$ emitted up to date. They define this relationship as the carbon-climate response (CCR). A surprising corollary of this concept is that if we were to stop all emissions tomorrow, temperature would remain at its current level. This has been confirmed by several climate models of intermediate$^{27}$ and full complexity.$^{28}$ Intuitively, once emissions stop, the CO$_2$ stock in the atmosphere decreases, which decreases radiative forcing. At the same time, the warming has not yet reached equilibrium with the current stock of CO$_2$, such that there is still some warming potential

$^{27}$ See Plattner et al. (2008); Solomon et al. (2009).
$^{28}$ See Lowe et al. (2009).
in the ‘pipeline’. These two effects cancel each other. Matthews et al. (2009) estimate that each cumulative emissions of 1000 GtC, warming is $1.0 - 2.1 \degree C$, with a best guess of $1.5 \degree C$ or $1.6 \degree C$ for observations and climate models, respectively. Note that this uncertainty range is because it is unclear how high the warming will turn out, but does not question the linear relations itself. This uncertainty range is not negligible. However, it supplants the much larger uncertainty related to the carbon cycle and the climate sensitivity, as it shortcuts the connection between emissions and climate change.

The carbon-climate response is a remarkably robust and easy-to-use concept. Thus, it is suitable for both economic modeling and communication to laymen. Translated into an economic framework it reads

$$\Delta T_t = CCR \times \sum_{i=0}^{t} E_i,$$

where $E_i$ are global emissions of CO$_2$ in period $i$. Put differently, the temperature increase in period $t + j$ caused by emissions in period $t$ for all $j \in [0, t^{CCR}]$ is

$$\frac{d \Delta T_{t+j}}{d E_t} = CCR,$$

where $t^{CCR} > 1000$ years is the time of validity of the CCR concept.

---

29 A more detailed sketch of the validity of the CCR runs along the following lines. Suppose there is instantaneous doubling of the stock of CO$_2$ in the atmosphere, caused by emitting the amount $C^{2 \times}$. First, the fraction of the totally-emitted CO$_2$ that remains in the atmosphere with time, $C_{Atm}^{t}(t)$, decreases in time according to a pattern that, once the fast-mixing reservoirs are in equilibrium, will be largely determined by the deep ocean’s turnover time. Second, the increase of the warming per unit of CO$_2$ that remains in the atmosphere, $\Delta T_{Atm}^{t}(t)$, is determined by the turn-over time of the ocean as well, as its heat capacity is much larger than the one of the atmosphere. Thus, as both the increase of $\Delta T_{Atm}^{t}(t)$ and the decrease of $C_{Atm}^{t}(t)$ are governed by the same timescale, $\Delta T_{Atm}^{t}(t)$ remains approximately constant with time. This is also true for more realistic emission scenarios, as warming seems to be insensitive to the emission pathway that led to the present cumulative emissions (Allen et al., 2009; Zickfeld et al., 2009).

30 Up to a total of $2000$ GtC.
6.3 Earth System Models Used in the Economic Literature

Having discussed the earth system, let us now address how the carbon cycle and global warming are modeled in the intertemporal cost-benefit analysis of climate change. We focus on simple models which can be solved analytically and on slightly more intricate ones that solve numerically with little effort. For those models, it is crucial that the representation of the earth system is as parsimonious as possible. On the other hand, this representation should reflect the properties of the earth system reasonably well. To strike the balance between these often-contradictory goals is an art in itself. As in all areas of economic inquiry, this requires that the modeler has a solid understanding of the system he wants to represent.

As it seems from an analysis of some examples, many models err on the side of simplicity. Acemoglu et al. (2009), for instance, state:

“While other papers in the environmental literature typically use more detailed descriptions of environmental dynamics, here we use a “reduced-form” approach and concentrate instead on identifying the new economic forces.”

We conjecture that economic modelers use the earth system as too flexible a tool, so that this tool is adapted to the results – and not the other way round –, at the expense of accuracy.

6.3.1 Carbon Cycle Models

Let us now examine several ways to depict the carbon cycle and find out under which circumstances they are appropriate and under which they lead to erroneous implications.

CO₂ Decay Model

As explained in the introduction, the CO₂ Decay Model was first prosed in Nordhaus (1980). It is a very simple model and its use is widespread. In fact, some papers use it even without any discussion or reference, very much like the widespread use of the capital accumulation equation in economic growth theory.

31 However, most often Nordhaus (1994) is cited.
According to the CO₂ Decay Model, the stock of CO₂ in the atmosphere, \( S_t \), accumulates according to

\[
S_{t+1} = (1 - \gamma)S_t + \theta E_t, \quad \gamma \in [0, 1],
\]

(6.6)

where \( 0 < \gamma < 1 \) denotes the constant and positive ‘natural decay rate’ of CO₂ in the atmosphere, and \( \theta \) is the ‘marginal atmospheric retention ratio’.\(^{32}\)

Note the close resemblance in terms of structure to the standard model of capital accumulation:

\[
K_{t+1} = (1 - \delta)K_t + sY_t,
\]

where \( K_t \) is the stock of capital in period \( t \), \( \delta \) is the capital depreciation rate, \( s \) is the savings rate and \( Y_t \) is the output. Using the CO₂ Decay Model thus allows economists to use their wonted toolkits, language and reasoning.

The CO₂ Decay Model model comprises two ideas. The one is that for each unit of emissions, the constant fraction \( 1 - \theta \) is immediately taken up by the short-term sinks. The rest remains in the atmosphere for the time being. The other is that the current stock ‘decays’ at the constant rate \( \gamma \), which ought to reflect the working of the deep ocean long-term sink. This approach implies that first, the deep ocean is an infinite sink, second, that in absolute terms, its sink strength is higher if the stock of carbon in the atmosphere is higher and third that the strength of the short-term and long-term sinks are independent of climate change or cumulative emissions.

To calibrate the model, Nordhaus (1994) used a deep ocean turn-over time of 120 years, which results in \( \gamma = \frac{1}{120} = 0.008 \). This value, together with emission data, allowed him to estimate \( \theta = 0.64 \) as a best guess. Calibrated this way, the model fits the historic emissions and the historic atmospheric CO₂ stock quite well. Nevertheless, this does not mean that the CO₂ Decay Model is robust and structural, as defined by Nordhaus and Boyer (2000).

The reason for this quite-good fit to past data is that the ‘airborne fraction’, defined as

\[
AF := \frac{S_{t+1} - S_t}{E_t}
\]

has been remarkably constant. However, for the airborne fraction to be constant, two conditions are crucial: emissions have to increase exponentially and CO₂ sinks have to be linear (Gloor et al., 2010). Both conditions have been approximately fulfilled in the past. Yet, this will change in the future. Emissions cannot increase exponentially forever and, as discussed in Section 6.2, the CO₂ sinks will not remain linear. This constancy is thus rather a coincidence than a rule.

\(^{32}\) For later calculation, we state

\[
\frac{d S_t}{d E_{t-h}} = \theta(1 - \gamma)^h.
\]
Given that the airborne fraction has been rather constant, it is easy to explain that Nordhaus (1994) found a reasonable fit between the CO$_2$ Decay Model and the data. The airborne fraction is closely connected to $\theta$, especially if $\gamma$ and $S_t$ are rather small, such that the term $-\gamma S_t$ has a minor influence. Historically this has been the case.$^{33}$

In addition, some take-up of anthropogenic carbon by the deep ocean has already taken place, which might be reflected by $-\gamma S_t$ to some degree. How correct this representation is, is hard to say. The calibration of the – as we will see – very important decay rate $\gamma$ is opaque in any case. According to Nordhaus (1994), deep-ocean turn-over times of 50 or 200 years yield fits that are as good as the 120 years used, according to the stated standard error of equation. This is no surprise, as this term only exerts a minor influence.

To summarize, as the airborne fraction has been approximately constant rather coincidentally, the $\theta E_t$-term is a reasonably well representation of the past data, given that the $-\gamma S_t$-term has the right direction and played a minor role.

Imagine we have an economist’s model in which future emission keep increasing exponentially and in which the near-term, were the sink pattern of the earth system remains rather unaffected by climate change, exerts a major influence. In such a model the CO$_2$ Decay Model is a reasonable approximation of the carbon cycle. This is essentially the realm of Nordhaus (1994). Because of the high discount rate he uses, the near-term future is overwhelmingly important. It is thus ‘optimal’ that emissions closely follow the exponentially increasing business-as-usual path.

Suppose, on the other hand, that an economist’s model takes the longer-term future into account to a significant extent. Then, when optimal emissions stop growing exponentially and sinks are not linear any more, the CO$_2$ Decay Model’s inadequacy will lead to erroneous results – a fact the modeler should be aware of. The inadequacy has several causes which can be easily explained by the description of the earth system provided in Section 6.2.

First, even if emissions keep increasing exponentially, there are convincing reasons to believe that, due to climate change and the decreasing chemical capacity of the ocean, the airborne fraction will not remain constant in the future. Instead, it might decrease considerably, and some claim that this has already happened (Canadell et al., 2007; $^{33}$Using $\gamma = 0.008$ and $\theta = 0.64$, in the period 2010-2015 and using values of DICE-NR-0124134-beta (which does not use the CO$_2$ Decay Model), it holds that $-\gamma S_t = -0.008 \times 200 \, GtC = -1.6 \, GtC$ and $\theta E_t = 0.64 \times 9.31 \, GtC = 5.95 \, GtC$.

In the period 1970-1975, according to own estimates, it holds that $-\gamma S_t = -0.008 \times 70 \, GtC = -0.56 \, GtC$ and $\theta E_t = 0.64 \times 4.5 \, GtC = 2.88 \, GtC$.}
The biosphere, which has contributed about a half to short-term sink in the past, may get weaker (Friedlingstein et al., 2006), and might even become a source of carbon (Cox et al., 2000). The surface ocean, which has contributed the other half of the short-term sink, remains a sink, but less so as a fraction of emissions.\(^{34}\) As explained in Section 6.2, this is because of the well-known change in the ocean’s chemistry and the warming and stratification of the ocean.

Second, the CO\(_2\) Decay Model implies that the deep ocean sink strength increases linearly in absolute terms with the stock of anthropogenic CO\(_2\) in the atmosphere. This might be true for small amounts of total emissions. However, it is almost certainly inappropriate for larger amounts of total emissions, because of the ocean chemistry and climate change. In addition, for a sufficiently high stock of CO\(_2\) and low current emissions, the CO\(_2\) Decay Model predicts an negative airborne fraction, a prediction never encountered in climate science.

Third, an important implication of the CO\(_2\) Decay Model is that once emissions cease, the stock of anthropogenic CO\(_2\) decays exponentially and eventually return to preindustrial values. In reality, however, a substantial fraction of the carbon in the atmosphere essentially will remain forever (Archer et al., 2009). Using the CO\(_2\) Decay Model, one is mislead into underestimating the persistence of CO\(_2\).\(^{35}\)

### Three-box CO\(_2\) Model

In this model the carbon cycle is depicted with the help of three reservoirs, the deep ocean, \(C^{\text{Deep}}\), the surface ocean/biosphere, \(C^{\text{OB}}\), and the atmosphere, \(S\):

\[
S_t = E_t + \phi_{11} S_{t-1} + \phi_{21} C^{\text{OC}}_{t-1},
\]

\[
C^{\text{OB}}_t = \phi_{12} S_{t-1} + \phi_{32} C^{\text{Deep}}_{t-1} + \phi_{22} C^{\text{OC}}_{t-1},
\]

\[
C^{\text{Deep}}_t = \phi_{33} C^{\text{Deep}}_{t-1} + \phi_{23} C^{\text{OC}}_{t-1},
\]

where the \(\phi\)s are transport coefficients. Compared to the CO\(_2\) Decay Model, this model is more accurate. Calibration is such that the deep ocean’s sink is not infinite, but finite – albeit vast.\(^{36}\) This model is too complicated to be of use in an analytical model, but

---

\(^{34}\) As emissions are expected to increase, the surface ocean sink, in absolute terms, will increase as well. The relative fraction of emissions that are taken up by the surface ocean, however, will decrease.

\(^{35}\) In the short-run, on the other hand, the CO\(_2\) Decay Model leads to slightly higher carbon stocks than estimated by more sophisticated carbon models.

\(^{36}\) DICE-NR-0124134-beta, for instance, is calibrated with

\[
\phi_{11} = 0.912, \phi_{12} = 0.088, \phi_{21} = 0.03833, \phi_{22} = 0.95917, \phi_{23} = 0.0025, \phi_{32} = 0.00034, \phi_{33} = 0.99966.
\]
is useful for numerical models, such as the newest versions of DICE (Nordhaus and Boyer, 2000; Nordhaus, 2008).

Already in Nordhaus (1980, 1994), William Nordhaus uses the Three-box CO$_2$ Model to explain the carbon cycle. Nevertheless, in these papers he uses the simpler CO$_2$ Decay Model for calculations, claiming that there is no big influence on the results with respect to the optimal emission path.

As a response to criticisms of the CO$_2$ Decay Model by Schultz and Kasting (1997), Nordhaus and Boyer (2000) use the more complex Three-box CO$_2$ Model, claiming it provides a more accurate depiction of the long-term behavior of CO$_2$ in the atmosphere. As can be seen in Figure 6.2 this is correct. However, the Three-box CO$_2$ Model misses the fact that the chemical capacity of the ocean decreases when cumulative emissions increase. According to the Three-box CO$_2$ Model, the steady-state shares of carbon in the reservoirs are independent of the cumulative emissions. Even though this is a major point in the argumentation of Schultz and Kasting (1997), this subject is not mentioned in Nordhaus and Boyer (2000).

**Separable CO$_2$ Decay Model**

Using a full-complexity carbon model of the ocean uptake, Maier-Reimer and Hasselman (1987) found that the unit decay function, $G_t$, can be approximated rather adequately as a superposition of a constant and of exponential functions

$$G_t = f_0 + \sum_{j=1}^{L} f_j \exp(-t/\tau),$$  \hspace{1cm} (6.7)

with $\sum_{j=0}^{L} f_j = 1$ and $\tau_j$ being timescales. $f_0$ reflects the fact that a fraction of the emissions stays forever and the exponentials reflect the fact that the sinks operate on different timescales. The Separable CO$_2$ Decay Model is an improvement of both the CO$_2$ Decay Model and the Tree-box CO$_2$ Model with regard to the accuracy of the long run properties of the carbon cycle.

Yet, if fixed values for the $f$’s and $\tau$’s are used, the unit decay function solely depends on time and the chemical capacity constraint of the ocean is not accounted for. This point has been made very clear in Maier-Reimer and Hasselman (1987). They use

\[
C_{t=2010}^{Atm} = 895.9 \text{ GtC} = 384.5 \text{ ppm}, \quad C_{t=2010}^{OB} = 1527 \text{ GtC} = 655.4 \text{ ppm}^{eq},
\]
\[
C_{t=2010}^{Deep} = 10010 \text{ GtC} = 4298 \text{ ppm}^{eq}.
\]
L=4, and calibrated Equation (6.7) for a sudden 25% increase, doubling and quadrupling of the preindustrial carbon stock in the atmosphere. They find that for the 25% increase and the doubling, the best-fit parameters \( \tau \) and \( f \) are rather similar, whereas for the quadrupling, the best-fit parameters differ substantially. In the latter case, the remaining fraction, \( f_0 \), increases, and longer response times gain importance. The reason is that the decrease in the chemical capacity plays an important role when the CO\(_2\) stock quadruples.

To use the Separable CO\(_2\) Decay Model in an economic model, one has to assume that each period’s emission decay according to Equation (6.7) – independent of past or future emissions – by using

\[
S_t = \sum_{i=1}^{t} G_i E_{t-i}, \tag{6.8}
\]

as in Harvey (1989) or Sarmiento et al. (1992), for instance.

**Separable CO\(_2\) Decay Model – Variants**

While the Separable CO\(_2\) Decay Model is more realistic than the CO\(_2\) Decay Model, the former is much less applicable in analytical models. Thus, more tractable versions have been developed. Farzin and Tahvonen (1996), for instance, use L=1 for their analytical model. They divide CO\(_2\), that is, into two stocks; one that stays that remains permanently and one that decays.

Golosov et al. (2011) use two such stocks as well and, instead of Equation (6.7), use\(^{37}\)

\[
\tilde{G}_h = \gamma_L + (1 - \gamma_L)\gamma_0(1 - \gamma_T)^h, \tag{6.9}
\]

where \( \tilde{G}_h \) is that fraction of emissions that remains in the atmosphere after \( h \) periods. \( \gamma_L = 0.2 \) is the fraction that remains forever. Of the remainder, the fraction \( \gamma_0 = 0.393 \) is immediately taken up, and the rest – of the remainder – decays at a constant rate according to \( \gamma_T = 0.0228 \). The values are calibrated such that within 10 years, 50% of initial emissions are taken up, and the half-life of the \( (1 - \gamma_L) \) fraction is 300 years.

\(^{37}\) Using this relationship yields

\[
\frac{d S_t}{d E_{t-h}} = \tilde{G}_h.
\]

This simplicity – and not the superiority of the carbon model per se – might have played a role in the authors’ choice of this particular carbon model.

Numerical Example

Figure 6.2 illustrates the differences of the unit decay functions for various carbon models. It shows that in particular the CO₂ Decay Model does not take the longevity of CO₂ in the atmosphere sufficiently into account.

![Unit Decay Function](image)

**Figure 6.2:** The unit decay function for various carbon models. The right-hand side depicts 1000 years. The left-hand side repeats the first 100 years to allow for a better comparison at short timescales. We use the parameters as stated in the respective sections. Note that at time zero, the CO₂ Decay Model and Golosov et al. (2011)’s model start at a fraction of \( \theta = 0.64 \) and \( \gamma_L + (1 - \gamma_L) \gamma_0 = 0.51 \), respectively.

6.3.2 Warming Models

Let us now discuss how the step from a given trajectory of the GHG stock in the atmosphere to the corresponding warming trajectory is modeled. We will stress the fact that for an increase of the GHG stock, the corresponding radiative forcing is almost instantaneous, whereas the temperature response follows with a considerable time-lag. There seems to be confusion in some parts of the literature about these two concepts, in the sense that instantaneous warming is assumed.

From the GHG Stock to the Radiative Forcing

A common approximation to model radiative forcing of CO₂ is to assume that a doubling of the stock of CO₂ leads to an increase in radiative forcing of \( \sim 4.0 \) W/m². Hence,

\[
RF(t) = 4.0 \text{ W/m}^2 \log_2 \left( \frac{S_t}{S_0} \right),
\]

(6.10)
where, with $S_0 = 280$ ppm, There is not much controversy on this issue, and this approximation is both simple and accurate.

**From the Radiative Forcing to the Temperature Increase**

*Instantaneous Warming Model*

This is the simplest form of a warming model. It uses a linear relationship and assumes that the temperature response immediately obtains the equilibrium value.

$$\Delta T_t = \lambda RF_t,$$

(6.11)

where $\lambda$ is the climate sensitivity parameter as defined in Equation (6.2). The implicit assumption behind this model is that the earth system is in equilibrium with respect to the forcing.

However, it might take centuries to reach the new equilibrium. In addition, the response time increases in $\lambda$ (Roe and Bauman, 2012). Thus, if periods are not extremely long, this model is quite misleading. Not only does it over-predict the temperature increase for rising temperatures, it also underestimates the time needed before temperatures start to decrease when radiative forcing decreases.

*Lagged Warming Model*

Nordhaus (1994) uses a Lagged Warming Model that correctly emphasizes the long time lags involved. It reads

$$\Delta T(t) = T(t - 1) + \frac{1}{R_1} \left\{ RF(t) - \lambda \Delta T(t - 1) - \frac{R_2}{\tau_{12}} (\Delta T(t - 1) - T^O(t - 1)) \right\},$$

(6.12)

$$T^O(t) = T^O(t - 1) + \frac{1}{R_1 \tau_{12}} \left( \Delta T(t - 1) - T^O(t - 1) \right),$$

(6.13)

where $T^O$ is the temperature of the ocean, the R’s are parameters and $\tau_{12}$ is a timescale. Due to its complexity, the Lagged Warming Model is of limited use in simple economic models.

**From the GHG Stock to the Temperature Increase**

Instead of taking the two steps just discussed, it is also quite common to directly determine the temperature increase for a given trajectory of the GHG stock – and omit the concept of radiative forcing. Whenever this is done, it is again important to
account for the lag in the warming response. Yet, this is often neglected and instead the Instantaneous Warming Model is used implicitly. An example is

\[
\Delta T_t = 3 \log_2 \left( \frac{S_t}{S_0} \right),
\]

which Acemoglu et al. (2012), for instance, use, commenting that

“[t]his equation implies that a doubling of atmospheric concentration in CO2 leads to a 3 °C increase in current temperature (see, e.g., IPCC 2007).”

This is a correct description of Equation (6.14), but does not correspond to the IPCC’s findings that the temperature response lags considerably.

**Numerical Examples**

For two emission scenarios, this section compares the CCR concept\(^{38}\) with the earth system model of the newest versions of DICE,\(^{39}\) which consists of the Lagged Warming Model in combination with the 3-Box CO\(_2\) Model.

Figure 6.3 depicts emission scenario 1,\(^{40}\) and the resulting temperature increase for the DICE model and according to the CCR concept. It shows that the two approaches are not too different here. DICE yields a larger warming at early periods, whereas CCR’s warming is slightly higher later on.

In emission scenario 2,\(^{41}\) emissions are much higher and the trajectories of the DICE model and the CCR concept are not in accordance any more. As Figure 6.4 shows, after 25 periods, the temperature increase is much higher for the CCR than for the DICE model.

### 6.4 Common Fallacies

Let us now address problematic implications of the CO\(_2\) Decay Model and the Instantaneous Warming Model, as they have a prominent role in the literature.

---

\(^{38}\) We use Matthews et al. (2012)’s best guess of 1.5 °C per 1000 GtC.

\(^{39}\) See Nordhaus (2008), for instance. We use the calibration of DICE-NR-0124134-beta.

\(^{40}\) In emission scenario 1, emissions start at approximately the present emission level of 21 ppm equivalent, for the period length of 5 years as used in DICE. In steps of 1 ppm, emissions thereafter shortly increase to 23 ppm and then decrease until they reach zero after 25 periods.

\(^{41}\) Emission scenario 2 corresponds to the business-as-usual path in DICE-NR-0124134-beta.
6.4.1 The ‘Steady State’ Fallacy

In general, the notion of a ‘steady-state’ equilibrium is rather deceptive in the context of climate change. First, if radiative forcing changes with time, an equilibrium might never be reached. Second, and more fundamentally, there are many little-known feedbacks – mainly long-term ones – that depend on the state of the climate in a possibly highly non-linear way. Simple linear warming models might give a misleading sense of stability.42 In relation with this illusion of stability, a third problem arises. If parts of the climate system cross certain thresholds for a time long enough, abrupt and irreversible changes to new equilibrium might occur (Alley et al., 2003; Lenton et al., 2008).

The CO₂ Decay Model, on the other hand, invites to calculate a steady state where \( S_{ss} = S_{t+1} = S_t \). From Equation (6.6), in steady state, the emissions that are not taken up by short term sinks \( \theta E_{ss} \) have to be equal to the carbon taken up by the long-term

42 As Walter Broecker puts it “The climate system is an angry beast, and we’re poking it with sticks”.

---

**Figure 6.3:** Comparison of DICE and CCR. The left figure shows the emission scenario 1. The right picture shows the temperature increases according to DICE and CCR.

**Figure 6.4:** Analogous to Figure 6.3, but for emission scenario 2.
sink.

\[ \theta E^{ss} = \gamma S^{ss} \]  

(6.15)

This equation is a misrepresentation of the climate system. It might indeed be possible to emit small amounts of carbon and keep the atmospheric concentration constant to the extent that the deep ocean acts as a sink. How much carbon emissions are permissible is a complicated issue, of course, but to represent this process by Equation (6.15) seems inappropriate. It neglects the dependence on past cumulative emissions, which cause climate change and decrease the chemical capacity of the ocean.

A model like Equation (6.15) gives also a dubious policy signal. If the damage function and the abatement cost function could be agreed upon, this would allow to determine the steady state sweet spot where the emissions equal the sinks as well as where the marginal abatement costs equal the – net present value of the future – marginal damages. In such a setting, higher emissions have, in fact, two advantages. Not only do they decrease the marginal abatement costs, but they also increases the long-term sink strength. While the former is undoubtedly correct, the latter is not.

### 6.4.2 Carbon Discounting Fallacy

The use of the CO\(_2\) Decay Model is also deceptive when considering the optimal abatement level in periods where the CO\(_2\) stock is not constant. As it depicts the current situation, this exercise has arguably more direct relevance for current policy recommendations, such as the optimal level of a carbon tax or the emission permit cap.

To assess the influence of the CO\(_2\) Decay Model, consider the analysis of Goulder and Mathai (1999). They use the equivalent of Equation (6.6) in a continuous-time framework and calculate an optimal growth path. In the optimum, the shadow price of carbon, \( p_t \), grows as follows:

\[ \frac{\dot{p}_t}{p_t} = \rho + \gamma. \]  

(6.16)

Interestingly, \( \gamma \) has the same significance as the discount rate \( \rho \). As the appropriate choice of the discount rate is subject to a lively discussion, the influence of the carbon model should be of interest as well. To see this more clearly, consider footnote 11 in Goulder and Mathai (1999), where the authors discuss the most cost-effective emission path leading to a predetermined steady state GHG stock.

“Consider an path leading to a given concentration \( S_t \) a time \( T \). Since CO\(_2\) is removed naturally, altering this path by increasing emissions slightly at time \( t \) and decreasing it
at time $\Delta T > t$ leads to a greater overall removal and thus leads to a concentration at time $T$ of less than $S_t$.”

According to the CO$_2$ Decay Model, leaving cumulative emissions constant, there is more decay if there is more carbon in the atmosphere earlier on, as the decay is proportional to the stock.

“Equivalently, $S_t$ can be achieved with less cumulative emission abatement if the path of abatement is oriented more toward the future. Hence there is value of postponing abatement beyond that implied by the interest rate $r$, captured by $\gamma$.”

Following this line of thought to the extreme, one could argue for the usefulness of emitting all available fossil carbon as soon as possible – assuming that it will be emitted anyhow –, such that the ocean sink is able to ‘work’ at maximum power. We conjecture that hardly any climate scientist would subscribe to this strategy.

### 6.4.3 Crisis Management Fallacy

Suppose an economist’s model, M, is such that current warming is closely related to the current stock of GHGs. In addition, CO$_2$ sinks are such that the anthropogenic stock of CO$_2$ in the atmosphere decreases rather rapidly if emissions are decreased substantially. The Instantaneous Warming Model, combined with the CO$_2$ Decay Model, would fulfill these requirements, for instance. Suppose further that future generations would experience climate change to such a degree that decision-makers around the globe would agree to stop global warming or even reverse it. According to model M, this would be possible by reducing emissions substantially, as the CO$_2$ stock would fall and temperature would decrease in lockstep.$^{43}$ A potential consequence of this possibility is that present decision-makers could be tempted to invest little in combating climate change, knowing that if climate change turns out to be worse than currently estimated, future decision-makers would still be able to manage such a crisis – and reverse global warming.

Yet, model M dangerously underestimates the irreversibility of climate change. If future generations were to stop emissions, state-of-the-art models show that they could merely stop further global warming (Solomon et al., 2009; Allen et al., 2009; Matthews et al., 2009). In addition, sea levels would continue to rise for a very long time, and if tipping point had been crossed, this would, per definition, result in an irreversible change of the corresponding part of the earth system (Meehl et al., 2005; Lenton et al., 2008).

$^{43}$ As an example, see Figure 1, panel E in Acemoglu et al. (2012).
To reverse the temperature increase, more drastic approaches like negative emission through some form of air capture or geo-engineering of some sort would be needed. Acknowledging this irreversibility might lead present policy-makers to take a more cautious approach to climate change.

### 6.5 Numerical Examples of the Optimal Emission Path

Using a generic intertemporal welfare model, we compare the optimal abatement path for some combinations of carbon and warming models, which we call ‘Earth System Models’ (ESM). Note that we did not calibrate the models consistently, as this is beyond the scope of this paper. Instead, we used the calibrations from the literature – as given in Section 6.3 – with regard to model specific calibrations and illustrative values for the abatement costs and damage functions. As a consequence, the different ESMs cannot be compared quantitatively and the figure have no units. It is, however, possible to show qualitative differences.

#### 6.5.1 Model Set-Up

A global social planner chooses an emission path $E^t = \{E_1, \ldots, E_t\}$ for the periods $t = 1, 2, \ldots, T$ over a finite time horizon $T$. The planner aims at minimizing the discounted stream of abatement costs and damages. Abatement costs in period $t$ are

$$\Phi \frac{(\bar{E} - E_t)^2}{2} \quad t = 1, \ldots, T,$$

where $\Phi > 0$ is the abatement costs parameter. $\bar{E}$ are the constant business-as-usual emissions. If emissions correspond to $\bar{E}$, abatement costs are zero. We assume that emission abatement is achieved through some national environmental policy, such as taxes or a permit market for instance and that abatement costs are quadratic.

In general, damages in period $t$ are a function of the emission path $E^t$. This will be specified in the respective ESM.

$$D_t(E^t) \quad t = 1, \ldots, T,$$
At time $t$, the global social planner minimizes

$$
\min_{E_t} \sum_{s=t}^{T} \delta^{s-t} \left( \frac{\Phi}{2} (\bar{E} - E_s)^2 + D_s(E^*) \right),
$$

subject to a carbon and warming model and with the discount factor $\delta = 0.95$.

### 6.5.2 ESM 1: The CO$_2$ Decay Model and The Instantaneous Warming Model

In ESM 1, warming is modeled instantaneously. Hence, damages in period $t$ are a function of the current stock of GHG in the atmosphere. A quadratic damage function is the most convenient choice

$$
D_t(E_t) = \frac{B}{2} S_t^2, \quad t = 1, \ldots, T,
$$

where $B > 0$ is the damage cost parameter. The global social planner’s optimization problem is

$$
\min_{E_t} \sum_{s=0}^{T-t} \delta^s \left( \frac{\Phi}{2} (\bar{E} - E_{t+s})^2 + \frac{B}{2} S_{t+s}^2 \right),
$$

subject to Equation (6.6). Thus, first order conditions for all $h = t, \ldots, T$ read

$$
-\Phi(\bar{E} - E_h) + B\theta \sum_{i=0}^{T-h} \delta^i (1 - \gamma)^i S_{h+i} = 0.
$$

The solution is depicted in Figure 6.5. It shows that after a while, a steady state is approached where emissions equal the sinks. As explained in Section 6.4.1, this is misleading.

![Figure 6.5](image)

**Figure 6.5:** Optimal emissions, atmospheric CO$_2$ and temperature increase for ESM 1.

---

44 Periods close to the final period $T$ are not shown.
6.5.3 ESM 2: Climate Carbon Response

ESM 2 uses the CCR such that Equation (6.5) translates the emissions into the temperature increase. As this concept leapfrogs the CO$_2$ stock, the damage function is written in terms of the temperature

$$D_t(E^t) = \frac{T}{2} T^2,$$

where $T$ is a parameter. The optimization problem at time $t$ is

$$\min_{E^t} \sum_{s=0}^{T-t} \delta^s \left( \frac{\Phi}{2} (\bar{E} - E_{t+s})^2 + \frac{T}{2} T^2 \right). \tag{6.21}$$

The first order conditions for all $h = t, \ldots, T$ read

$$-\Phi(\bar{E} - E_h) + T \times CCR \times \sum_{i=0}^{T-h} \delta^i T_{h+i} = 0.$$

The solution is depicted in Figure 6.6. The optimal emissions approach zero. The intuition is straightforward. As long as emissions are positive, the current value of the marginal damages increases. In the optimum, the current value of marginal damage has to be equal to the current marginal abatement costs, which increase with decreasing emissions. As a consequence, emissions have to decline continuously up to the point where the current value of marginal damages are as high as the current marginal abatement costs of complete abatement.

![Figure 6.6: Optimal emissions and temperature increase for ESM 2.](image)


Barnosky, A. D., Hadly, E. A., Bascompte, J., Berlow, E. L., Brown, J. H., Fortelius,
Bibliography


Bibliography


Bibliography


194
Bibliography


Bibliography


Curriculum Vitae

Name: Quirin Oberpriller
Mail: quirinoberpriller@yahoo.de
Born: July 6, 1980 in Munich
Nationality: German

Education

June 2009 - May 2013
Doctor of Sciences, ETH Zürich

September 2007 - May 2009
Master in Atmospheric and Climate Science, ETH Zürich
Master thesis: Long Term Trends in the Ocean Carbon Sink

September 1999 - May 2005
Dipl.-Ing. Chemical Engineer, TU München
Diploma thesis: Cloud Feedbacks and Variable Climate Sensitivity

September 1990 - July 1999
Abitur (high school diploma), Gymnasium Dorfen
Intensive courses: Physics and Chemistry