Welfare-Maximizing Climate Policy and the Role of Climate Finance

Simon Lang[*]

May 23, 2024

[Click here for the latest version]

Abstract

This paper examines how optimal carbon prices that ignore intratemporal inequality compare to optimal carbon prices that account for inequality and the distribution of the costs and benefits of reducing emissions. Using a theoretical model, I identify plausible conditions under which accounting for inequality increases optimal emission reductions in the absence of international transfers. In numerical simulations with the integrated assessment model RICE, I find that accounting for inequality results in lower optimal global emissions, both if carbon prices are allowed to be regionally differentiated and if they are constrained to be globally uniform. I then assess how the Paris Agreement transfer of $100 billion per year affects optimal carbon prices. I find that optimal emission reductions increase considerably if the transfer is used to finance mitigation projects in developing countries.

[*] Yale University, School of the Environment (email: simon.lang@yale.edu).

I thank my advisors Kenneth Gillingham, William Nordhaus, Matthew Kotchen, David Anthoff, and Narasimha Rao for their invaluable feedback and support. I also thank Eli Fenichel, Mirco Dinelli, Matthew Gordon, Ethan Addicott, Andrew Vogt, Lisa Rennels, Francis Dennig, Paul Kelleher, Antony Millner, Kevin Kuruc, Itai Sher, Maya Eden, Eric Donald and participants of the Yale ENRE seminar, the EQWL lab group, the IAMC Annual Meeting, the Penn State Science and Values in Climate Risk Management webinar, the Oxford Global Priorities Fellowship, the Swiss workshop on Environmental, Resource and Energy Economics, and the Columbia IPWSD workshop for valuable discussions and comments.

1 Introduction

The distributional effects of climate change and climate policies are at the heart of international climate change negotiations. Central to these debates are inequalities in the impacts of climate change, the responsibilities for causing it, and the capabilities to mitigate and adapt to it—aspects that are all interlinked with global wealth inequality (Chancel et al., 2023). International agreements suggest that there is a political consensus to account for these inequalities in international climate policy. This is exemplified by the principle of "common but differentiated responsibilities and respective capabilities, in the light of different national circumstances" of the United Nations Framework Convention on Climate Change (UNFCCC, 2015). Moreover, the Paris Agreement highlights that developed countries should take the lead in reducing emissions and support developing countries in their transitions to low-carbon economies, emphasizing the necessity of incorporating the principle of equity and the goal of poverty eradication into climate policy (UNFCCC, 2015). In this context, international climate finance has become a central topic in recent meetings of the Conference of the Parties (UNFCCC, 2015; UNFCCC, 2023).

Despite the importance of inequalities and climate finance in international climate policy, the conventional, positive approach to optimal climate policy effectively ignores intratemporal inequality by maximizing a social welfare function (SWF) with Negishi welfare weights, making the distribution of the costs and benefits across countries irrelevant and leaving no role for climate finance (Yang and Nordhaus, 2006). In contrast, an alternative, normative approach focuses on maximizing global welfare (Budolfson et al., 2021). Invoking the ethical principle of impartiality, this approach commonly maximizes the equally-weighted utilitarian SWF (subject to constraints on international transfers), which accounts for the diminishing marginal utility of consumption and intratemporal inequality.

The important implication is that the distribution of the burden of abatement costs and climate damages across rich and poor countries matters under the normative approach, while it is irrelevant under the positive approach. This difference is particularly important since poor countries tend to be disproportionately harmed by climate change (Burke et al., 2015; Hallegatte et al., 2014; Kalkuhl and Wenz, 2020; Mendelsohn et al., 2006; Oppenheimer et al., 2014).

This paper asks how optimal carbon prices that ignore intratemporal inequality compare to optimal carbon prices that account for inequality and the distribution of the cost burden of mitigation and climate damages. I address this question first in the absence of international transfers, before allowing for international transfers to finance mitigation in recipient regions, thereby affecting the distribution of the cost burden of mitigation and, consequently, optimal carbon prices. Seeking to inform ongoing international climate change negotiations, I focus on examining the effect of a total transfer of \$100 billion per year by 2025, and rising thereafter, which developed countries have committed to mobilize at the Conferences of the Parties in Copenhagen and Paris (UNFCCC, 2009; UNFCCC, 2015). I study these questions both theoretically using an analytical model as well as through numerical simulations with the integrated assessment model (IAM) RICE.

I structure the analysis into two parts. First, I ignore international transfers and explore how the optimal carbon prices under the Negishi approach compare to the optimal carbon prices under the utilitarian approach with constraints on transfers. I refer to the carbon prices under the utilitarian solution as welfare-maximizing carbon prices to highlight that they maximize the (unweighted) sum of individuals' utilities[1]. I start by imposing the same two constraints on the utilitarian optimization that are implicit in the Negishi solution: no international transfers and uniform carbon prices. This constrains the utilitarian problem to an identical choice of policy instrument - a globally uniform carbon price in each period - allowing for a direct comparison with the Negishi solution. Subsequently, I remove the uniform carbon price constraint, allowing for differentiated carbon prices.

Using a theoretical model, I show that optimal carbon prices and aggregate abatement may be higher or lower in the utilitarian solutions than in the Negishi solution and that this depends on the distribution of the marginal climate damages and the burden of the abatement costs on different countries. In numerical simulations with RICE, I find that the optimal abatement is greater in the welfare-maximizing solutions than in the Negishi solution. This shows that accounting for background inequality and the distribution of costs and benefits of abatement results in higher optimal carbon prices than when inequality is ignored. This is the case even when carbon prices are constrained to be uniform across regions. A main part of the intuition is that Negishi weights place a lower weight on the welfare of poor countries, which tend to be disproportionately affected by climate change, resulting in lower carbon prices in the Negishi solution than in the welfare-maximizing solution.

Second, I introduce international transfers for mitigation to study how climate finance affects optimal carbon prices in the welfare maximization framework. I focus on examining the effect of the "Paris Agreement transfer" of \$100 billion per year by 2025 and rising thereafter. I find that financial support for mitigation in developing countries considerably increases the stringency of the welfare-maximizing climate policy under both the uniform and the differentiated carbon price solutions. For instance, under the default discounting parameters in RICE, the welfare-maximizing uniform carbon price in 2025 almost doubles, from \$29/tCO_2 to \$54/tCO_2, if the Paris Agreement transfer is used to finance additional

[1] Notice that I use the terms "welfare" and "utility" interchangeably throughout this paper.

mitigation in developing countries. Moreover, compared with the Negishi solution, optimal global cumulative emissions are 31% lower in the utilitarian solution with differentiated carbon prices and international mitigation finance; this reduces the optimal peak temperature increase from about 3°C to around 2.4°C.

This paper makes several contributions to the literature on optimal carbon prices with heterogeneous regions (Nordhaus and Yang, 1996). First, it provides novel theoretical results on how optimal carbon prices depend on intratemporal welfare weights and the distribution of costs and benefits of abatement. In particular, I show under which conditions welfare-maximizing carbon prices (and global emission reductions) are greater or lower than the conventional efficient carbon prices, in the absence of transfers. These results build on an influential paper by Chichilnisky and Heal (1994), which shows that a globally uniform carbon price is optimal if, and only if, distributional issues are ignored (through the choice of particular welfare weights) or lump-sum transfers are made between countries. Related papers examined aspects of efficiency and equity in emission permit markets (Chichilnisky and Heal, 2000; Shiell, 2003; Sandmo, 2007; Borissov and Bretschger, 2022), the importance of accounting for inequalities at a fine-grained level (Dennig et al., 2015; Schumacher, 2018), and how optimal carbon taxes, under arbitrary welfare weights, depend on distortionary fiscal policy (Barrage, 2020) and inequality within and between countries (Kornek et al., 2021).

Second, this paper adds to a small literature that numerically investigates the role of intratemporal welfare weights in IAMs. The studies most closely related to this research are those by Anthoff (2011) and Budolfson and Dennig (2019), which also compare the Negishi solution[2] to utilitarian solutions, although using different models (the FUND and NICE model, respectively). Another relevant study, Adler et al. (2017), compares the social cost of carbon estimates (along a business-as-usual emissions trajectory) derived from the standard discounted-utilitarian SWF with those obtained using a prioritarian SWF without time discounting. The present paper builds upon these studies but, due to the new theoretical results, allows for an improved understanding of important drivers of the numerical optimization results. Furthermore, I extend the analysis by examining the distributional implications and I provide additional insights on heterogeneous climate policy preferences by computing regions' preferred uniform carbon prices, a notion that was first introduced by Weitzman (2014) and Kotchen (2018).

Third, this paper makes contributions to the literature studying climate policy in conjunction with transfers. To my knowledge, it is the first to examine how limited international

[2] Note that Budolfson and Dennig (2019) do not technically use Negishi weights but a model version in which all individuals in all regions consume the global average consumption.

climate finance, specifically the "Paris Agreement transfer" of $100 billion per year, impacts welfare-maximizing carbon prices. Furthermore, it estimates the optimal allocation of transfers for mitigation. The study most closely related to the present paper is Yang and Nordhaus (2006), which examines optimal unrestricted transfers for mitigation under different social welfare weights, showing that zero (large) transfers take place if Negishi (utilitarian) weights are used. Another relevant paper, Kornek et al. (2021), focuses on how national redistribution impacts optimal carbon prices. In an extension, the authors also theoretically explore how unrestricted international lump-sum transfers impact optimal carbon prices, and provide a brief qualitative discussion of the effects of restricted transfers. Other papers in this broader literature have explored the potential of transfers to facilitate international cooperation (Hoel, 1994; Hoel et al., 2019; Kotchen, 2020; Hillebrand and Hillebrand, 2023), how the intended effects of mitigation and adaptation transfers can be achieved (Eyckmans et al., 2016), and how transfers and differentiated carbon prices may be combined to equalize mitigation costs as a share of income across countries (Bauer et al., 2020).

The remainder of this paper is structured as follows. Section 2 provides conceptual background on positive and normative optimization approaches. In Section 3, a theoretical model is introduced and key analytical results are derived. Section 4 describes modifications to the IAM RICE and presents and discusses the results from the numerical simulations. Section 5 concludes.

2 Conceptual Background

This section provides conceptual background on positive and normative optimization approaches. In Section 2.1, the difference between the two approaches is introduced. Section 2.2 examines the positive approach, focusing on the rationale for and critiques of Negishi weights. Finally, Section 2.3 provides a welfare-economic conceptualization of the normative optimization approach.

2.1 Positive versus normative optimizations

The purpose of optimization is a main source of debate in IAMs and there are two main approaches: positive and normative optimization. An instructive discussion of these two approaches is provided by Nordhaus (2013, p. 1081) who notes that: "the use of optimization can be interpreted in two ways: they can be seen both, from a positive point of view, as a means of simulating the behavior of a system of competitive markets and, from a normative point of view, as a possible approach to comparing the impact of alternative paths or policies on economic welfare." In brief, the positive approach seeks to identify the competitive equilibrium, while the normative approach aims at maximizing social welfare.

The issue of discounting, which determines the intertemporal weighting of consumption and welfare, has received much attention in the debate on positive versus normative optimization approaches (Arrow et al., 2013; Azar and Sterner, 1996; Beckerman and Hepburn, 2007; Dasgupta, 2008; Dietz and Stern, 2008; Nordhaus, 2007). Under the positive approach, the discount rate is determined based on market observations. In contrast, under the normative approach, ethical reasoning is used to determine the discount rate.

However, the difference between positive and normative optimization approaches extends to the intratemporal weighting of welfare. The typical positive approach relies on Negishi welfare weights – which assign higher welfare weights to rich individuals – to identify the competitive equilibrium. In contrast, under the normative approach, uniform welfare weights, which are also called utilitarian welfare weights, are most commonly used, weighting everybody's welfare equally[3]. The remainder of this section, and this paper in general, focuses on intratemporal welfare weights.

2.2 The positive approach: Background on Negishi weights

Negishi welfare weights are commonly used in regionally disaggregated integrated assessment models of climate change. Popular IAMs that use Negishi weights include RICE (Nordhaus and Yang, 1996), which this paper focuses on, MERGE (Manne and Richels, 2005), REMIND (Leimbach et al., 2010) and WITCH (Bosetti et al., 2012). This section outlines the rationale for using Negishi weights in IAMs and presents some critiques that have been raised. It finishes with a welfare economics perspective on the positive optimization approach.

2.2.1 Rationale for using Negishi weights in IAMs

The theoretical basis for the use of Negishi weights in IAMs is a theorem of Negishi (1960). Negishi proved that a competitive equilibrium can be found by maximizing a social welfare function in which the welfare of each agent is appropriately weighted so that each agent's budget constraint (consistent with their initial endowments) is satisfied at the equilibrium (Nordhaus and Yang, 1996). The Negishi-weighted social welfare function is given by a weighted sum of agents' utilities, where the weights are inversely proportional to the marginal utility of consumption.

[3] Note that other normatively-founded SWFs have been used in the climate economics literature, including the prioritarian SWF (Adler et al., 2017) and variants of the Rawlsian SWF (Roemer, 2011; Llavador et al., 2010; Llavador et al., 2011).

In addition to their theoretical foundation, a main motivation for the use of Negishi weights in regionally disaggregated IAMs is to prevent large capital flows between regions, which may be deemed politically infeasible or unrealistic (Nordhaus and Yang, 1996). Without Negishi weights, social welfare could be increased by redistributing capital or consumption from rich to poor regions in models that maximize the unweighted sum of agents' utilities, if utility is an increasing concave function of consumption, which is commonly assumed (i.e., the utility function features a diminishing marginal utility of consumption, reflecting that an additional unit of consumption is of greater value to a poor person than to a rich person).

However, the use of the time-invariant welfare weights proposed by Negishi (1960) did not solve the problem of unrealistically large transfers in intertemporal optimization models over the entire time horizon (i.e., in all model periods) (Nordhaus and Yang, 1996). This motivated Nordhaus and Yang (1996) to undertake refinements to what they call the "pure Negishi solution" of using time-invariant welfare weights with no additional constraints. The first approach taken by Nordhaus and Yang to avoid unrealistically large transfers "was to impose certain flow and stock constraints on debt and current accounts to ensure that net foreign investment does not exceed certain limits" (Nordhaus and Yang, 1996, p. 747). While this solved the problem of large transfers between regions, it did not yield a globally uniform carbon price.

To ensure that the carbon price is equalized, Nordhaus and Yang (1996) adjust the Negishi weights in each time period such that the weighted marginal utility of consumption is equalized in each period (Stanton, 2011). This approach thus yields time-variant Negishi weights and accomplishes the goal of equalizing the carbon price across regions in every period. Moreover, these weights ensure that each region's budget constraint is satisfied in each period, incorporating the constraint of no cross-regional capital flows (Nordhaus and Yang, 1996). Hence, the constraints of equalized carbon prices and no transfers of capital or consumption are effectively incorporated in the time-variant Negishi weights used in RICE.

Formally, the social welfare function with time-variant Negishi welfare weights is

$$W = \sum_t \sum_i \beta^t L_{it} \alpha_{it} u(x_{it}), \tag{1}$$

where i and t are the region and time indices, L_{it} is the population, x_{it} is the per-capita consumption, u is the utility function which is typically assumed to be increasing and concave, β^t is the utility discount factor, and α_{it} are the time-variant Negishi weights. The time-variant Negishi weights are proportional to the inverse of the marginal utility of consumption, that is, $\alpha_{it} \propto 1/u'(x_{it})$, in each period (for more details on the Negishi weights used in RICE, see Section 4.1.2).

To summarize, the rationale for using time-variant Negishi weights in regionally disaggregated welfare maximizing IAMs is twofold: (1) to prevent transfers across regions in every period, and (2) to equalize the carbon price across regions in every period.

2.2.2 Critiques of using Negishi weights in IAMs

Negishi weights are criticized on both ethical and theoretical grounds (Anthoff et al., 2021; Dennig and Emmerling, 2019; Stanton, 2011; Stanton et al., 2009). This section provides a summary of main critiques.

The main criticism from an ethical perspective is that Negishi weights assign a greater weight to the welfare of people in rich countries than in poor countries. This is the case because Negishi weights are inversely proportional to the marginal utility of consumption and the utility function is typically assumed to be concave (i.e., richer regions are assigned a greater Negishi weight since their marginal utility of consumption is smaller). Models with Negishi weights are thus "acting as if human welfare is more valuable in the richer parts of the world" (Stanton et al., 2009, p. 176). Moreover, because Negishi weights equalize the weighted marginal utility of consumption, aspects of interregional equity are effectively ignored and the reality of global inequality is neglected (Stanton, 2011; Stanton et al., 2009). As a result, it is irrelevant whether poor or rich countries are affected by climate change and climate policies (Dennig et al., 2015).

Moreover, Stanton (2011) notes that models with Negishi weights have an inherent inconsistency: the diminishing marginal utility of consumption is embraced intertemporally, but suppressed interregionally. This leads to the controversial result that transfers from richer to poorer individuals is desired in an intertemporal context but rejected in an interregional context.

Another criticism from a theoretical perspective is provided by Dennig and Emmerling (2019) and Anthoff et al. (2021). In a simple analytical model, these authors show that the time-variant Negishi weights, used for example in the RICE model (Nordhaus and Yang, 1996), distort the time-preferences of agents and result in different saving rates than those implied by the underlying preference parameters. Furthermore, they note that the time-invariant weights proposed by Negishi (1960) do not have this problem because they only consist of one weight per agent, and thus only affect the distribution between agents, but leave the intertemporal choices of each agent unaffected. Because of the distorting effect of time-variant Negishi weights, Dennig and Emmerling (2019) argue that they should no longer be used. Instead, they propose that modellers should accept that optimality will not yield uniform carbon prices if international capital flows are limited.

A final criticism of Negishi weights concerns the manner in which Negishi weights are often introduced – if discussed at all – which is frequently rather technical with no or little transparent discussion of the ethical implications (Abbott and Fenichel, 2014; Stanton, 2011). Given that Negishi weights contain a key ethical assumption, Stanton et al. (2009) highlight the importance of making their ethical implications transparent to be visible for debate, rather than presenting them in an opaque technical manner, thereby discouraging discussion.

2.2.3 Welfare economics perspective on the positive approach

This section provides a discussion of the positive optimization approach from the perspective of welfare economics.

From the first fundamental theorem of welfare economics, it is known that, under certain conditions, the competitive equilibrium is Pareto-efficient (Sen, 1985). That is, no one can be made better off without making someone else worse off. The maximization of a Negishi-weighted SWF in IAMs seeks to identify the competitive equilibrium with a Pareto-efficient level of abatement[4]. I will refer to this solution as the "Negishi solution". The Negishi solution is one particular point (among infinitely many points) on the Pareto frontier in a first-best setting in which only resource and technology constraints are present (assuming that the conditions for the first fundamental theorem of welfare economics hold otherwise; for a discussion of first- and second-best settings, see Section 2.3). Notably, it is the only Pareto-efficient allocation in a first-best setting that does not require transfers (Shiell, 2003). In the absence of abatement, the competitive equilibrium is not efficient due to the climate externality. This is illustrated in Figure 1a, which shows the utility possibility set and the Pareto frontier for a simple case of two regions: a relatively rich Global North, and a comparatively poor Global South.

While the Negishi solution is Pareto-efficient, it cannot generally be considered to maximize social welfare. This is because the Negishi-weighted SWF is not intended to measure social welfare. Instead, it is calibrated such that a Pareto-efficient allocation which does not require transfers is obtained. In contrast, normative optimizations rely on SWFs that are rooted in theories of social welfare. The most common theory of social welfare in economics is utilitarianism, which places equal weight on the welfare of all individuals. Importantly, the Negishi solution does not maximize aggregate welfare if the welfare of all people is weighted equally. Maximizing a utilitarian SWF maximizes the (equally weighted) sum of the welfare of all individuals. This is illustrated in Figure 1b.

[4] However, Anthoff et al. (2021) show that the time-variant Negishi weights used in IAMs do not, in fact, yield a Pareto-efficient solution. This is because of a time-preference altering effect of time-variant Negishi weights. In this section, I focus on a static setting in which this issue does not arise.

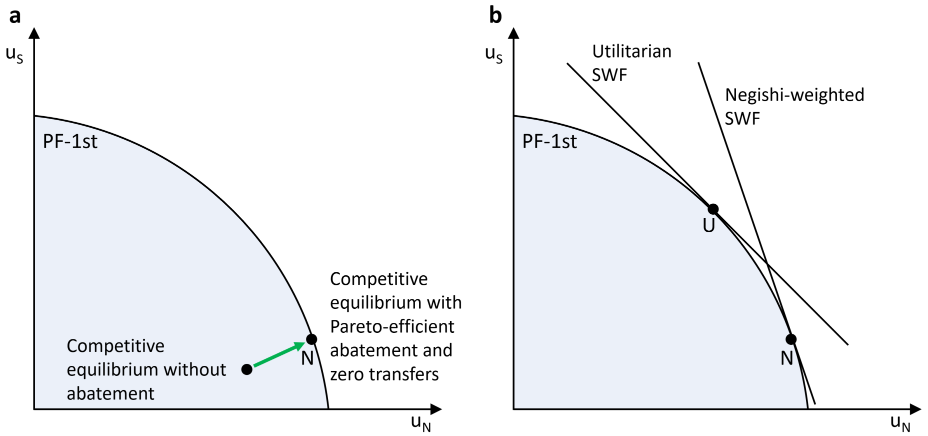
a
u_S
PF-1st
Competitive equilibrium without abatement
N
Competitive equilibrium with Pareto-efficient abatement and zero transfers
u_N
b
u_S
Utilitarian SWF
Negishi-weighted SWF
PF-1st
U
N
u_N

Figure 1: Illustration of the welfare outcomes under the Negishi and utilitarian solutions in a two-region world. (a) The Negishi solution (N) is an efficiency improvement relative to the competitive equilibrium without abatement. (b) Comparison of the Negishi (N) and utilitarian (U) solutions. u_N and u_S are the utilities (i.e., welfare levels) of representative agents in the Global North and Global South, respectively. PF-1st is Pareto frontier in a first-best setting. The graphs are adapted from Shiell (2003).

Given that the Negishi solution does not maximize aggregate (unweighted) welfare, how may the use of Negishi weights in IAMs be justified? There are at least two possible lines of argument. First, it may be argued that the Negishi solution has no normative but only a positive interpretation; that it is merely a procedure to identify the competitive equilibrium with Pareto-efficient abatement and zero transfers. For example, Nordhaus (2013, p. 1111) notes that "if the distribution of endowments across individuals, nations, or time is ethically unacceptable, then the "maximization" is purely algorithmic and has no compelling normative properties". Moreover, Nordhaus and Sztorc (2013, p. 20) clarify: "We do not view the solution as one in which a world central planner is allocating resources in an optimal fashion".

A second line of argument relies on the second fundamental theorem of welfare economics, which states that any point on the Pareto frontier can be supported as a competitive equilibrium if unrestricted lump-sum transfers can be made. Moreover, this argument is sometimes used to argue that the issues of equity and efficiency can be separated. However, it has been noted that distributional issues and Pareto-efficiency are not separable in the case of climate policy, as the Pareto-efficient abatement level generally depends on the distribution of wealth. This is because the marginal willingness to pay for abatement generally varies with income (Shiell, 2003). Therefore, the Negishi solution only identifies a Pareto-efficient abatement level if no transfers occur. Moreover, the practical relevance of the second fundamental theorem of welfare economics has been questioned. For instance, Sen (1985, p. 12) notes that "if there is an absence of – or reluctance to use – a political mechanism that would actually redistribute resource-ownership and endowments appropriately, then the practical relevance of the converse theorem [the second fundamental theorem of welfare economics] is severely limited".

To summarize, the abatement under the Negishi solution is not equal to the abatement that maximizes utilitarian social welfare[5], regardless of whether unrestricted transfers are feasible or not.

2.3 The normative approach: Welfare-economic conceptualization

This section provides a conceptualization of the normative optimization approach, grounded in welfare economics. In doing so, the objective of this section is to clarify the fundamental distinction between positive and normative optimization approaches in climate economics.

In Section 2.2.1, I have argued that constraints are implicitly incorporated in the welfare weights under the positive approach. Here, I emphasize that this constitutes a key difference to the normative approach, where constraints and welfare weights are determined separately.

I propose to conceptualize the normative optimization approach as consisting of two steps. First, the social welfare function is defined based on ethical principles. Second, potential constraints on the optimization affect the feasible set of allocations. Importantly, these two steps are separate under the normative approach. It is worth elaborating on each step.

The first step is standard in normative analyses; it is the specification of the SWF based on ethical principles. Such SWFs have a long tradition in normative economics and are referred to as Bergson-Samuelson SWFs, since they were introduced by Bergson (1938) and prominently adopted by Samuelson (1947). The Bergson-Samuelson SWF is used to produce an ethical ordering of possible societal outcomes. Common Bergson-Samuelson SWFs include the utilitarian, prioritarian and Rawlsian SWFs (Mas-Colell et al., 1995). In this paper, I focus on the utilitarian SWF, which is most commonly used in the climate economics literature.

The second step is to carefully consider and explicitly account for real-world constraints in the optimization. This step is often less thoroughly addressed in the existing literature. It is of course challenging to determine and formalize plausible real-world constraints (especially in stylized IAMs). It therefore seems valuable to explore a plausible range of constraints. Conceptually, such constraints affect the feasible set of allocations, which, in turn, determines the utility possibility set (UPS), which was introduced by Samuelson (1947). Ultimately, we are interested in the Pareto frontier, which is defined as the upper frontier of the UPS[6].

[5] Or, more generally, social welfare measured with any other linear SWF with welfare weights that are not equal to Negishi weights.

[6] Economists sometimes use the term efficiency to simply mean outcomes that maximize the total mon-

In short, the Pareto frontier is defined on the set of feasible allocations. Finally, the social optimum is simply the point in the UPS that maximizes the SWF. This allocation is, of course, a point on the Pareto frontier.

Depending on the constraints imposed on the optimization, a conceptual distinction between first-best and second-best settings is frequently made (Mas-Colell et al., 1995). Typically, a first-best setting is considered to be a setting in which only resource and technology constraints are present, but otherwise the social planner has access to any policy instrument, including unrestricted lump-sum transfers. In contrast, the notion of second-best settings is used when additional constraints are present, such as constraints on transfers and the set of policy instruments that may be used.

It is instructive to illustrate how the normative optimization approach works in the context of this paper. This is shown in Figure 2 for optimization problems considered in this paper. In the first step, the utilitarian SWF is specified (which has linear social indifference curves with slope -1). In the second step, potential constraints are specified. Of particular relevance in the context of international climate policy are constraints on international transfers and whether carbon prices are constrained to be uniform across countries.

In the first-best setting, there are no constraints (apart from the usual resource and technology constraints). In particular, unrestricted lump-sum transfers can be made. In this setting, the social planner uses cost-effective uniform carbon prices to internalize the climate externality and lump-sum transfers to address distributional issues. With identical and concave utility functions, large transfers are made to equalize per-capita consumption across regions (Dennig et al., 2015), eliminating inequality. This results in the highest utilitarian welfare; the outermost social indifference curve, W_{1st}, is achieved.

However, as discussed above, such a first-best setting with large international transfers may be politically infeasible. As Shiell (2003, p. 43) puts it, “Unrestricted lump-sum transfers are a useful construct which scarcely exist outside the confines of economic theory”. As discussed in Section 2.2.1, the political infeasibility of large transfers motivated, in part, the specification of the welfare weights under the positive optimization approach. In contrast, under the normative optimization approach, welfare weights are independently specified of constraints since they are set in accordance with ethical principles. Instead, political constraints on transfers affect the feasible set of allocations.

etary sum (for short, maximizing dollars). In a first-best setting in which unrestricted lump-sum transfers are feasible, maximizing dollars is necessary and sufficient for Pareto efficiency. Importantly, however, in a second-best setting in which unrestricted lump-sum transfers are infeasible, maximizing dollars is not necessary for Pareto efficiency (nonetheless, maximizing dollars is, of course, one Pareto efficient outcome on the Pareto frontier among infinitely many other points on the Pareto frontier that do not maximise dollars). Throughout this paper, I use the standard definition of Pareto efficiency that no one can be made better off without making someone else worse off, given the constraints of the problem.

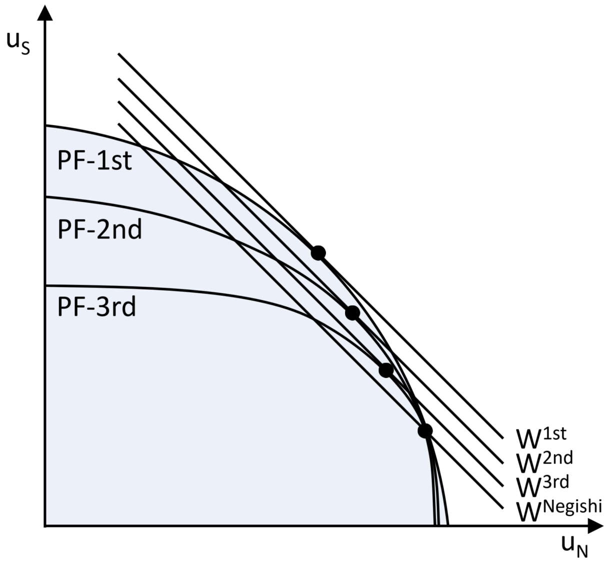

Figure 2: Illustration of the normative optimization approach under first- second- and third-best settings. The figure also shows a comparison to the utilitarian welfare level of the Negishi solution. u_N and u_S are the utilities (i.e., welfare levels) of representative agents in the Global North and Global South, respectively. PF-x is the Pareto frontier in the x^{th}-best setting. W^x is the social indifference curve that corresponds to the social optima in the x^{th}-best setting or the Negishi solution.

Hence, let us consider a second-best setting in which international lump-sum transfers are infeasible[7]. The lack of this policy option reduces the feasible set of allocations, the UPS gets smaller, and the Pareto frontier moves inward (except for one point on the frontier, which corresponds to the Negishi solution, which does not require transfers). Consequently, the social optimum lies on a lower social indifference curve, W_{2nd}. In the absence of the option to eliminate inequality with lump-sum transfers, the social planner accounts for background inequality in the climate policy design. Specifically, differentiated carbon prices that are higher in rich regions and lower in poor regions are used to reduce the welfare cost of abating emissions (Chichilnisky and Heal, 1994) (also see Section 3.2.4). It should be noted that a potential problem with differentiated carbon prices is carbon leakage – an increase in emissions in countries with laxer climate policies as a result of stricter climate policies elsewhere. However, additional policies such as carbon border adjustments and binding emission targets can avert the issue of carbon leakage. For a more detailed discussion, see (Budolfson et al., 2021) and Appendix C.4.

Finally, consider a third-best setting in which the policy instruments the social plan-

[7] I intentionally focus on the case of no transfers here to keep the discussion simple. In reality, however, some transfers are feasible (e.g., international aid or climate finance). I consider the effect of climate finance in Section 4.3.

ner can use are restricted even further to a globally uniform carbon price (in addition to a constraint of no transfers). It should be emphasized that this is not a plausible constraint in reality, as evidenced by widely different empirical carbon prices across countries (World Bank, 2023a). Nevertheless, it provides a useful comparison to the solution under the positive optimization approach, as it constrains the utilitarian problem to an identical choice of policy instrument - a globally uniform carbon price and no transfers. Yet, an important difference remains. The utilitarian uniform carbon price accounts for background inequality, while the optimal carbon price under the positive optimization approach ignores background inequality through the specification of the Negishi weights which equalize the weighted marginal utility across regions. As a consequence, the utilitarian uniform carbon price is generally different from the uniform carbon price under the Negishi solution. Importantly, the utilitarian uniform carbon price solution is weakly better, from the perspective of utilitarian welfare, than the Negishi solution (compare social indifference curves W_{3rd} and $W_{Negishi}$). This is simply because the utilitarian uniform carbon price is, by construction, the uniform carbon price that maximizes utilitarian welfare in a setting in which transfers are infeasible. Thus, any other uniform carbon price can only be weakly worse in terms of utilitarian welfare.

It is worth highlighting how the different solutions respond to background inequality. The spectrum ranges from completely solving inequality through lump-sum transfers in the first-best utilitarian setting to ignoring inequality altogether in the Negishi solution. While, the social optima in the second- and third-best settings do not solve inequality through transfers, they account for inequality to different degrees in the carbon pricing policy. In the second-best setting, inequality is accounted for in the *level* and *differentiation* of carbon prices across regions. In contrast, in the third-best setting, inequality is only accounted for in the *level* of the carbon price, thereby reducing the degree to which inequality is taken into account.

To which extent inequality is ultimately accounted for in international climate policy is decided by policymakers and international negotiations. However, international agreements indicate that there is a political consensus to account for inequality to some extent. This is evidenced, for example, by the UNFCCC principle of "common but differentiated responsibilities and respective capabilities, in the light of different national circumstance" and a general recognition that developed countries have an obligation to reduce their emissions faster and support developing countries in their transitions toward low-carbon economies, which is also reflected in the respective nationally determined contributions (NDCs) under the Paris Agreement (UNFCCC, 2015; Climate Watch, 2022). More broadly, the Paris Agreement underscores the necessity of incorporating the principle of equity and the goal of poverty eradication into climate policy, indicating that countries have agreed to account for inequality in international climate policy (UNFCCC, 2015). Hence, policymakers may be interested in socially optimal climate policies that take inequality into account. The present study seeks to identify such policies and contrasts them with the conventional, positive approach that neglects inequality.

3 Theory

This section introduces a theoretical model to study how optimal carbon prices depend on welfare weights, in the absence of international transfers. The model setup is outlined in Section 3.1. Results are derived and discussed in Section 3.2.

3.1 Model setup

The general model setup builds on Chichilnisky and Heal (1994) and Dennig and Emmerling (2017). The intention is to construct the simplest models possible to generate key insights and to provide conceptual underpinnings for important drivers of the simulation results in Section 4.

There are two regions $i \in \{N, S\}$ and a single period. Let $\mathcal{I} = \{N, S\}$ denote the set of regions; for intuition, consider the regions as the Global North (N) and Global South (S). I normalize the population size in both regions to unity. Thus, aggregate variables equal per-capita variables.

Abatement costs, $C_i(A_i)$, are a function of the abatement $A_i \geq 0$ in a region. The abatement cost function differs by region and is assumed to be smooth, strictly increasing, $\frac{dC_i}{dA_i} > 0$, and strictly convex, $\frac{d^2C_i}{dA_i^2} > 0$. Moreover, to keep the exposition simple, I assume that $\frac{d^2C_i}{dA_i^2}$ is constant but region-specific; that is, $\frac{d^3C_i}{dA_i^3} = 0$ for all A_i[8]. This is the case for the commonly assumed quadratic abatement cost function. I define the aggregate global abatement as $A \equiv \sum_i A_i$ (note the missing i subscript on the aggregate abatement). Region-specific climate damages, $D_i(A)$, are a function of the aggregate abatement. The damage function is assumed to be smooth, strictly decreasing, $\frac{dD_i}{dA} < 0$, and strictly convex in abatement, $\frac{d^2D_i}{dA^2} > 0$, reflecting the idea of convex damages as a function of emissions.

There is a representative agent in each region, who derives utility, $u(X_i)$, from consumption, X_i. The utility function is assumed to be the same for all individuals, strictly increasing,

[8] Note, however, that some of the main results do not require the assumption of constant second derivatives of the abatement cost functions (e.g., Propositions 1 and 2). However, to keep the exposition simple, I assume constant second derivatives of the abatement cost functions throughout.

strictly concave, and smooth. Thus, $\frac{du}{dX_i} > 0$ and $\frac{d^2u}{dX_i^2} < 0$. Regional consumption, X_i is given by the endowment, W_i, net of abatement costs and climate damages. That is,

$$X_i = W_i - C_i(A_i) - D_i(A_{\bar{t}}), \tag{2}$$

I assume throughout that the Global North is richer than the Global South, both in terms of endowment and consumption. Thus, we have $W_N > W_S$ and $X_N > X_S$. The implicit assumption is that the endowment difference between the Global North and the Global South is sufficiently large such that the Global North still always remains richer after abatement costs and climate damages are subtracted. From the concavity of the utility function, it follows that $u'(X_N) < u'(X_S)$.

3.2 Optimal carbon prices in the absence of transfers

I this section, I establish how the optimal carbon tax depends on the welfare weights in the absence of interregional transfers. Note that this implicitly imposes a constraint of no transfers in the optimization problem. I start with deriving results for arbitrary welfare weights before determining the results for specific choices of welfare weights that are commonly used.

3.2.1 Solutions for arbitrary welfare weights

The optimal carbon prices are determined through the maximization of a social welfare function (SWF) with welfare weights, $\alpha_i \geq 0$, which I do not further define at this point. Conceptually, varying the welfare weights across their full range traces out the *constrained* Pareto frontier. The resulting carbon prices are thus *constrained* Pareto-efficient, where the notion of constrained Pareto-efficiency depends on the constraints imposed on the optimization problem.

I consider two general optimization problems, reflecting the optimizations that are most commonly performed in the literature on optimal carbon prices (e.g., in Nordhaus and Yang (1996), Dennig et al. (2015), Budolfson et al. (2021)), . The first allows (but does not require) differentiated carbon prices and the second requires uniform carbon prices. The objective is to choose the carbon prices that maximize the SWF subject to regional budget constraints, reflecting a constraint of no interregional transfers. An additional constraint of uniform marginal abatement costs is imposed in the case that requires uniform carbon prices.

Formally, the *differentiated carbon price optimization problem* is

$$\max_{X_i,A_i} \sum_i \alpha_i u\left(X_i\right) \tag{3}$$

$$\text{subject to: } X_i = W_i - C_i(A_i) - D_i(A), \quad \forall i. \tag{4}$$

The *uniform carbon price optimization problem* is

$$\max_{X_i,A_i} \sum_i \alpha_i u\left(X_i\right) \tag{5}$$

$$\begin{aligned} \text{subject to: } X_i &= W_i - C_i(A_i) - D_i(A), \quad \forall i \\ C'_N(A_N) &= C'_S(A_S), \end{aligned} \tag{6}$$

where the additional constraint of uniform marginal abatement costs is imposed [9].

Solving the optimization problems yields expressions for the optimal marginal abatement costs. Optimal carbon prices, τ_i, equal the optimal marginal abatement costs, $C_i'^*$, because regions optimally respond to a carbon price by abating until their marginal abatement cost equals the carbon price; that is, $C_i'^*(A_i^*(\tau_i)) = \tau_i$. I record these results in the following two definitions. The derivations are provided in Appendix A.1.

Definition 1. *The* ***optimal differentiated carbon price*** *(for arbitrary welfare weights) for region i is implicitly defined by*

$$\tau_i^{diff} = C_i'^*(A_i^*) = -\frac{1}{\alpha_i u'(X_i^*)} \sum_{j\in\mathcal{I}} \alpha_j u'(X_j^*) D_j'(A^*). \tag{7}$$

In words, the optimal differentiated carbon price is equal to the sum of the avoided weighted marginal welfare damages divided by the weighted marginal utility. Thus, the optimal differentiated carbon price is inversely proportional to the *weighted* marginal utility, $\alpha_i u_i'$. Consequently, the optimal differentiated carbon price is lower in the region with the higher weighted marginal utility. This result has first been established by Chichilnisky and Heal (1994). Note that, if the weighted marginal utilities are equal across regions (i.e., $\alpha_S u_S' = \alpha_N u_N'$), we obtain the knife-edge result that the optimal "differentiated" carbon price is in fact uniform. This is the case if the weights are the Negishi weights. I return to this below.

[9] Note that I am using prime notation for some derivatives. The derivatives are defined as $C_i'(A_i) \equiv \frac{dC_i(A_i)}{dA_i}$, $D_i'(A) \equiv \frac{dD_i(A)}{dA}$, $u'(X_i) \equiv \frac{du(X_i)}{dX_i}$.

It is insightful to rearrange Equation (7) to

$$\alpha_i u'(X_i^*) C_i'^* = -\sum_{j \in \mathcal{I}} \alpha_j u'(X_j^*) D_j'(A^*). \tag{8}$$

Since the right-hand side is the same for all regions, we know that $\alpha_N u'(X_N^*) C_N'^* = \alpha_S u'(X_S^*) C_S'^*$. That is, the weighted marginal welfare cost of abatement (rather than the marginal abatement cost in monetary terms) is equalized across regions.

Definition 2. *The* ***optimal uniform carbon price*** *(for arbitrary welfare weights) is implicitly defined by*

$$\tau^{uni} = C_i'^*(A_i^*) = -\sum_i \alpha_i u'(X_i^*) D_i'(A^*) \frac{C_S''(A_S^*) + C_N''(A_N^*)}{\alpha_N u'(X_N^*) C_S''(A_S^*) + \alpha_S u'(X_S^*) C_N''(A_N^*)}. \tag{9}$$

The optimal uniform carbon price again depends on the sum of the avoided weighted marginal welfare damages. However, it also depends on a second factor which contains the second derivatives of the abatement cost functions. To gain some intuition, we can note that the expression collapses to the expression for the optimal differentiated carbon price if one of the regions has a linear abatement cost function[10]; that is, $C_i'' = 0$ for one i. Specifically, if the Global North has a linear abatement cost function, then the expression collapses to the differentiated carbon price expression for the Global North[11]; and vice-versa for the Global South. The intuition is that if one region has a linear abatement cost function, and thus constant marginal abatement costs, then the only way to equalize marginal abatement costs across regions is to adjust the marginal abatement cost of the other region. Unsurprisingly, this provides the intuition that the optimal uniform carbon price lies in between the two optimal differentiated prices. Moreover, whether the uniform carbon price is closer to one or the other differentiated carbon prices depends on the relative convexities of the abatement cost functions, the welfare weights, and the relative marginal utilities at the optimal solution.

Equation (9) shows that the optimal uniform carbon price generally depends on the welfare weights. However, we may wonder if the optimal uniform carbon price does not depend on the welfare weights under certain conditions.

Proposition 1. *The optimal uniform carbon price does not depend on the welfare weights,* α_i*, if and only if* $\frac{D_S'}{D_N'} = \frac{C_N''}{C_S''}$ *holds at the optimal solution.*

[10] Note that I am here, for a moment, relaxing the assumption of strictly convex abatement cost functions.

[11] However, note that while the algebraic expression is the same as for the optimal differentiated carbon price, the values of the arguments, and thus the optimal carbon prices, are not. This is because the aggregate abatement would be different from the differentiated carbon price optimum since the optimal carbon price in both regions is given by this expression under the uniform carbon price solution.

Proof. See Appendix A.3.

Thus, while it is possible that the optimal uniform carbon price does not depend on the welfare weights, this is only the case if the ratio of marginal damages equals the inverse ratio of the convexities of the abatement cost function. Otherwise, the optimal uniform carbon price does depend on the welfare weights. Consequently, in the absence of international transfers, the choice of intratemporal welfare weights affects the optimal uniform carbon price. I will discuss this in more detail in the context of the comparison of Negishi and utilitarian weights in Section 3.2.3.

Having established expressions for the optimal carbon prices with arbitrary welfare weights, it is straightforward to obtain the results for specific choices of welfare weights. This is the focus of the next subsections.

3.2.2 The Negishi solution

I begin with Negishi weights, which are inversely proportional to a region's marginal utility of consumption at the optimal solution that was obtained with the Negishi weights[12]. I normalize the weights such that they sum to unity. Formally, I define the Negishi weights as follows:

$$\tilde{\alpha}_i = \frac{\frac{1}{u'(\tilde{X}_i)}}{\sum_{j \in \mathcal{I}} \frac{1}{u'(\tilde{X}_j)}}. \tag{10}$$

Note that I am using "tilde" to indicate the Negishi solution. From the assumptions of a strictly concave utility function and higher consumption levels in the Global North than the Gobal South, it follows that the welfare weight for the North is greater than for the South; i.e., $\alpha_N > \alpha_S$.

To obtain the Negishi solution, we plug the Negishi weights into the definition for the optimal differentiated carbon price.

Definition 3. *The* ***Negishi-weighted carbon price*** *is implicitly defined by*

$$\tilde{\tau} = C_i'(\tilde{A}_i) = -\sum_i D_i'(\tilde{A}). \tag{11}$$

The Negishi-weighted carbon price is simply equal to the sum of marginal benefits of abatement (which are the reduced marginal damages) in monetary terms. This condition is similar to the Samuelson condition for the optimal provision of public goods (Samuelson,

[12] Note that the Negishi weights that satisfy this are obtained by iteratively updating the weights until convergence.

1954), but it additionally requires that the marginal cost of abatement is equalized across regions.

Thus, we have obtained the knife-edge result that the Negishi-weighted carbon price is uniform even though we allowed for differentiated carbon prices by solving the differentiated carbon price optimization problem. The uniform carbon price arises from the specification of the Negishi weights, which equalize weighted marginal utilities across regions, making uniform carbon prices optimal, rather than from imposing a uniform carbon price constraint[13]. Notably, equalized weighted marginal utilities also render no transfers between regions optimal.

It is insightful to also characterize the optimality conditions in terms of the derivatives with respect to carbon prices. Rewriting the optimality condition in Equation (11), we can see that the Negishi-weighted carbon price equalizes the sum of the marginal abatement costs and benefits (in terms of reduced damages) from marginally increasing the carbon price. (see Appendix A.2.1 for a derivation):

$$\sum_i \frac{d\tilde{C}_i(\tilde{A}_i(\tilde{\tau}))}{d\tilde{\tau}} = -\sum_i \frac{dD_i(\tilde{A}(\tilde{\tau}))}{d\tilde{\tau}}. \tag{12}$$

3.2.3 The utilitarian solution with uniform carbon prices

Next, I define the utilitarian welfare weights as uniform weights that sum to unity; that is, $\alpha_i^U = \frac{1}{|\mathcal{I}|}$. To highlight that the maximization of the utilitarian SWF maximizes the (unweighted) sum of utilities, I also refer to the utilitarian solutions as *welfare-maximizing* solutions. Plugging the utilitarian weights into Equation (9) yields the utilitarian uniform carbon prices.

Definition 4. *The* ***utilitarian uniform carbon price*** *is implicitly defined by*

$$\check{\tau} = C_i'(\check{A}_i) = -\sum_i u'(\check{X}_i) D_i'(\check{A}) \frac{C_S''(\check{A}_S) + C_N''(\check{A}_N)}{u'(\check{X}_N) C_S''(\check{A}_S) + u'(\check{X}_S) C_N''(\check{A}_N)}. \tag{13}$$

Note that the utilitarian uniform carbon price is a function of the sum of the avoided marginal damages in welfare terms rather than monetary terms, as it is the case for the Negishi-weighted carbon price. Moreover, it depends on a second factor which contains the second derivatives of the abatement cost functions, which govern the abatement changes in response to a marginal change in carbon prices. Specifically, the change in abatement

[13] Note that we would obtain the same result if we plug the Negishi weights into the expression for the optimal uniform carbon price, but the key point is that the uniform carbon price constraint is not needed to obtain uniform carbon prices if Negishi weights are used.

in response to a marginal change in the carbon price is given by the inverse of the second derivative of the abatement cost functions $\frac{dA_i(\tau_i)}{d\tau_i} = \frac{1}{C_i''(A_i(\tau_i))}$. Thus, marginally increasing the carbon price increases abatement more in the region with the flatter marginal abatement cost curve.

As before, it is instructive to rewrite the optimality condition in Equation (9) in terms of the derivatives with respect to the carbon price (see Appendix A.2.2:

$$\sum_i u'(\check{X}_i)\frac{dC_i(\check{A}_i(\check{\tau}))}{d\check{\tau}} = -\sum_i u'(\check{X}_i)\frac{dD_i(\check{A}(\check{\tau}))}{d\check{\tau}}. \tag{14}$$

The utilitarian uniform carbon price equalizes the sum of the marginal *welfare* costs and benefits of abatement from marginally increasing the carbon price. This can be contrasted with the Negishi-weighted carbon price, which equalizes the sum of the marginal *monetary* costs and benefits of abatement from marginally increasing the carbon price (Equation (12)).

By construction, the utilitarian uniform carbon price is the uniform carbon price that maximizes global (utilitarian) welfare, while the Negishi-weighted carbon price maximizes global consumption in monetary terms. The central question is how these two uniform carbon prices compare. The following proposition and corollary establish the conditions under which the utilitarian uniform carbon price is greater than the Negishi-weighted uniform carbon price.

Proposition 2. *The utilitarian uniform carbon price is greater than the Negishi-weighted carbon price, that is $\check{\tau} > \tilde{\tau}$, if and only if $\frac{\check{D}_S'}{\check{D}_N'} > \frac{\check{C}_N''}{\check{C}_S''}$ holds at the utilitarian solution*[14].

Proof. See Appendix A.4.

Corollary 1. *The utilitarian uniform carbon price is greater than the Negishi-weighted carbon price, that is $\check{\tau} > \tilde{\tau}$, if and only if $\frac{-\frac{d\check{D}_S}{d\check{\tau}}}{\frac{d\check{C}_S}{d\check{\tau}}} > 1 > \frac{-\frac{d\check{D}_N}{d\check{\tau}}}{\frac{d\check{C}_N}{d\check{\tau}}}$ holds at the utilitarian solution.*

Proof. See Appendix A.5.

Proposition 2 states that, in the absence of international transfers, the uniform carbon price that maximizes global welfare is greater than the Negishi-weighted carbon price if and only if the ratio of marginal damages is greater than the inverse ratio of the second derivatives of the abatement cost functions. Intuitively, this depends on the relative benefits and costs of increasing the carbon price to the two regions. The left-hand side, $\frac{\check{D}_S'}{\check{D}_N'}$, is the

[14] Note that I use the notation $\check{D}_i'$ as a short-hand for $D_i'(\check{A}_i)$ (i.e., the marginal damage function evaluated at the utilitarian uniform carbon price solution). This general notational system also applies to other functions and solutions.

relative benefit of an extra unit of aggregate abatement A. The right-hand side, $\frac{\check{C}''_N}{\check{C}''_S}$ is the relative cost of an extra unit of aggregate abatement. Since the marginal abatement cost (MAC) is equal across regions, the relative cost of an extra unit of aggregate abatement is determined by the relative fractions of that unit of aggregate abatement that are provided by each region, which in turn is determined by the ratio of the inverse of the slopes of the MAC function. To see this, notice that

$$\frac{\check{C}''_N}{\check{C}''_S} = \frac{\frac{d\check{A}_S}{d\check{\tau}}}{\frac{d\check{A}_N}{d\check{\tau}}} = \frac{\frac{d\check{A}_S}{d\check{A}}}{\frac{d\check{A}_N}{d\check{A}}} = \frac{\frac{d\check{C}_S}{d\check{A}}}{\frac{d\check{C}_N}{d\check{A}}}, \tag{15}$$

where $\frac{dA_i}{d\tau_i} = \frac{1}{C''_i}$[15], and the third equality follows from $\frac{d\check{C}_S}{d\check{A}_S} = \frac{d\check{C}_N}{d\check{A}_N}$. Thus, a relatively greater slope of the MAC function results in a relatively smaller abatement increase, and therefore a relatively smaller increase in abatement costs.

Proposition 2 thus establishes that the welfare-maximizing uniform carbon is greater than the Negishi-weighted carbon price if and only if the benefits to the Global South relative to the Global North, exceed the relative costs of increasing abatement.

Corollary 1 provides an additional piece to understand the condition under which the utilitarian uniform carbon price exceeds the Negishi-weighted carbon price. It states that this is the case if and only if, at the utilitarian uniform carbon price, the ratio of the marginal benefits of abatement to the marginal costs of abatement from marginally increasing the carbon price is greater than one for the South and less than one for the North. Intuitively, this implies that the South would benefit from further increasing the carbon price while the North would be worse off. The corollary shows that this is necessary and sufficient for the utilitarian uniform carbon price to be greater than the Negishi-weighted carbon price.

Since I use the RICE model for numerical simulations, it is useful to briefly examine its functional forms of the damage and abatement cost functions. They are defined as follows:

$$C_{it} = \frac{b_{it} A_{it}^{\theta}}{\theta(\sigma_{it} Y_{it})^{\theta-1}}, \tag{16}$$

$$D_{it} = Y_{it}\mathcal{D}_{it}. \tag{17}$$

Here, Y_{it} is the GDP gross of damages and abatement costs for region i in period t, b_{it} is the price of a backstop technology (i.e., the MAC at which emissions can be abated completely), σ_{it} is the baseline emissions intensity (emissions per GDP) of the economy in the absence of abatement, $\theta > 1$ is a parameter that governs the convexity of the abatement cost function

[15] To see this, note that $C''_i = \frac{dC'_i}{dA_i}$ which can be rearranged to $\frac{dA_i}{dC'_i} = \frac{1}{C''_i}$.

(in RICE, $\theta = 2.8$), and $\mathcal{D}_{it}$ denotes the climate damage as a fraction of GDP.

Using these functional forms and the uniform carbon price condition, we can write $\frac{\check{D}'_S}{\check{D}'_N} > \frac{\check{C}''_N}{\check{C}''_S}$ as follows[16]:

$$\frac{\check{\mathcal{D}}'_S}{\check{\mathcal{D}}'_N} > \left(\frac{b_N}{b_S}\right)^{\frac{1}{\theta-1}} \frac{\sigma_S}{\sigma_N}. \tag{18}$$

The inequality is more likely to hold if the Global South has relatively (1) higher marginal damages as a fraction of GDP, (2) a higher backstop technology price, and (3) a lower baseline emissions intensity[17]. In the RICE model, the baseline emissions intensity tends to be higher in countries belonging to the Global South (with the important exception of the poorest region, Africa), but there is no clear pattern for backstop technology prices (see Figure A2 in the appendix). Importantly, however, climate damages, as well as marginal damages, tend to be higher in the RICE model in poorer countries of the Global South (with the exception of Eurasia) than in rich countries of the Global North (see Figure A3). This is in line with the consensus of the climate impacts literature that poor countries generally are, and are likely continue to be, disproportionately harmed by climate change (Ahmed et al., 2009; Burke et al., 2015; Diffenbaugh and Burke, 2019; Hallegatte et al., 2014; Kalkuhl and Wenz, 2020; Mendelsohn et al., 2006; Oppenheimer et al., 2014). If this effect dominates, then welfare-maximizing uniform carbon prices exceed the Negishi-weighted carbon prices. Indeed, this is the case in the numerical results (see Section 4.2). Proposition 2 provides the theoretical underpinnings for this result. The main intuition is that Negishi weights tend to down-weight the welfare of countries with are hit the hardest by climate change, resulting in lower carbon prices in the Negishi solution than the welfare-maximizing solution.

3.2.4 The utilitarian solution with differentiated carbon prices

Plugging the utilitarian weights into Equation (7) yields the utilitarian differentiated carbon prices, which I record in the following definition:

Definition 5. *The **utilitarian differentiated carbon price** for region i is implicitly*

[16] To see this, derive

$$A_i = \left(\frac{C'_i(\sigma_i Y_i)^{\theta-1}}{b_i}\right)^{\frac{1}{\theta-1}},$$

plug it into C''_i and use the uniform carbon price condition.

[17] The effect of θ depends on whether $\frac{b_N}{b_S} \lesseqgtr 1$. For $\frac{b_N}{b_S} > 1$, condition (18) is more likely to be satisfied if θ is large; and vice versa for $\frac{b_N}{b_S} < 1$.

defined by

$$\hat{\tau}_i = C_i'(\hat{A}_i) = -\frac{1}{u'(\hat{X}_i)} \sum_{j \in \mathcal{I}} u'(\hat{X}_j) D_j'(\hat{A}). \tag{19}$$

The utilitarian differentiated carbon prices equalize the marginal *welfare* cost of abatement[18] (as opposed to the marginal *monetary* cost of abatement in the Negishi solution), which, in turn, is equal to the marginal welfare benefit of abatement (the summation term in Equation (19)). This can be interpreted as a form of equal burden sharing, a common concept in international climate negotiations and the related literature (e.g., Bretschger (2013) and Rao (2014)); specifically, it would equalize the marginal welfare burden of abatement.

Equation (19) also shows that the welfare-maximizing differentiated carbon price is higher in the richer region, as it is inversely proportional to the marginal utility of consumption (Chichilnisky and Heal, 1994). Of course, this means that emissions are not reduced at the lowest *monetary* cost, (i.e., emission reductions are not cost-effective). Importantly, however, by equalizing the marginal *welfare* cost of abatement, utilitarian differentiated carbon prices achieve emission reductions at the lowest possible *welfare* cost (in the absence of transfers). Thus, I propose to classify these emission reductions as *welfare-cost-effective*, contrasting it with the concept of (monetary) cost-effectiveness. The concept of welfare-cost-effectiveness may also offer a useful perspective in other public policy contexts[19].

A second important point is that the utilitarian differentiated carbon prices are Pareto efficient if international transfers cannot be made[20]. This point requires elaboration. It is well known that cost-effective emission reductions are necessary to achieve Pareto efficiency if unrestricted lump-sum transfers can be made (Shiell, 2003). However, this is no longer the case when transfers are not feasible. In such a constrained, second-best setting, the set of feasible allocations becomes smaller and the Pareto frontier moves inward (except for one point on the frontier, which corresponds to the Negishi solution, which does not require transfers). If transfers cannot be made, the only way to move from one Pareto efficient allocation to another is through changing the differentiation of carbon prices. In

[18] To see this, rearrange Equation (19) to

$$u'(\hat{X}_i) C_i'(\hat{A}_i) = -\sum_{j \in \mathcal{I}} u'(\hat{X}_j) D_j'(\hat{A}),$$

and notice that the right-hand side is the same for both regions. Thus, $u'(\hat{X}_N) C_N'(\hat{A}_N) = u'(\hat{X}_S) C_S'(\hat{A}_S)$.

[19] It seems especially useful in contexts in which transfers by other means are not feasible.

[20] Sometimes the notion of *constrained* Pareto efficiency is used to refer to Pareto efficiency in settings with additional constraints (beyond the usual resource and technology constraints), particularly constraints on lump-sum transfers (Chichilnisky et al., 2000; Shiell, 2003). Instead, I opt to be explicit about the setting, and the corresponding constraints, which determine the Pareto frontier.

fact, all points on the Pareto frontier require differentiated carbon prices, except for one point, which corresponds to the Negishi solution (see Section 3.2.1 and Equation (7)). The utilitarian differentiated carbon price yields a particular point on the Pareto frontier which maximizes (unweighted) global welfare.

As before, we ask whether the utilitarian differentiated carbon price solution results in more or less global emissions than the Negishi solution. We have already established that the carbon price in the North is greater than in the South under the utilitarian differentiated carbon price solution. We may also intuit that the abatement in the South (North) is lower (higher) in the utilitarian differentiated carbon price solution than in the Negishi solution. I show that this intuition is correct in the proof of Lemma 1 below.

Lemma 1. *South's (North's) carbon price under the utilitarian differentiated carbon price solution is less (greater) than the Negishi-weighted carbon price. That is, $\hat{\tau}_S < \tilde{\tau} < \hat{\tau}_N$. Consequently, South's (North's) abatement level is lower (higher) in the utilitarian differentiated carbon price solution than in the Negishi solution; that is $\hat{A}_S < \tilde{A}_S$ and $\hat{A}_N > \tilde{A}_N$.*

Proof. See Appendix A.6. □

Therefore, whether global abatement is higher or lower in the utilitarian differentiated carbon price solution than in the Negishi solution depends on whether the additional abatement in the North outweighs the reduced abatement in the South. Proposition 3 establishes the condition under which this is the case. Deriving this result requires additional functional form assumptions on the abatement cost function (since the proof involves inverting the abatement cost function). To obtain clean results, while maintaining the assumption of strict convexity, I assume that the abatement cost function is quadratic. Specifically, $C_i(A_i) = k_i A_i^2$, where k_i is a region-specific constant that depends on regional characteristics. To give an idea about what affects this constant, the characteristics that determine k_i in the RICE model are the size of the economy, the baseline emissions intensity, the price of a backstop technology, and the parameter that determines the convexity of the abatement cost function (see Equation 16).

Proposition 3. *The aggregate abatement under the utilitarian differentiated carbon prices is greater than under the Negishi-weighted carbon price, that is $\hat{A} > \tilde{A}$, if and only if $\frac{\hat{u}'_S}{\hat{u}'_N}\frac{\hat{D}'_S}{\hat{D}'_N} > \frac{C''_N}{C''_S}$ holds at the utilitarian solution.*

Proof. See Appendix A.7. □

The first thing to notice is the similarity of this condition with the corresponding condition for the comparison between the utilitarian uniform carbon price and the Negishi solution detailed in Proposition 2. The aggregate abatement is again more likely to be higher under the utilitarian solution if the South has relatively high marginal damages and a steep marginal abatement cost curve, compared to the North.

However, there is an additional term in the condition of Proposition 3; the ratio of marginal utilities of consumption, $\frac{\hat{u}'_S}{\hat{u}'_N}$. Thus, the marginal damages in the two regions are weighted by their respective marginal utilities, reflecting marginal damages in welfare terms (as opposed to monetary terms). For a poorer South, $\hat{u}'_S > \hat{u}'_N$ and hence $\frac{\hat{u}'_S}{\hat{u}'_N} > 1$, which can also be interpreted as the factor that up-weights monetary damages in the South to reflect the greater welfare loss of a given damage in dollar terms in the South compared to the richer North. The important implication is that the aggregate abatement in the utilitarian differentiated carbon price solution is more likely to be greater than in the Negishi solution if the inequality in consumption is large.

The attentive reader may wonder why the marginal utilities only appear on the left-hand side of the inequality (representing the relative benefits of abatement), but not on the right-hand side (concerning the costs of abatement). The intuition for this is as follows. The difference in marginal utilities is already accounted for in the region-specific carbon prices which equalize the marginal welfare costs of abatement (i.e., $\hat{u}'_N \hat{C}'_N = \hat{u}'_S \hat{C}'_S$). Consequently, the carbon price in the poorer region is lower because of its higher marginal utility. The term on the right-hand side, $\frac{C''_N}{C''_S}$, simply determines how much the abatement decreases in the South and increases in the North (relative to the Negishi solution). A steep marginal abatement cost curve in the South means that abatement in the South does not decrease greatly under its lower carbon price of the utilitarian differentiated carbon price solution. Conversely, a flat marginal abatement cost curve in the North means that abatement in the North increases substantially under its higher carbon price. Thus, a relatively steeper marginal abatement cost in the South and a flatter one in the North make it more likely for the aggregate abatement to increase." It is also worth noting the subtle, but important, difference in intuition behind the $\frac{C''_N}{C''_S}$ term in Propositions 2 and 3. In Proposition 2, this term reflects the relative abatement cost increases to the two regions as a result of a marginal increase in a uniform carbon price. In contrast, in Proposition 3, it reflects how much abatement in the South decreases and how much it increases in the North as we allow for differentiated carbon prices.

3.2.5 Regions' preferred uniform carbon prices

To obtain additional insights into how heterogeneous climate policy preferences affect the optimal carbon prices under different SWFs, I derive regions' preferred globally uniform carbon prices. In doing so, I establish connections to Weitzman (2014) and Kotchen (2018), who introduced the notions of preferred uniform carbon prices and the preferred social cost of carbon, respectively.

The preferred uniform carbon price of a region is simply obtained by maximizing a social welfare function that puts full weight on that region and zero weight on the other region. Thus, to determine the preferred uniform carbon price of region i, I use the welfare weights $\alpha_i = 1$ and $\alpha_{-i} = 0$. Plugging these weights into Equation (9) yields regions' preferred uniform carbon prices.

Definition 6. *The* ***preferred uniform carbon price of region*** i *is implicitly defined by*

$$\mathring{\tau}^i = C'_{-i}(\mathring{A}^i_{-i}) = C'_i(\mathring{A}^i_i) = -D'_i(\mathring{A}^i)\frac{C''_i(\mathring{A}^i_i) + C''_{-i}(\mathring{A}^i_{-i})}{C''_{-i}(\mathring{A}^i_{-i})}, \tag{20}$$

where the superscript i indicates that the functions are evaluated at the solution under the preferred uniform carbon price of region i (for example, $\mathring{A}^N_S$ is the abatement in the South under the preferred uniform carbon price of the North)[21]. Equation (20) reveals that a region's preferred uniform carbon price depends on its marginal benefit of abatement, $-D'_i$ and the relative convexities of the abatement cost functions of the two regions. Unsurprisingly, a region's preferred uniform carbon price is higher if its marginal benefit of abatement is greater (i.e., the region is vulnerable to climate change). The role of the convexities of the abatement cost functions warrants further discussion.

The first aspect to notice is that, for strictly convex abatement cost functions, a region's preferred uniform carbon price is greater than its own marginal benefit of abatement, $-D'_i$. This is a key difference to the Nash equilibrium (with voluntary abatement provision), in which a region's optimal abatement level equalizes its own marginal costs and benefits of abatement (i.e., $C'_i = -D'_i$). The reason for this result is intuitive. Raising a uniform carbon price globally does not only result in additional abatement in one's own region, but also everywhere else. When choosing its preferred globally uniform carbon price, region i accounts for this effect of additional abatement in the other region. This is represented in the ratio term in Equation (20), indicating by which factor a region's uniform carbon price exceeds its own marginal benefit of abatement[22]. Using $A'_i(\tau_i) \equiv \frac{dA_i(\tau_i)}{d\tau_i} = \frac{1}{C''_i(A_i(\tau_i))}$, it is insightful to rewrite this term as follows:

$$\frac{C''_i(\mathring{A}^i_i) + C''_{-i}(\mathring{A}^i_{-i})}{C''_{-i}(\mathring{A}^i_{-i})} = 1 + \frac{A'_{-i}(\mathring{\tau}^i)}{A'_i(\mathring{\tau}^i)} > 1. \tag{21}$$

[21] For clarity, and to highlight that marginal abatement costs are equalized across regions under the preferred uniform carbon prices, I have included the equality $C'_{-i}(\mathring{A}^i_{-i}) = C'_i(\mathring{A}^i_i)$ in the definition.

[22] Weitzman (2014) thus refers to such a factor as the *externality-internalizing multiplier*.

For strictly convex abatement cost functions, this term is larger than one. Moreover, it is greater if the other region's change in abatement to a change in the uniform carbon price is relatively greater than region i's change in abatement. This is the case if region i's marginal abatement cost curve is relatively steeper. Thus, the convexity of the abatement cost functions plays a crucial role for regions' preferred uniform carbon prices, a fact that has been underappreciated in the existing literature[23]. To summarize, a region's preferred uniform carbon price is higher when its marginal benefit of abatement is large and when its abatement cost function exhibits greater convexity compared to the other region. Put simply, this is the case if a region is particularly vulnerable to climate change and if the cost burden of raising a uniform carbon price falls predominantly on the other region.

It is again instructive to rewrite the optimality condition in Equation (20) in terms of the derivatives with respect to the uniform carbon price (see Appendix A.2.3):

$$\frac{dC_i(\mathring{A}_i(\mathring{\tau}^i))}{d\mathring{\tau}^i} = -\frac{dD_i(\mathring{A}(\mathring{\tau}^i))}{d\mathring{\tau}^i}. \tag{22}$$

Intuitively, the preferred uniform carbon price of region i equalizes the cost and benefits to region i from marginally increasing the uniform carbon price.

Next, we ask how the preferred uniform carbon prices relate to the optimal uniform carbon prices under the utilitarian solution and the Negishi solution.

I start by establishing the following lemma, which helps to build intuition and acts as a building block towards proving the proposition that follows.

Lemma 2. *The utilitarian uniform carbon price ($\check{\tau}$) and the Negishi-weighted carbon price ($\tilde{\tau}$) are in between the preferred uniform carbon prices of the Global North ($\mathring{\tau}^N$) and the Global South ($\mathring{\tau}^S$), unless they all coincide.*

Proof. See Appendix A.8.

The intuition for the result of Lemma 2 is as follows. Regions' preferred uniform carbon prices are obtained by using "edge weights" in the SWF, giving full weight to one region and zero weight to the other. The utilitarian weights and the Negishi weights are linear combinations of these edge weights, giving a positive weight to both regions. Given the assumptions on the damage and abatement cost functions, it should thus not be surprising that maximizing a SWF with "edge weights" results in the highest and lowest uniform

[23] Importantly, however, Weitzman (2017a) allows for different convexities of the abatement cost function across regions, but he does not highlight the role of the convexity of the abatement cost function. Weitzman (2014) and Weitzman (2017b) and Kotchen (2018) assume uniform convexities of the abatement cost function across regions.

carbon prices, while using "more balanced" welfare weights results in uniform carbon prices in between those two extremes.

Using Lemma 2, I establish the following insightful relationship between regions' preferred uniform carbon prices and the main result detailed in Proposition 2.

Proposition 4. *The utilitarian uniform carbon price is greater than the Negishi-weighted carbon price, that is $\check{\tau} > \tilde{\tau}$, if and only if the preferred uniform carbon price of the Global South is greater than the preferred uniform carbon price of the Global North, that is $\mathring{\tau}^S > \mathring{\tau}^N$.*

Proof. See Appendix A.9.

The intuition for the result of Proposition 4 builds on the logic behind Lemma 2. Giving a positive weight to both regions, the utilitarian uniform carbon price and the Negishi-weighted carbon price can be understood as "weighted averages" of regions' preferred uniform carbon prices, where the welfare weights determine the relative weight given to the preferences of the two regions. Negishi weights are lower for the poor Global South than the rich Global North. In contrast, utilitarian weights place equal weight on the welfare of both regions. Thus, the Negishi-weighted SWF gives less weight to the preferences of the South than the utilitarian SWF. If the South prefers a higher uniform carbon price than the North, it is intuitive that the utilitarian uniform carbon price is greater than the Negishi-weighted carbon price, which downweights the preferences of the South. Roughly speaking, the Negishi solution is closer to the preferences of the North, while the utilitarian solution is closer to the preferences of the South (compared to the Negishi solution)[24]. This result provides perhaps the clearest intuition for the conditions under which the utilitarian uniform carbon price is higher or lower than the Negishi-weighted carbon price: it depends on whether South's preferred uniform carbon price is greater or lower than North's. This, in turn, depends on the functional forms of the damage and abatement cost functions, as shown in Equation (20) and discussed above.

4 Simulations

This section presents the results from simulations using the IAM RICE. Section 4.1 introduces the RICE model and describes the optimizations that I perform. Sections 4.2 and 4.3 discuss how optimal carbon prices are affected by the choice of welfare weights and international climate finance, respectively.

[24] Note that this does not necessarily imply that the utilitarian uniform carbon price is closer to the preferred uniform carbon price of the South than the preferred uniform carbon price of the North.

4.1 Method

4.1.1 Model

To provide simulation-based empirical evidence, I use the IAM Mimi-RICE-2010 (Anthoff et al., 2019), which is an implementation of the RICE-2010 model (Nordhaus, 2010) in the Julia programming language using the modular modeling framework Mimi. RICE is the regional variant of the Dynamic Integrated model of Climate and the Economy (DICE), disaggregating the world into 12 regions (see Figure A1 for a map showing the region classification) (Nordhaus and Sztorc, 2013). It is based on a neoclassical optimal growth model, known as the Ramsey model, which is linked to a simple climate model. Economic production is determined by a Cobb-Douglas production function and results in industrial CO_2 emissions. The relationship between economic production and emissions depends on the emissions intensity of an economy which can be reduced by investments in abatement. Emissions then translate to atmospheric CO_2 concentrations, radiative forcing, atmospheric and oceanic warming, and finally economic damages resulting from atmospheric temperature changes and sea-level rise.

4.1.2 Optimizations

Two main modifications were made to the Mimi-RICE-2010 model: (1) three different optimization problems were implemented along with numerical optimization algorithms to solve them, and (2) interregional transfers were incorporated. The modifications are described below. The final model that includes these modifications is referred to as *Mimi-RICE-plus*.

Optimization problems. The following three optimization problems are implemented:

1. Maximization of the discounted Negishi-weighted SWF with no constraints on the marginal abatement costs and the interregional transfers[25].

2. Maximization of the discounted utilitarian SWF with a constraint on the total level of interregional transfers, but with no constraint on the marginal abatement costs.

3. Maximization of the discounted utilitarian SWF with a constraint on the total level of interregional transfers, and an additional constraint of equalized marginal abatement costs across regions in each period.

[25] Note that regions are autarkic in the RICE model. Thus, the model implicitly contains a constraint of zero transfers. This is also the case in the optimization using the Negishi-weighted objective, even though in this case, zero transfers are also optimal under the Negishi-weighted SWF.

I refer to the solutions of these optimization problems as (1) the Negishi solution, (2) the utilitarian differentiated carbon price solution, and (3) the utilitarian uniform carbon price solution. In addition, I also compute regions' preferred uniform carbon prices by maximizing the respective regional SWFs (with welfare weights that equal unity for one region, and zero for all other regions) subject to a zero transfer constraint and a constraint of equalized marginal abatement costs across regions.

There are two sets of choice variables[26]: The emissions control rate (which implies carbon prices), and the allocation shares of the total international transfer quantity. Both are described in more detail below.

Social welfare functions. The first optimization problem is the maximization of the *discounted Negishi-weighted SWF*

$$\mathcal{W}^N = \sum_{t\in\mathfrak{T}} \sum_{i\in\mathfrak{I}} L_{it}\beta^t \alpha_{it} u\left(x_{it}\right) \tag{23}$$

where $\mathfrak{I}$ denotes the set of the 12 RICE regions, and $\mathfrak{T} = \{0, 1, 2, ..., 590\}$ is the time horizon of the RICE model[27], corresponding to the model years 2005 to 2595, L_{it} is the population, x_{it} is the per-capita consumption, β^t is the utility discount factor (given by $\beta^t = (1+\rho)^{-t}$, where ρ is the utility discount rate), and α_{it} are the time-variant Negishi welfare weights.

The utility function is given by

$$u\left(x_{it}\right) = \begin{cases} \ln\left(x_{it}\right) & \text{for } \eta = 1 \\ \frac{{x_{it}}^{1-\eta}}{1-\eta} + 1 & \text{for } \eta \neq 1 \end{cases} \tag{24}$$

where η is the elasticity of marginal utility of consumption, which is set to 1.5 to conform with the value used in the original RICE model.

The time-variant Negishi weights used in this study are obtained directly from the Mimi-RICE-2010 model (which, in turn, is based on the original RICE-2010 model (Nordhaus, 2010)). They are given by

$$\alpha_{it} = \frac{1}{u'\left(x_{it}\right)} v_t, \tag{25}$$

[26] Note that I do not optimize the saving rates because optimizing emission control rates and transfers in each period already results in long convergence times. Moreover, assuming fixed saving rates is relatively common in the climate economics literature (see Golosov et al. (2014), Dennig et al. (2015), and Budolfson et al. (2021) for more information). I use the saving rates from the base scenario of the original RICE model.

[27] For the clarity of this exposition, I am omitting the detail that one time period in RICE represents 10 years.

where v_t is the wealth-based component of the social discount factor [28]. In the RICE-2010 model, it is defined as the capital-weighted average of the regional wealth-based discount factors (see Nordhaus (2010) and Appendix C.1 for more details).

The second and third optimization problems maximize the (unweighted) *discounted utilitarian SWF*

$$\mathcal{W}^U = \sum_{t \in \mathfrak{T}} \sum_{i \in \mathfrak{I}} L_{it} \beta^t u\left(x_{it}\right). \tag{26}$$

Carbon prices. In optimization problems (1) and (2), carbon prices are allowed to be differentiated across regions. However, in the Negishi solution, uniform carbon prices are optimal by the construction of the Negishi weights (see Section 3.2.2). In the third optimization problem, a constraint of equal marginal abatement costs across regions is implemented. The marginal abatement costs are equalized by requiring the carbon price to be equal across regions until the region-specific backstop prices (i.e., the prices at which complete mitigation is achieved) are reached. The source code for the implementation of this constraint was adopted from the Mimi-NICE model (Dennig et al., 2017).

International transfers. First, I solve the three optimization problems with no international transfers to examine the role of welfare weights in the absence of transfers (Section 4.2). Second, I implement a conditional transfer for mitigation in recipient regions in the utilitarian optimization problems (Section 4.3). This allows me to study how welfare-maximizing carbon prices are affected by international transfers for mitigation, which is a main type of climate finance agreed on in international climate change negotiations (UNFCCC, 2015). Finally, I also consider non-conditional transfers that are not earmarked and could be interpreted as compensatory payments, for example for Loss and Damage (see Appendix C.3 for more information and results).

I implement the conditional transfer for mitigation as follows (for additional details, see Appendix C.2). The transfer is levied in the richest four regions of the RICE model (US, Other High Income countries, Japan, and EU), with each region contributing in proportion to its net output[29]. The total (potential) transfer quantity is set to \$100 billion per year in 2025 (in 2025 dollars) and increases over time with the aggregate net output in the donor regions[30]. While highly stylized, this implementation reflects the developed countries' goal

[28] For a model with a single representative agent, the wealth-based component of the social discount factor is approximated by $\frac{1}{1+\eta g}$, where g is the growth rate in per-capita consumption. Note that ηg is the wealth-based component of the social discount rate (SDR) in the Ramsey Rule, $SDR \approx \rho + \eta g$, reflecting the rationale for discounting future consumption if future generations are richer.

[29] Net output is the gross output/production minus climate damages.

[30] Note that the transfer potential might not be exhausted if not all of it is needed to fully abate emissions

of jointly mobilizing $100 billion per year by 2020, which was first agreed upon in 2009 at the Conference of the Parties (COP) 15 in Copenhagen (UNFCCC, 2009), and extended through 2025 at COP21 in Paris, after which a new collective goal shall be set of at least $100 billion per year (UNFCCC, 2015). I refer to this trajectory of the total transfer quantity as the *Paris Agreement transfer*.

The total transfer is then allocated optimally, in terms of maximizing the utilitarian SWF, toward abatement in the remaining eight regions. An important question is whether this internationally financed abatement (hereafter "foreign abatement") is additional to the domestic abatement that would have taken place in the absence of the transfer. In essence, this depends on the conditions that donor countries impose on the transfer provision, in particular, with respect to its additional effect on emission reductions. I consider both cases: the presence and absence of an "additionality" condition, and I refer to the foreign abatement as either *additional* or *non-additional*[31]. In the case of the former, I impose additional constraints on the optimization problem so that the domestic abatement costs cannot fall below their optimal level in the absence of the transfer. For the uniform carbon price solution, I set the constraint equal to either the domestic abatement costs of the uniform or the differentiated carbon price solution. I consider additionality relative to the differentiated carbon price solution as the main scenario, given that the differentiated carbon price solution, which equalizes the marginal welfare cost of abatement across regions, is most straightforwardly in accordance with the principle of "common but differentiated responsibilities" of the UNFCCC (Budolfson and Dennig, 2019). It may thus be considered closest to the actual political constraints imposed on the transfer provision. Conceptually, the constraint on the domestic abatement costs reduces the feasible set among which the social planner can choose, (weakly) reducing global welfare.

Optimization algorithms. The optimization problems are solved with the numerical optimization algorithm "`NLOPT_LN_SBPLX`" which is an implementation of the Subplex algorithm (Rowan, 1990) in the NLopt (nonlinear-optimization) package (Johnson, 2020). For the implementation of the transfer constraints, I use the augmented Lagrangian algorithm "`NLOPT_AUGLAG`", which is an implementation of the algorithm by Birgin and Martínez (2008). Some parts of the source code for the implementation of the optimization algorithms were adopted from the mimi-NICE model (Dennig et al., 2017) and the RICEupdate model (Dennig et al., 2019), which is based on Mimi-RICE-2010.

in recipient regions.

[31] Note that "non-additional" here merely means the absence of an additionality condition. It does not necessarily mean that the transfer for mitigation does not yield additional abatement.

4.2 The role of welfare weights

4.2.1 The effect on optimal climate policy

This section investigates how optimal carbon prices depend on the choice of welfare weights in the absence of international transfers. Moreover, I distinguish between two utilitarian solutions depending on whether carbon prices are constrained to be uniform.

It is useful to first examine the overall stringency of the optimal climate policy paths. To this end, Figure 3 shows the respective optimal atmospheric temperature trajectories for different optimization problems and different utility discount rates (also referred to as the pure rate of time preference in the literature); specifically, I compare the results for the utility discount rates used by Nordhaus (1.5%) and Stern (0.1%), respectively (Nordhaus, 2011; Stern et al., 2006)[32].

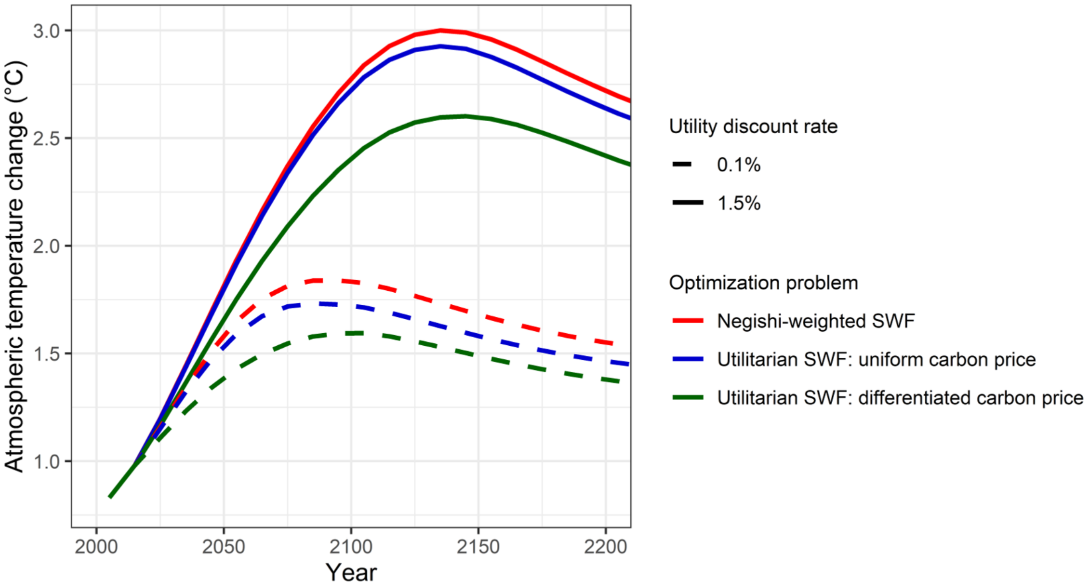

Figure 3: Optimal atmospheric temperature trajectories conditional on the optimization problem and the utility discount rate. The Negishi-weighted solutions (red) are compared to the solutions under the utilitarian objective with (green) and without (blue) the additional constraint of equalized regional carbon prices for the Nordhaus (solid lines) and Stern (dashed lines) utility discount rates (Nordhaus, 2011; Stern et al., 2006). Temperature changes are relative to 1900.

The utilitarian solutions yield lower optimal temperature trajectories than the Negishi

[32] Like Negishi weights, the utility discount rate also places different weights on the welfare of different people. However, it does so on the basis of time – giving lower weight to the welfare of future generations – rather than on the basis of the wealth (or, more precisely, the consumption level) of an individual. The issue of discounting future utilities is heavily debated among economists and has received much more attention than the use of Negishi weights.

solution for both utility discount rates. Allowing for differentiated carbon prices in the utilitarian optimization results in the lowest warming trajectories. Figure 3 also shows the well-known large sensitivity of optimal climate policy to the utility discount rate. Specifically, peak warming is 3.00°C (1.84°C) in the Negishi solution, 2.93°C (1.73°C) in the utilitarian solution with uniform carbon prices, and 2.60°C (1.59°C) in the utilitarian solution with differentiated carbon prices for the 1.5% (0.1%) utility discount rate.

The corresponding cumulative global industrial[33] carbon dioxide emissions for the entire model horizon from 2005-2595 are shown in Table 1. The effect of increased optimal abatement in the utilitarian solutions relative to the Negishi solution is larger for the lower utility discount rate, when the welfare impact of future damages is given comparatively greater weight. Specifically, relative to the Negishi solution, the optimal cumulative global industrial CO_2 emissions are around 5% (13%) lower for the utilitarian solution with the additional constraint of uniform carbon prices, and 21% (27%) lower for the utilitarian differentiated carbon price solution, using the 1.5% (0.1%) utility discount rate.

Table 1: Cumulative global industrial CO_2 emissions ($GtCO_2$) depending on the optimization problem and the utility discount rate.

Utility discount rate	Optimization problem: Negishi SWF	Optimization problem: Utilitarian SWF: Uniform carbon price	Optimization problem: Utilitarian SWF: Differentiated carbon price
1.5%	3,815	3,629	3,032
0.1%	1,373	1,199	1,005

The trajectories of the optimal carbon prices[34] and the corresponding industrial emissions for a utility discount rate of 1.5% are shown in Figure 4. The utilitarian solution with uniform carbon prices yields similar trajectories as the Negishi solution, albeit a slightly increased mitigation effort, with most regions reaching zero (industrial) carbon emissions in the first half of the 22nd century. The utilitarian solution with differentiated carbon prices presents vastly different results. The carbon price trajectories in rich regions (US, EU, Japan, and Other High Income countries) are much higher compared to the path of the optimal uniform carbon price, while the carbon prices in poor regions start low and rise relatively slowly. The four rich regions (and Russia) reach zero carbon emissions within this century, while carbon emissions in poorer regions continue throughout the 22nd century.

[33] There are two sources of emissions in RICE: endogenous region-level industrial emissions and exogenous emissions from land use change. Industrial emissions constitute the bulk of total emissions. The cumulative exogenous emissions from land use change over the entire model horizon from 2005-2595 are 29 Gt CO_2 globally.

[34] Note that all dollar values are 2022 USD. I convert the 2005 USD values of the RICE model to 2022

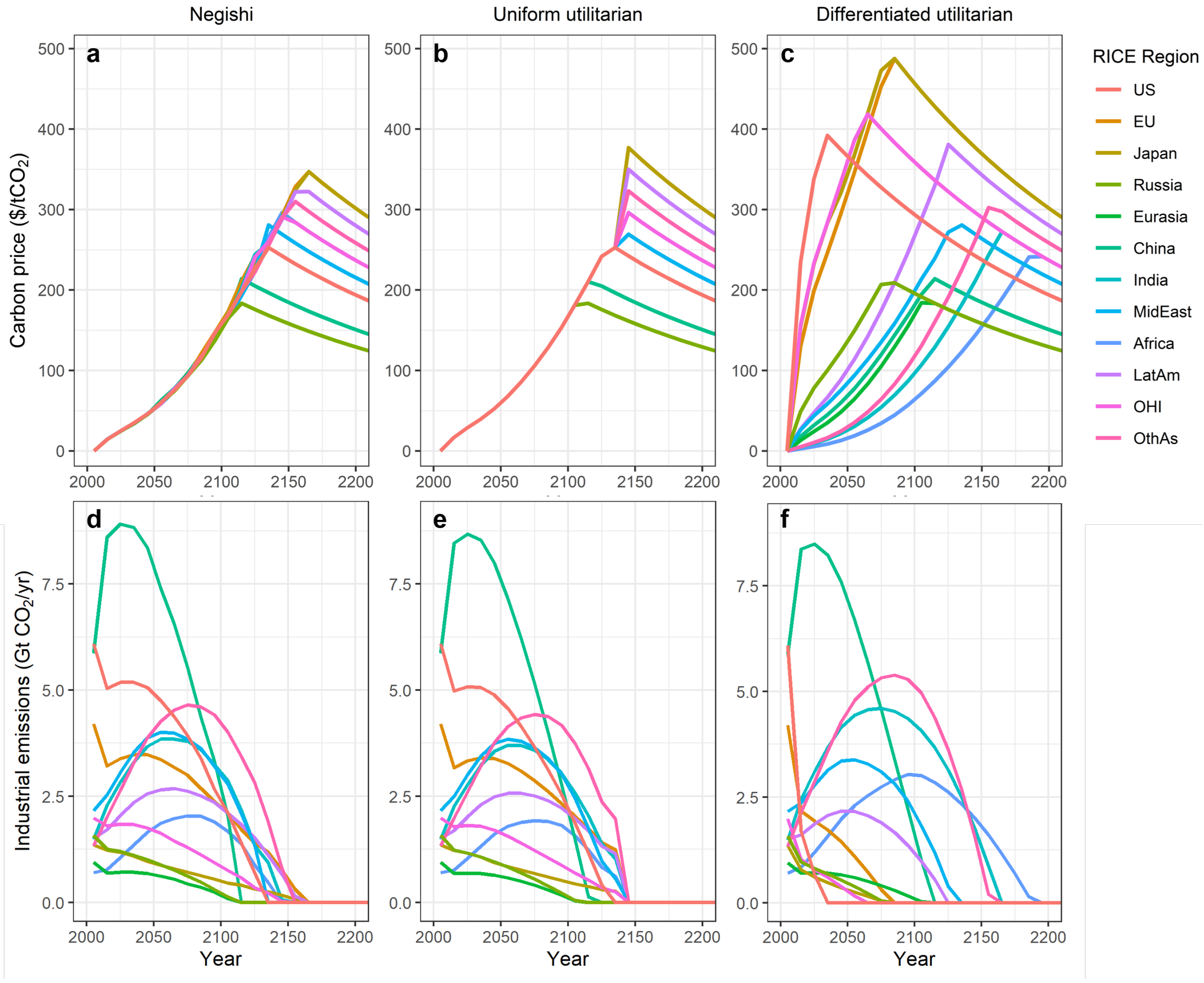

Figure 4: Optimal trajectories for carbon prices and industrial emissions conditional on the optimization problem. Results are for the utility discount rate of 1.5%. **a**, **b**, **c**, the optimal carbon price trajectories under the Negishi solution (a) and the utilitarian solution with (b) and without (c) the additional constraint of equalized carbon prices. **d**, **e**, **f**, the corresponding industrial emission trajectories. Note that the carbon price decreases once it reaches the region-specific backstop price. Also note that Mimi-RICE-plus only yields an approximately equalized carbon price for the Negishi solution.

Table 2: Optimal carbon price in 2025 (in 2022 $/tCO$_2$) depending on the optimization problem and the utility discount rate (ρ).

Optimization problem	Utility discount rate	
	$\rho = 1.5\%$	$\rho = 0.1\%$
A) *Negishi-weighted SWF*	25	100*
B) *Utilitarian SWF: uniform carbon price*	29	121
C) *Utilitarian SWF: differentiated carbon price*		
US	338	> 410
Other High Income	233	> 501
Japan	232	> 638
EU	199	> 638
Russia	78	> 273
Latin America	48	202
Middle East	44	182
China	32	134
Eurasia	24	103
Other Asia	10	44
India	10	41
Africa	5	23

Note: Mimi-RICE-plus only yields an approximately equalized carbon price for the Negishi solution. In this case (*), it varied between 98 and 102 $/tCO$_2$ across regions. The ">" sign indicates that the regional backstop price has been reached. Thus, any price above the backstop price is optimal as complete abatement is required.

The optimal carbon prices in 2025 are shown in Table 2 for both utility discount rates. As noted above, the uniform carbon price is somewhat higher (about 20%) for the utilitarian solution compared to the Negishi solution. More striking, however, are the large differences in optimal carbon prices across regions when the constraint of equalized carbon prices is not imposed. This yields very substantial carbon prices in rich regions – exceeding ~$200/tCO$_2$, even for the high utility discount rate – and much lower carbon prices in poor regions. For the lower Stern utility discount rate, the richest five regions have already reached their backstop price in 2025, yielding zero carbon emission.

Finally, regional cumulative industrial emissions depending on the optimization problem are shown in Figure 5 (for the high utility discount rate). The utilitarian solution with uniform carbon prices yields somewhat lower emissions than the Negishi solution. The utilitarian solution with differentiated carbon prices requires much higher abatement in most

USD values using the World Bank GDP deflator (World Bank, 2023b).

regions (but particularly in the richest regions) due to the overall increased stringency of climate policy. However, the cumulative carbon emissions of the poorest three regions (Africa, India, and Other Asia) increase relative to the Negishi optimum.

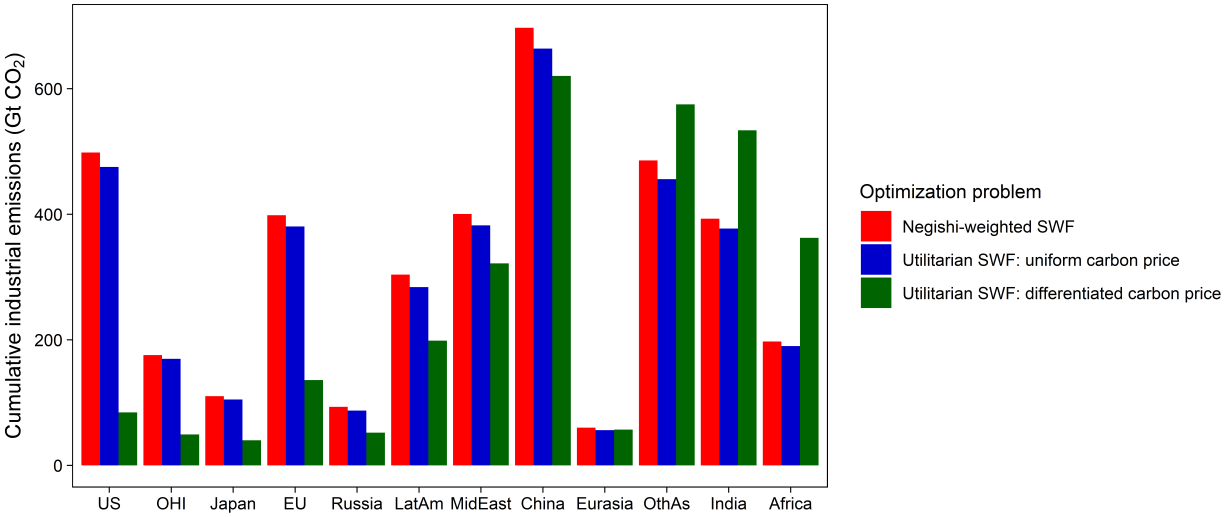

Figure 5: Optimal cumulative industrial emissions depending on the optimization problem. Note that these are the results for the utility discount rate of 1.5%.

The underlying reason for a more stringent climate policy under the utilitarian SWF is that, in contrast to the Negishi-weighted objective, the utilitarian objective reflects that a certain change in consumption has a greater welfare effect for a poor person than for a rich person. The utilitarian objective is thus more sensitive to the interests of the poor, whose welfare is harmed the most by a given climate damage[35] and who tend to bear a disproportionately large share of the damages.

Allowing for differentiated carbon prices under the utilitarian objective further increases the stringency of optimal climate policy relative to the utilitarian uniform carbon price optimum. In addition to its sensitivity to the welfare of the poor, this solution also reflects that the disutility of a unit of abatement cost is lower in regions with higher consumption levels. Welfare maximization thus requires higher marginal abatement costs (and implied carbon prices) in richer regions and lower marginal abatement costs in the poorest regions. This is the result that was theoretically demonstrated by Chichilnisky and Heal (1994) and in Section 3.2.4 of this paper. The additional abatement in rich regions outweighs the reduced abatement in the poorest regions, resulting in increased emission reductions overall. The differentiated carbon price optimum is discussed in more detail in Appendix C.4.

[35] This is the case even for a certain percentage damage relative to one's consumption if the elasticity of the marginal utility of consumption, η, is greater than 1. For most of the analysis of the present paper, and unless noted otherwise, η is set to 1.5, matching the value used in the original RICE model.

4.2.2 Regions' preferred uniform carbon prices

Section 3.2.3 derived the condition under which the utilitarian uniform carbon price exceeds the Negishi-weighted carbon price for a simple static two-region model. The previous section documented that this is the case for the RICE model. Moreover, from the stylized theoretical model in Section 3.2.5 we know that this is the case if and only if the poorer region prefers a higher uniform carbon price than the richer region. The aim of this section is to gain additional intuition for this result by examining regions' preferred uniform carbon prices in the RICE model. Moreover, regions' preferences regarding the level of a uniform carbon price are informative in their own right, for example, to understand which regions may be inclined to push for more or less stringent global climate policy in international negotiations[36].

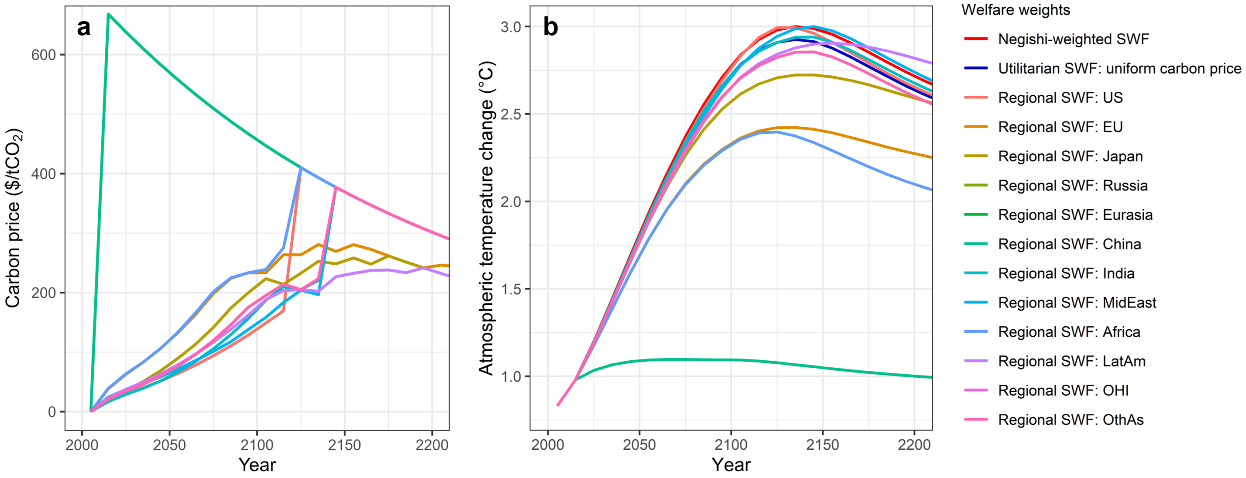

Figure 6: Region's preferred uniform carbon prices and corresponding temperature trajectories. Note that these are the results for the utility discount rate of 1.5%.

Figure 6 shows each region's preferred uniform carbon price trajectory as well as the resulting trajectory for the atmospheric temperature change. Three regions (Russia, Eurasia, China) prefer carbon prices that yield complete abatement in the first period[37]. The underlying reason is that these are the regions with the lowest backstop technology prices, which place an upper bound on the marginal abatement costs. The results indicate that the regions with the cheapest backstop technologies prefer to set the carbon price at a level that forces all regions to completely abate emissions, knowing that their own marginal abatement cost is bounded at a lower level than for other regions[38]. However, this result should perhaps

[36] However, it is important to note that within the framework of international negotiations under the Paris Agreement, which emphasizes nationally determined contributions, a globally uniform carbon price has not been the central focus.

[37] That is, the first period in which carbon prices are set in the model, which is assumed to be 2015. Carbon prices in 2005 are assumed to be zero.

[38] Indeed, these three regions have the highest preferred uniform carbon prices solely because of the effect

not be taken too seriously since it depends on the assumed large differences in backstop technology prices in the RICE model.

Importantly, however, the region with the next highest preferred uniform carbon prices is Africa, the poorest region in the RICE model, mainly due to its disproportionately large climate damages (see Figure A3). On the other side of the spectrum, the richest region in the model, the US, has among the lowest preferred uniform carbon prices until 2115. While the overall effect depends on all regions, this gives rise to the following intuition for higher carbon prices in the utilitarian solution with uniform carbon prices than in the Negishi solution. Negishi weights give a lower weight to the welfare in poor regions (like Africa), which tend to prefer higher carbon prices, and a higher weight to rich regions (like the US), which tend to prefer lower carbon prices. Consequently, the Negishi solution yields lower carbon prices than the utilitarian solution which places an equal weight on the welfare of all regions. Indeed, the large welfare gains from higher uniform carbon prices in Africa are the primary reason for higher carbon prices in the utilitarian uniform carbon price solution compared to the Negishi solution (see also Figure 8b).

4.2.3 Distributional effects

Different intratemporal welfare weights lead to different distributional outcomes. This section examines which regions are better-off, and which regions are worse-off, under the utilitarian solutions compared to the Negishi solution.

The higher carbon prices in the utilitarian uniform carbon price solution result in consumption losses from increased abatement costs in all regions until 2055, and consumption gains from lower climate damages later on (see Figure 7b). Yet, there is a large heterogeneity in consumption changes across regions. Notably, Africa has the smallest consumption losses (in percentage as well as absolute terms) in the first half of this century and starts to experience consumption gains after 2055. On the contrary, consumption losses in Russia, Eurasia and China are much larger and consumption gains start several decades later. In general, however, consumption changes are relatively small and do not exceed ±0.25% of the consumption level in the Negishi solution (for consumption changes in absolute terms, see Figure A4 in the appendix).

Consumption changes between the utilitarian differentiated carbon price solution and the Negishi solution are considerably larger (see Figure 7a). Moreover, the poorest four regions are now better-off in all periods, due to both lower abatement costs in their regions and

of setting the carbon price above their own backstop price. That these regions do not prefer higher uniform carbon prices in the absence of this effect can be seen in Figure 8, which shows that they incur the greatest relative present value consumption losses resulting from the higher carbon prices in the utilitarian solution compared to the Negishi solution.

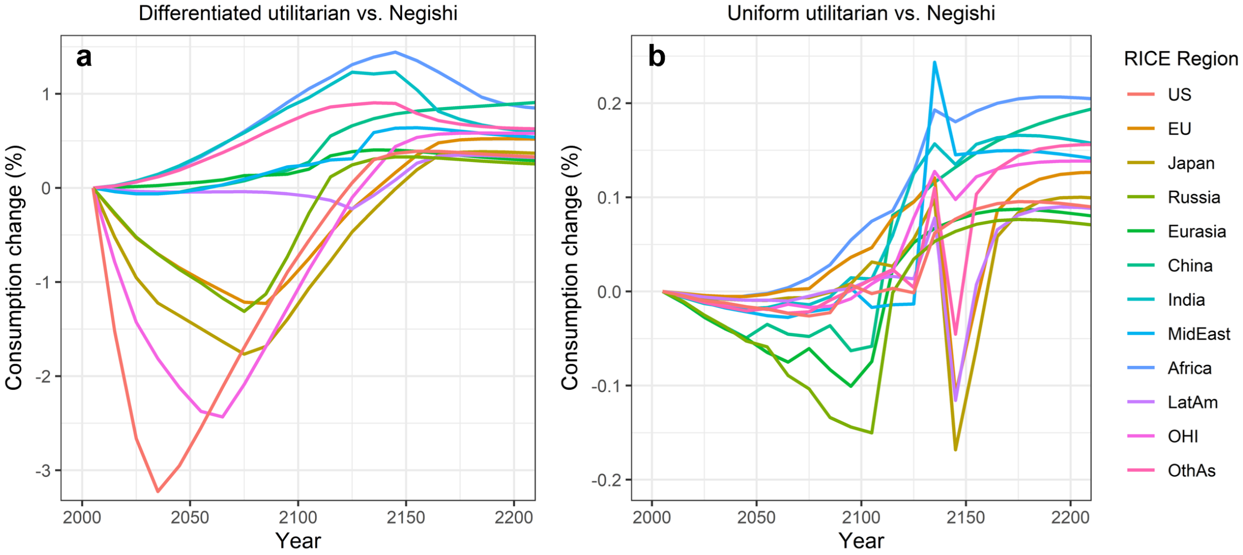

Figure 7: Relative regional consumption changes between the Negishi solution and the utilitarian solutions. Consumption changes are percentage changes relative to the consumption level in the Negishi solution. Positive values indicate a higher consumption level in the utilitarian solutions. Note that these are the results for the utility discount rate of 1.5%.

lower global emissions. Consumption in the richer regions is lower initially due to the higher carbon prices in these regions under the utilitarian solution with differentiated carbon prices. However, after 2150, all regions enjoy consumption gains. Thus, the increased abatement in rich regions does not lead to persistently lower consumption trajectories. In addition, it is worth noting that the consumption losses in rich regions do not imply negative consumption per capita growth rates. More generally, the consumption per capita trajectories of all regions are not heavily altered, especially compared to the magnitude of the inequality across regions. Thus, while the utilitarian solutions result in greater global welfare by accounting for the background inequality in setting the carbon prices, they do not solve the inequality issue. The consumption per capita trajectories for the regions with the largest positive and negative consumption changes, Africa and the US, respectively, are shown in Figure A5 in the appendix.

I also calculate the net present value (NPV) of consumption changes to understand how regions' aggregate intertemporal welfare changes (see Figure 8). More specifically, I compute the consumption changes in the initial period (2005) that would yield a welfare change (in utility terms) that is equivalent to the welfare difference between each of the utilitarian solutions and the Negishi solution. The details of this calculation are provided in Appendix A.10.

Unsurprisingly, the utilitarian differentiated carbon price solution results in NPV con-

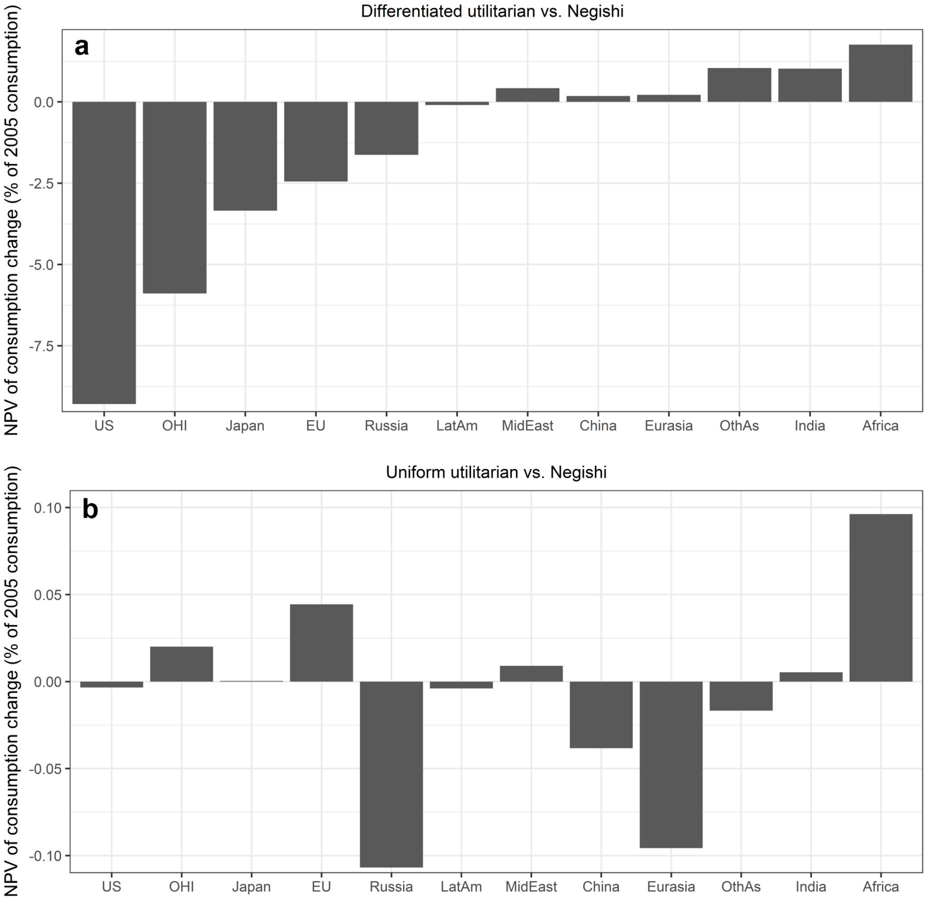

Figure 8: Net present value of consumption changes. The values show the welfare-equivalent consumption change in 2005, as a percentage of the consumption in 2005 (for details, see Appendix A.10). Note that these are the results for the utility discount rate of 1.5%.

sumption losses (relative to the Negishi solution) in rich regions, where carbon prices are high, and NPV consumption gains in poor regions, where carbon prices are low (see Figure 8a). The region that benefits the most from higher carbon prices in the utilitarian uniform carbon price solution, relative to the Negishi solution, is the poorest region, Africa (see Figure 8b). This also strengthens the intuition that it is primarily the down-weighting of Africa in the Negishi-weighted SWF that results in lower carbon prices in the Negishi solution. The relatively cold regions of Russia and Eurasia experience the greatest consumption losses.

4.3 The role of climate finance

4.3.1 The effect on optimal climate policy

This section evaluates how conditional transfers for mitigation affect welfare-maximizing (utilitarian) carbon prices. As described in Section 4.1.2, the total transfer quantity increases with the GDP of donor regions from a baseline of $100 billion per year in 2025.

I start again by examining how the transfer affects the overall stringency of the optimal climate policy paths. Figure 9 shows the optimal temperature trajectories in the presence of transfers that finance mitigation in recipient regions. Under both the uniform and the differentiated carbon price solutions, the availability of foreign-funded abatement considerably increases the welfare-maximizing climate policy stringency, resulting in lower optimal warming. In particular, foreign abatement reduces peak temperatures by around 0.17°C (0.18°C) under the utilitarian differentiated (uniform) carbon price solutions if the transfer is used for additional abatement (relative to the domestic abatement level of the differentiated carbon price solution without transfers). This corresponds to a 14% (12%) reduction in cumulative global industrial emissions (see Figure 10).

The differentiated carbon price solution with additional foreign abatement may be particularly relevant (from a normative perspective) as it may be considered closest to the welfare-maximizing climate policy conditional on plausible real-world constraints on transfers; namely, restricted general redistribution and an additionality condition on the provision of transfers for mitigation. It is thus especially interesting to compare it to the conventional Negishi solution. Cumulative global emissions are 31% lower under the differentiated carbon price solution with additional foreign abatement (see Figure 10), resulting in a reduction of peak warming by 0.57°C, from 3.00°C to 2.43°C (see Figure 9).

Transfers for mitigation also substantially affect carbon prices. Figure 11 shows the utilitarian differentiated carbon prices conditional on the transfer scenario. It is first worth noting that the marginal abatement costs of the foreign abatement are lower than the marginal abatement costs in donor regions. Thus, foreign abatement is inframarginal from the per-

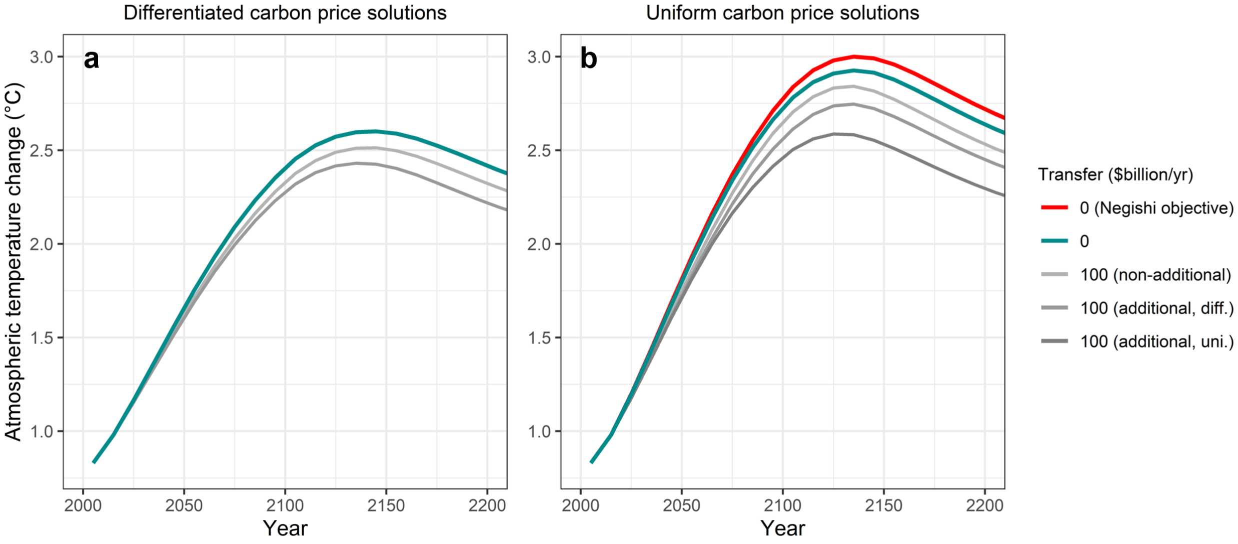

Figure 9: Optimal atmospheric temperature trajectories conditional on the optimization problem and the transfer scenario. The Negishi-weighted solution (red) is compared to the utilitarian solutions without (teal) and with foreign abatement (grey). The different shades of grey indicate the transfer scenario regarding the additionality condition of foreign abatement (*diff./uni.:* foreign abatement funding is additional to the domestic abatement spending in the utilitarian no-transfer differentiated/uniform carbon price solution. Temperature changes are relative to 1900. Note that these are the results for the utility discount rate of 1.5%.

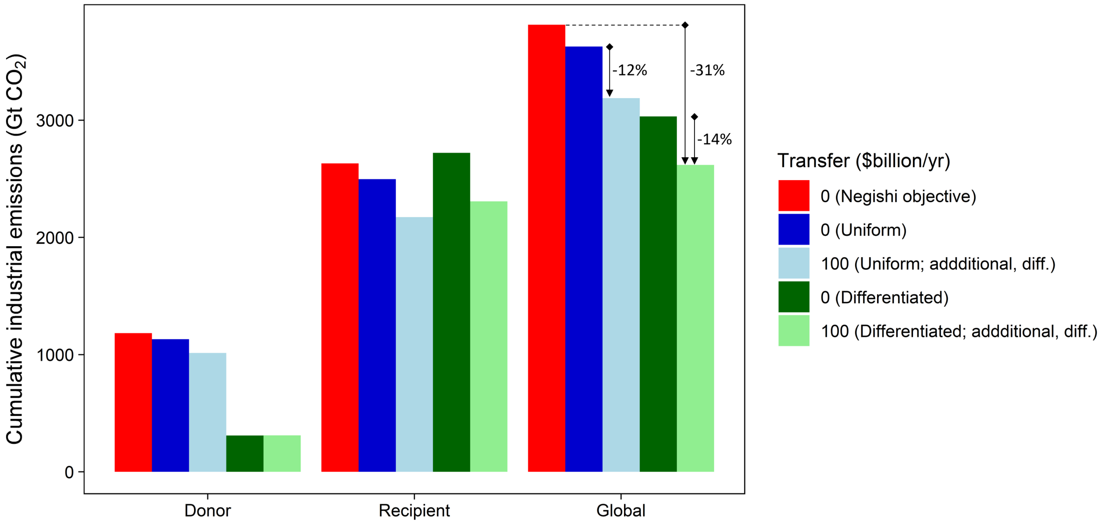

Figure 10: Optimal cumulative industrial emissions depending on the optimization problem and the transfer scenario. This figure shows the results for foreign abatement that is additional relative to the utilitarian no-transfer differentiated carbon price solutions. Note that these are the results for the utility discount rate of 1.5%.

spective of the donor regions; that is, international mitigation finance is used for relatively cheap abatement options in recipient regions. This also indicates that the constraint on the total level of mitigation finance is binding[39]. In other words, the social planner would prefer to relax this constraint and increase mitigation finance to enhance global welfare.

Moreover, two features of the optimal transfer allocation are worth highlighting: (1) transfers are allocated to the poorest regions first (since this reduces the welfare cost of abatement the most) and (2) transfers are allocated cost-effectively (i.e., lowest cost abatement options are pursued first). The second feature can be seen in Figure 11. Carbon prices are lifted to a certain level in all regions (compare Panels (b) and (c) with foreign abatement to Panel (a) without foreign abatement) [40]. For example, carbon prices in 2025 are lifted to at least $\$44/tCO_2$ in all regions in the main transfer scenario with additional foreign abatement; a substantial increase from as low as $\$5/tCO_2$ in Africa without foreign abatement. In the scenario with non-additional foreign abatement, the stepwise trajectory emerges from the social planner's decision on how to crowd out domestic abatement in recipient regions. The lower bound of marginal abatement cost increases as abatement in some of the recipient regions starts to increasingly be funded from domestic sources, releasing funds that are then redirected to other (poorer) recipient regions.

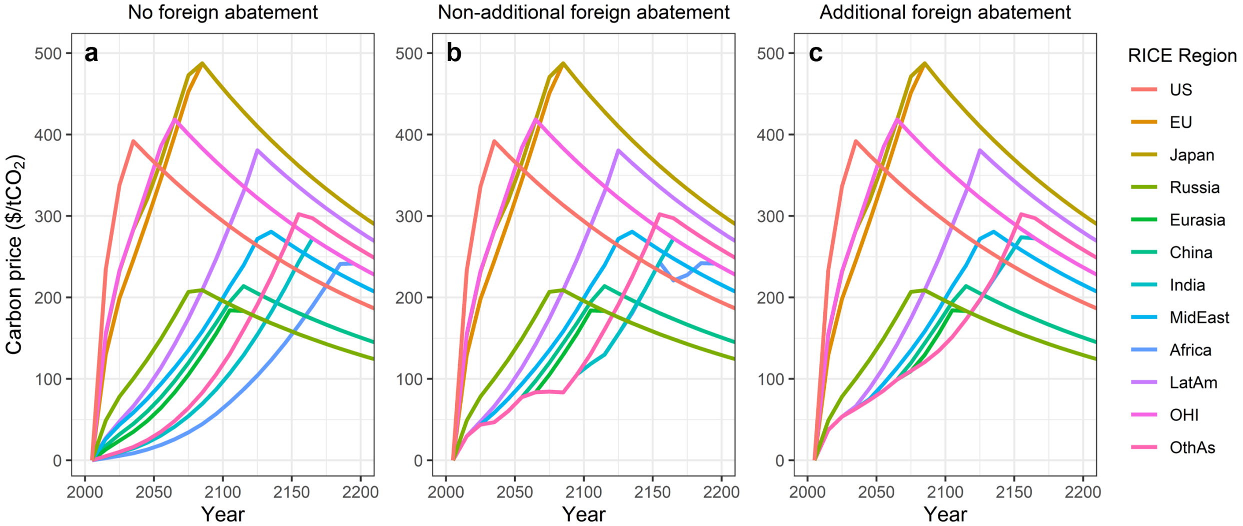

Figure 11: Optimal differentiated carbon price trajectories conditional on the optimization problem and the transfer scenario. The figure shows the utilitarian differentiated carbon prices under (a) no foreign abatement, (b) non-additional foreign abatement, (c) additional foreign abatement. Note that the carbon price decreases once it reaches the region-specific backstop price. Results are for the utility discount rate of 1.5%.

Foreign abatement also has a large effect on optimal uniform carbon prices (see Table

[39] Note that, in some scenarios, the constraint stops binding in the second half of the 22^{nd} century.

[40] Note, however, that this price level may be exceeded through domestically-funded abatement.

3). Importantly, the welfare-maximizing uniform carbon price in 2025 roughly doubles from \$29/tCO$_2$ without transfers to \$54/tCO$_2$ when the "Paris Agreement transfer" is used to finance additional abatement in developing countries; a price more than twice as high as the carbon price of the conventional Negishi solution of \$25/tCO$_2$. The important policy implication is that the availability of international climate finance for mitigation considerably increases the carbon prices that maximize global welfare.

Table 3: Optimal uniform carbon prices (\$/tCO$_2$) depending on the optimization problem and the transfer scenario.

SWF:		Negishi	Utilitarian	Utilitarian	Utilitarian	Utilitarian
Foreign abatement:		No	No	Yes	Yes	Yes
Additionality:		N/A	N/A	No	Differentiated	Uniform
Year:	2025	25	29	45	54	58
	2035	34	39	51	64	68
	2045	46	52	65	78	88
	2055	60	68	85	98	116
	2095	142	153	170	180	189

Note: The table shows the results for a utility discount rate of 1.5%. The third row ("Additionality") specifies whether the foreign abatement funding is required to be additional to the domestic abatement spending in the utilitarian no-transfer differentiated/uniform carbon price solution. The total transfer quantity, in the scenarios with transfers, is \$100 billion per year in 2025 and increases over time with the aggregate net output (GDP) of the donor regions.

4.3.2 Optimal transfer allocation

The optimal allocation of international mitigation finance is shown in Figure 12, for the case of additional foreign abatement relative to the differentiated carbon price solution without transfers. The pattern of the optimal allocation over time is similar under both the differentiated and uniform carbon price solutions. The region that receives most of the financial support is China in the next couple of decades, followed by Other Asia and India in the second half of this century. The poorest region in the model, Africa, requires the most support in the twenty-second century. The reason that China receives most of the transfer in the near-term is because of its large economy and associated abatement opportunities. The logic is that abatement in China absorbs most of the transfer as the marginal abatement costs are increased uniformly across recipient regions to allocate mitigation finance cost-effectively. The richest recipient region, Russia, is the only region that does not receive any foreign funding.

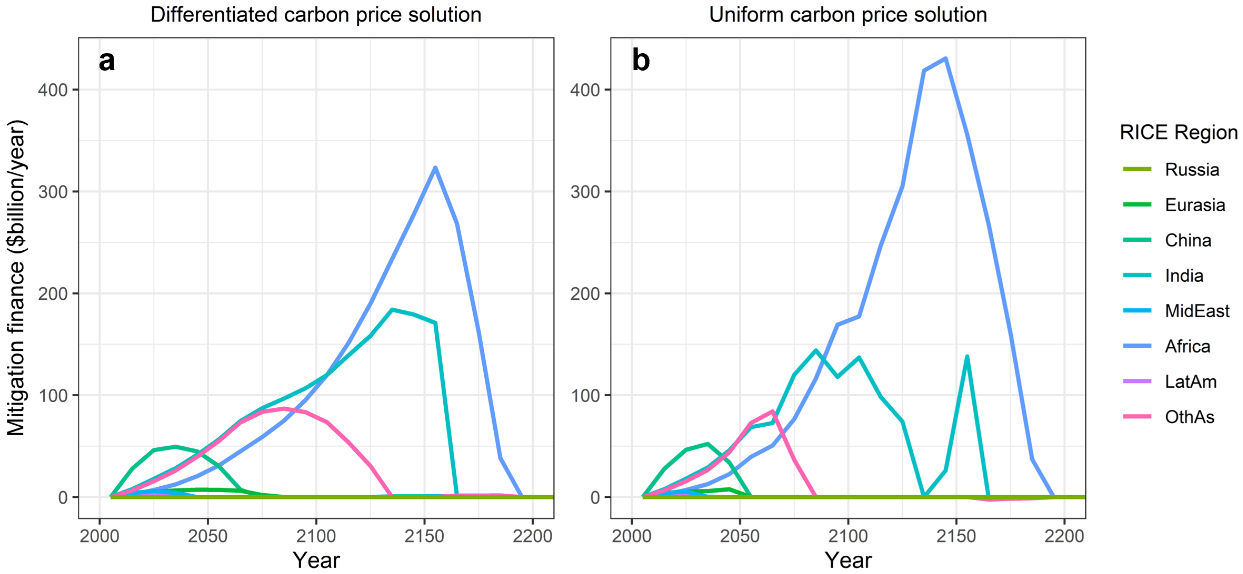

Figure 12: Optimal allocation of international mitigation finance over time. The optimal transfer allocation trajectories are shown under (a) the differentiated carbon price solution, and (b) the uniform carbon price solution. In both cases, foreign abatement funding is required to be additional to the domestic abatement spending in the utilitarian no-transfer differentiated carbon price solution. Results are for the utility discount rate of 1.5%.

5 Conclusion

This paper explores the differences between optimal carbon prices that ignore intratemporal inequality and those that consider inequality, along with the distribution of costs and benefits associated with emission reductions. Specifically, it compares the optimal carbon prices under two main optimization approaches: the conventional, positive approach which maximizes the Negishi-weighted social welfare function (SWF), and a normative approach which focuses on maximizing global welfare, relying on constrained maximizations of the utilitarian SWF.

Using a theoretical model, I show that, in the absence of international transfers, accounting for inequality may result in higher or lower optimal carbon prices and that this depends on the distribution of the marginal climate damages and the burden of the abatement costs on different countries. Intuitively, global welfare maximization warrants more stringent climate policy if poor populations face disproportionately high marginal climate damages and low abatement costs, highlighting the importance of accurately accounting for the regional heterogeneity in the damage and abatement cost functions. In numerical simulations with the integrated assessment model RICE, I find that accounting for inequality results in lower optimal global emissions, both if carbon prices are allowed to be regionally differentiated and if they are constrained to be globally uniform.

Moreover, focusing on the “Paris Agreement transfer” of $100 billion per year, I find that financial support for mitigation in developing countries considerably increases the stringency of welfare-maximizing climate policy under both the uniform and the differentiated carbon price solutions. For instance, the welfare-maximizing uniform carbon price in 2025 almost doubles, from $29/tCO2 to $54/tCO2, under the default discounting parameters in RICE.

To summarize, the important policy implication is that international mitigation finance and accounting for inequality both increase the stringency of the climate policy that maximizes global welfare. There are some limitations of this study which are left for future research. First, the RICE model masks inequality within its twelve regions. Thus, a valuable avenue for future research would be to account for inequality at a finer resolution and examine how the quantitative results change. Existing modifications of the RICE model may be used for this analysis, including NICE and RICE50+ (Dennig et al., 2015; Gazzotti et al., 2021).

Second, the numerical simulations of this study are performed with a single integrated assessment model (IAM). As different IAMs are known to produce different results, it would be worthwhile to replicate the analysis with other IAMs to assess the robustness of the findings of this paper. Another insightful exercise would be to modify certain model components of the RICE model and explore how the results change. This may be guided by the theoretical results of this paper, which identify the functional forms of the regional damage and abatement cost functions as important elements determining the results. Given the lack of consensus in empirical studies regarding the forms of damage and abatement cost functions, it would be valuable to assess how the results change when these functions are substituted with alternative formulations from the literature.

References

Abbott, J. K. and E. P. Fenichel (2014). “Following the golden rule: Negishi welfare weights without apology”. *Arizona State University & Yale University.*

Adler, M., D. Anthoff, V. Bosetti, G. Garner, K. Keller, and N. Treich (2017). “Priority for the worse-off and the social cost of carbon”. *Nature Climate Change* 7.6, p. 443.

Ahmed, S. A., N. S. Diffenbaugh, and T. W. Hertel (2009). “Climate volatility deepens poverty vulnerability in developing countries”. *Environmental Research Letters* 4.3, p. 034004.

Anthoff, D. (2011). *Optimal Global Dynamic Carbon Abatement.* URL: `https://www.david-anthoff.com/AnthoffJobmarketPaper.pdf`.

Anthoff, D., F. Dennig, and J. Emmerling (2021). *On Differentiated Carbon Prices and Discount Rates.* SSRN Scholarly Paper ID 3906413. Rochester, NY: Social Science Research Network. URL: `https://papers.ssrn.com/abstract=3906413` (visited on 09/18/2021).

Anthoff, D., F. Errickson, and L. Rennels (2019). *Mimi-RICE-2010.jl - a julia implementation of the RICE 2010 model.* Berkeley, CA, United States. URL: `https://github.com/anthofflab/mimi-rice-2010.jl` (visited on 02/22/2019).

Arrow, K. J., M. Cropper, C. Gollier, B. Groom, G. M. Heal, R. G. Newell, W. D. Nordhaus, R. S. Pindyck, W. A. Pizer, P. Portney, T. Sterner, R. S. J. Tol, and M. Weitzman (2013). *How Should Benefits and Costs Be Discounted in an Intergenerational Context? The Views of an Expert Panel.* SSRN Scholarly Paper ID 2199511. Rochester, NY: Social Science Research Network. URL: `https://papers.ssrn.com/abstract=2199511` (visited on 03/31/2019).

Azar, C. and T. Sterner (1996). “Discounting and distributional considerations in the context of global warming”. *Ecological Economics* 19.2, pp. 169–184.

Barrage, L. (2020). “Optimal Dynamic Carbon Taxes in a Climate–Economy Model with Distortionary Fiscal Policy”. *The Review of Economic Studies* 87.1, pp. 1–39.

Bauer, N., C. Bertram, A. Schultes, D. Klein, G. Luderer, E. Kriegler, A. Popp, and O. Edenhofer (2020). “Quantification of an efficiency–sovereignty trade-off in climate policy”. *Nature* 588.7837, pp. 261–266.

Beckerman, W. and C. Hepburn (2007). “Ethics of the Discount Rate in the Stern Review on the Economics of Climate Change”. 8.1, p. 26.

Bergson, A. (1938). “A Reformulation of Certain Aspects of Welfare Economics”. *The Quarterly Journal of Economics* 52.2, pp. 310–334.

Birgin, E. G. and J. M. Martínez (2008). “Improving ultimate convergence of an augmented Lagrangian method”. *Optimization Methods and Software* 23.

Borissov, K. and L. Bretschger (2022). “Optimal carbon policies in a dynamic heterogeneous world”. *European Economic Review* 148, p. 104253.

Bosetti, V., C. Carraro, E. De Cian, E. Massetti, and M. Tavoni (2012). *Incentives and Stability of International Climate Coalitions: An Integrated Assessment.* Rochester, NY. DOI: 10.2139/ssrn.1991775. URL: https://papers.ssrn.com/abstract=1991775 (visited on 01/30/2024).

Bretschger, L. (2013). “Climate policy and equity principles: fair burden sharing in a dynamic world”. *Environment and Development Economics* 18.5, pp. 517–536.

Budolfson, M. and F. Dennig (2019). *Optimal global climate policy and regional carbon prices.*

Budolfson, M. B., D. Anthoff, F. Dennig, F. Errickson, K. Kuruc, D. Spears, and N. K. Dubash (2021). “Utilitarian benchmarks for emissions and pledges promote equity, climate and development”. *Nature Climate Change*, pp. 1–7.

Burke, M., S. M. Hsiang, and E. Miguel (2015). “Global non-linear effect of temperature on economic production”. *Nature* 527.7577, pp. 235–239.

Chancel, L., P. Bothe, and T. Voituriez (2023). “Climate Inequality Report: Fair Taxes for a Sustainable Future in the Global South”.

Chichilnisky, G. and G. Heal (1994). “Who should abate carbon emissions?: An international viewpoint”. *Economics Letters* 44.4, pp. 443–449.

— eds. (2000). *Environmental Markets: Equity and Efficiency.* Columbia University Press. ISBN: null. DOI: 10.7312/chic11588. URL: www.jstor.org/stable/10.7312/chic11588 (visited on 05/31/2020).

Chichilnisky, G., G. Heal, and D. Starrett (2000). “Equity and efficiency in environmental markets: global trade in carbon dioxide emissions”. *Environmental markets: equity and efficiency.* Vol. 15.

Climate Watch (2022). *Explore Nationally Determined Contributions (NDCs).* Washington, DC. URL: https://www.climatewatchdata.org.

Dasgupta, P. (2008). “Discounting climate change”. *Journal of Risk and Uncertainty* 37.2, pp. 141–169.

Dennig, F., D. Anthoff, F. Errickson, and L. Rennels (2019). *RICEupdate.* URL: `https://github.com/fdennig/RICEupdate` (visited on 05/03/2020).

Dennig, F., M. B. Budolfson, M. Fleurbaey, A. Siebert, and R. H. Socolow (2015). “Inequality, climate impacts on the future poor, and carbon prices”. *Proceedings of the National Academy of Sciences*, p. 201513967.

Dennig, F. and J. Emmerling (2017). “A Note on Optima with Negishi Weights”. *Working Paper.*

— (2019). “A note on optima with time-varying Negishi weights”. *Working Paper.*

Dennig, F., F. Errickson, and D. Anthoff (2017). *mimi_NICE.* URL: `https://github.com/fdennig/mimi_NICE` (visited on 05/02/2020).

Dietz, S. and N. Stern (2008). “Why Economic Analysis Supports Strong Action on Climate Change: A Response to the Stern Review’s Critics”. *Review of Environmental Economics and Policy* 2.1, pp. 94–113.

Diffenbaugh, N. S. and M. Burke (2019). “Global warming has increased global economic inequality”. *Proceedings of the National Academy of Sciences*, p. 201816020.

Eyckmans, J., S. Fankhauser, and S. Kverndokk (2016). “Development Aid and Climate Finance”. *Environmental and Resource Economics* 63.2, pp. 429–450.

Flannery, B., J. Hillman, J. Mares, and M. C. Porterfield (2018). *Framework Proposal for a US Upstream Greenhouse Gas Tax with WTO-Compliant Border Adjustments.* SSRN Scholarly Paper ID 3148213. Rochester, NY: Social Science Research Network. URL: `https://papers.ssrn.com/abstract=3148213` (visited on 12/21/2021).

Gazzotti, P., J. Emmerling, G. Marangoni, A. Castelletti, K.-I. v. d. Wijst, A. Hof, and M. Tavoni (2021). “Persistent inequality in economically optimal climate policies”. *Nature Communications* 12.1, p. 3421.

Golosov, M., J. Hassler, P. Krusell, and A. Tsyvinski (2014). “Optimal Taxes on Fossil Fuel in General Equilibrium”. *Econometrica* 82.1, pp. 41–88.

Hallegatte, S., M. Bangalore, L. Bonzanigo, M. Fay, U. Narloch, J. Rozenberg, and A. Vogt-Schilb (2014). *Climate Change and Poverty—An Analytical Framework.* Policy Research

Working Papers. The World Bank. URL: http://elibrary.worldbank.org/doi/book/10.1596/1813-9450-7126 (visited on 02/15/2019).

Hillebrand, E. and M. Hillebrand (2023). "Who pays the bill? Climate change, taxes, and transfers in a multi-region growth model". *Journal of Economic Dynamics and Control* 153, p. 104695.

Hoel, M. (1994). "Efficient Climate Policy in the Presence of Free Riders". *Journal of Environmental Economics and Management* 27.3, pp. 259–274.

Hoel, M. O., S. A. C. Kittelsen, and S. Kverndokk (2019). "Correcting the Climate Externality: Pareto Improvements Across Generations and Regions". *Environmental and Resource Economics* 74.1, pp. 449–472.

Johnson, S. G. (2020). *The NLopt nonlinear-optimization package.* URL: http://github.com/stevengj/nlopt.

Kalkuhl, M. and L. Wenz (2020). "The impact of climate conditions on economic production. Evidence from a global panel of regions". *Journal of Environmental Economics and Management* 103, p. 102360.

Kornek, U., D. Klenert, O. Edenhofer, and M. Fleurbaey (2021). "The social cost of carbon and inequality: When local redistribution shapes global carbon prices". *Journal of Environmental Economics and Management* 107, p. 102450.

Kotchen, M. J. (2018). "Which Social Cost of Carbon? A Theoretical Perspective". *Journal of the Association of Environmental and Resource Economists* 5.3, pp. 673–694.

— (2020). "On the scope of climate finance to facilitate international agreement on climate change". *Economics Letters* 190, p. 109070.

Leimbach, M., N. Bauer, L. Baumstark, and O. Edenhofer (2010). "Mitigation Costs in a Globalized World: Climate Policy Analysis with REMIND-R". *Environmental Modeling & Assessment* 15.3, pp. 155–173.

Llavador, H., J. E. Roemer, and J. Silvestre (2010). "Intergenerational justice when future worlds are uncertain". *Journal of Mathematical Economics*. Mathematical Economics: Special Issue in honour of Andreu Mas-Colell, Part 1 46.5, pp. 728–761.

— (2011). ""A dynamic analysis of human welfare in a warming planet"". *Journal of Public Economics*. Special Issue: International Seminar for Public Economics on Normative Tax Theory 95.11, pp. 1607–1620.

Manne, A. S. and R. G. Richels (2005). "Merge: An Integrated Assessment Model for Global Climate Change". *Energy and Environment*. New York: Springer-Verlag, pp. 175–189. ISBN: 978-0-387-25351-0. DOI: 10.1007/0-387-25352-1_7. URL: http://link.springer.com/10.1007/0-387-25352-1_7 (visited on 01/30/2024).

Mas-Colell, A., M. D. Whinston, and J. R. Green (1995). *Microeconomic theory.* Vol. 1. Oxford university press New York.

Mendelsohn, R., A. Dinar, and L. Williams (2006). "The distributional impact of climate change on rich and poor countries". *Environment and Development Economics* 11.2, pp. 159–178.

Negishi, T. (1960). "Welfare Economics and Existence of an Equilibrium for a Competitive Economy". *Metroeconomica* 12.2, pp. 92–97.

Nordhaus, W. (2010). "Economic aspects of global warming in a post-Copenhagen environment". *Proceedings of the National Academy of Sciences* 107.26, pp. 11721–11726.

— (2011). *Estimates of the Social Cost of Carbon: Background and Results from the RICE-2011 Model.* Working Paper 17540. National Bureau of Economic Research. URL: http://www.nber.org/papers/w17540 (visited on 02/22/2019).

— (2013). "Integrated Economic and Climate Modeling". *Handbook of Computable General Equilibrium Modeling.* Ed. by Dixon, P. B. and Jorgenson, D. W. Vol. 1. Handbook of Computable General Equilibrium Modeling SET, Vols. 1A and 1B. Elsevier, pp. 1069–1131. DOI: 10.1016/B978-0-444-59568-3.00016-X. URL: http://www.sciencedirect.com/science/article/pii/B978044459568300016X (visited on 09/17/2020).

Nordhaus, W. and P. Sztorc (2013). "DICE 2013R: Introduction and user's manual". Second edition.

Nordhaus, W. and Z. Yang (1996). "A Regional Dynamic General-Equilibrium Model of Alternative Climate-Change Strategies". *The American Economic Review* 86.4, pp. 741–765.

Nordhaus, W. D. (2007). "A Review of the Stern Review on the Economics of Climate Change". *Journal of Economic Literature*, p. 49.

Nordhaus, W. D. and J. Boyer (2000). *Warming the world: economic models of global warming.* MIT press. ISBN: 0-262-64054-6.

OECD (2019). *Economic Outlook No 106 - November 2019. : Nominal GDP growth, forecast.* URL: `https://stats.oecd.org/Index.aspx?QueryId=51654` (visited on 05/03/2020).

Oliver, P., A. Clark, and C. Meattle (2018). *Global Climate Finance: An Updated View 2018.* Climate Policy Initiative.

Oppenheimer, M., M. Campos, R. Warren, J. Birkmann, G. Luber, B. O'Neill, and K. Takahashi (2014). "Emergent risks and key vulnerabilities. Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change".

Rao, N. D. (2014). "International and intranational equity in sharing climate change mitigation burdens". *International Environmental Agreements: Politics, Law and Economics* 14.2, pp. 129–146.

Roemer, J. E. (2011). "The Ethics of Intertemporal Distribution in a Warming Planet". *Environmental and Resource Economics* 48.3, pp. 363–390.

Rowan, T. H. (1990). "Functional Stability Analysis of Numerical Algorithms". PhD thesis. Department of Computer Sciences, University of Texas at Austin.

Samuelson, P. A. (1954). "The Pure Theory of Public Expenditure". *The Review of Economics and Statistics* 36.4, pp. 387–389.

Samuelson, P. A. (1947). *Foundations of economic analysis.* Cambridge: Harvard University Press.

Sandmo, A. (2007). "Global public economics: Public goods and externalities". *Économie publique/Public economics* 18.

Schumacher, I. (2018). "The aggregation dilemma in climate change policy evaluation". *Climate Change Economics* 09.3, p. 1850008.

Sen, A. (1985). "The Moral Standing of the Market". *Social Philosophy and Policy* 2.2, pp. 1–19.

Shiell, L. (2003). "Equity and efficiency in international markets for pollution permits". *Journal of Environmental Economics and Management* 46.1, pp. 38–51.

Stanton, E. A. (2011). "Negishi welfare weights in integrated assessment models: the mathematics of global inequality". *Climatic Change* 107.3, pp. 417–432.

Stanton, E. A., F. Ackerman, and S. Kartha (2009). "Inside the integrated assessment models: Four issues in climate economics". *Climate and Development* 1.2, pp. 166–184.

Stern, N., S. Peters, V. Bakhshi, A. Bowen, C. Cameron, S. Catovsky, D. Crane, S. Cruickshank, S. Dietz, and N. Edmonson (2006). *Stern Review: The economics of climate change*. Vol. 30. HM treasury London.

UNFCCC (1992). *United Nations framework convention on climate change.*

— (2009). *Report of the Conference of the Parties on its fifteenth session, held in Copenhagen from 7 to 19 December 2009. Part Two: Action taken by the Conference of the Parties at its fifteenth session.* URL: https://unfccc.int/resource/docs/2009/cop15/eng/11a01.pdf.

— (2015). *Report of the Conference of the Parties on its twenty-first session, held in Paris from 30 November to 13 December 2015 Addendum Contents Part two: Action taken by the Conference of the Parties at its twenty-first session, Decision 1/CP.21 Adoption of the Paris Agreement.*

— (2023). *Report of the Conference of the Parties on its twenty-seventh session, held in Sharm el-Sheikh from 6 to 20 November 2022. Addendum. Part two: Action taken by the Conference of the Parties at its twentyseventh session.*

Weitzman, M. L. (2014). "Can Negotiating a Uniform Carbon Price Help to Internalize the Global Warming Externality?" *Journal of the Association of Environmental and Resource Economists* 1.1, pp. 29–49.

— (2017a). "On a World Climate Assembly and the Social Cost of Carbon". *Economica* 84.336, pp. 559–586.

— (2017b). "Voting on prices vs. voting on quantities in a World Climate Assembly". *Research in Economics* 71.2, pp. 199–211.

World Bank (2023a). *State and Trends of Carbon Pricing 2023.* DOI: 10.1596/39796. URL: https://openknowledge.worldbank.org/handle/10986/39796 (visited on 01/27/2024).

— (2023b). *World Bank Open Data. GDP deflator: linked series (base year varies by country) - United States.* World Bank Open Data. URL: https://data.worldbank.org/indicator/NY.GDP.DEFL.ZS.AD?locations=US (visited on 11/18/2023).

Yang, Z. and W. D. Nordhaus (2006). “Magnitude and direction of technological transfers for mitigating GHG emissions”. *Energy Economics*. Modeling Technological Change in Climate Policy Analyses 28.5, pp. 730–741.

Appendix A: Proofs and Derivations

A.1 Derivation of optimal carbon prices

This section shows the derivation of the optimal uniform carbon price (Equation (9) in the main text). As discussed briefly below, the derivation of the optimal differentiated carbon prices is largely analogous (and, in fact, simpler) and therefore not explicitly shown.

The Lagrangian of the uniform carbon price optimization problem is

$$\begin{aligned}\mathcal{L} = \alpha_i u_i\left(X_i\right) - \sum_i \lambda_i\left(X_i - W_i + C_i(A_i) + D_i(A)\right) \\ - \mu\left(C'_N(A_N) - C'_S(A_S)\right),\end{aligned} \tag{A1}$$

where λ_i and μ are Lagrange multipliers.

The first-order conditions (FOCs) are:

$$\begin{aligned}&[X_i] : \alpha_i u'_i(X_i) = \lambda_i, \quad \forall i \\ &[A_N] : \lambda_N C'_N(A_N) = -\sum_i \lambda_i D'_i(A) - \mu C''_N(A_N) \\ &[A_S] : \lambda_S C'_S(A_S) = -\sum_i \lambda_i D'_i(A) + \mu C''_S(A_S) \\ &[\lambda_i] : X_i = W_i - C_i(A_i) - D_i(A), \quad \forall i \\ &[\mu] : C'_N(A_N) = C'_S(A_S).\end{aligned} \tag{A2}$$

Combining the FOCs, we get the following two optimality conditions:

$$\alpha_N u'(X_N)C'_N(A_N) = -\sum_i \alpha_i u'_i D'_i(A) - \mu C''_N(A_N), \tag{A3}$$

$$\alpha_S u'(X_S)C'_S(A_S) = -\sum_i \alpha_i u'_i D'_i(A) + \mu C''_S(A_S). \tag{A4}$$

We can now solve these two equations for the optimal uniform carbon price, noting that $C'_N(A_N) = C'_S(A_S)$ by the uniform carbon price constraint. Eliminating μ by dividing Equation (A3) by Equation (A4), and simple manipulations yield the optimal uniform carbon price (Equation (9) in the main text),

$$\tau^{uni} = C_i'^{*}(A_i^*) = -\sum_i \alpha_i u'(X_i^*)D'_i(A^*)\frac{C''_S(A_S^*) + C''_N(A_N^*)}{\alpha_N u'(X_N^*)C''_S(A_S^*) + \alpha_S u'(X_S^*)C''_N(A_N^*)}. \tag{A5}$$

The derivation of the optimal differentiated carbon price is largely analogous. The main difference is that the uniform carbon price constraint is missing (i.e., $\mu(C'_N(A_N) - C'_S(A_S))$ is missing in the Lagrangian). As a result, we (generally) get different optimal carbon prices in the two regions.

A.2 Derivation of optimality conditions

A.2.1 Negishi solution

We start with Equation (11),

$$\frac{dC_i(\tilde{A}_i)}{d\tilde{A}_i} = -\sum_i \frac{dD_i(\tilde{A})}{d\tilde{A}}. \tag{A6}$$

Multiplying both sides by $\frac{d\tilde{A}}{d\tilde{\tau}}$, and using $d\tilde{A} = d\tilde{A}_S + d\tilde{A}_N$, yields

$$\frac{dC_i(\tilde{A}_i)}{d\tilde{A}_i}\frac{d\tilde{A}_S + d\tilde{A}_N}{d\tilde{\tau}} = -\sum_i \frac{dD_i(\tilde{A})}{d\tilde{\tau}}. \tag{A7}$$

Using $\tilde{\tau} = \frac{dC_N(\tilde{A}_N)}{d\tilde{A}_N} = \frac{dC_S(\tilde{A}_S)}{d\tilde{A}_S}$, we obtain

$$\sum_i \frac{dC_i(\tilde{A}_i)}{d\tilde{\tau}} = -\sum_i \frac{dD_i(\tilde{A})}{d\tilde{\tau}}. \tag{A8}$$

A.2.2 Utilitarian uniform carbon price solution

We start with Equation (13),

$$\frac{dC_i(\check{A}_i)}{d\check{A}_i} = -\sum_i u'(\check{X}_i)\frac{dD_i(\check{A})}{d\check{A}}\frac{C''_S(\check{A}_S) + C''_N(\check{A}_N)}{u'(\check{X}_N)C''_S(\check{A}_S) + u'(\check{X}_S)C''_N(\check{A}_N)}. \tag{A9}$$

Using $\check{C}''_i = \frac{d\check{C}'_i}{d\check{A}_i} = \frac{d\check{\tau}}{d\check{A}_i}$, multiplying both sides by $\frac{d\check{A}}{d\check{\tau}}$ and rearranging, we have

$$\frac{dC_i(\check{A}_i)}{d\check{A}_i}\frac{u'(\check{X}_N)\frac{d\check{A}}{d\check{A}_S} + u'(\check{X}_S)\frac{d\check{A}}{d\check{A}_N}}{\frac{d\check{\tau}}{d\check{A}_S} + \frac{d\check{\tau}}{d\check{A}_N}} = -\sum_i u'(\check{X}_i)\frac{dD_i(\check{A})}{d\check{\tau}}. \tag{A10}$$

Using $\frac{dC_S(\check{A}_S)}{d\check{A}_S} = \frac{dC_N(\check{A}_N)}{d\check{A}_N}$, equalizing the denominators of the ratios in the denominator and rearranging yields

$$\begin{aligned} u'(\check{X}_N)\frac{d\check{A}}{d\check{A}_S d\check{\tau}}\frac{dC_S(\check{A}_S)dC_N(\check{A}_N)}{dC_S(\check{A}_S)+dC_N(\check{A}_N)} + u'(\check{X}_S)\frac{d\check{A}}{d\check{A}_N d\check{\tau}}\frac{dC_S(\check{A}_S)dC_N(\check{A}_N)}{dC_S(\check{A}_S)+dC_N(\check{A}_N)} \\ = -\sum_i u'(\check{X}_i)\frac{dD_i(\check{A})}{d\check{\tau}}. \end{aligned} \tag{A11}$$

Using $\check{\tau} = \frac{d\check{C}_S}{d\check{A}_S} = \frac{d\check{C}_N}{d\check{A}_N}$, and thus $\check{\tau} d\check{A}_i = d\check{C}_i$ for all i, we rewrite the previous equation as

$$\check{\tau}\frac{d\check{A}_S + d\check{A}_N}{\check{\tau} d\check{A}_S + \check{\tau} d\check{A}_N}\left(u'(\check{X}_N)\frac{dC_N(\check{A}_N)}{d\check{\tau}} + u'(\check{X}_S)\frac{dC_S(\check{A}_S)}{d\check{\tau}}\right) = -\sum_i u'(\check{X}_i)\frac{dD_i(\check{A})}{d\check{\tau}}, \tag{A12}$$

which simplifies to

$$\sum_i u'(\check{X}_i)\frac{dC_i(\check{A}_i)}{d\check{\tau}} = -\sum_i u'(\check{X}_i)\frac{dD_i(\check{A})}{d\check{\tau}}. \tag{A13}$$

A.2.3 Regions' preferred uniform carbon price

We start with Equation (20),

$$\frac{dC_i(\mathring{A}^i_i)}{d\mathring{A}^i_i} = -\frac{dD_i(\mathring{A}^i)}{d\mathring{A}^i}\frac{C''_i(\mathring{A}^i_i) + C''_{-i}(\mathring{A}^i_{-i})}{C''_{-i}(\mathring{A}^i_{-i})}, \tag{A14}$$

Using $C''_i(\mathring{A}^i_i) = \frac{dC'_i(\mathring{A}^i_i)}{d\mathring{A}^i_i} = \frac{d\mathring{\tau}^i}{d\mathring{A}^i_i}$ and $C''_i(\mathring{A}^i_i) = \frac{dC'_{-i}(\mathring{A}^i_{-i})}{d\mathring{A}^i_{-i}} = \frac{d\mathring{\tau}^i}{d\mathring{A}^i_{-i}}$, multiplying both sides by $\frac{d\check{A}}{d\check{\tau}}$ and rearranging, we have

$$\begin{aligned} \frac{dC_i(\mathring{A}^i_i)}{d\mathring{A}^i_i} &= -\frac{dD_i(\mathring{A}^i)}{d\mathring{A}^i}\frac{\frac{d\mathring{\tau}^i}{d\mathring{A}^i_i} + \frac{d\mathring{\tau}^i}{d\mathring{A}^i_{-i}}}{\frac{d\mathring{\tau}^i}{d\mathring{A}^i_{-i}}} \\ &= -\frac{dD_i(\mathring{A}^i)}{d\mathring{A}^i}\left(\frac{d\mathring{A}^i_{-i}}{d\mathring{A}^i_i} + 1\right) \\ &= -\frac{dD_i(\mathring{A}^i)}{d\mathring{A}^i}\left(\frac{d\mathring{A}^i_{-i} + d\mathring{A}^i_i}{d\mathring{A}^i_i}\right) \\ &= -\frac{dD_i(\mathring{A}^i)}{d\mathring{A}^i_i}. \end{aligned} \tag{A15}$$

Multiplying both sides by $\frac{d\mathring{A}_i^i}{d\mathring{\tau}^i}$, we obtain

$$\frac{dC_i(\mathring{A}_i^i)}{d\mathring{\tau}^i} = -\frac{dD_i(\mathring{A}^i)}{d\mathring{\tau}^i}. \tag{A16}$$

A.3 Proof of Proposition 1

Proof. Formally, the optimal uniform carbon price does not depend on welfare weights if $\frac{d\tau^{uni}}{d\alpha_i} = 0$ for all i, or equivalently, $\frac{dC'}{d\alpha_i} = 0$ for all i. We thus seek to identify the conditions under which $\frac{d\tau^{uni}}{d\alpha_i} = 0$.

First, we compute the derivative $\frac{d\tau^{uni}}{d\alpha_N}$ (the derivation for $\frac{d\tau^{uni}}{d\alpha_S}$ is analogous):

$$\begin{gathered}\frac{d\tau^{uni}}{d\alpha_N} = \Bigg[\Bigg\langle\Bigg\{-u'_N + \alpha_N u''_N \left[C'_N \frac{dA_N}{d\alpha_N} + D'_N \frac{dA}{d\alpha_N}\right]\Bigg\} D'_N - \alpha_N u'_N D''_N \frac{dA}{d\alpha_N}\Bigg\rangle \\ + \Bigg\langle\Bigg\{\alpha_S u''_S \left[C'_S \frac{dA_S}{d\alpha_N} + D'_S \frac{dA}{d\alpha_N}\right]\Bigg\} D'_S - \alpha_S u'_S D''_S \frac{dA}{d\alpha_N}\Bigg\rangle\Bigg] \\ \frac{C''_S + C''_N}{\alpha_N u'_N C''_S + \alpha_S u'_S C''_N} \\ -(\alpha_N u'_N D'_N + \alpha_S u'_S D'_S) \\ \frac{\left\langle C'''_S \frac{dA_S}{d\alpha_N} + C'''_N \frac{dA_N}{d\alpha_N}\right\rangle (\alpha_N u'_N C''_S + \alpha_S u'_S C''_N)}{(\alpha_N u'_N C''_S + \alpha_S u'_S C''_N)^2} \\ +(\alpha_N u'_N D'_N + \alpha_S u'_S D'_S) \\ \frac{(C''_S + C''_N)\left(\left\langle\left\{u'_N + \alpha_N u''_N \left[-C'_N \frac{dA_N}{d\alpha_N} - D'_N \frac{dA}{d\alpha_N}\right]\right\} C''_S + \alpha_N u'_N C'''_S \left(\frac{dA_S}{d\alpha_N}\right)\right\rangle + \left\langle\left\{\alpha_S u''_S \left[-C'_S \frac{dA_S}{d\alpha_N} - D'_S \frac{dA}{d\alpha_N}\right]\right\} C''_N + \alpha_S u'_S C'''_N \left(\frac{dA_N}{d\alpha_N}\right)\right\rangle\right)}{(\alpha_N u'_N C''_S + \alpha_S u'_S C''_N)^2}.\end{gathered} \tag{A17}$$

Next, we note that we can simplify this expression substantially by utilizing the fact that we are looking for conditions under which $\frac{d\tau^{uni}}{d\alpha_i} = 0$. For strictly convex abatement cost functions, we have the following implications:

$$\frac{d\tau^{uni}}{d\alpha_N} = 0 \implies \frac{dC'}{d\alpha_N} = 0 \implies \frac{dA_i}{d\alpha_N} = 0, \forall i \implies \frac{dC_i}{d\alpha_N} = 0, \frac{dD_i}{d\alpha_N} = 0, \frac{du_i}{d\alpha_N} = 0, \forall i. \tag{A18}$$

Equation (A17) then simplifies to

$$\frac{d\tau^{uni}}{d\alpha_N} = -\frac{u'_N D'_N (C''_S + C''_N)}{\alpha_N u'_N C''_S + \alpha_S u'_S C''_N} + \frac{(\alpha_N u'_N D'_N + \alpha_S u'_S D'_S)(C''_S + C''_N) u'_N C''_S}{(\alpha_N u'_N C''_S + \alpha_S u'_S C''_N)^2} = 0, \tag{A19}$$

which I have now set to zero because we have used $\frac{d\tau^{uni}}{d\alpha_i} = 0$.

Note that we can now cancel $(C''_S + C''_N)$. Next, we multiply and divide the first term by its denominator to obtain common denominators, which we then cancel, yielding

$$-u'_N D'_N (\alpha_N u'_N C''_S + \alpha_S u'_S C''_N) + (\alpha_N u'_N D'_N + \alpha_S u'_S D'_S) u'_N C''_S = 0. \tag{A20}$$

Canceling u'_N, multiplying out and rearranging yields

$$\alpha_N u'_N (C''_S D'_N - C''_S D'_N) = \alpha_S u'_S (C''_N D'_N - C''_S D'_S). \tag{A21}$$

Notice that the left-hand side is zero

$$0 = \alpha_S u'_S (C''_N D'_N - C''_S D'_S). \tag{A22}$$

Divide by $\alpha_S u'_S$ and rearrange to obtain

$$\frac{C''_S}{C''_N} = \frac{D'_N}{D'_S}. \tag{A23}$$

Thus, we have shown that

$$\frac{d\tau^{uni}}{d\alpha_N} = 0 \iff \frac{C''_S}{C''_N} = \frac{D'_N}{D'_S}. \tag{A24}$$

The derivations are analogous for $\frac{dC'}{d\alpha_S}$. Together this proves that

$$\frac{d\tau^{uni}}{d\alpha_i} = 0 \iff \frac{C''_S}{C''_N} = \frac{D'_N}{D'_S}, \quad \forall i. \tag{A25}$$

We have thus shown that the optimal uniform carbon price does not depend on the welfare weights α_i if and only if $\frac{D'_S}{D'_N} = \frac{C''_N}{C''_S}$. □

A.4 Proof of Proposition 2

Proof. We split the proof into the forward and backward implications.

Proof of forward implication: $\tilde{\tau} < \check{\tau} \implies \frac{\check{D}'_S}{\check{D}'_N} > \frac{\check{C}''_N}{\check{C}''_S}$.

Let us ask under which conditions $\tilde{\tau} < \check{\tau}$, or equivalently, $\tilde{C}' < \check{C}'$. First note that, for strictly convex abatement cost functions, $\tilde{C}' < \check{C}'$ implies $\tilde{A}_i < \check{A}_i$ for all i, and thus $\tilde{A} < \check{A}$. For strictly convex damage functions, this implies $\tilde{D}'_i < \check{D}'_i$ (note that marginal damages of abatement are negative) for all i.

We have $\tilde{C}' < \check{C}'$ if and only if

$$-\tilde{D}'_N - \tilde{D}'_S < (-\check{u}'_N \check{D}'_N - \check{u}'_S \check{D}'_S) \frac{\check{C}''_S + \check{C}''_N}{\check{u}'_N \check{C}''_S + \check{u}'_S \check{C}''_N}. \tag{A26}$$

Multiplying both sides by the denominator on the right-hand side (which is positive), and rearranging, we get

$$(\breve{u}'_N \breve{D}'_N + \breve{u}'_S \breve{D}'_S)(\breve{C}''_S + \breve{C}''_N) < (\tilde{D}'_N + \tilde{D}'_S)(\breve{u}'_N \breve{C}''_S + \breve{u}'_S \breve{C}''_N). \tag{A27}$$

Multiplying out and collecting terms, we have

$$\breve{C}''_N\Big(\breve{u}'_N\breve{D}'_N + \breve{u}'_S\breve{D}'_S - \breve{u}'_S\tilde{D}'_N - \breve{u}'_S\tilde{D}'_S\Big) < \breve{C}''_S\Big(\breve{u}'_N\tilde{D}'_N + \breve{u}'_N\tilde{D}'_S - \breve{u}'_N\breve{D}'_N - \breve{u}'_S\breve{D}'_S\Big). \tag{A28}$$

We know that $\Big(\breve{u}'_N\breve{D}'_N + \breve{u}'_S\breve{D}'_S - \breve{u}'_S\tilde{D}'_N - \breve{u}'_S\tilde{D}'_S\Big) > 0$ if $\breve{u}'_S > \breve{u}'_N$ and $\tilde{D}'_i < \breve{D}'_i$ because $\breve{u}'_S\breve{D}'_S - \breve{u}'_S\tilde{D}'_S > 0$ since $\tilde{D}'_i < \breve{D}'_i$ and $\breve{u}'_N\breve{D}'_N - \breve{u}'_S\tilde{D}'_N > 0$ since $\breve{u}'_S > \breve{u}'_N$ and $\tilde{D}'_i < \breve{D}'_i$. Hence, we can divide by it and the sign of the inequality does not flip. Moreover, note that we must also have

$$\Big(\breve{u}'_N\tilde{D}'_N + \breve{u}'_N\tilde{D}'_S - \breve{u}'_N\breve{D}'_N - \breve{u}'_S\breve{D}'_S\Big) > 0 \tag{A29}$$

for the inequality in Equation (A28) to hold, since $\breve{C}''_i > 0$ for all i.

Cross-division, collecting common terms, and rearranging yields

$$\frac{\breve{C}''_N}{\breve{C}''_S} < \frac{\breve{u}'_N(\tilde{D}'_N - \breve{D}'_N + \tilde{D}'_S) - \breve{u}'_S\breve{D}'_S}{\breve{u}'_N\breve{D}'_N - \breve{u}'_S(\tilde{D}'_S - \breve{D}'_S + \tilde{D}'_N)}. \tag{A30}$$

Note that

$$\underbrace{\tilde{D}'_N - \breve{D}'_N}_{<0} + \tilde{D}'_S < \tilde{D}'_S \tag{A31}$$

and

$$\underbrace{\tilde{D}'_S - \breve{D}'_S}_{<0} + \tilde{D}'_N < \tilde{D}'_N \tag{A32}$$

Thus, for the numerator we have

$$\begin{aligned}\overbrace{\breve{u}'_N\underbrace{(\tilde{D}'_N - \breve{D}'_N + \tilde{D}'_S)}_{<\tilde{D}'_S} - \breve{u}'_S\breve{D}'_S}^{>0 \text{ by Equation (A29)}} &< \breve{u}'_N\tilde{D}'_S - \breve{u}'_S\breve{D}'_S \\ &< \breve{u}'_N\breve{D}'_S - \breve{u}'_S\breve{D}'_S \\ &> 0,\end{aligned} \tag{A33}$$

where the second inequality holds since $\tilde{D}'_S < \check{D}'_S$ and the last inequality holds since $\check{u}'_S > \check{u}'_N$.

Similarly, for the denominator we have

$$\begin{aligned} \check{u}'_N \check{D}'_N - \check{u}'_S \underbrace{(\tilde{D}'_S - \check{D}'_S + \tilde{D}'_N)}_{<\tilde{D}'_N} &> \check{u}'_N \check{D}'_N - \check{u}'_S \tilde{D}'_N \\ &> \check{u}'_N \check{D}'_N - \check{u}'_S \check{D}'_N \\ &> 0, \end{aligned} \tag{A34}$$

where the second inequality holds since $\tilde{D}'_S < \check{D}'_S$ and the last inequality holds since $\check{u}'_S > \check{u}'_N$.

Compared to the case "Negishi = Utilitarian", we have a greater (positive) denominator, and a smaller (positive, by Equation (A29)) numerator.

Putting this together we have

$$\begin{aligned} \frac{\check{C}''_N}{\check{C}''_S} &< \frac{\check{u}'_N(\tilde{D}'_N - \check{D}'_N + \tilde{D}'_S) - \check{u}'_S \check{D}'_S}{\check{u}'_N \check{D}'_N - \check{u}'_S(\tilde{D}'_S - \check{D}'_S + \tilde{D}'_N)} \\ &< \frac{\check{u}'_N \check{D}'_S - \check{u}'_S \check{D}'_S}{\check{u}'_N \check{D}'_N - \check{u}'_S \check{D}'_N} \\ &= \frac{\check{D}'_S}{\check{D}'_N}. \end{aligned} \tag{A35}$$

We have thus shown that $\tilde{C}' < \check{C}' \implies \frac{\check{D}'_S}{\check{D}'_N} > \frac{\check{C}''_N}{\check{C}''_S}$.

Proof of backward implication: $\frac{\check{D}'_S}{\check{D}'_N} > \frac{\check{C}''_N}{\check{C}''_S} \implies \tilde{\tau} < \check{\tau}$.

In order to derive a contradiction, suppose that $\frac{\check{D}'_S}{\check{D}'_N} > \frac{\check{C}''_N}{\check{C}''_S} \implies \tilde{\tau} \geq \check{\tau}$.

We start by establishing the implications of $\tilde{\tau} \geq \check{\tau}$, or equivalently, $\tilde{C}' \geq \check{C}'$. First note that, for strictly convex abatement cost functions, $\tilde{C}' \geq \check{C}'$ implies $\tilde{A}_i \geq \check{A}_i$ for all i, and thus $\tilde{A} \geq \check{A}$. For strictly convex damage functions, this implies $\tilde{D}'_i \geq \check{D}'_i$ (note that marginal damages of abatement are negative) for all i.

Next, note that $\tilde{C}' \geq \check{C}'$ if and only if

$$-\tilde{D}'_N - \tilde{D}'_S \geq (-\check{u}'_N \check{D}'_N - \check{u}'_S \check{D}'_S) \frac{\check{C}''_S + \check{C}''_N}{\check{u}'_N \check{C}''_S + \check{u}'_S \check{C}''_N}. \tag{A36}$$

Multiplying both sides by the denominator on the right-hand side (which is positive), and rearranging, we get

$$(\check{u}'_N \check{D}'_N + \check{u}'_S \check{D}'_S)(\check{C}''_S + \check{C}''_N) \geq (\tilde{D}'_N + \tilde{D}'_S)(\check{u}'_N \check{C}''_S + \check{u}'_S \check{C}''_N). \tag{A37}$$

Multiplying this out and collecting common terms gives

$$\check{C}''_N\Big(\check{u}'_N\check{D}'_N + \check{u}'_S\check{D}'_S - \check{u}'_S\tilde{D}'_N - \check{u}'_S\tilde{D}'_S\Big) \geq \check{C}''_S\Big(\check{u}'_N\tilde{D}'_N + \check{u}'_N\tilde{D}'_S - \check{u}'_N\check{D}'_N - \check{u}'_S\check{D}'_S\Big). \tag{A38}$$

We know that

$$\Big(\check{u}'_N\tilde{D}'_N + \check{u}'_N\tilde{D}'_S - \check{u}'_N\check{D}'_N - \check{u}'_S\check{D}'_S\Big) > 0 \tag{A39}$$

because $\check{u}'_N\tilde{D}'_S - \check{u}'_S\check{D}'_S > 0$ since $\tilde{D}'_i \geq \check{D}'_i$ and $\check{u}'_S > \check{u}'_N$ and $\check{u}'_N\tilde{D}'_N - \check{u}'_N\check{D}'_N \geq 0$ since $\tilde{D}'_i \geq \check{D}'_i$.

Moreover, note that Equations (A38) and (A39) imply

$$\Big(\check{u}'_N\check{D}'_N + \check{u}'_S\check{D}'_S - \check{u}'_S\tilde{D}'_N - \check{u}'_S\tilde{D}'_S\Big) > 0 \tag{A40}$$

since $\check{C}''_i > 0$ for all i. Hence, we can divide by it and the sign of the inequality does not flip.

Cross-division and collecting common terms yields

$$\frac{\check{C}''_N}{\check{C}''_S} \geq \frac{\check{u}'_N(\tilde{D}'_N - \check{D}'_N + \tilde{D}'_S) - \check{u}'_S\check{D}'_S}{\check{u}'_N\check{D}'_N - \check{u}'_S(\tilde{D}'_S - \check{D}'_S + \tilde{D}'_N)}. \tag{A41}$$

It is worthwhile to take stock at this point. So far, we have established that

$$\tilde{C}' \geq \check{C}' \iff \frac{\check{C}''_N}{\check{C}''_S} \geq \frac{\check{u}'_N(\tilde{D}'_N - \check{D}'_N + \tilde{D}'_S) - \check{u}'_S\check{D}'_S}{\check{u}'_N\check{D}'_N - \check{u}'_S(\tilde{D}'_S - \check{D}'_S + \tilde{D}'_N)}. \tag{A42}$$

Next, we show that $\frac{\check{D}'_S}{\check{D}'_N} > \frac{\check{C}''_N}{\check{C}''_S} \implies \tilde{C}' \geq \check{C}'$ yields a contradiction:

$$\begin{aligned}
\frac{\check{C}''_N}{\check{C}''_S} &< \frac{\check{D}'_S}{\check{D}'_N} \\
&= \frac{\check{u}'_N\check{D}'_S - \check{u}'_S\check{D}'_S}{\check{u}'_N\check{D}'_N - \check{u}'_S\check{D}'_N} \\
&\leq \frac{\check{u}'_N\tilde{D}'_S - \check{u}'_S\check{D}'_S}{\check{u}'_N\check{D}'_N - \check{u}'_S\tilde{D}'_N} \\
&\leq \frac{\check{u}'_N(\tilde{D}'_N - \check{D}'_N + \tilde{D}'_S) - \check{u}'_S\check{D}'_S}{\check{u}'_N\check{D}'_N - \check{u}'_S(\tilde{D}'_S - \check{D}'_S + \tilde{D}'_N)} \\
&\leq \frac{\check{C}''_N}{\check{C}''_S},
\end{aligned} \tag{A43}$$

where the second and third inequalities follow from $\tilde{D}'_i \geq \check{D}'_i$ for all i and the fact that the denominator (and, trivially, the numerator) is positive by Equation (A40)[41]. The last inequality follows from the implication of $\tilde{C}' \geq \check{C}'$ documented in Equation (A42).

We have reached the contradiction $\frac{\check{C}''_N}{\check{C}''_S} < \frac{\check{C}''_N}{\check{C}''_S}$. Hence, $\frac{\check{D}'_S}{\check{D}'_N} > \frac{\check{C}''_N}{\check{C}''_S} \implies \tilde{C}' \geq \check{C}'$ is incorrect, and we have thus shown that we must have $\frac{\check{D}'_S}{\check{D}'_N} > \frac{\check{C}''_N}{\check{C}''_S} \implies \tilde{C}' < \check{C}'$.

Together, the proofs of the forward and backward implications yield the equivalence

$$\check{\tau} > \tilde{\tau} \iff \frac{\check{D}'_S}{\check{D}'_N} > \frac{\check{C}''_N}{\check{C}''_S}. \tag{A44}$$

□

A.5 Proof of Corollary 1

Proposition 2 establishes that $\check{\tau} > \tilde{\tau}$, if and only if $\frac{\check{D}'_S}{\check{D}'_N} > \frac{\check{C}''_N}{\check{C}''_S}$. We can rewrite this condition as

$$\frac{\frac{d\check{D}_S}{d\check{\tau}}}{\frac{d\check{D}_S}{d\check{\tau}}} = \frac{\frac{d\check{D}_S}{d\check{A}}}{\frac{d\check{D}_S}{d\check{A}}} > \frac{\frac{d\check{C}'_N}{d\check{A}_N}}{\frac{d\check{C}'_S}{d\check{A}_S}} = \frac{\frac{d\check{A}_S}{d\check{\tau}}}{\frac{d\check{A}_N}{d\check{\tau}}} = \frac{\frac{d\check{A}_S}{d\check{A}}}{\frac{d\check{A}_N}{d\check{A}}} = \frac{\frac{d\check{C}_S}{d\check{A}}}{\frac{d\check{C}_N}{d\check{A}}} = \frac{\frac{d\check{C}_S}{d\check{\tau}}}{\frac{d\check{C}_N}{d\check{\tau}}}, \tag{A45}$$

where $\frac{dA_i}{d\tau_i} = \frac{1}{C''_i}$, and the third equality on the right-hand side follows from $\frac{d\check{C}_S}{d\check{A}_S} = \frac{d\check{C}_N}{d\check{A}_N}$.

This establishes that

$$\frac{-\frac{d\check{D}_S}{d\check{\tau}}}{\frac{d\check{C}_S}{d\check{\tau}}} > \frac{-\frac{d\check{D}_N}{d\check{\tau}}}{\frac{d\check{C}_N}{d\check{\tau}}}. \tag{A46}$$

It remains to be shown that the left-hand side is greater than one, while the right-hand side is less than one. I utilize Proposition 4 and Lemma 2 to show this.

From Lemma 2 we know that the utilitarian uniform carbon price ($\check{\tau}$) and the Negishi-weighted carbon price ($\tilde{\tau}$) are in between the preferred uniform carbon prices of the Global North ($\mathring{\tau}^N$) and the Global South ($\mathring{\tau}^S$), unless they all coincide. Moreover, from Proposition 4 we know that the utilitarian uniform carbon price ($\check{\tau}$) is greater than the Negishi-weighted carbon price ($\tilde{\tau}$) if and only if the preferred uniform carbon price of the Global South ($\mathring{\tau}^S$) is

[41] Note that the denominator in the third line is positive because it is greater than the positive denominator in the fourth line. This can be seen as follows:

$$\check{u}'_N \check{D}'_N - \check{u}'_S \tilde{D}'_N > \check{u}'_N \check{D}'_N - \check{u}'_S (\underbrace{\tilde{D}'_S - \check{D}'_S}_{\geq 0} + \tilde{D}'_N) > 0$$

greater than the preferred uniform carbon price of the Global North ($\mathring{\tau}^N$). Therefore, $\check{\tau} > \tilde{\tau}$ implies $\mathring{\tau}^S > \check{\tau} > \tilde{\tau} > \mathring{\tau}^N$.

From Equation (22) we know that the marginal benefit-cost ratio with respect to the uniform carbon price equals one at the preferred uniform carbon price. That is,

$$\frac{-\frac{dD_i(\mathring{A}(\mathring{\tau}^i))}{d\mathring{\tau}^i}}{\frac{dC_i(\mathring{A}_i(\mathring{\tau}^i))}{d\mathring{\tau}^i}} = 1. \tag{A47}$$

We can relate this to the marginal benefit-cost ratios at the utilitarian uniform carbon price.

For the North, we have $-\frac{dD_N(\mathring{A}(\mathring{\tau}^N))}{d\mathring{\tau}^N} > -\frac{dD_N(\check{A}(\check{\tau}))}{d\check{\tau}}$. To see this, note that we have $\frac{d\mathring{\tau}^N}{d\mathring{A}(\mathring{\tau}^N)} = \frac{d\check{\tau}}{d\check{A}(\check{\tau})}$ since $\frac{d^3C(A_i)}{dA_i^3} = 0$ for all A_i[42] . Therefore, $-\frac{dD_N(\mathring{A}(\mathring{\tau}^N))}{d\mathring{\tau}^N} > -\frac{dD_N(\check{A}(\check{\tau}))}{d\check{\tau}}$ if and only if $-\frac{dD_N(\mathring{A}(\mathring{\tau}^N))}{d\mathring{\tau}^N}\frac{d\mathring{\tau}^N}{d\mathring{A}(\mathring{\tau}^N)} > -\frac{dD_N(\check{A}(\check{\tau}))}{d\check{\tau}}\frac{d\check{\tau}}{d\check{A}(\check{\tau})}$, which simplifies to $-\frac{dD_N(\mathring{A}(\mathring{\tau}^N))}{d\mathring{A}(\mathring{\tau}^N)} > -\frac{dD_N(\check{A}(\check{\tau}))}{d\check{A}(\check{\tau})}$. We know that this inequality holds from the strict convexity of the damage function and since $\check{\tau} > \mathring{\tau}^N$ implies $\check{A}_i(\check{\tau}) > \mathring{A}_i(\mathring{\tau}^N)$ for all i for strictly convex abatement cost functions.

Moreover, we have $\frac{dC_N(\mathring{A}(\mathring{\tau}^N))}{d\mathring{\tau}^N} < \frac{dC_N(\check{A}(\check{\tau}))}{d\check{\tau}}$ for the North. To see this, note that we can write this as $\frac{dC_N(\mathring{A}(\mathring{\tau}^N))}{d\mathring{A}(\mathring{\tau}^N)}\frac{d\mathring{A}(\mathring{\tau}^N)}{d\mathring{\tau}^N} < \frac{dC_N(\check{A}(\check{\tau}))}{d\check{A}(\check{\tau})}\frac{d\check{A}(\check{\tau})}{d\check{\tau}}$, which in turn can be rewritten as $\mathring{\tau}^N\frac{1}{C_N''(\mathring{A}(\mathring{\tau}^N))} < \check{\tau}\frac{1}{C_N''(\check{A}(\check{\tau}))}$. This inequality holds since $\check{\tau} > \mathring{\tau}^N$ and $\frac{d^3C(A_i)}{dA_i^3} = 0$ for all A_i.

Together, this establishes the following inequalities for the North:

$$1 = \frac{-\frac{dD_N(\mathring{A}(\mathring{\tau}^N))}{d\mathring{\tau}^N}}{\frac{dC_N(\mathring{A}_N(\mathring{\tau}^N))}{d\mathring{\tau}^N}} > \frac{-\frac{dD_N(\check{A}(\check{\tau}))}{d\check{\tau}}}{\frac{dC_N(\mathring{A}_N(\mathring{\tau}^N))}{d\mathring{\tau}^N}} > \frac{-\frac{dD_N(\check{A}(\check{\tau}))}{d\check{\tau}}}{\frac{dC_N(\check{A}_N(\check{\tau}))}{d\check{\tau}}}. \tag{A48}$$

Conversely, for the South, we have $-\frac{dD_S(\mathring{A}(\mathring{\tau}^S))}{d\mathring{\tau}^S} < -\frac{dD_S(\check{A}(\check{\tau}))}{d\check{\tau}}$. To see this, note that we have $\frac{d\mathring{\tau}^S}{d\mathring{A}(\mathring{\tau}^S)} = \frac{d\check{\tau}}{d\check{A}(\check{\tau})}$ since $\frac{d^3C(A_i)}{dA_i^3} = 0$ for all A_i. Therefore, $-\frac{dD_S(\mathring{A}(\mathring{\tau}^S))}{d\mathring{\tau}^S} < -\frac{dD_S(\check{A}(\check{\tau}))}{d\check{\tau}}$ if and only if $-\frac{dD_N(\mathring{A}(\mathring{\tau}^S))}{d\mathring{\tau}^S}\frac{d\mathring{\tau}^S}{d\mathring{A}(\mathring{\tau}^S)} < -\frac{dD_S(\check{A}(\check{\tau}))}{d\check{\tau}}\frac{d\check{\tau}}{d\check{A}(\check{\tau})}$, which simplifies to $-\frac{dD_S(\mathring{A}(\mathring{\tau}^S))}{d\mathring{A}(\mathring{\tau}^S)} < -\frac{dD_S(\check{A}(\check{\tau}))}{d\check{A}(\check{\tau})}$. We know that this inequality holds from the strict convexity of the damage function and since $\check{\tau} < \mathring{\tau}^S$ implies $\check{A}_i(\check{\tau}) < \mathring{A}_i(\mathring{\tau}^S)$ for all i for strictly convex abatement cost functions.

Moreover, we have $\frac{dC_S(\mathring{A}(\mathring{\tau}^S))}{d\mathring{\tau}^S} > \frac{dC_S(\check{A}(\check{\tau}))}{d\check{\tau}}$ for the South. To see this, note that we can write this as $\frac{dC_S(\mathring{A}(\mathring{\tau}^S))}{d\mathring{A}(\mathring{\tau}^S)}\frac{d\mathring{A}(\mathring{\tau}^S)}{d\mathring{\tau}^S} > \frac{dC_S(\check{A}(\check{\tau}))}{d\check{A}(\check{\tau})}\frac{d\check{A}(\check{\tau})}{d\check{\tau}}$, which in turn can be rewritten as $\mathring{\tau}^S\frac{1}{C_S''(\mathring{A}(\mathring{\tau}^S))} > \check{\tau}\frac{1}{C_S''(\check{A}(\check{\tau}))}$. This inequality holds since $\check{\tau} < \mathring{\tau}^S$ and $\frac{d^3C(A_i)}{dA_i^3} = 0$ for all A_i.

[42] This can be seen as follows:

$$\frac{d\tau}{dA(\tau)} = \frac{d\tau}{dA_N(\tau) + dA_S(\tau)} = \frac{d\tau}{\frac{1}{C_N''(A_N(\tau))}d\tau + \frac{1}{C_S''(A_S(\tau))}d\tau} = \frac{1}{\frac{1}{C_N''(A_N(\tau))} + \frac{1}{C_S''(A_S(\tau))}},$$

where the second equality holds since $C_i''(A_i(\tau)) = \frac{d\tau}{dA_i(\tau)}$. Notice that the last term is constant since $\frac{d^3C(A_i)}{dA_i^3} = 0$.

Together, this establishes the following inequalities for the South:

$$1 = \frac{-\frac{dD_S(\mathring{A}(\mathring{\tau}^S))}{d\mathring{\tau}^S}}{\frac{dC_S(\mathring{A}_S(\mathring{\tau}^S))}{d\mathring{\tau}^S}} > \frac{-\frac{dD_S(\check{A}(\check{\tau}))}{d\check{\tau}}}{\frac{dC_S(\mathring{A}_S(\mathring{\tau}^S))}{d\mathring{\tau}^S}} > \frac{-\frac{dD_S(\check{A}(\check{\tau}))}{d\check{\tau}}}{\frac{dC_S(\check{A}_S(\check{\tau}))}{d\check{\tau}}}. \tag{A49}$$

We have thus shown that

$$\frac{-\frac{d\check{D}_S}{d\check{\tau}}}{\frac{d\check{C}_S}{d\check{\tau}}} > 1 > \frac{-\frac{d\check{D}_N}{d\check{\tau}}}{\frac{d\check{C}_N}{d\check{\tau}}}. \tag{A50}$$

A.6 Proof of Lemma 1

Proof. We start by showing that $\hat{\tau}_S < \tilde{\tau}$. Suppose, towards a contradiction, that $\hat{\tau}_S \geq \tilde{\tau}$, which is the case if and only if

$$-\tilde{D}'_N - \tilde{D}'_S \leq -\hat{D}'_S - \frac{\hat{u}'_N}{\hat{u}'_S}\hat{D}'_N. \tag{A51}$$

Since $\frac{\hat{u}'_N}{\hat{u}'_S} < 1$ this inequality is satisfied if and only if $\hat{A} < \tilde{A}$, and thus $\hat{D}'_i < \tilde{D}'_i$ for all i. From the definition of the utilitarian differentiated carbon price, we know that $\hat{\tau}_S < \hat{\tau}_N$. However, $\tilde{\tau} \leq \hat{\tau}_S < \hat{\tau}_N$ implies $\hat{A} > \tilde{A}$, and we have thus arrived at a contradiction. Therefore, we must have $\hat{\tau}_S < \tilde{\tau}$. $\hat{A}_S < \tilde{A}_S$ follows from the strict convexity of the abatement cost function (and the definitions of the optimal carbon prices).

Next, we show that $\tilde{\tau} < \hat{\tau}_N$. Suppose, towards a contradiction, that $\tilde{\tau} \geq \hat{\tau}_N$, which is the case if and only if

$$-\tilde{D}'_N - \tilde{D}'_S \geq -\frac{\hat{u}'_S}{\hat{u}'_N}\hat{D}'_S - \hat{D}'_N. \tag{A52}$$

Since $\frac{\hat{u}'_S}{\hat{u}'_N} > 1$ this inequality is satisfied if and only if $\hat{A} > \tilde{A}$, and thus $\hat{D}'_i > \tilde{D}'_i$ for all i. From the definition of the utilitarian differentiated carbon price, we know that $\hat{\tau}_S < \hat{\tau}_N$. However, $\tilde{\tau} \geq \hat{\tau}_N > \hat{\tau}_S$ implies $\hat{A} < \tilde{A}$, and we have thus arrived at a contradiction. Therefore, we must have $\tilde{\tau} < \hat{\tau}_N$. $\hat{A}_N > \tilde{A}_N$ follows from the strict convexity of the abatement cost function (and the definitions of the optimal carbon prices). □

A.7 Proof of Proposition 3

Proof. We first need to obtain expressions for the abatement as a function of the marginal abatement cost. As stated in the main text, the proof of this proposition makes use of the following functional form assumption on the abatement cost function: $C_i(A_i) = k_i A_i^2$. The marginal abatement cost is thus $C'_i(A_i) = 2k_i A_i$ and the second derivative is $C''_i(A_i) = 2k_i$.

Therefore, $k_i = \frac{C_i''}{2}$.

We invert the marginal abatement cost function to obtain an expression for the abatement:

$$A_i = \frac{1}{2} C_i' k_i^{-1}. \tag{A53}$$

We split the proof into the forward and backward implications.

Proof of forward direction: $\hat{A} > \tilde{A} \implies \frac{\hat{u}_S'}{\hat{u}_N'} \frac{\hat{D}_S'}{\hat{D}_N'} > \frac{C_N''}{C_S''}$.

We start by establishing the implications of $\hat{A} > \tilde{A}$. First, $\hat{A} > \tilde{A}$ implies $\hat{A}_N + \hat{A}_S > \tilde{A}_N + \tilde{A}_S$. Therefore, $\hat{A} > \tilde{A}$ if and only if

$$\tilde{C}_N' k_N^{-1} + \tilde{C}_S' k_S^{-1} < \hat{C}_N' k_N^{-1} + \hat{C}_S' k_S^{-1}. \tag{A54}$$

Plugging in the expressions for the marginal abatement costs detailed in Definitions 11 and 19, and rewriting, we get

$$\left(\hat{D}_N' - \tilde{D}_N' + \frac{\hat{u}_S'}{\hat{u}_N'} \hat{D}_S' - \tilde{D}_S' \right) k_N^{-1} < \left(\tilde{D}_S' - \hat{D}_S' + \tilde{D}_N' - \frac{\hat{u}_N'}{\hat{u}_S'} \hat{D}_N' \right) k_S^{-1}. \tag{A55}$$

Next, note that $\tilde{A} < \hat{A}$ implies $\tilde{D}_i' < \hat{D}_i'$, for all i. Therefore, the previous inequality implies[43]

$$\left(\frac{\hat{u}_S'}{\hat{u}_N'} \hat{D}_S' - \hat{D}_S' \right) k_N^{-1} < \left(\hat{D}_N' - \frac{\hat{u}_N'}{\hat{u}_S'} \hat{D}_N' \right) k_S^{-1}, \tag{A56}$$

which we can rewrite as (recall that $\hat{D}_i' < 0$ so the inequality flips)

$$\frac{\hat{u}_S'}{\hat{u}_N'} \frac{\hat{D}_S'}{\hat{D}_N'} > \frac{k_N}{k_S}. \tag{A57}$$

[43] To see this, note that $\tilde{D}_i' < \hat{D}_i'$ implies the following inequalities:

$$\begin{aligned} \left(\frac{\hat{u}_S'}{\hat{u}_N'} \hat{D}_S' - \hat{D}_S' \right) k_N^{-1} &< \left(\frac{\hat{u}_S'}{\hat{u}_N'} \hat{D}_S' - \tilde{D}_S' \right) k_N^{-1} < \left(\hat{D}_N' - \tilde{D}_N' + \frac{\hat{u}_S'}{\hat{u}_N'} \hat{D}_S' - \tilde{D}_S' \right) k_N^{-1} \\ &< \left(\tilde{D}_S' - \hat{D}_S' + \tilde{D}_N' - \frac{\hat{u}_N'}{\hat{u}_S'} \hat{D}_N' \right) k_S^{-1} < \left(\tilde{D}_N' - \frac{\hat{u}_N'}{\hat{u}_S'} \hat{D}_N' \right) k_S^{-1} < \left(\hat{D}_N' - \frac{\hat{u}_N'}{\hat{u}_S'} \hat{D}_N' \right) k_S^{-1}. \end{aligned}$$

Using $k_i = \frac{C_i''}{2}$, we get

$$\frac{\hat{u}_N'}{\hat{u}_S'}\frac{\hat{D}_S'}{\hat{D}_N'} > \frac{C_N''}{C_S''}. \tag{A58}$$

Proof of backward direction: $\frac{\hat{D}_S'}{\hat{D}_N'} > \frac{\hat{u}_N'}{\hat{u}_S'}\frac{C_N''}{C_S''} \implies \hat{A} > \tilde{A}$.

In order to derive a contradiction, suppose that $\frac{\hat{D}_S'}{\hat{D}_N'} > \frac{\hat{u}_N'}{\hat{u}_S'}\frac{C_N''}{C_S''} \implies \hat{A} \leq \tilde{A}$. We start by establishing the implications of $\hat{A} \leq \tilde{A}$. $\hat{A} \leq \tilde{A}$ implies $\hat{A}_N + \hat{A}_S \leq \tilde{A}_N + \tilde{A}_S$. Therefore, $\hat{A} \leq \tilde{A}$ if and only if

$$\tilde{C}_N' k_N^{-1} + \tilde{C}_S' k_S^{-1} \geq \hat{C}_N' k_N^{-1} + \hat{C}_S' k_S^{-1}. \tag{A59}$$

Plugging in the expressions for the marginal abatement costs detailed in Definitions 11 and 19, and rewriting, we get

$$\left(\hat{D}_N' - \tilde{D}_N' + \frac{\hat{u}_S'}{\hat{u}_N'}\hat{D}_S' - \tilde{D}_S'\right) k_N^{-1} \geq \left(\tilde{D}_S' - \hat{D}_S' + \tilde{D}_N' - \frac{\hat{u}_N'}{\hat{u}_S'}\hat{D}_N'\right) k_S^{-1}. \tag{A60}$$

Using $k_i = \frac{C_i''}{2}$, we get

$$\left(\hat{D}_N' - \tilde{D}_N' + \frac{\hat{u}_S'}{\hat{u}_N'}\hat{D}_S' - \tilde{D}_S'\right) \frac{1}{C_N''} \geq \left(\tilde{D}_S' - \hat{D}_S' + \tilde{D}_N' - \frac{\hat{u}_N'}{\hat{u}_S'}\hat{D}_N'\right) \frac{1}{C_S''}. \tag{A61}$$

Next, note that $\tilde{A} \geq \hat{A}$ implies $\tilde{D}_i' \geq \hat{D}_i'$, for all i. $\tilde{D}_i' \geq \hat{D}_i'$ for all i and $\hat{u}_S' > \hat{u}_N'$ imply

$$\hat{D}_N' - \tilde{D}_N' + \frac{\hat{u}_S'}{\hat{u}_N'}\hat{D}_S' - \tilde{D}_S' < 0. \tag{A62}$$

Moreover, Equations (A61) and (A62) imply

$$\tilde{D}_S' - \hat{D}_S' + \tilde{D}_N' - \frac{\hat{u}_N'}{\hat{u}_S'}\hat{D}_N' < 0, \tag{A63}$$

since $C_i'' > 0$ for all i.

We therefore know that the inequality flips upon cross-division, yielding

$$\frac{C_N''}{C_S''} \geq \frac{\hat{D}_N' - \tilde{D}_N' + \frac{\hat{u}_S'}{\hat{u}_N'}\hat{D}_S' - \tilde{D}_S'}{\tilde{D}_S' - \hat{D}_S' + \tilde{D}_N' - \frac{\hat{u}_N'}{\hat{u}_S'}\hat{D}_N'}. \tag{A64}$$

Multiplying both sides by $\frac{\hat{u}'_N}{\hat{u}'_S}$ and collecting common terms, we have

$$\frac{\hat{u}'_N}{\hat{u}'_S}\frac{C''_N}{C''_S} \geq \frac{\hat{u}'_N\left(\hat{D}'_N - \tilde{D}'_N - \tilde{D}'_S\right) + \hat{u}'_S\hat{D}'_S}{\hat{u}'_S\left(\tilde{D}'_S - \hat{D}'_S + \tilde{D}'_N\right) - \hat{u}'_N\hat{D}'_N}. \tag{A65}$$

So far, we have established that

$$\tilde{A} \geq \hat{A} \iff \frac{\hat{u}'_N}{\hat{u}'_S}\frac{C''_N}{C''_S} \geq \frac{\hat{u}'_N\left(\hat{D}'_N - \tilde{D}'_N - \tilde{D}'_S\right) + \hat{u}'_S\hat{D}'_S}{\hat{u}'_S\left(\tilde{D}'_S - \hat{D}'_S + \tilde{D}'_N\right) - \hat{u}'_N\hat{D}'_N}. \tag{A66}$$

Next, we show that $\frac{\hat{u}'_S}{\hat{u}'_N}\frac{\hat{D}'_S}{\hat{D}'_N} > \frac{C''_N}{C''_S} \implies \hat{A} \leq \tilde{A}$ yields a contradiction. We start by rearranging $\frac{\hat{u}'_S}{\hat{u}'_N}\frac{\hat{D}'_S}{\hat{D}'_N} > \frac{C''_N}{C''_S}$ to $\frac{\hat{D}'_S}{\hat{D}'_N} > \frac{\hat{u}'_N}{\hat{u}'_S}\frac{C''_N}{C''_S}$. We then obtain the following contradiction:

$$\begin{aligned}
\frac{\hat{u}'_N}{\hat{u}'_S}\frac{C''_N}{C''_S} &< \frac{\hat{D}'_S}{\hat{D}'_N} \\
&= \frac{\hat{u}'_S\hat{D}'_S - \hat{u}'_N\hat{D}'_S}{\hat{u}'_S\hat{D}'_N - \hat{u}'_N\hat{D}'_N} \\
&\leq \frac{\hat{u}'_S\hat{D}'_S - \hat{u}'_N\tilde{D}'_S}{\hat{u}'_S\tilde{D}'_N - \hat{u}'_N\hat{D}'_N} \\
&\leq \frac{\hat{u}'_N\left(\hat{D}'_N - \tilde{D}'_N - \tilde{D}'_S\right) + \hat{u}'_S\hat{D}'_S}{\hat{u}'_S\left(\tilde{D}'_S - \hat{D}'_S + \tilde{D}'_N\right) - \hat{u}'_N\hat{D}'_N} \\
&\leq \frac{\hat{u}'_N}{\hat{u}'_S}\frac{C''_N}{C''_S}.
\end{aligned} \tag{A67}$$

where the second and third inequalities follow from $\tilde{D}'_i \geq \check{D}'_i$ for all i and the fact that the denominator and the numerator are negative by Equations (A62) and (A63) [44]. The last inequality follows from the implication of $\tilde{A} \geq \hat{A}$ documented in Equation (A66).

We have reached the contradiction $\frac{\hat{u}'_N}{\hat{u}'_S}\frac{C''_N}{C''_S} < \frac{\hat{u}'_N}{\hat{u}'_S}\frac{C''_N}{C''_S}$. Hence, $\frac{\hat{u}'_S}{\hat{u}'_N}\frac{\hat{D}'_S}{\hat{D}'_N} > \frac{C''_N}{C''_S} \implies \hat{A} \leq \tilde{A}$ is incorrect, and we have thus shown that we must have $\frac{\hat{u}'_S}{\hat{u}'_N}\frac{\hat{D}'_S}{\hat{D}'_N} > \frac{C''_N}{C''_S} \implies \hat{A} > \tilde{A}$.

[44] Note that the denominator in the third line is negative because it is less than the negative denominator in the fourth line. This can be seen as follows:

$$\hat{u}'_S\tilde{D}'_N - \hat{u}'_N\hat{D}'_N < \hat{u}'_S(\underbrace{\tilde{D}'_S - \hat{D}'_S}_{\geq 0} + \tilde{D}'_N) - \hat{u}'_N\hat{D}'_N < 0$$

Together, the proofs of the forward and backward implications yield the equivalence

$$\hat{A} > \tilde{A} \iff \frac{\hat{u}'_S}{\hat{u}'_N}\frac{\hat{D}'_S}{\hat{D}'_N} > \frac{C''_N}{C''_S}. \tag{A68}$$

□

A.8 Proof of Lemma 2

I first prove foundational lemmas which act as building blocks to prove Lemma 2.

Lemma 3. *North's preferred uniform carbon price is less than the utilitarian uniform carbon price, that is $\mathring{\tau}^N < \check{\tau}$, if and only if $\frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S}$.*

Proof. We split the proof into the forward and backward implications.

Proof of forward direction: $\check{\tau} > \mathring{\tau}^N \implies \frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S}$.

I start by establishing the conditions under which $\check{\tau} > \mathring{\tau}^N$, or equivalently, $\check{C}' > \mathring{C}^{N\prime}$[45]. First note that, for strictly convex abatement cost functions, $\check{C}' > \mathring{C}^{N\prime}$ implies $\check{A}_i > \mathring{A}^N_i$ for all i, and thus $\check{A} > \mathring{A}^N$. For strictly convex damage functions, this implies $\check{D}'_i > \mathring{D}^{N\prime}_i$ (note that marginal damages of abatement are negative) for all i.

We have $\check{C}' > \mathring{C}^{N\prime}$ if and only if

$$(-\check{u}'_N\check{D}'_N - \check{u}'_S\check{D}'_S)\frac{C''_S + C''_N}{\check{u}'_N C''_S + \check{u}'_S C''_N} > -\mathring{D}^{N\prime}_N \frac{C''_S + C''_N}{C''_S}, \tag{A69}$$

which can be rewritten as

$$-\frac{\check{u}'_S}{\check{u}'_N}\check{D}'_S > -\mathring{D}^{N\prime}_N\left(1 + \frac{\check{u}'_S C''_N}{\check{u}'_N C''_S}\right) + \check{D}'_N. \tag{A70}$$

Using $\check{D}'_i > \mathring{D}^{N\prime}_i$, we have

$$\begin{aligned} -\frac{\check{u}'_S}{\check{u}'_N}\check{D}'_S &> -\mathring{D}^{N\prime}_N\left(1 + \frac{\check{u}'_S C''_N}{\check{u}'_N C''_S}\right) + \check{D}'_N \\ &> -\check{D}'_N\left(1 + \frac{\check{u}'_S C''_N}{\check{u}'_N C''_S}\right) + \check{D}'_N \\ &= -\check{D}'_N\left(\frac{\check{u}'_S C''_N}{\check{u}'_N C''_S}\right). \end{aligned} \tag{A71}$$

[45] While this notation is somewhat cumbersome, I use the notation $\mathring{C}^{i\prime}$ for clarity and conciseness as a short-hand for $C'_i(\mathring{A}^i_i)$, and I drop the subscript to reflect that $C'_i(\mathring{A}^i_i) = C'_{-i}(\mathring{A}^i_{-i})$ under uniform carbon prices.

Rewriting yields

$$\frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S}. \tag{A72}$$

We have thus shown that $\check{C}' > \mathring{C}^{N\prime} \implies \frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S}$.

<u>Proof of backward direction</u>: $\frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S} \implies \check{\tau} > \mathring{\tau}^N$.

In order to derive a contradiction, suppose that $\frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S} \implies \check{\tau} \leq \mathring{\tau}^N$.

We start by establishing the implications of $\check{\tau} \leq \mathring{\tau}^N$, or equivalently, $\check{C}' \leq \mathring{C}^{N\prime}$. First note that, for strictly convex abatement cost functions, $\check{C}' \leq \mathring{C}^{N\prime}$ implies $\check{A}_i \leq \mathring{A}^N_i$ for all i, and thus $\check{A} \leq \mathring{A}^N$. For strictly convex damage functions, this implies $\check{D}'_i \leq \mathring{D}^{N\prime}_i$ (note that marginal damages of abatement are negative) for all i.

Next, note that $\check{C}' \leq \mathring{C}^{N\prime}$ if and only if

$$(-\check{u}'_N\check{D}'_N - \check{u}'_S\check{D}'_S)\frac{C''_S + C''_N}{\check{u}'_N C''_S + \check{u}'_S C''_N} \leq -\mathring{D}^{N\prime}_N\frac{C''_S + C''_N}{C''_S}, \tag{A73}$$

which can be rewritten as

$$-\frac{\check{u}'_S}{\check{u}'_N}\check{D}'_S \leq -\mathring{D}^{N\prime}_N\left(1 + \frac{\check{u}'_S C''_N}{\check{u}'_N C''_S}\right) + \check{D}'_N. \tag{A74}$$

Let us temporarily define $\delta_N \equiv \mathring{D}^{N\prime}_N - \check{D}'_N$[46]. We know that $\delta_N \geq 0$ since $\check{D}'_N \leq \mathring{D}^{N\prime}_N$. Substitute $\check{D}'_N = \mathring{D}^{N\prime}_N - \delta_N$ into the previous expression to obtain

$$-\frac{\check{u}'_S}{\check{u}'_N}\check{D}'_S \leq -\mathring{D}^{N\prime}_N\left(1 + \frac{\check{u}'_S C''_N}{\check{u}'_N C''_S}\right) + \mathring{D}^{N\prime}_N - \delta_N. \tag{A75}$$

Rewriting yields

$$\frac{\check{D}'_S}{\mathring{D}^{N\prime}_N} \leq \frac{C''_N}{C''_S} - \underbrace{\frac{\delta_N}{-\mathring{D}^{N\prime}_N}\frac{\check{u}'_N}{\check{u}'_S}}_{\geq 0}. \tag{A76}$$

So far, we have established that

$$\check{C}' \leq \mathring{C}^{N\prime} \iff \frac{\check{D}'_S}{\mathring{D}^{N\prime}_N} \leq \frac{C''_N}{C''_S} - \underbrace{\frac{\delta_N}{-\mathring{D}^{N\prime}_N}\frac{\check{u}'_N}{\check{u}'_S}}_{\geq 0}. \tag{A77}$$

[46] Note that I redefine δ_i below, keeping the same notation for simplicity.

Next, we show that $\frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S} \implies \check{C}' \leq \mathring{C}^{N\prime}$ yields a contradiction:

$$\frac{C''_N}{C''_S} < \frac{\check{D}'_S}{\check{D}'_N} \leq \frac{\check{D}'_S}{\mathring{D}^{N\prime}_N} \leq \frac{C''_N}{C''_S} - \underbrace{\frac{\delta_N}{-\mathring{D}^{N\prime}_N}\frac{\check{u}'_N}{\check{u}'_S}}_{\geq 0} \leq \frac{C''_N}{C''_S}, \tag{A78}$$

where the second inequality follows from $\check{D}'_i \leq \mathring{D}^{N\prime}_i$, and the third inequality follows from the implication of $\check{C}' \leq \mathring{C}^{N\prime}$ documented in Equation (A77).

We have reached the contradiction $\frac{C''_N}{C''_S} < \frac{C''_N}{C''_S}$. Hence, $\frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S} \implies \check{C}' \leq \mathring{C}^{N\prime}$ is incorrect, and we have thus shown that we must have $\frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S} \implies \check{C}' > \mathring{C}^{N\prime}$.

Together, the proofs of the forward and backward directions yield the equivalence

$$\check{\tau} > \mathring{\tau}^N \iff \frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S}. \tag{A79}$$

□

Lemma 4. *South's preferred uniform carbon price is greater than the utilitarian uniform carbon price, that is $\mathring{\tau}^S > \check{\tau}$, if and only if $\frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S}$.*

Proof. We split the proof into the forward and backward implications.

Proof of forward direction: $\mathring{\tau}^S > \check{\tau} \implies \frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S}$.

I start by establishing the conditions under which $\mathring{\tau}^S > \check{\tau}$, or equivalently, $\check{C}' < \mathring{C}^{S\prime}$. First note that, for strictly convex abatement cost functions, $\check{C}' < \mathring{C}^{S\prime}$ implies $\check{A}_i < \mathring{A}^S_i$ for all i, and thus $\check{A} < \mathring{A}^S$. For strictly convex damage functions, this implies $\check{D}'_i < \mathring{D}^{S\prime}_i$ (note that marginal damages of abatement are negative) for all i.

We have $\check{C}' < \mathring{C}^{S\prime}$ if and only if

$$(-\check{u}'_N \check{D}'_N - \check{u}'_S \check{D}'_S)\frac{C''_S + C''_N}{\check{u}'_N C''_S + \check{u}'_S C''_N} < -\mathring{D}^{S\prime}_S \frac{C''_S + C''_N}{C''_N}, \tag{A80}$$

which can be rewritten as

$$-\frac{\check{u}'_N}{\check{u}'_S}\check{D}'_N < -\mathring{D}^{S\prime}_S\left(1 + \frac{\check{u}'_N C''_S}{\check{u}'_S C''_N}\right) + \check{D}'_S. \tag{A81}$$

Using $\check{D}'_i < \mathring{D}^{S\prime}_i$, we have

$$\begin{aligned} -\frac{\check{u}'_N}{\check{u}'_S}\check{D}'_N &< -\mathring{D}^{S\prime}_S\left(1+\frac{\check{u}'_N C''_S}{\check{u}'_S C''_N}\right)+\check{D}'_S \\ &< -\check{D}'_S\left(1+\frac{\check{u}'_N C''_S}{\check{u}'_S C''_N}\right)+\check{D}'_S \\ &= -\check{D}'_S\left(\frac{\check{u}'_N C''_S}{\check{u}'_S C''_N}\right). \end{aligned} \tag{A82}$$

Rewriting yields

$$\frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S}. \tag{A83}$$

We have thus shown that $\check{C}' < \mathring{C}^{S\prime} \implies \frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S}$.

Proof of backward direction: $\frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S} \implies \check{\tau} < \mathring{\tau}^S$.

In order to derive a contradiction, suppose that $\frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S} \implies \check{\tau} \geq \mathring{\tau}^S$.

We start by establishing the implications of $\check{\tau} \geq \mathring{\tau}^S$, or equivalently $\check{C}' \geq \mathring{C}^{S\prime}$. First note that, for strictly convex abatement cost functions, $\check{C}' \geq \mathring{C}^{S\prime}$ implies $\check{A}_i \geq \mathring{A}^S_i$ for all i, and thus $\check{A} \geq \mathring{A}^S$. For strictly convex damage functions, this implies $\check{D}'_i \geq \mathring{D}^{S\prime}_i$ (note that marginal damages of abatement are negative) for all i.

Next, note that $\check{C}' \geq \mathring{C}^{S\prime}$ if and only if

$$(-\check{u}'_N\check{D}'_N - \check{u}'_S\check{D}'_S)\frac{C''_S + C''_N}{\check{u}'_N C''_S + \check{u}'_S C''_N} \geq -\mathring{D}^{S\prime}_S\frac{C''_S + C''_N}{C''_N}, \tag{A84}$$

which can be rewritten as

$$-\frac{\check{u}'_N}{\check{u}'_S}\check{D}'_N \geq -\mathring{D}^{S\prime}_S\left(1+\frac{\check{u}'_N C''_S}{\check{u}'_S C''_N}\right)+\check{D}'_S. \tag{A85}$$

Let us temporarily define $\delta_S \equiv \check{D}'_S - \mathring{D}^{S\prime}_S$. We know that $\delta_S \geq 0$ since $\check{D}'_S \geq \mathring{D}^{S\prime}_S$. Substitute $\check{D}'_S = \mathring{D}^{S\prime}_S + \delta_S$ into the previous expression to obtain

$$-\frac{\check{u}'_N}{\check{u}'_S}\check{D}'_N \geq -\mathring{D}^{S\prime}_S\left(1+\frac{\check{u}'_N C''_S}{\check{u}'_S C''_N}\right)+\mathring{D}^{S\prime}_i + \delta_i. \tag{A86}$$

Simplifying and rearranging yields

$$\frac{\check{D}'_N}{\mathring{D}^{S\prime}_S} \geq \frac{C''_S}{C''_N} + \underbrace{\frac{\delta_S}{-\mathring{D}^{S\prime}_S}\frac{\check{u}'_S}{\check{u}'_N}}_{\geq 0}. \tag{A87}$$

So far, we have established that

$$\check{C}' \geq \mathring{C}^{S\prime} \iff \frac{\check{D}'_N}{\mathring{D}^{S\prime}_S} \geq \frac{C''_S}{C''_N} + \underbrace{\frac{\delta_S}{-\mathring{D}^{S\prime}_S}\frac{\check{u}'_S}{\check{u}'_N}}_{\geq 0}. \tag{A88}$$

Next, we show that $\frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S} \implies \check{C}' \geq \mathring{C}^{S\prime}$ yields a contradiction. We start by rearranging $\frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S}$ to $\frac{\check{D}'_N}{\check{D}'_S} < \frac{C''_S}{C''_N}$. We then obtain the following contradiction:

$$\frac{C''_S}{C''_N} > \frac{\check{D}'_N}{\check{D}'_S} \geq \frac{\check{D}'_N}{\mathring{D}^{S\prime}_S} \geq \frac{C''_S}{C''_N} + \underbrace{\frac{\delta_S}{-\mathring{D}^{S\prime}_S}\frac{\check{u}'_S}{\check{u}'_N}}_{\geq 0} \geq \frac{C''_S}{C''_N}. \tag{A89}$$

where the second inequality follows from $\check{D}'_i \geq \mathring{D}^{S\prime}_i$, and the third inequality follows from the implication of $\check{C}' \geq \mathring{C}^{S\prime}$ documented in Equation (A88).

We have reached the contradiction $\frac{C''_S}{C''_N} > \frac{C''_S}{C''_N}$. Hence, $\frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S} \implies \check{C}' \geq \mathring{C}^{S\prime}$ is incorrect, and we have thus shown that we must have $\frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S} \implies \check{C}' < \mathring{C}^{S\prime}$.

Together, the proofs of the forward and backward directions yield the equivalence

$$\check{\tau} < \mathring{\tau}^S \iff \frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S}. \tag{A90}$$

□

Lemma 5. *North's preferred uniform carbon price is less than the Negishi-weighted carbon price, that is* $\mathring{\tau}^N < \tilde{\tau}$, *if and only if* $\frac{\tilde{D}'_S}{\tilde{D}'_N} > \frac{C''_N}{C''_S}$.

Proof. We split the proof into the forward and backward implications.

Proof of forward direction: $\tilde{\tau} > \mathring{\tau}^N \implies \frac{\tilde{D}'_S}{\tilde{D}'_N} > \frac{C''_N}{C''_S}$.

I start by establishing the conditions under which $\tilde{\tau} > \mathring{\tau}^N$, or equivalently, $\tilde{C}' > \mathring{C}^{N\prime}$. First note that, for strictly convex abatement cost functions, $\tilde{C}' > \mathring{C}^{N\prime}$ implies $\tilde{A}_i > \mathring{A}^N_i$ for all i, and thus $\tilde{A} > \mathring{A}^N$. For strictly convex damage functions, this implies $\tilde{D}'_i > \mathring{D}^{N\prime}_i$ (note that marginal damages of abatement are negative) for all i.

We have $\tilde{C}' > \mathring{C}^{N\prime}$ if and only if

$$-\tilde{D}'_N - \tilde{D}'_S > -\mathring{D}^{N\prime}_N \frac{C''_S + C''_N}{C''_S}, \tag{A91}$$

which we can rewrite as (note that $\mathring{D}^{N\prime}_N$ is negative so the sign of the inequality flips)

$$\frac{C''_S + C''_N}{C''_S} < \frac{\tilde{D}'_N}{\mathring{D}^{N\prime}_N} + \frac{\tilde{D}'_S}{\mathring{D}^{N\prime}_N}. \tag{A92}$$

Let us temporarily define $\delta_N \equiv \mathring{D}^{N\prime}_N - \tilde{D}'_N$. We know that $\delta_N < 0$ since $\tilde{D}'_N > \mathring{D}^{N\prime}_N$. Substitute $\tilde{D}'_N = \mathring{D}^{N\prime}_N - \delta_N$ into the previous expression to obtain

$$\frac{C''_S + C''_N}{C''_S} < \frac{\mathring{D}^{N\prime}_N - \delta_N}{\mathring{D}^{N\prime}_N} + \frac{\tilde{D}'_S}{\mathring{D}^{N\prime}_N}, \tag{A93}$$

which simplifies to

$$\frac{C''_N}{C''_S} < \frac{\tilde{D}'_S}{\mathring{D}^{N\prime}_N} + \frac{-\delta_N}{\mathring{D}^{N\prime}_N}. \tag{A94}$$

We can now establish the following inequalities:

$$\frac{C''_N}{C''_S} < \frac{\tilde{D}'_S}{\mathring{D}^{N\prime}_N} + \underbrace{\frac{-\delta_N}{\mathring{D}^{N\prime}_N}}_{<0} < \frac{\tilde{D}'_S}{\mathring{D}^{N\prime}_N} < \frac{\tilde{D}'_S}{\tilde{D}'_N}. \tag{A95}$$

where the last inequality follows from $\tilde{D}'_N > \mathring{D}^{N\prime}_N$.

We have thus shown that $\tilde{C}' > \mathring{C}^{N\prime} \implies \frac{\tilde{D}'_S}{\tilde{D}'_N} > \frac{C''_N}{C''_S}$.

Proof of backward direction: $\frac{\tilde{D}'_S}{\tilde{D}'_N} > \frac{C''_N}{C''_S} \implies \tilde{\tau} > \mathring{\tau}^N$.

In order to derive a contradiction, suppose that $\frac{\tilde{D}'_S}{\tilde{D}'_N} > \frac{C''_N}{C''_S} \implies \tilde{\tau} \leq \mathring{\tau}^N$.

We start by establishing the implications of $\tilde{\tau} \leq \mathring{\tau}^N$, or equivalently, $\tilde{C}' \leq \mathring{C}^{N\prime}$. First note that, for strictly convex abatement cost functions, $\tilde{C}' \leq \mathring{C}^{N\prime}$ implies $\tilde{A}_i \leq \mathring{A}^N_i$ for all i, and thus $\tilde{A} \leq \mathring{A}^N$. For strictly convex damage functions, this implies $\tilde{D}'_i \leq \mathring{D}^{N\prime}_i$ (note that marginal damages of abatement are negative) for all i.

Next, note that $\tilde{C}' \leq \mathring{C}^{N\prime}$ if and only if

$$-\tilde{D}'_N - \tilde{D}'_S \leq -\mathring{D}^{N\prime}_N \frac{C''_S + C''_N}{C''_S}, \tag{A96}$$

which we can rewrite as (note that $\mathring{D}_N^{N\prime}$ is negative so the sign of the inequality flips)

$$\frac{C_S'' + C_N''}{C_S''} \geq \frac{\tilde{D}_N'}{\mathring{D}_N^{N\prime}} + \frac{\tilde{D}_S'}{\mathring{D}_N^{N\prime}}. \tag{A97}$$

Let us temporarily define $\delta_N \equiv \mathring{D}_N^{N\prime} - \tilde{D}_N'$. We know that $\delta_N \geq 0$ since $\tilde{D}_N' \leq \mathring{D}_N^{N\prime}$. Substitute $\tilde{D}_N' = \mathring{D}_N^{N\prime} - \delta_N$ into the previous expression to obtain

$$\frac{C_S'' + C_N''}{C_S''} \geq \frac{\mathring{D}_N^{N\prime} - \delta_N}{\mathring{D}_N^{N\prime}} + \frac{\tilde{D}_S'}{\mathring{D}_N^{N\prime}}, \tag{A98}$$

which simplifies to

$$\frac{C_N''}{C_S''} \geq \frac{\tilde{D}_S'}{\mathring{D}_N^{N\prime}} + \underbrace{\frac{-\delta_N}{\mathring{D}_N^{N\prime}}}_{\geq 0}. \tag{A99}$$

So far, we have established that

$$\tilde{C}' \leq \mathring{C}^{N\prime} \iff \frac{C_N''}{C_S''} \geq \frac{\tilde{D}_S'}{\mathring{D}_N^{N\prime}} + \underbrace{\frac{-\delta_N}{\mathring{D}_N^{N\prime}}}_{\geq 0}. \tag{A100}$$

Next, we show that $\frac{\tilde{D}_S'}{\tilde{D}_N'} > \frac{C_N''}{C_S''} \implies \tilde{C}' \leq \mathring{C}^{N\prime}$ yields a contradiction.

$$\frac{C_N''}{C_S''} < \frac{\tilde{D}_S'}{\tilde{D}_N'} \leq \frac{\tilde{D}_S'}{\mathring{D}_N^{N\prime}} \leq \frac{\tilde{D}_S'}{\mathring{D}_N^{N\prime}} + \underbrace{\frac{-\delta_N}{\mathring{D}_N^{N\prime}}}_{\geq 0} \leq \frac{C_N''}{C_S''}. \tag{A101}$$

where the second and third inequalities follow from $\tilde{D}_i' \leq \mathring{D}_i^{N\prime}$ for all i, and the last inequality follows from the implication of $\tilde{C}' \leq \mathring{C}^{N\prime}$ documented in Equation (A100).

We have reached the contradiction $\frac{C_N''}{C_S''} < \frac{C_N''}{C_S''}$. Hence, $\frac{\tilde{D}_S'}{\tilde{D}_N'} > \frac{C_N''}{C_S''} \implies \tilde{C}' \leq \mathring{C}^{N\prime}$ is incorrect, and we have thus shown that we must have $\frac{\tilde{D}_S'}{\tilde{D}_N'} > \frac{C_N''}{C_S''} \implies \tilde{C}' > \mathring{C}^{N\prime}$.

Together, the proofs of the forward and backward directions yield the equivalence

$$\tilde{\tau} > \mathring{\tau}^N \iff \frac{\tilde{D}_S'}{\tilde{D}_N'} > \frac{C_N''}{C_S''}. \tag{A102}$$

□

Lemma 6. *South's preferred uniform carbon price is greater than the Negishi-weighted carbon* *price, that is* $\mathring{\tau}^S > \tilde{\tau}$*, if and only if* $\frac{\tilde{D}'_S}{\tilde{D}'_N} > \frac{C''_N}{C''_S}$*.*

Proof. We split the proof into the forward and backward implications.

Proof of forward direction: $\tilde{\tau} < \mathring{\tau}^S \implies \frac{\tilde{D}'_S}{\tilde{D}'_N} > \frac{C''_N}{C''_S}$.

I start by establishing the conditions under which $\tilde{\tau} < \mathring{\tau}^S$, or equivalently, $\tilde{C}' < \mathring{C}^{S\prime}$. First note that, for strictly convex abatement cost functions, $\tilde{C}' < \mathring{C}^{S\prime}$ implies $\tilde{A}_i < \mathring{A}_i^S$ for all i, and thus $\tilde{A} < \mathring{A}^S$. For strictly convex damage functions, this implies $\tilde{D}'_i < \mathring{D}_i^{S\prime}$ (note that marginal damages of abatement are negative) for all i.

We have $\tilde{C}' < \mathring{C}^{S\prime}$ if and only if

$$-\tilde{D}'_N - \tilde{D}'_S < -\mathring{D}_S^{S\prime} \frac{C''_S + C''_N}{C''_N}, \tag{A103}$$

which we can rewrite as (note that $\mathring{D}_S^{S\prime}$ is negative so the sign of the inequality flips)

$$\frac{C''_S + C''_N}{C''_N} > \frac{\tilde{D}'_N}{\mathring{D}_S^{S\prime}} + \frac{\tilde{D}'_S}{\mathring{D}_S^{S\prime}}. \tag{A104}$$

Let us temporarily define $\delta_S \equiv \mathring{D}_S^{S\prime} - \tilde{D}'_S$. We know that $\delta_S > 0$ since $\tilde{D}'_S < \mathring{D}_S^{S\prime}$. Substitute $\tilde{D}'_S = \mathring{D}_S^{S\prime} - \delta_S$ into the previous expression to obtain

$$\frac{C''_S + C''_N}{C''_N} > \frac{\tilde{D}'_N}{\mathring{D}_S^{S\prime}} + \frac{\mathring{D}_S^{S\prime} - \delta_S}{\mathring{D}_S^{S\prime}}, \tag{A105}$$

which simplifies to

$$\frac{C''_S}{C''_N} > \frac{\tilde{D}'_N}{\mathring{D}_S^{S\prime}} + \frac{-\delta_S}{\mathring{D}_S^{S\prime}}. \tag{A106}$$

We can now establish the following inequalities:

$$\frac{C''_S}{C''_N} > \frac{\tilde{D}'_N}{\mathring{D}_S^{S\prime}} + \underbrace{\frac{-\delta_S}{\mathring{D}_S^{S\prime}}}_{>0} > \frac{\tilde{D}'_N}{\mathring{D}_S^{S\prime}} > \frac{\tilde{D}'_N}{\tilde{D}'_S}. \tag{A107}$$

where the last inequality follows from $\tilde{D}'_S < \mathring{D}_S^{S\prime}$.

We have thus shown that $\tilde{C}' < \mathring{C}^{S\prime} \implies \frac{\tilde{D}'_S}{\tilde{D}'_N} > \frac{C''_N}{C''_S}$.

Proof of backward direction: $\frac{\tilde{D}'_S}{\tilde{D}'_N} > \frac{C''_N}{C''_S} \implies \tilde{\tau} < \mathring{\tau}^S$.

In order to derive a contradiction, suppose that $\frac{\tilde{D}'_S}{\tilde{D}'_N} > \frac{C''_N}{C''_S} \implies \tilde{\tau} \geq \mathring{\tau}^S$.

We start by establishing the implications of $\tilde{\tau} \geq \mathring{\tau}^S$, or equivalently, $\tilde{C}' \geq \mathring{C}^{S\prime}$. First note that, for strictly convex abatement cost functions, $\tilde{C}' \geq \mathring{C}^{S\prime}$ implies $\tilde{A}_i \geq \mathring{A}_i^S$ for all i, and thus $\tilde{A} \geq \mathring{A}^S$. For strictly convex damage functions, this implies $\tilde{D}_i' \geq \mathring{D}_i^{S\prime}$ (note that marginal damages of abatement are negative) for all i.

Next, note that $\tilde{C}' \geq \mathring{C}^{S\prime}$ if and only if

$$-\tilde{D}_N' - \tilde{D}_S' \geq -\mathring{D}_S^{S\prime} \frac{C_S'' + C_N''}{C_N''}, \tag{A108}$$

which we can rewrite as (note that $\mathring{D}_S^{S\prime}$ is negative so the sign of the inequality flips)

$$\frac{C_S'' + C_N''}{C_N''} \leq \frac{\tilde{D}_N'}{\mathring{D}_S^{S\prime}} + \frac{\tilde{D}_S'}{\mathring{D}_S^{S\prime}}. \tag{A109}$$

Let us temporarily define $\delta_S \equiv \mathring{D}_S^{S\prime} - \tilde{D}_S'$. We know that $\delta_S \leq 0$ since $\tilde{D}_S' \geq \mathring{D}_S^{S\prime}$. Substitute $\tilde{D}_S' = \mathring{D}_S^{S\prime} - \delta_S$ into the previous expression to obtain

$$\frac{C_S'' + C_N''}{C_N''} \leq \frac{\tilde{D}_N'}{\mathring{D}_S^{S\prime}} + \frac{\mathring{D}_S^{S\prime} - \delta_S}{\mathring{D}_S^{S\prime}}, \tag{A110}$$

which simplifies to

$$\frac{C_S''}{C_N''} \leq \frac{\tilde{D}_N'}{\mathring{D}_S^{S\prime}} + \underbrace{\frac{-\delta_S}{\mathring{D}_S^{S\prime}}}_{\leq 0}. \tag{A111}$$

So far, we have established that

$$\tilde{C}' \geq \mathring{C}^{S\prime} \iff \frac{C_S''}{C_N''} \leq \frac{\tilde{D}_N'}{\mathring{D}_S^{S\prime}} + \underbrace{\frac{-\delta_S}{\mathring{D}_S^{S\prime}}}_{\leq 0}. \tag{A112}$$

Next, we show that $\frac{\tilde{D}_S'}{\tilde{D}_N'} > \frac{C_N''}{C_S''} \implies \tilde{C}' \geq \mathring{C}^{S\prime}$ yields a contradiction. We start by rearranging $\frac{\tilde{D}_S'}{\tilde{D}_N'} > \frac{C_N''}{C_S''}$ to $\frac{\tilde{D}_N'}{\tilde{D}_S'} < \frac{C_S''}{C_N''}$. We then obtain the following contradiction:

$$\frac{C_S''}{C_N''} > \frac{\tilde{D}_N'}{\tilde{D}_S'} \geq \frac{\tilde{D}_N'}{\mathring{D}_S^{S\prime}} \geq \frac{\tilde{D}_N'}{\mathring{D}_S^{S\prime}} + \underbrace{\frac{-\delta_S}{\mathring{D}_S^{S\prime}}}_{\leq 0} \geq \frac{C_S''}{C_N''}, \tag{A113}$$

where the second and third inequalities follow from $\tilde{D}_i' \geq \mathring{D}_i^{S\prime}$ for all i, and the last inequality follows from the implication of $\tilde{C}' \geq \mathring{C}^{S\prime}$ documented in Equation (A112).

We have reached the contradiction $\frac{C''_S}{C''_N} > \frac{C''_S}{C''_N}$. Hence, $\frac{\tilde{D}'_S}{\tilde{D}'_N} > \frac{C''_N}{C''_S} \implies \tilde{C}' \geq \mathring{C}^{S\prime}$ is incorrect, and we have thus shown that we must have $\frac{\tilde{D}'_S}{\tilde{D}'_N} > \frac{C''_N}{C''_S} \implies \tilde{C}' < \mathring{C}^{S\prime}$.

Together, the proofs of the forward and backward directions yield the equivalence

$$\tilde{\tau} < \mathring{\tau}^S \iff \frac{\tilde{D}'_S}{\tilde{D}'_N} > \frac{C''_N}{C''_S}. \tag{A114}$$

□

Using Lemmas 3-6, we can now prove Lemma 2, which is restated below.

Lemma 2. *The utilitarian uniform carbon price ($\check{\tau}$) and the Negishi-weighted carbon price ($\tilde{\tau}$) are in between North's ($\mathring{\tau}^N$) and South's ($\mathring{\tau}^S$) preferred uniform carbon prices, unless they all coincide.*

Proof. Let us begin by showing that the utilitarian uniform carbon price lies between North's and South's preferred uniform carbon prices, unless they coincide. Lemma 3 and 4 imply that $\mathring{\tau}^S > \tilde{\tau} > \mathring{\tau}^N$ if and only if $\frac{\tilde{D}'_S}{\tilde{D}'_N} > \frac{C''_N}{C''_S}$. From Proposition 1, we know that the optimal uniform carbon price does not depend on the welfare weights if and only if $\frac{\tilde{D}'_S}{\tilde{D}'_N} = \frac{C''_N}{C''_S}$. Therefore, $\mathring{\tau}^S = \tilde{\tau} = \mathring{\tau}^N$ if and only if $\frac{\tilde{D}'_S}{\tilde{D}'_N} = \frac{C''_N}{C''_S}$. This suffices to know that $\mathring{\tau}^S < \tilde{\tau} < \mathring{\tau}^N$ if and only if $\frac{\tilde{D}'_S}{\tilde{D}'_N} < \frac{C''_N}{C''_S}$, as this is the only remaining possibility for each inequality.

Analogously, we can show that the Negishi-weighted uniform carbon price lies between North's and South's preferred uniform carbon prices, unless they coincide. Lemma 5 and 6 imply that $\mathring{\tau}^S > \tilde{\tau} > \mathring{\tau}^N$ if and only if $\frac{\hat{D}'_S}{\hat{D}'_N} > \frac{C''_N}{C''_S}$. From Proposition 1, we know that $\mathring{\tau}^S = \tilde{\tau} = \mathring{\tau}^N$ if and only if $\frac{\hat{D}'_S}{\hat{D}'_N} = \frac{C''_N}{C''_S}$. This is again sufficient to know that $\mathring{\tau}^S < \tilde{\tau} < \mathring{\tau}^N$ if and only if $\frac{\hat{D}'_S}{\hat{D}'_N} < \frac{C''_N}{C''_S}$, as this is the only remaining possibility for each inequality. □

A.9 Proof of Proposition 4

Proof. We split the proof into the forward and backward implications.

Proof of forward direction: $\mathring{\tau}^S > \mathring{\tau}^N \implies \check{\tau} > \tilde{\tau}$.

South's preferred uniform carbon price is greater than North's preferred uniform carbon price if and only if

$$-\mathring{D}^{S\prime}_S \frac{C''_S + C''_N}{C''_N} > -\mathring{D}^{N\prime}_N \frac{C''_S + C''_N}{C''_S}. \tag{A115}$$

Simplifying and rearranging yields

$$\frac{\mathring{D}_S^{S\prime}}{\mathring{D}_N^{N\prime}} > \frac{C_N''}{C_S''}. \tag{A116}$$

From Lemma 2, we know that the utilitarian uniform carbon price lies between the two preferred uniform carbon prices. For strictly convex abatement cost functions, we know that if South's preferred uniform carbon price is greater than North's preferred uniform carbon price, then $\mathring{A}_i^S > \mathring{A}_i^N$ for all i, and thus $\mathring{A}^S > \mathring{A}^N$. For strictly convex damage functions, and recalling that marginal damages of abatement are negative, this implies

$$\mathring{D}_i^{S\prime} > \tilde{D}_i' > \mathring{D}_i^{N\prime}, \quad \forall i. \tag{A117}$$

We thus have

$$\frac{\tilde{D}_S'}{\tilde{D}_N'} > \frac{\mathring{D}_S^{S\prime}}{\mathring{D}_N^{N\prime}} > \frac{C_N''}{C_S''}. \tag{A118}$$

We have thus shown that

$$\mathring{\tau}^S > \mathring{\tau}^N \implies \check{\tau} > \tilde{\tau}. \tag{A119}$$

Proof of backward direction: $\check{\tau} > \tilde{\tau} \implies \mathring{\tau}^S > \mathring{\tau}^N$.

Proposition 2 establishes that $\check{\tau} > \tilde{\tau}$ if and only if $\frac{\tilde{D}_S'}{\tilde{D}_N'} > \frac{C_N''}{C_S''}$. From Lemma 2, we know that $\frac{\tilde{D}_S'}{\tilde{D}_N'} > \frac{C_N''}{C_S''}$ implies $\mathring{\tau}^S > \tilde{\tau} > \mathring{\tau}^N$. Therefore,

$$\check{\tau} > \tilde{\tau} \implies \mathring{\tau}^S > \mathring{\tau}^N. \tag{A120}$$

Together, the proofs of the forward and backward directions yield the equivalence

$$\check{\tau} > \tilde{\tau} \iff \mathring{\tau}^S > \mathring{\tau}^N. \tag{A121}$$

□

A.10 Calculation of welfare-equivalent consumption changes

The aim is to calculate the consumption changes in the initial period (2005), ΔX_{i0} (where $t = 0$ corresponds to the year 2005), that would yield a welfare change (in utility terms) that is equivalent to the intertemporal welfare difference between each of the utilitarian solutions and the Negishi solution. I start by computing the net present value (NPV) of the utilitarian welfare changes across two solutions for each region (the numerator in Equation

(A122)) [47], and divide that by the population size in 2005 to obtain the required per capita welfare change in 2005. I then set the NPV of the per-capita welfare change equal to a counterfactual per-capita welfare change in the initial period:

$$\frac{\sum_t L_{it}\beta^t u(x_{it}^{Util}) - \sum_t L_{it}\beta^t u(x_{it}^{Neg})}{L_{i0}} = u(x_{i0}^{cf}) - u(x_{i0}^{Neg}), \tag{A122}$$

where β^t is the utility discount factor ($\beta^t = (1+\rho)^{-t}$, where ρ is the utility discount rate), and the superscripts on x_{it} indicate whether this is the per-capita consumption of one of the two utilitarian solutions ($Util$), the Negishi solution (Neg), or a counterfactual (cf) consumption which we compute. The remaining notation is the same as in the main text.

Using the isoelastic specification of the utility function in the RICE model, $u(x_{it}) = \frac{x_{it}^{1-\eta}}{1-\eta}$ (where η is the elasticity of the marginal utility of consumption), we can solve for the counterfactual per-capita consumption in the initial period:

$$x_{i0}^{cf} = \left[(1-\eta)\frac{\sum_t L_{it}\beta^t u(x_{it}^{Util}) - \sum_t L_{it}\beta^t u(x_{it}^{Neg})}{L_{i0}} + (x_{i0})^{1-\eta}\right]^{\frac{1}{1-\eta}}. \tag{A123}$$

Finally, the aggregate welfare-equivalent consumption change is calculated as

$$\Delta X_{i0} = L_{i0}\left(x_{i0}^{cf} - x_{i0}^{Neg}\right). \tag{A124}$$

[47] I use this approach, rather than calculating the NPV by discounting the consumption changes with fixed discount rates, to account for the fact that the social discount rates (SDR) are different across regions and change over time due to different economic growth rates. To see this, note that the SDR is approximated by the Ramsey Rule, $SDR \approx \rho + \eta g$, where g is the growth rate in per-capita consumption, which differs across regions and over time.

Appendix B: Additional Figures

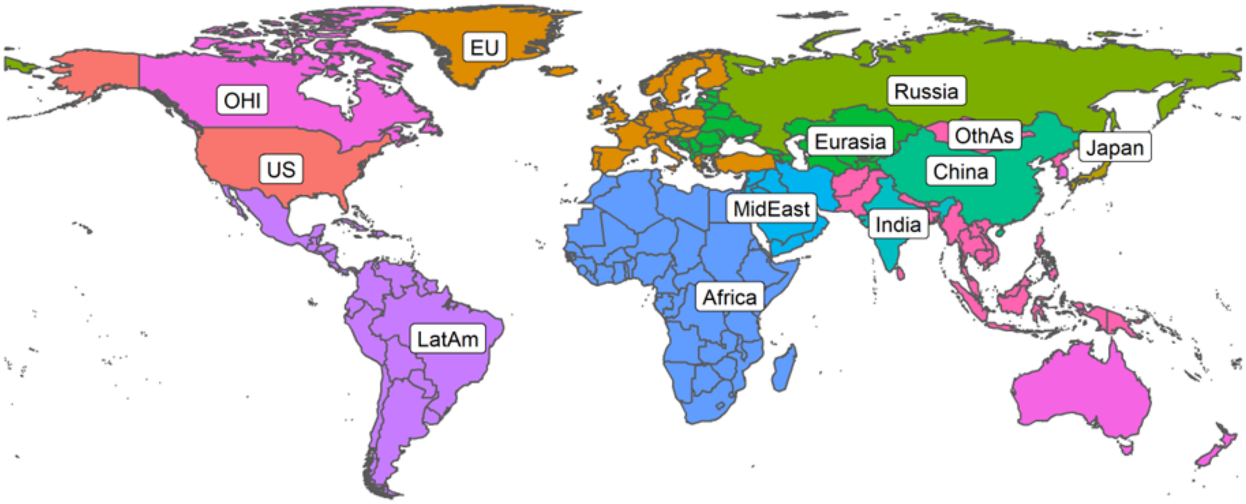

Figure A1: Regions of the RICE model. Countries of the same color belong to the same region, which is labeled at the largest country (OHI = Other High Income countries, OthAs = Other Asia).

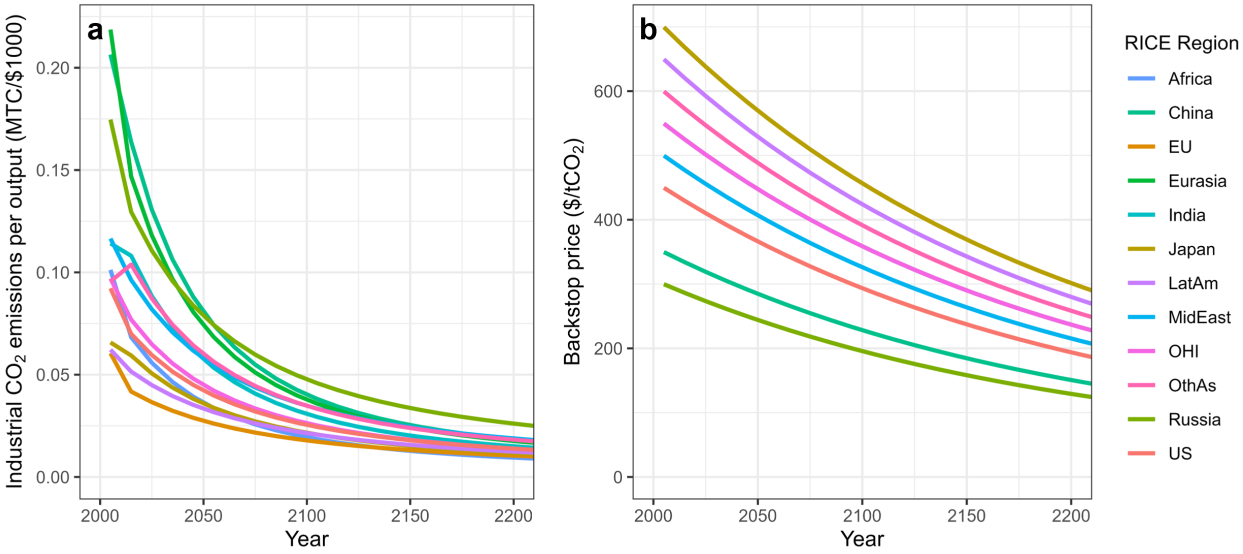

Figure A2: Regional baseline carbon intensities (a) and backstop technology prices (b) in the RICE model. The carbon intensity is given by the industrial CO_2 emissions per economic output. The backstop technology price corresponds to the marginal abatement cost at which all emissions are abated.

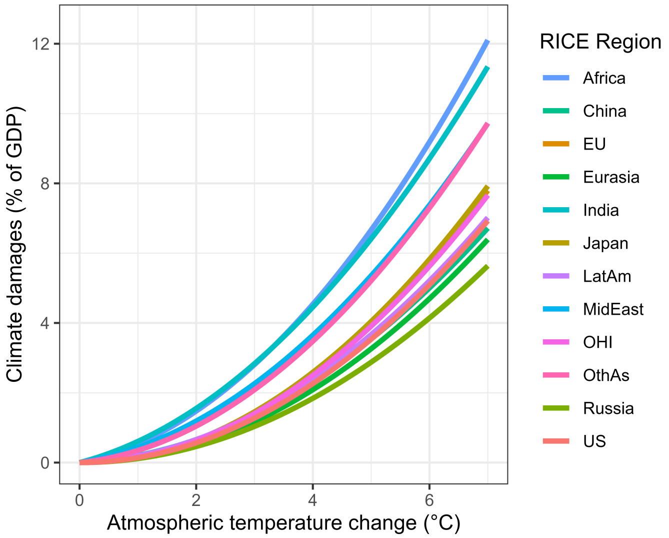

Figure A3: Regional damage functions for atmospheric temperature changes in the RICE model. Temperature changes are relative to temperatures in 1900.

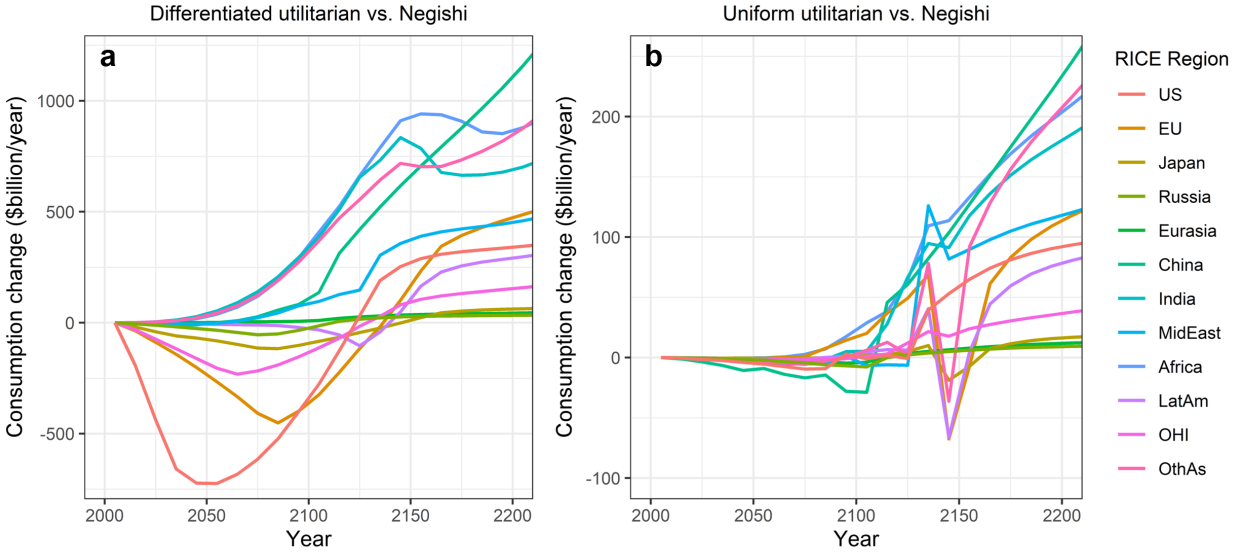

Figure A4: Regional consumption changes between the Negishi solution and the utilitarian solutions. Positive values indicate a higher consumption level in the utilitarian solutions. Note that these are the results for the utility discount rate of 1.5%.

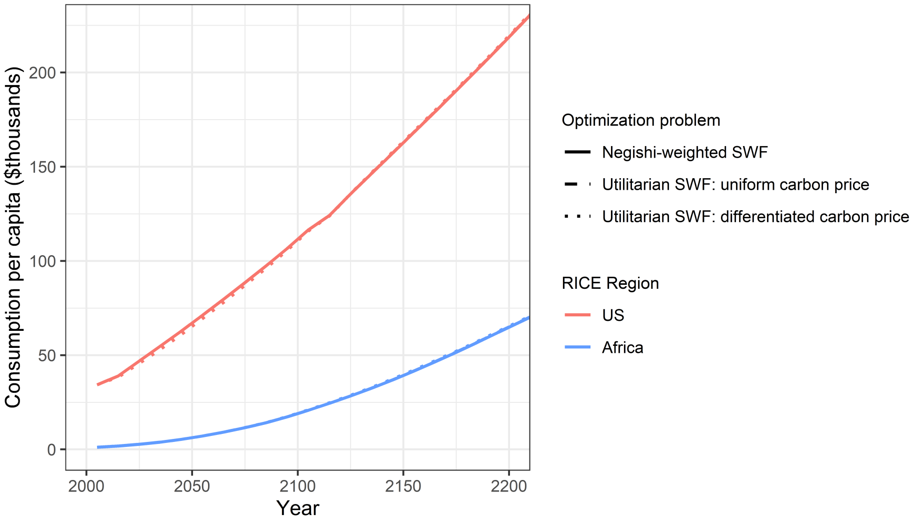

Figure A5: Consumption per capita trajectories for Africa and the US. Note that these are the results for the utility discount rate of 1.5%.

Appendix C: Supplementary Information

C.1 Time-variant Negishi weights

The time-variant Negishi welfare weights are given by

$$\alpha_{it} = \frac{1}{u'\left(x_{it}\right)} v_t, \tag{A125}$$

where v_t is the wealth-based component of the social discount factor. In the RICE-2010 model (Nordhaus, 2010), it is defined as the capital-weighted average of the regional wealth-based discount factors:

$$v_t = \frac{u'_{US,t}}{u'_{US,0}} \sqrt{\frac{\sum_{i \in \mathfrak{I}} \left(\frac{\frac{u'_{US,0}}{u'_{i0}}}{\frac{u'_{US,t}}{u'_{it}}} \frac{K_{it}}{\sum_{j \in \mathfrak{I}} K_{jt}} \right)}{\sum_{i \in \mathfrak{I}} \left(\frac{\frac{u'_{US,t}}{u'_{it}}}{\frac{u'_{US,0}}{u'_{i0}}} \frac{K_{it}}{\sum_{j \in \mathfrak{I}} K_{jt}} \right)}}, \tag{A126}$$

where K_{it} is the capital stock.

Note that $\frac{1}{u'(x_{it})} v_t$ equalizes the weighted marginal utility across regions. To obtain equalized weighted marginal utilities in each period, the discount factor needs to be equal across regions. Thus, v_t is not region-specific and it pins down the wealth-based component of the world discount factor (Nordhaus and Boyer, 2000).

C.2 Modeling of international transfers

Three different transfer scenarios are considered: (1) no transfers, which is standard in most IAMs, (2) non-conditional transfers (e.g., for loss and damage), and (3) conditional transfers for mitigation ("abatement abroad"). Transfer scenarios (2) and (3) reflect the types of transfers that are discussed in international climate change negotiations.

Interregional transfers were implemented from 2015 until the end of the model horizon (2595) [48]. The transfer quantities were defined as exogenous baseline transfers in 2025, which increase over time with the GDP of donor regions. In the main scenarios, I set the baseline transfer in 2025 to \$100 billion per year, which developed countries agreed to provide through 2025 (UNFCCC, 2015). In addition, I consider baseline transfers of \$1 trillion and \$10 trillion

[48] Note that the total interregional transfers are set to 0 and \$37 billion in the (historic) first two model periods (2005 and 2015). The \$37 billion figure is the annual average climate finance from OECD to non-OECD countries in 2015 and 2016 according to Oliver et al. (2018) (converted to 2005 USD, since the RICE model is in 2005 USD).

per year for the case of the non-conditional transfer to evaluate whether noticeable effects occur at larger transfer quantities (since the \$100 billion transfer did not markedly affect optimal climate policy trajectories)[49].

More specifically, the total transfer quantity, T_t^{tot}, increases from its baseline value in 2025, T_{2025}^{tot} [50], in proportion to the GDP increase in the richest four regions (US, EU, Japan, and Other High Income Countries), which are the donor regions (denoted by $\mathcal{D}$). The total interregional transfer in period t is thus

$$T_t^{\text{tot}} \ = T_{2025}^{\text{tot}} \frac{\sum_{j \in \mathcal{D}} Y_{jt}^{net}}{\sum_{j \in \mathcal{D}} Y_{j2025}^{\text{net}}}, \tag{A127}$$

where Y_{it}^{net} is the net output of a region after accounting for damages but before subtracting abatement costs.

The total redistribution quantity is levied in the donor regions in proportion to their regional net output in the previous model period. Thus, a region's contribution to the total interregional transfer is

$$T_{it} = -T_t^{tot} \frac{Y_{r(t-1)}^{net}}{\sum_{j \in \mathcal{D}} Y_{j(t-1)}^{net}}, \quad \forall i \in \mathcal{D}, \tag{A128}$$

where T^{tot} is the total transfer quantity, and Y_{it}^{net} is the net output of a region after accounting for damages but before subtracting abatement costs.

The total transfer quantity is redistributed to the remaining eight regions. In the case of non-conditional transfers, it is redistributed in proportion to the population size, L_{it}, of the recipient regions:

$$T_{it} = T_t^{tot} \frac{L_{it}}{\sum_{j \notin \mathcal{D}} L_{it}}, \quad \forall i \notin \mathcal{D}. \tag{A129}$$

In the case of the condition transfer, the total transfer is allocated optimally by choosing the redistribution shares, s_{it}, that maximize the utilitarian SWF:

$$T_{it} = s_{it} T_t^{tot}, \quad \forall i \notin \mathcal{D}. \tag{A130}$$

Under the non-conditional transfer, the region-specific transfer is then added to a region's net output to attain the post-transfer net output. Under the conditional transfer, the transfer quantity of the recipient regions is allocated toward their abatement costs.

[49] It should be noted, however, that the transfers of \$1 and \$10 trillion per year may be well outside the realm of what is politically realistic, at least in the near term – for comparison, the total nominal GDP of OECD countries was about \$60 trillion in 2018 (OECD, 2019). These large transfer scenarios were thus only included to assess whether such large transfers would substantially alter the optimal climate policy path, not because they are considered realistic.

[50] For clarity, I use the calendar year as a subscript here, which corresponds to $t = 20$ in the model.

C.3 The role of non-conditional transfers

The optimal temperature trajectories in the presence of interregional non-conditional transfers are shown in Figure A6. I find that non-conditional interregional transfers play a minor role and have virtually no effect on the optimal climate policy up to a total transfer quantity of at least $1 trillion per year. Thus, an important conclusion is that politically realistic levels of redistribution do not considerably alter the stringency of optimal climate policy. In particular, the optimal policy path under the redistribution quantity of $100 billion per year consistent with the Paris Agreement is practically identical to the optimal policy paths without any interregional transfers. It is also worth noting that realistic redistribution quantities do not bring the optimal cumulative emissions back up to the optimal level under the Negishi solution. Hence, the increased optimal mitigation effort under the utilitarian approaches is not obviated in the presence of transfer policies. Indeed, it is ambiguous whether the stringency of optimal climate policy increases or decreases in the presence of interregional transfers.

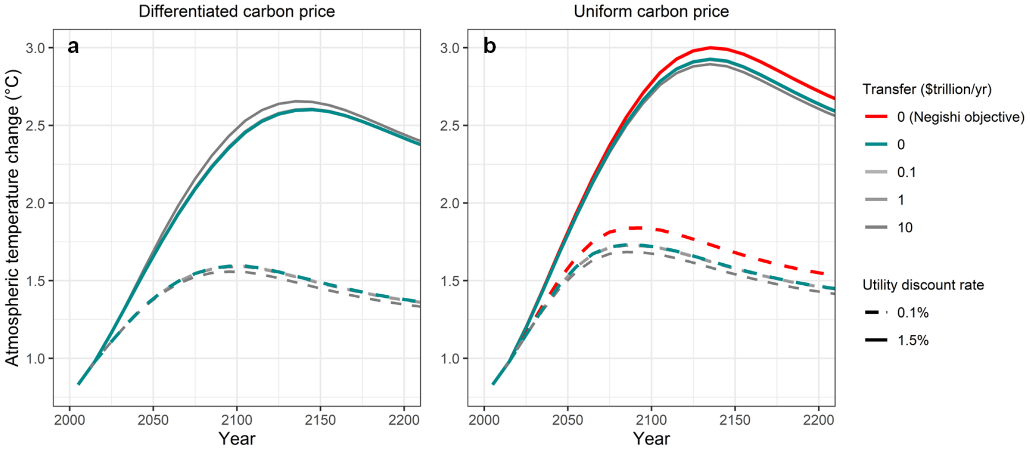

Figure A6: Optimal atmospheric temperature trajectories conditional on the optimization problem, the total non-conditional interregional transfer, and the utility discount rate. The Negishi-weighted solutions (red) are compared to the utilitarian solutions with (b) and without (a) the additional constraint of equalized regional carbon prices under a variable interregional transfer for the Nordhaus (solid lines) and Stern (dashed lines) utility discount rates (Nordhaus, 2011; Stern et al., 2006). Temperature changes are relative to 1900.

While politically realistic non-conditional transfers do not have a large quantitative effect on the optimal climate policy, it is still interesting to note the direction of the effect. If the constraint of a uniform carbon price is imposed, transfers from rich to poor regions result in an increased mitigation effort under both high and low utility discount rates. At least part of the intuition for this result is that the utilitarian solution is particularly sensitive to consumption changes of the poor due to the diminishing marginal utility of consumption. The optimal carbon price balances the discounted marginal welfare costs and benefits of mitigation. The welfare costs of mitigation are particularly high in poor regions, so a uniform carbon price needs to be kept quite low in order to prevent large welfare reductions in those regions. By making poor regions richer, redistribution makes it possible to increase the uniform carbon price at a lower welfare cost. To put it simply, poor regions can afford a higher uniform carbon price after they have received transfers. The effect of redistribution under differentiated carbon prices is ambiguous. Under the lower utility discount rate, the carbon prices in rich regions reach the corresponding backstop prices (implying complete abatement) early in the 21st century, even under the highest redistribution scenario. Once this is the case, the transfer increases the abatement effort in poor regions without decreasing the abatement effort in rich regions, resulting in an increased overall abatement level. In contrast, under the higher utility discount rate, which places relatively more weight on the present, the backstop price is reached much later in rich regions. In this case, the decreased mitigation in rich donor regions outweighs the increased mitigation in poor recipient regions, thus decreasing the overall abatement level.

C.4 Discussion of the differentiated carbon price optimum

The welfare maximizing policy that allows for differentiated carbon prices requires much higher carbon prices in rich regions than in poor regions (see Figure 4 and Table 2). This result warrants a discussion of several issues. First, the differentiated carbon price optimum may be opposed by rich nations as it results in an implicit transfer from rich to poor regions. It should be noted, however, that the uniform carbon price optimum is welfare inferior to the differentiated carbon price optimum, as it imposes an additional constraint (Budolfson and Dennig, 2019). Importantly, the differentiated carbon price optimum is also in accordance with the principle of "common but differentiated responsibilities and respective capabilities" of the United Nations Framework Convention on Climate Change (UNFCCC, 1992). As such, Budolfson and Dennig (2019) argue that the differentiated carbon price optimum is a natural focal point for international climate policy and for evaluating the adequacy of the nationally determined contributions (NDCs), which are at the heart of the Paris Agreement. A more recent study by Budolfson et al. (2021) provides this comparison of the NDCs to implied carbon budgets under the differentiated carbon price optimum. Second, since differentiated carbon prices are not cost-effective, it should be reemphasized that a further welfare improvement over the differentiated price optimum could be achieved by establishing an international emission trading scheme. This would allow regions with higher carbon prices to buy emission permits from poorer regions where the carbon price is lower, implying a transfer from the rich to the poor. Due to the differential carbon prices, mutual gains can be achieved by such a trading scheme (Budolfson and Dennig, 2019). If the permit market is fully competitive, this would result in a globally harmonized carbon price. However, as Budolfson and Dennig (2019) point out, this outcome would be different from the uniform carbon price optimum discussed above, where an a priori constraint of equalized carbon prices was imposed; total emissions will be reduced and the poorest countries will bear a lower burden under the harmonized carbon price attained by the emission trading scheme. Chichilnisky and Heal (1994) thus propose that the efficient allocation of emission permits is established by the differentiated carbon price optimum, and once the optimal allocation of permits is found, these permits are then traded internationally to achieve further welfare gains. The emission budgets shown in Figure 4 can thus be understood as providing the first step of this process. Third, a potential problem with differentiated carbon prices is carbon leakage – an increase in carbon emissions in a country with comparatively laxer climate policies as a result of stricter climate policies in another country (e.g., due to a relocation of carbon-intensive industries to countries with laxer climate policies). The problem of carbon leakage, if it is not addressed, may thus undermine the policy. Budolfson et al. (2021) provide a brief discussion of the issue of carbon leakage and how it may be addressed. They note that there are two channels for carbon leakage: (1) competitiveness differences resulting from carbon price differences, and (2) lower fossil fuel prices due to decreased global demand. Budolfson et al. (2021) argue that the competitiveness channel can be addressed with border tax adjustments, such as those proposed by Flannery et al. (2018). The second channel is shut down if countries commit to a global emissions cap (Budolfson et al., 2021). Of course, there is also no carbon leakage if each region commits to its own regional carbon budget.