Welfare-Maximizing Carbon Prices and the Role of International Climate Finance

Simon F. Lang*

February 28, 2025

[Click here for the latest version]

Abstract

How should nations price carbon? This paper examines how interregional equity considerations and the availability of international climate finance affect optimal carbon prices—two key policy aspects not reflected in standard estimates. I develop theory to identify the conditions under which accounting for differences in marginal utilities of consumption across countries leads to more stringent global climate policy in the absence of international transfers. In calibrated simulations, I find that this inequality-sensitive approach reduces 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 the impact of the Paris Agreement's $100 billion annual transfer on optimal carbon prices and emissions, finding that it further reduces global emissions if directed toward mitigation projects in developing countries. Accounting for inequality and transfers reduces optimal global emissions by 31% compared to a policy that excludes these factors.

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

I am grateful to my advisors Kenneth Gillingham, William Nordhaus, Matthew Kotchen, David Anthoff, and Narasimha Rao for their invaluable feedback and support. I also thank Eli Fenichel, François Salanié, Francis Dennig, Frikk Nesje, Antony Millner, Kevin Kuruc, Itai Sher, Maya Eden, Paul Kelleher, Mirco Dinelli, Andrew Vogt, Matthew Gordon, Ethan Addicott, Matthew Ashenfarb, Andie Creel, Kenneth Jung, Jillian Stallman, Lisa Rennels, Enrico Salonia, Eric Donald, Gustav Fredriksson, Costanza Tomaselli, and participants of the Yale ENRE seminar, the EQWL lab group, the CER-CEPE ETH Brown Bag seminar, the EAERE Annual Conference, the AERE Summer Conference, the Toulouse Conference on the Economics of Energy and Climate, SURED, the Swiss workshop on Environmental, Resource and Energy Economics, the Oxford Global Priorities Fellowship, the Columbia IPWSD workshop, the IAMC Annual Meeting, and the Penn State Science and Values in Climate Risk Management webinar 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 climate impacts, responsibilities for emissions, and capabilities to mitigate and adapt—aspects that are all interlinked with global wealth inequality (Chancel et al., 2023). These inequalities are recognized in international climate agreements, as exemplified by the principle of "common but differentiated responsibilities and respective capabilities" of the United Nations Framework Convention on Climate Change (UNFCCC, 1992). The Paris Agreement further emphasizes that developed countries should lead in reducing emissions and support developing nations' transitions to low-carbon economies, stressing equity considerations (UNFCCC, 2015). In this context, international climate finance—financial transfers from developed to developing countries for climate mitigation and adaptation—has become a major topic in climate summits (UNFCCC, 2009; UNFCCC, 2015; UNFCCC, 2023).

Despite the importance of inequalities and transfers in international climate policy, the standard approach to estimating optimal carbon prices concentrates solely on efficiency and effectively disregards global inequality and international climate finance (Nordhaus, 2010; Yang and Nordhaus, 2006). It does so by maximizing a social welfare function (SWF) with Negishi welfare weights, which are higher for wealthier countries, offsetting differences in marginal utilities of consumption across countries. In contrast, an alternative approach focuses on maximizing global welfare, subject to constraints on international transfers (Budolfson et al., 2021). A common version of this approach maximizes the utilitarian SWF, which assigns equal weight to the welfare of all individuals. Crucially, the utilitarian SWF accounts for global inequality in that it considers differences in marginal utilities of consumption across wealthier and poorer countries. These differences in welfare weights may be particularly important, as the costs and benefits of emission reductions are unevenly distributed, with poorer countries often disproportionately impacted by climate change (Burke et al., 2015; Carleton et al., 2022; Kalkuhl and Wenz, 2020; Méjean et al., 2024). Given that optimal carbon prices are well-known to be highly sensitive to the utility discount rate (Stern, 2007; Nordhaus, 2007)—a form of temporal welfare weighting—it may be surprising that comparatively little research has explored the role of regional welfare weights.

In this paper, I examine how optimal carbon prices are influenced by two central considerations in international climate policy: global inequality and the availability of international climate finance. I study these questions theoretically, to uncover insights into the underlying economic forces, and through numerical simulations with the integrated assessment model RICE (Nordhaus, 2010), to evaluate the quantitative implications for climate policy. I begin by exploring how accounting for inequality affects optimal carbon prices in the absence of transfers. I then investigate the impact of international transfers on optimal carbon prices, focusing on conditional transfers for emission reductions, known as mitigation finance, the main type of international climate finance under the Paris Agreement (OECD, 2024). I concentrate on evaluating the effect of an annual transfer of $100 billion by 2025, set to increase over time, as committed by developed countries at the United Nations climate conferences in Copenhagen and Paris (UNFCCC, 2009; UNFCCC, 2015). Furthermore, I examine the jointly optimal policy package of carbon prices and transfers in a setting in which the transfer quantity is endogenous to climate policy.

I organize the inquiry into two parts based on whether transfers are available. First, I assume no international transfers and compare optimal carbon prices under the Negishi-weighted SWF to those under the utilitarian SWF. I begin 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 policy instrument—a globally uniform carbon price in each period—enabling a direct comparison with the Negishi solution. Next, I remove the uniform carbon price constraint, allowing for differentiated carbon prices, which can improve utilitarian welfare by shifting some of the abatement cost burden from poorer to wealthier countries. I refer to carbon prices in the utilitarian solution as welfare-maximizing carbon prices to highlight that they maximize the (unweighted or equally-weighted) sum of individuals' utilities.

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 abatement cost burden across countries. Specifically, the utilitarian climate policy is more stringent if poorer nations have comparatively high marginal climate damages and a relatively steep marginal abatement cost function, resulting in smaller changes in abatement when carbon prices are altered. In a dynamic extension of the model, I show that regional differences in population growth and economic growth critically influence how regional welfare weights affect optimal carbon prices by impacting the relative importance that regions place on the future, when most climate impacts occur, versus the present, when abatement efforts take place. Moreover, I establish a novel and intuitive connection between the uniform carbon prices under utilitarian and Negishi weights and nations' preferred uniform carbon prices that maximize national welfare, a notion that was established by Weitzman (2014) and Kotchen (2018): The utilitarian uniform carbon price exceeds the Negishi-weighted carbon price if and only if poorer nations prefer higher uniform carbon prices than wealthier nations. At a conceptual level, I introduce the concept of *welfare-cost-effectiveness*, which refers to emission reductions that are achieved at the lowest possible welfare (utility) cost, contrasting it with the prevalent concept of cost-effectiveness, which focuses on minimizing monetary costs. I demonstrate that regionally differentiated utilitarian carbon prices are welfare-cost-effective, yielding higher carbon prices in wealthier countries, while Negishi-weighted carbon prices are cost-effective.

To address the theoretical ambiguity regarding whether accounting for inequality raises or lowers optimal carbon prices, I employ numerical simulations with RICE to explore the direction and magnitude of this effect. I find that the welfare-maximizing solutions yield more stringent climate policies than the Negishi solution. Simply put, accounting for inequality means stronger climate action. Specifically, the utilitarian uniform carbon price in 2025 is around 15% higher than the Negishi-weighted carbon price, under default discounting parameters. The utilitarian solution that allows for differentiated carbon prices features higher carbon prices in rich countries and lower carbon prices in poor countries; globally, cumulative emissions are 21% lower compared to the Negishi solution.

I leverage the theoretical insights to uncover the key drivers of the numerical results. The utilitarian differentiated carbon price solution results in reduced global emissions because the additional abatement in developed countries outweighs the reduced abatement in the poorest regions. Higher uniform carbon prices in the utilitarian solution, compared to the Negishi solution, are primarily driven by the poorest region, Africa, which is most impacted by climate change for two main reasons: it has the highest marginal damages and the fastest population growth. Using the theoretical link to regions' preferred uniform carbon prices to strengthen the intuition, I find that Africa also favors the highest uniform carbon price—more than twice the preferred price of the US and the Negishi-weighted carbon price. Thus, the main intuition for lower carbon prices in the Negishi solution is as follows: by assigning lower weight to the welfare of poorer regions, Negishi weights effectively also downweight the region most impacted by climate change, leading to less stringent climate policy.

In the second part of the paper, I introduce international transfers to examine how welfare-maximizing carbon prices are affected by the availability of transfers from rich to poor countries. Focusing on conditional transfers to finance emission reductions, I theoretically show how an exogenous quantity of such transfers affects welfare-maximizing carbon prices. Furthermore, I consider a setting in which the transfer quantity is endogenous to climate policy. Specifically, I explore the jointly welfare-maximizing policy package of carbon prices and transfers, contingent on the feasibility of redistributing gains in wealthy countries resulting from changes in climate policy. The main finding here is that uniform carbon prices are welfare-maximizing only if demanding transfer conditions are satisfied.

In numerical simulations, I focus on examining the effect of the "Paris Agreement transfer" of $100 billion per year by 2025, set to increase thereafter. I find that financial support for mitigation in developing countries substantially increases the stringency of welfare-maximizing climate policy by lowering the welfare cost of abatement. Notably, under default discounting parameters, 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 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. In the optimal allocation, the majority of mitigation finance is directed to Africa, India, China, and other Asian countries, the regions with the largest untapped low-cost abatement opportunities in the absence of transfers.

This paper makes several contributions to the literature on optimal carbon prices with heterogeneous regions. First, it provides a set of novel theoretical results on how optimal carbon prices depend on regional welfare weights. To my knowledge, it is the first paper to derive the conditions under which accounting for global inequality increases the optimal global climate policy stringency in the absence of transfers. In doing so, I identify factors that have previously been underappreciated in this context: regional heterogeneities in the convexity of the abatement cost function, population growth, and economic growth. These results build on influential papers by Chichilnisky and Heal (1994) and Eyckmans et al. (1993), which show that globally uniform carbon prices are optimal if and only if distributional issues are ignored (through the choice of welfare weights) or unrestricted lump-sum transfers can be made between countries. I contribute to this discussion by providing an additional rationale for uniform carbon prices when transfers are endogenous to climate policy. I also offer a new perspective to the literature on regions' preferred uniform carbon prices (Weitzman, 2014; Kotchen, 2018). Instead of focusing on voting mechanisms, I explore an aggregation of preferences rooted in welfare-economic theory. A related strand of literature explored aspects of efficiency and equity in emission permit markets (Chichilnisky and Heal, 2000; Shiell, 2003; Sandmo, 2007; Borissov and Bretschger, 2022). In contrast, this paper focuses on a setting without international permit markets. Other related papers examined the importance of accounting for inequalities at a fine-grained level (Dennig et al., 2015; Schumacher, 2018), and how optimal carbon taxes, under different welfare weights, depend on distortionary fiscal policy (Barrage, 2020; d'Autume et al., 2016; Douenne et al., 2023) and inequality within and between countries (Kornek et al., 2021). However, unlike the present paper, these studies do not theoretically explore how utilitarian and Negishi weights shape the optimal stringency of climate policy.

Second, this paper adds to a body of work that numerically investigates the role of regional welfare weights in IAMs. The study most closely related to this research is Anthoff (2011), which also compares the Negishi solution to a utilitarian solution, although using a different integrated assessment model[1]. The present paper expands upon this study in multiple ways. First, it offers a deeper understanding of the key drivers behind the results by linking the new theoretical insights to the numerical findings and regional characteristics. Furthermore, I extend the analysis by examining the distributional implications, the utilitarian uniform carbon price, and regions' preferred uniform carbon prices. This offers additional insights into heterogeneous climate policy preferences and further strengthens the intuition behind the numerical results. A related but distinct literature estimates the equity-weighted social cost of carbon (SCC), a measure of the marginal damages of carbon emissions that places more weight on the costs and benefits in poor countries. The key difference between this literature and the present paper is that the equity-weighted SCC typically estimates the marginal damages along *non-optimal* emissions pathways (Azar and Sterner, 1996; Anthoff et al., 2009; Adler et al., 2017; Anthoff and Emmerling, 2018; Prest et al., 2024), while this paper investigates how regional welfare weights affect optimal carbon prices.

Third, this paper makes contributions to the literature that examines climate policy in conjunction with transfers. To the best of my knowledge, this is the first study to theoretically and numerically assess how limited international climate finance impacts welfare-maximizing carbon prices, and to estimate the optimal allocation of mitigation finance. Moreover, it is the first paper to theoretically explore the jointly welfare-maximizing policy package of carbon prices and transfers in a setting in which the transfer quantity is endogenous to climate policy. A related study in this literature, Yang and Nordhaus (2006), examines optimal unrestricted transfers for mitigation under different social welfare weights, showing that zero transfers occur with Negishi weights, while large transfers take place with utilitarian weights. Another related study by 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 have explored the potential of transfers to facilitate Pareto improvements and international cooperation (Hoel, 1994; Hoel et al., 2019; Kotchen, 2020; Hillebrand and Hillebrand, 2023; Kotlikoff et al., 2024), the required transfers to obtain globally uniform carbon prices under different normative criteria (Landis and Bernauer, 2012), 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 organized as follows. Section 2 provides conceptual back-

[1]Budolfson and Dennig (2020) also compare utilitarian and Negishi solutions. However, they do not technically use Negishi weights but a model in which all individuals consume the global average consumption.

ground on the different optimization approaches. In Section 3, a theoretical model is introduced and key analytical results are derived. Section 4 describes modifications to the RICE model and presents simulation results. Section 5 concludes.

2 Conceptual Background

This section provides conceptual background on different optimization approaches in order to lay the foundation for my own analysis. In Section 2.1, the difference between positive and normative approaches is introduced. Section 2.2 discusses the positive approach. Finally, Section 2.3 provides a welfare-economic conceptualization of normative optimizations.

2.1 Positive and normative optimizations

The purpose of optimization is a main source of debate in climate economics, and two approaches are sometimes distinguished: positive and normative optimizations (Kelleher, 2019). Nordhaus (2013, p. 1081) provides an instructive discussion of these two approaches, noting 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. Which approach is taken depends on the welfare weights in the SWF[2].

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; Gollier, 2012; Nordhaus, 2007). Under the positive approach, the discount rate is determined based on market observations. In contrast, the normative approach relies on ethical reasoning to determine the discount rate.

However, the difference between positive and normative optimization approaches extends to the interregional weighting of welfare. The typical positive approach relies on Negishi welfare weights, which are higher for 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 within

[2]The positively and normatively determined welfare weights coincide for a specific normative stance, but in general they are different.

a time period equally[3]. This paper focuses on interregional welfare weights and how they influence optimal carbon prices.

While the distinction between positive and normative optimizations is useful to clarify the different purposes of optimization, this distinction is not always clear-cut in climate economics. In particular, it has been questioned whether it is possible to interpret the modeling choices that are typically made under the positive optimization approach as purely positive (see Chawla (2023), for a discussion)[4]. Keeping this caveat in mind, I use the labels "positive" and "normative" to highlight the conceptual difference underlying these optimization approaches: whether the optimization seeks to simulate markets or to maximize social welfare.

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 and critiques of using Negishi weights in IAMs. 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 is a theorem by 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 such that each agent's budget constraint is satisfied at the equilibrium (Nordhaus and Yang, 1996). The Negishi-weighted SWF is given by a weighted sum of agents' utilities, where the weights are inversely proportional to the marginal utility of consumption. For identical and concave utility functions, which are commonly assumed, this implies higher welfare weights for wealthy individuals, with a low marginal utility of consumption, than for poorer individuals. The appeal of the Negishi-weighted SWF is that it provides a computationally convenient method to identify

[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).

[4]An additional confusion sometimes arises when "positive" optimization results appear to be used to suggest how policies *should* be designed (Kelleher, 2019). While possible in principle, a normative justification of positively calibrated welfare weights would be required to draw normative conclusions from a positive analysis.

the competitive equilibrium, which is Pareto efficient if the conditions of the first fundamental theorem of welfare economics are satisfied.

Besides this theoretical foundation, there were two additional motivations for the use of Negishi weights in IAMs: (1) to prevent transfers across regions, which were considered politically infeasible or unrealistic (Nordhaus and Yang, 1996), and (2) to obtain a uniform carbon price in all regions, ensuring that global emissions are reduced at the lowest possible cost. Indeed, to achieve these two objectives in every period of the RICE model, Nordhaus and Yang (1996) made refinements to what they call the "pure Negishi solution" that relies on time-invariant welfare weights. Specifically, 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 yields time-variant Negishi weights and accomplishes the goal of equalizing the carbon price across regions in every period. Moreover, these weights ensure that no cross-regional transfers take place, since such transfers do not increase the objective value of the Negishi-weighted SWF. Notably, 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, as long as utility is an increasing concave function of consumption, which is commonly assumed. Hence, the constraints of equalized carbon prices and no transfers are effectively incorporated in the time-variant Negishi weights used in RICE.

2.2.2 Critiques of using Negishi weights in IAMs

While Negishi weights are commonly used in IAMs, the use of such weights has been 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.

From an ethical perspective, a main critique is that Negishi weights assign 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. 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 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 conceptual inconsistency: the diminishing marginal utility of consumption is embraced intertemporally, but suppressed interregionally. Consequently, transfers from richer to poorer individuals are 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.

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).

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). The maximization of a Negishi-weighted SWF in IAMs seeks to identify the competitive equilibrium with a Pareto efficient level of abatement[5]. I 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 of the first fundamental theorem of welfare economics hold). 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[6]. The Pareto efficiency of the Negishi solution, and the inefficiency of no abatement, is illustrated in Figure 1a, which shows the Pareto frontier for a simple example with two regions: a 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 that does not require transfers is obtained. In contrast, normative optimizations use SWFs that are rooted

[5]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.

[6]This is also the case if abatement is inefficiently low, as it is the case in the Nash equilibrium.

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, which shows the social indifference curves of the utilitarian and Negishi-weighted SWFs, and the points that maximize these SWFs[7].

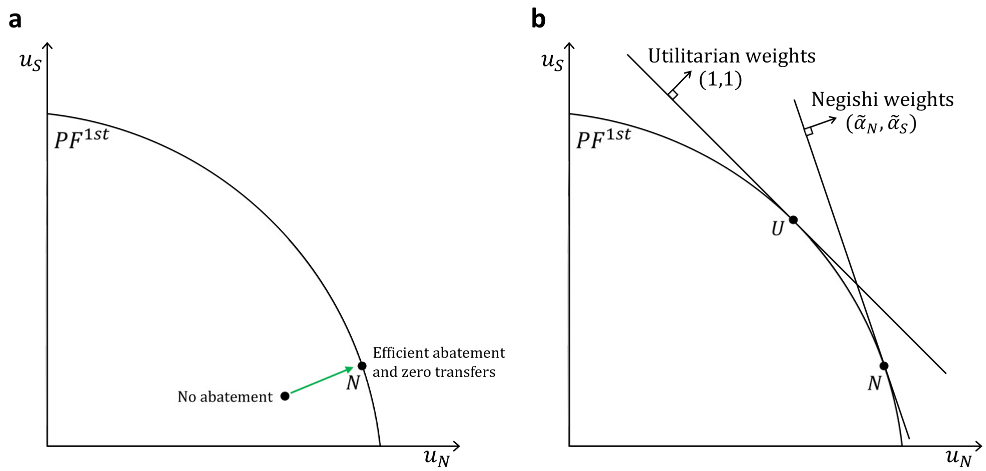

Figure 1: Illustrative two-region example of the welfare outcomes under the Negishi and utilitarian solutions.

Notes: Panel (a) shows that the Negishi solution (N) is Pareto efficient. Panel (b) shows an illustrative comparison of the Negishi (N) and utilitarian (U) solutions. The utilities of representative agents in the Global North and Global South are denoted by u_N and u_S, respectively. PF^{1st} is the Pareto frontier in a first-best setting. The welfare weights vectors are the gradient vectors of the SWFs, which are perpendicular to the linear social indifference curves. The Negishi weights are denoted by $\tilde{\alpha}_i$.

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

[7]To choose among different points on the Pareto frontier (or, more generally, any vector of utilities), interpersonal utility comparisons are often made. However, the admissibility of such comparisons is a longstanding point of contention in welfare economics (Robbins, 1935; Harsanyi, 1955; Sen, 1970; Stiglitz, 1987). Indeed, contemporary welfare economics is split into two branches: one dismisses interpersonal utility comparisons, while the other branch relies on such comparisons and uses SWFs to determine socially preferable outcomes (Fleurbaey and Maniquet, 2008). This paper belongs to the second branch.

the solution as one in which a world central planner is allocating resources in an optimal fashion".

A second line of argument used to support employing Negishi weights 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. This is sometimes used to argue that the issues of equity and efficiency can be separated. However, this typically is not the case for climate policy; 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 welfare theorem 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 welfare theorem] is severely limited".

To summarize, the abatement in the Negishi solution generally differs from the abatement that maximizes global utilitarian welfare (hereafter, simply "global welfare"), regardless of whether unrestricted transfers are feasible.

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 marks a key difference from 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 are specified which affect the feasible set of allocations. The first step—the specification of the SWF based on ethical principles—is common in normative analyses. Such SWFs have a long tradition in public economics, and particularly in the optimal income taxation literature (Mirrlees, 1971; Piketty and Saez, 2013; Fleurbaey and Maniquet, 2018). They are referred to as Bergson-Samuelson SWFs (Bergson, 1938; Samuelson, 1947) and produce an ethical ordering of 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[8]. Finally, the social optimum is the point on the Pareto frontier that maximizes the SWF.

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.

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 and efficient 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

[8]Economists sometimes use the term efficiency to simply mean outcomes that maximize the total monetary 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 maximize 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.

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 was one of the reasons that motivated the use of Negishi weights under the positive optimization approach. In contrast, under the normative optimization approach, political transfer constraints affect the feasible set of allocations while welfare weights remain unchanged.

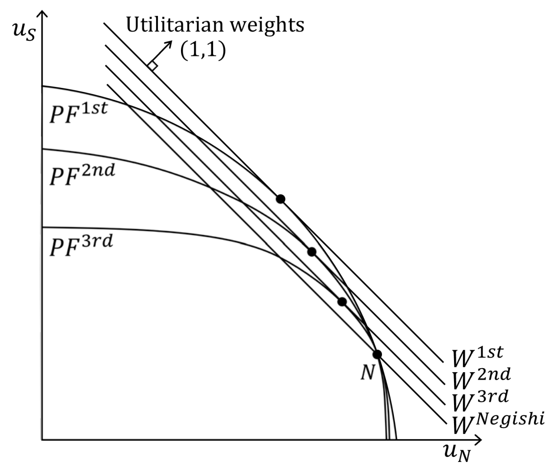

Figure 2: Illustration of the utilitarian social welfare outcomes in first-, second-, and third-best settings, and the Negishi solution.

Notes: The figure shows the utilitarian optima in first-, second-, and third-best settings and the utilitarian welfare level of the Negishi solution. The utilities of representative agents in the Global North and Global South are denoted by u_N and u_S, respectively. PF^x is the Pareto frontier in the x^{th}-best setting. W^x is the utilitarian social indifference curve that corresponds to the social optimum in the x^{th}-best setting or the Negishi solution. The utilitarian weights vector is the gradient vector of the utilitarian SWF.

Let us now consider such a second-best setting in which international lump-sum transfers are infeasible[9]. 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 address inequality with lump-sum transfers, the utilitarian social planner accounts for global 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 3.2.2). It should be noted that a potential problem with differentiated carbon prices is carbon leakage—an increase

[9] 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 Sections 3.3 and 4.3.

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.6.

Finally, consider a third-best setting in which the policy instruments the social planner can use are restricted even further to a globally uniform carbon price (in addition to a constraint of no transfers). I would argue 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 policy instrument—a globally uniform carbon price and no transfers. Yet, an important difference remains. The utilitarian uniform carbon price accounts for inequality in the carbon price level, while the positive optimization approach ignores inequality altogether through the specification of the Negishi weights, which equalize the weighted marginal utility across regions. Consequently, 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.

It is worth highlighting how the different solutions respond to global 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.

The extent to which 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 provides a theoretical analysis of how regional welfare weights and international transfers affect optimal carbon prices, demonstrating the important role both factors play.

3.1 Model setup

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

There are two regions $i \in \{N, S\}$ and a single period (a two-period model will be considered in Section 3.2.5). Let $\mathcal{I} = \{N, S\}$ denote the set of regions; for intuition, consider the regions as the Global North (N) and Global South (S). The population of each region i is exogenous and denoted by L_i. Uppercase letters are used for aggregate variables at the region level, while lowercase letters are used for per capita variables and, in some cases, per endowment variables.

Abatement costs, $C_i(A_i)$, are a function of the abatement $A_i \geq 0$ in region i. 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. This is the case for the commonly assumed quadratic abatement cost function. The aggregate global abatement is given by $A \equiv \sum_i A_i$. Region-specific climate damages, $D_i(A)$, are a function of the global 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. Regional consumption, X_i, is given by the exogenous endowment, W_i, net of abatement costs and climate damages: $X_i = W_i - C_i(A_i) - D_i(A)$. There is a representative agent in each region, who derives utility, $u(x_i)$, from per capita consumption, $x_i = X_i/L_i$. The utility function is assumed to be identical for all individuals, strictly increasing, strictly concave, and smooth. Thus, $\frac{du}{dx_i} > 0$ and $\frac{d^2u}{dx_i^2} < 0$.

I assume throughout that the Global North is richer than the Global South, both in terms of per capita endowment and consumption[10]. Thus, we have $w_N > w_S$ and $x_N > x_S$. The implicit assumption is that the difference in endowment per capita between the Global North and the Global South is sufficiently large such that individuals in the Global North remain richer after abatement costs and climate damages are subtracted. From the concavity of the utility function, it follows that $u'(x_N) < u'(x_S)$.

While I derive the theoretical results for general functional forms, it is useful to put more structure on the abatement cost and damage functions to closely link theory and simulation results. To do this, I use simplified versions of the functions employed in the RICE model[11], capturing their key characteristics. I define these "simplified RICE functions" as follows:

$$D_i(A) = L_i w_i d_i(A), \tag{1}$$

$$C_i(A_i) = L_i w_i c_i \left(\frac{A_i}{L_i w_i} \right), \tag{2}$$

where w_i is the endowment per capita and d_i and c_i denote the damages and abatement costs per aggregate regional endowment, respectively. Note that c_i is a function of abatement relative to the size of the economy, reflecting that larger economies have more abatement opportunities of a certain type and cost.

3.2 Optimal carbon prices in the absence of transfers

In this section, I establish how optimal carbon prices are influenced by regional welfare weights in the absence of interregional transfers. I begin by presenting the optimization problems and deriving the optimal carbon prices under both utilitarian and Negishi weights. Following this, I compare the stringency of the resulting climate policies.

3.2.1 Optimization problems

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 for (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, with welfare weights $\alpha_i \geq 0$, subject to regional budget constraints, reflecting a constraint of no interregional transfers.

[10] There is one exception to this: In Section 3.3, I consider the possibility of transfers that are sufficient to equalize consumption across regions.

[11] See Appendix C.2 for the abatement cost and damage functions of the RICE model.

Formally, the *differentiated carbon price optimization problem* is

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

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

The *uniform carbon price optimization problem* is identical except that an additional constraint of uniform marginal abatement costs is imposed[12],

$$C'_N(A_N) = C'_S(A_S). \tag{5}$$

3.2.2 Optimal carbon prices under different welfare weights

Solving the optimization problems yields expressions for the optimal marginal abatement costs. Optimal carbon prices, τ_i, are equal to the optimal marginal abatement costs, $C_i'^*$, because regions are assumed to 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 focus on the optimal carbon prices under the welfare weights that are most commonly used in climate economics—Negishi weights and utilitarian weights. Optimal carbon prices under arbitrary welfare weights are shown in Appendix C.1. Derivations are provided in Appendix A.1.

The Negishi solution

I begin with the Negishi solution. Negishi weights, $\tilde{\alpha}_i$, are inversely proportional to a region's marginal utility of consumption at the optimal solution that was obtained with the Negishi weights[13]; that is, $\tilde{\alpha}_i = 1/u'(\tilde{x}_i)$. I use "tilde" to indicate the Negishi solution. Since we assume that consumption is higher in the North and the utility function is concave, it follows that the Negishi weight is greater for the North than the South: $\alpha_N > \alpha_S$.

Solving the differentiated carbon price optimization problem with Negishi weights yields the Negishi solution. For reference, I record the optimal carbon prices in definitions.

Definition 1. *The* ***Negishi-weighted carbon price*** *is implicitly defined by*

$$\tilde{\tau} = C'_i(\tilde{A}_i) = -\sum_i D'_i(\tilde{A}). \tag{6}$$

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

[12] Note that I am using prime notation for derivatives: $C'_i(A_i) \equiv \frac{dC_i(A_i)}{dA_i}$, $D'_i(A) \equiv \frac{dD_i(A)}{dA}$, $u'(x_i) \equiv \frac{du(x_i)}{dx_i}$.

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

1954). We have thus obtained the knife-edge result that the Negishi-weighted carbon price is uniform even though we allowed for differentiated carbon prices. Uniform carbon prices arise from the specification of the Negishi weights, which equalize weighted marginal utilities across regions. Notably, this also renders 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 Equation (6), we can see that the Negishi-weighted carbon price equalizes the sum of the marginal abatement costs and benefits from marginally increasing the carbon price (see Appendix A.2.1 for a derivation):

$$\sum_i \frac{dC_i(\tilde{A}_i(\tilde{\tau}))}{d\tilde{\tau}} = -\sum_i \frac{dD_i(\tilde{A}(\tilde{\tau}))}{d\tilde{\tau}}. \tag{7}$$

The utilitarian solution with uniform carbon prices

Next, I turn to the optimal carbon prices under the utilitarian SWF. Utilitarian welfare weights are uniform across regions. Without loss of generality, I set them equal to unity: $\alpha_i^U = 1$. To highlight that the maximization of the utilitarian SWF maximizes the (equally-weighted/unweighted) sum of utilities, I refer to the utilitarian solutions as *welfare-maximizing* solutions.

First, I solve the uniform carbon price optimization problem to determine the uniform carbon price that maximizes global welfare.

Definition 2. *The* ***utilitarian uniform carbon price*** *is implicitly defined by*

$$\check{\tau} = C_i'(\check{A}_i) = -\sum_i u'(\check{x}_i) D_i'(\check{A}) \frac{C_S'' + C_N''}{u'(\check{x}_N)C_S'' + u'(\check{x}_S)C_N''}. \tag{8}$$

The utilitarian uniform carbon price is a function of the sum of the avoided marginal damages in welfare terms rather than monetary terms, which 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, $\frac{dA_i(\tau_i)}{d\tau_i} = \frac{1}{C_i''}$. Thus, raising 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 (8) in terms of the derivatives with respect to the carbon price (see Appendix A.2.2):

$$\sum_i u'(\check{x}_i) \frac{dC_i(\check{A}_i(\check{\tau}))}{d\check{\tau}} = -\sum_i u'(\check{x}_i) \frac{dD_i(\check{A}(\check{\tau}))}{d\check{\tau}}. \tag{9}$$

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.

The utilitarian solution with differentiated carbon prices

I now relax the constraint of uniform carbon prices and solve the differentiated carbon price optimization problem.

Definition 3. *The* ***utilitarian differentiated carbon price*** *for region i is implicitly defined by*

$$\hat{\tau}_i = C_i'(\hat{A}_i) = -\frac{1}{u'(\hat{x}_i)} \sum_{j \in \mathcal{I}} u'(\hat{x}_j) D_j'(\hat{A}). \tag{10}$$

The utilitarian differentiated carbon prices equalize the marginal *welfare* costs of abatement across regions (as opposed to the marginal *monetary* costs of abatement in the Negishi solution), which, in turn, are equal to the marginal welfare benefits of abatement:

$$u'(\hat{x}_N) C_N'(\hat{A}_N) = u'(\hat{x}_S) C_S'(\hat{A}_S) = -\sum_{j \in \mathcal{I}} u'(\hat{x}_j) D_j'(\hat{A}). \tag{11}$$

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)).

Thus, the welfare-maximizing differentiated carbon price is higher in the richer region, as it is inversely proportional to the marginal utility of consumption—a result that was first established by Eyckmans et al. (1993) and Chichilnisky and Heal (1994). This implies that emissions are not reduced at the lowest *monetary* cost, and emission reductions are therefore 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[14], particularly in the context of the new regulatory impact analysis guidelines in the US (Circular A-4), which allow for distributional weighting in cost-benefit analyses (US Office of Management and Budget, 2023).

A second important point is that the utilitarian differentiated carbon prices are Pareto efficient if international transfers cannot be made[15]. This point requires elaboration. It is

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

[15] 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.

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 infeasible. In such a constrained, second-best setting, the set of feasible allocations shrinks and the Pareto frontier moves inward (except for one point that does not require transfers, which is the Negishi solution). 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 fact, in this setting, all points on the Pareto frontier require differentiated carbon prices, except for one point, which corresponds to the Negishi solution (see Equation (A18) in the appendix). The utilitarian differentiated carbon price yields the point on the Pareto frontier that maximizes global welfare.

3.2.3 Comparison of optimal climate policy stringency

I now address the central question of this section: How does the optimal climate policy stringency depend on regional welfare weights?

Utilitarian uniform versus Negishi

I first compare the uniform carbon prices under utilitarian and Negishi weights. By construction, the utilitarian carbon price maximizes global utilitarian welfare, while the Negishi-weighted carbon price maximizes global consumption in monetary terms. The following proposition and corollary establish the conditions under which one is greater than the other.

Proposition 1. *The utilitarian uniform carbon price is greater than the Negishi-weighted carbon price, that is $\check{\tau} > \tilde{\tau}$, if and only if $\frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S}$*[16].

Proof. See Appendix A.3.

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

Proof. See Appendix A.4.

Proposition 1 establishes that the welfare-maximizing uniform carbon is greater than the Negishi-weighted carbon price if and only if the relative benefits of increasing global abatement, for the Global South compared to the Global North, exceed the relative costs. The left-hand side, $\frac{\check{D}'_S}{\check{D}'_N}$, is the relative benefit of an extra unit of global abatement A. The right-hand side, $\frac{C''_N}{C''_S}$ is the relative cost of an extra unit of global abatement. Since the marginal abatement cost (MAC) is equal across regions, the relative cost of an extra unit of

[16] I use $\check{D}'_i$ as a short-hand for $D'_i(\check{A}_i)$. This notation also applies to other functions and solutions.

aggregate abatement is determined by the relative fractions of that unit of global abatement that are provided by each region, which in turn is determined by the relative slopes of the MAC function. A steeper MAC function results in a smaller abatement increase, and therefore a smaller increase in abatement costs[17].

Using the simplified RICE functions, Proposition 1 can also be expressed in terms of the damage and abatement cost functions per endowment, allowing for a more straightforward comparison of economies of different sizes[18]:

$$\check{\tau} > \tilde{\tau} \iff \frac{\check{d}'_S}{\check{d}'_N} > \frac{c''_N}{c''_S}.$$

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 made 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.

We may also be interested in how different factors affect the magnitude of the difference between the two carbon prices. To this end, it is useful to define the *utilitarian-Negishi uniform carbon price ratio*, $\check{\tau}/\tilde{\tau}$. Using the simplified RICE functions (which allow for an easier interpretation), this ratio is given by

$$\begin{aligned} \frac{\check{\tau}}{\tilde{\tau}} &= \frac{\check{u}'_N L_N w_N \check{d}'_N + \check{u}'_S L_S w_S \check{d}'_S}{L_N w_N \tilde{d}'_N + L_S w_S \tilde{d}'_S} \frac{c''_S \frac{1}{L_S w_S} + c''_N \frac{1}{L_N w_N}}{\check{u}'_N c''_S \frac{1}{L_S w_S} + \check{u}'_S c''_N \frac{1}{L_N w_N}} \\ &\approx \frac{\frac{L_S}{L_N}\left(\frac{w_S}{w_N}\right)^{1-\eta}\frac{d'_S}{d'_N} + 1}{\frac{L_S}{L_N}\frac{w_S}{w_N}\frac{d'_S}{d'_N} + 1} \frac{\frac{L_S}{L_N}\frac{w_S}{w_N}\frac{c''_N}{c''_S} + 1}{\frac{L_S}{L_N}\left(\frac{w_S}{w_N}\right)^{1-\eta}\frac{c''_N}{c''_S} + 1}, \end{aligned} \tag{12}$$

where the second line assumes that the utility function is isoelastic, $u(x) = \frac{x^{1-\eta}}{1-\eta}$ (for $\eta = 1$, $u(x) = \log(x)$, where η is the elasticity of the marginal utility of consumption), marginal damages are approximately equal in the utilitarian and Negishi solutions, $\check{d}'_i \approx \tilde{d}'_i$, and the per capita consumption and endowment ratios are approximately equal, $\frac{x_S}{x_N} \approx \frac{w_S}{x_N}$. The latter

[17]To see this, notice that $\frac{C''_N}{C''_S} = \frac{\frac{d\check{A}_S}{d\check{\tau}}}{\frac{d\check{A}_N}{d\check{\tau}}} = \frac{\frac{d\check{A}_S}{d\check{A}}}{\frac{d\check{A}_N}{d\check{A}}} = \frac{\frac{d\check{C}_S}{d\check{A}}}{\frac{d\check{C}_N}{d\check{A}}}$, where $\frac{dA_i}{d\tau_i} = \frac{1}{C''_i}$, and the third equality follows from $\frac{d\check{C}_S}{d\check{A}_S} = \frac{d\check{C}_N}{d\check{A}_N}$.

[18]Here, $d'_i = \frac{dd_i(A)}{dA}$ and $c''_i = \frac{d^2 c_i\left(\frac{A_i}{W_i}\right)}{d\left(\frac{A_i}{W_i}\right)^2}$. Also note that $D'_i = W_i d'_i$ and $C''_i = c''_i \frac{1}{W_i}$.

two approximations are useful because they allow us to write the utilitarian-Negishi uniform carbon price ratio simply as a function of the ratios of variable values in the South compared to the North.

Using these approximations, Table 1 illustrates how the carbon price ratio is affected by the abatement cost and damage functions, inequality and inequality aversion. The default values of the population and endowment per capita ratios are $\frac{L_S}{L_N} = 3.7$ and $\frac{w_N}{w_S} = 3.2$, respectively, which are calibrated to empirical data in 2023 (World Economics, 2024)[19].

Table 1: Utilitarian-Negishi uniform carbon price ratio ($\check{\tau}/\tilde{\tau}$). Static model.

	$\eta = 1$			$\eta = 1.5$		
d'_S/d'_N:	0.5	1	2	0.5	1	2
A) Abatement costs						
$c''_N/c''_S = 0.5$	1.00	1.21	1.40	1.00	1.29	1.57
$c''_N/c''_S = 1$	0.83	1.00	1.16	0.77	1.00	1.22
$c''_N/c''_S = 2$	0.71	0.86	1.00	0.64	0.82	1.00
B) Inequality						
$w_N/w_S = 1$	1.00	1.00	1.00	1.00	1.00	1.00
$w_N/w_S = 3.2$	0.83	1.00	1.16	0.77	1.00	1.22
$w_N/w_S = 6.4$	0.74	1.00	1.31	0.67	1.00	1.39

Notes: The carbon price ratios are approximations based on Equation (12). Variable values that are not shown are set as follows: In both panels, $L_S/L_N = 3.7$. In panel A, $w_N/w_S = 3.2$. In panel B, $c''_N/c''_S = 1$.

Panel A of Table 1 shows that relatively higher marginal damages and a more convex abatement cost function in the South increase the carbon price ratio. Confirming the insight from Proposition 1, carbon prices are equal if $\frac{d'_S}{d'_N} = \frac{c''_N}{c''_S}$. Panel B demonstrates that greater inequality amplifies the difference between the utilitarian and Negishi-weighted uniform carbon prices[20], as does a more concave utility function, which implies a higher inequality aversion. Furthermore, there is no difference between the carbon prices if there is no inequality or if the utility function is linear (i.e., $\eta = 0$).

Utilitarian differentiated versus Negishi

I now turn to the utilitarian differentiated carbon price solution. I explore when the utilitar-

[19] The endowment per capita ratio is calibrated to the GDP per capita ratio (in PPP terms).

[20] This also suggests that accounting for inequality at a more granular resolution (e.g., across countries) may increase the carbon price ratio.

ian differentiated carbon price solution leads to higher or lower global emissions compared to the Negishi solution. I begin by establishing the following lemma.

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.5.

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 2 establishes the condition under which this is the case.

Proposition 2. *The global abatement under utilitarian differentiated carbon prices is greater than under the Negishi-weighted carbon price, that is $\hat{A} > \tilde{A}$, if and only if $\frac{\hat{u}'_S}{\hat{u}'_N}\frac{\hat{D}'_S}{\hat{D}'_N} > \frac{C''_N}{C''_S}$.*

Proof. See Appendix A.6.

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 1. 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 2; 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 $\frac{\hat{u}'_S}{\hat{u}'_N} > 1$. 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 relatively steeper marginal abatement cost in the South and a flatter one in the North make it more likely that the aggregate abatement increases. It is also worth noting the subtle, but important, difference in intuition behind the $\frac{C''_N}{C''_S}$ term in Propositions 1 and 2. In Proposition 1, 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 2, it reflects how much abatement in the South decreases and how much it increases in the North when we allow for differentiated carbon prices.

3.2.4 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 for a region is obtained by solving the uniform carbon price optimization problem with welfare weights fully assigned to that region; thus, $\alpha_i = 1$ and $\alpha_{-i} = 0$.

Definition 4. *The **preferred uniform carbon price of region i** is implicitly defined by*

$$\mathring{\tau}^i = C'_i(\mathring{A}^i_i) = C'_{-i}(\mathring{A}^i_{-i}) = -D'_i(\mathring{A}^i)\frac{C''_i + C''_{-i}}{C''_{-i}}, \tag{13}$$

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).

Equation (13) reveals that a region's preferred uniform carbon price is higher when its marginal benefit of abatement is large and when its abatement cost function is more convex 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[21]. The crucial role of the relative convexities of the abatement cost functions for region's preferred uniform carbon prices is underappreciated in the existing literature[22]. It is also worth noting that a region's preferred uniform carbon price is greater

[21]To see this, note that $\frac{C''_i + C''_{-i}}{C''_{-i}} = 1 + \frac{\frac{dA_{-i}}{d\mathring{\tau}^i}}{\frac{dA_i}{d\mathring{\tau}^i}} = 1 + \frac{\frac{d\mathring{C}_{-i}}{d\mathring{\tau}^i}}{\frac{d\mathring{C}_i}{d\mathring{\tau}^i}}$, where the second equality follows from $\frac{d\mathring{C}_{-i}}{dA_{-i}} = \frac{d\mathring{C}_i}{dA_i}$.

[22]Most studies assume uniform convexities of the abatement cost function across regions (Weitzman, 2014; Weitzman, 2017b; Kotchen, 2018). Weitzman (2017a) allows for different convexities of the abatement cost function across regions, but does not highlight their role.

than its own marginal benefit of abatement, $-D_i'$. The is because region i accounts for the fact that increasing a uniform carbon price results in additional abatement in the other region. This is represented by the term $\frac{C_i''+C_{-i}''}{C_{-i}''} = 1 + \frac{A_{-i}'(\mathring{\tau}^i)}{A_i'(\mathring{\tau}^i)} > 1$, where $A_i'(\tau_i) \equiv \frac{dA_i(\tau_i)}{d\tau_i}$.

It is again instructive to rewrite the optimality condition in Equation (13) in terms of the derivatives with respect to the uniform carbon price (see Appendix A.2.3):

$$\frac{dC_i(\mathring{A}_i(\mathring{\tau}^i))}{d\mathring{\tau}^i} = -\frac{dD_i(\mathring{A}(\mathring{\tau}^i))}{d\mathring{\tau}^i}. \tag{14}$$

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 begin 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.7.

The intuition behind 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. Utilitarian and Negishi weights are linear combinations of these edge weights, giving a positive weight to both regions. It is therefore not surprising that "edge weights" results in more extreme carbon prices than "more balanced" welfare weights.

Using Lemma 2, I establish the following relationship between regions' preferred uniform carbon prices and the main result detailed in Proposition 1.

Proposition 3. *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.8.

The intuition for the result of Proposition 3 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. Since Negishi weights downweight the South, it is intuitive that the utilitarian uniform carbon price is greater than the Negishi-weighted carbon price if the South prefers a higher uniform carbon price than the North. 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 simply depends on whether South's preferred uniform carbon price is greater or lower than North's.

3.2.5 Extension: Dynamic model

An important aspect of climate change is that emission reductions today reduce the impact of climate change in the future. To capture this temporal dimension, I consider a two-period model in this section, which I refer to as "dynamic". I focus on uniform carbon prices in this extension to illustrate how welfare weights affect optimal carbon prices in a dynamic setting, even when the policy instrument is identical—a globally uniform carbon price.

Model modifications

The objective of the dynamic model is to account for the fact that the benefits of abatement come with a delay. To capture this in the simplest way, I assume that abatement occurs in the first period and climate damages in the second period. Aggregate regional consumption is thus given by $X_{i1} = W_{i1} - C_{i1}(A_i)$ and $X_{i2} = W_{i2} - D_{i2}(A)$, where the second subscript denotes the period $t \in \{1, 2\}$.

Optimal carbon prices

The optimal uniform carbon prices for the dynamic model are obtained by solving the following optimization problem:

$$\max_{X_{it}, A_{i1}} \sum_t \sum_i \beta^{t-1} L_{it} \alpha_{it} u_{it} \left(\frac{X_{it}}{L_{it}} \right) \tag{15}$$

$$\begin{aligned} \text{subject to: } & X_{i1} = W_{i1} - C_{i1}(A_i), \quad \forall i, \\ & X_{i2} = W_{i2} - D_{i2}(A), \quad \forall i, \\ & C'_{N1}(A_N) = C'_{S1}(A_S), \end{aligned} \tag{16}$$

where β^{t-1} is the utility discount factor (given by $\beta^{t-1} = (1+\rho)^{1-t}$, where ρ is the utility discount rate or pure rate of time preference).

The welfare weights are defined as follows. Utilitarian weights are uniform across regions and periods and set to unity; that is, $\alpha_{it}^U = 1$. Negishi weights are time-variant and defined in accordance with the RICE model: $\tilde{\alpha}_{i1} = \frac{1}{\tilde{u}'_{i1}}$ and $\tilde{\alpha}_{i2} = v \frac{1}{\tilde{u}'_{i2}}$, where $v = \pi \frac{\tilde{u}'_{N2}}{\tilde{u}'_{N1}} + (1-\pi) \frac{\tilde{u}'_{S2}}{\tilde{u}'_{S1}}$ is the wealth-based component of the social discount factor[23], which is pinned down as a weighted average of the regional wealth-based discount factors. The discounting weights $\pi \in (0,1)$ and $(1-\pi)$ are given by the regional capital or output shares in previous versions of the RICE model (Nordhaus and Boyer, 2000; Nordhaus, 2010). I consider general discounting weights, unless explicitly specified.

Solving the optimization problem above yields the following optimal carbon prices[24].

Definition 5. *The* ***dynamic Negishi-weighted carbon price*** *is implicitly defined by*

$$\tilde{\tau} = C'_{i1}(\tilde{A}_i) = -\upsilon\beta \sum_i D'_{i2}(\tilde{A}). \tag{17}$$

Definition 6. *The* ***dynamic utilitarian uniform carbon price*** *is implicitly defined by*

$$\check{\tau} = \tilde{C}'_{i1} = -\beta \sum_i u'(\check{x}_{i2}) D'_{i2}(\check{A}) \frac{C''_{S1} + C''_{N1}}{u'(\check{x}_{N1})C''_{S1} + u'(\check{x}_{S1})C''_{N1}} \tag{18}$$

These expressions are similar to the ones of the static model with the important difference that damages occur in the second period (and are discounted) while abatement takes place in the first period. Consequently, optimal carbon prices are generally affected by the developments of endowment, consumption per capita, and population, which is the focus of the comparative analysis below. Moreover, discounting is affected by the choice of welfare weights; while the utility discount factor is assumed to be the same, the wealth-based component of the social discount factor differs. Under Negishi weights, the wealth-based component of the social discount factor is υ, which is uniform across regions and given by the weighted average of the regional wealth-based discount factors[25]. In contrast, under utilitarian weights, it is simply the regional wealth-based discount factor, $u'(x_{i2})/u'(x_{i1})$, for each region. Notably, utilitarian weights value consumption across regions in the same fashion as across periods.

Comparative results

As before, the central question is how the utilitarian uniform carbon price compares to the Negishi-weighted carbon price. The following proposition establishes this relationship.

[23]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 η is the elasticity of the marginal utility of consumption and 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.

[24]The derivation is largely analogous to the static model (see Appendix A.1).

[25]Dennig and Emmerling (2019) and Anthoff et al. (2021) show that this distorts regional time-preferences.

Proposition 4. *The dynamic utilitarian uniform carbon price is greater than the dynamic Negishi-weighted carbon price, that is $\check{\tau} > \tilde{\tau}$, if and only if*

$$\frac{\check{u}'_{N2}\check{D}'_{N2} + \check{u}'_{S2}\check{D}'_{S2}}{\check{D}'_{N2} + \check{D}'_{S2}} > v\frac{\check{u}'_{N1}C''_{S1} + \check{u}'_{S1}C''_{N1}}{C''_{S1} + C''_{N1}}, \tag{19}$$

where $v = \pi\frac{\check{u}'_{N2}}{\check{u}'_{N1}} + (1-\pi)\frac{\check{u}'_{S2}}{\check{u}'_{S1}}$.

Proof. See Appendix A.9.

As in the static model, this condition is more likely to be satisfied if the South has relatively higher marginal damages and a more convex abatement cost function. All else equal, this is also the case for a lower wealth-based component of the social discounting factor under the Negishi-weighted SWF, v.

Crucially, the damage and abatement cost functions generally depend on the economy size, which in turn depends on the population size[26]. Since the costs and benefits of abatement occur in different periods, economic and population growth affect the relative regional costs and benefits of abatement. Using the simplified RICE functions, we can rewrite the condition in Proposition 4 as

$$\frac{\check{u}'_{N2}L_{N1}g^L_N w_{N1}g^w_N \check{d}'_{N2} + \check{u}'_{S2}L_{S1}g^L_S w_{S1}g^w_S \check{d}'_{S2}}{L_{N1}g^L_N w_{N1}g^w_N \check{d}'_{N2} + L_{S1}g^L_S w_{S1}g^w_S \check{d}'_{N2}} > v\frac{\check{u}'_{N1}\frac{1}{L_{S1}w_{S1}}c''_{S1} + \check{u}'_{S1}\frac{1}{L_{N1}w_{N1}}c''_{N1}}{\frac{1}{L_{S1}w_{S1}}c''_{S1} + \frac{1}{L_{N1}w_{N1}}c''_{N1}}, \tag{20}$$

where I used $L_{i2} = L_{i1}g^L_i$ and $w_{i2} = w_{i1}g^w_i$, with g^L_i and g^w_i denoting the population and economic growth factors, respectively.

Equation (20) yields an important insight: if population growth is faster in the South than the North, then the utilitarian uniform carbon price is more likely to exceed the Negishi-weighted carbon price. The intuition is that relatively faster population growth in the South increases the relative damages of climate change in the South, as they manifest in the future, which are given comparatively less weight under the Negishi-weighted SWF. Simply put, climate change is a bigger problem for the South if its population is growing faster, as this results in more people being harmed by climate change[27].

The role of economic growth is more complicated. This is because economic growth simultaneously affects climate damages and the development of marginal utilities of consumption,

[26] To keep the exposition simple, I assume that the endowment per capita is exogenous and does not depend on the population size.

[27] Equivalently, a larger population results in a bigger economy, thereby increasing aggregate marginal damages (which are assumed to be proportional to the economy size). To see the different interpretations formally, note that $D' = Lwd' = Wd'$. Population growth effectively plays an analogous (but opposite) role to time discounting, a point that was formally made by Budolfson et al. (2018).

which affect discounting under both SWFs[28] (note that it also affects v). However, we can gain traction on the role of economic growth with additional assumptions. To obtain intuition for the role of economic growth, it is again useful to define the utilitarian-Negishi uniform carbon price ratio. Using the simplified RICE functions, this ratio is given by

$$
\begin{aligned}
\frac{\check{\tau}}{\tilde{\tau}} &= \frac{1}{\pi\frac{\tilde{u}'_{N2}}{\tilde{u}'_{N1}} + (1-\pi)\frac{\tilde{u}'_{S2}}{\tilde{u}'_{S1}}} \frac{\check{u}'_{N2} L_{N2} w_{N2} \check{d}'_{N2} + \check{u}'_{S2} L_{S2} w_{S2} \check{d}'_{S2}}{L_{N2} w_{N2} \tilde{d}'_{N2} + L_{S2} w_{S2} \tilde{d}'_{S2}} \frac{c''_{S1}\frac{1}{L_{S1}w_{S1}} + c''_{N1}\frac{1}{L_{N1}w_{N1}}}{\check{u}'_{N1} c''_{S1}\frac{1}{L_{S1}w_{S1}} + \check{u}'_{S1} c''_{N1}\frac{1}{L_{N1}w_{N1}}} \\
&\approx \frac{\frac{L_{S1}}{L_{N1}}\frac{g^L_S}{g^L_N}\frac{w_{S1}}{w_{N1}}\frac{g^w_S}{g^w_N} + 1}{\frac{L_{S1}}{L_{N1}}\frac{g^L_S}{g^L_N}\frac{w_{S1}}{w_{N1}}\left(\frac{g^w_S}{g^w_N}\right)^{1-\eta} + 1} \frac{\frac{L_{S1}}{L_{N1}}\frac{g^L_S}{g^L_N}\left(\frac{w_{S1}}{w_{N1}}\frac{g^w_S}{g^w_N}\right)^{1-\eta}\frac{d'_{S2}}{d'_{N2}} + 1}{\frac{L_{S1}}{L_{N1}}\frac{g^L_S}{g^L_N}\frac{w_{S1}}{w_{N1}}\frac{g^w_S}{g^w_N}\frac{d'_{S2}}{d'_{N2}} + 1} \frac{\frac{c''_{N1}}{c''_{S1}}\frac{L_{S1}}{L_{N1}}\frac{w_{S1}}{w_{N1}} + 1}{\left(\frac{w_{S1}}{w_{N1}}\right)^{-\eta}\frac{c''_{N1}}{c''_{S1}}\frac{L_{S1}}{L_{N1}}\frac{w_{S1}}{w_{N1}} + 1}.
\end{aligned}
\tag{21}
$$

The second line utilizes the following assumptions and approximations: (1) the utility function is isoelastic, $u(x) = \frac{x^{1-\eta}}{1-\eta}$ (for $\eta = 1$, $u(x) = \log(x)$), (2) the discounting weights are given by the regional endowment shares[29], $\pi = \frac{W_{N2}}{\sum_i W_{i2}}$, (3) per capita consumption and endowment growth are approximately equal, $g^x_i \approx g^w_i$, and the per capita consumption and endowment ratios are approximately equal, $\frac{x_{St}}{x_{Nt}} \approx \frac{w_{St}}{x_{Nt}}$, (4) per capita consumption growth and marginal damages are approximately equal in the utilitarian and Negishi solutions, $\check{g}^x_i \approx \tilde{g}^x_i$ and $\check{d}'_i \approx \tilde{d}'_i$. These assumptions will generally not hold precisely but can be expected to be good approximations, serving the purpose of obtaining clean intuitions for the role of economic growth.

Using these approximations, the utilitarian-Negishi uniform carbon price ratio only depends on ratios of variable values in the South compared to the North. To demonstrate the role of population and economic growth, an illustrative numerical example of carbon price ratios is shown in Table 2. For these calculations, I assume that the first and second periods are 50 years apart and the growth factors are given by $g^y_i = (1 + \bar{g}^y_i)^t$, where $y \in \{L, w\}$, $\bar{g}^y_i$ are the annual growth rates and $t = 50$.

Panel A of Table 2 confirms that faster population growth in the South increases the carbon price ratio. Importantly, this holds even if marginal damages per endowment are homogeneously distributed across regions (i.e., $d'_{S2} = d'_{N2}$). Panel B examines the effect of faster economic growth in the South in terms of endowment per capita. The first thing to note is that economic growth plays no role if marginal damages per endowment are evenly distributed and $\eta = 1$. However, economic growth reduces the carbon price ratio if either

[28] For a thorough examination of how interregional inequality and heterogeneous economic growth impact the discount rate under the utilitarian SWF, see Gollier (2015).

[29] Both endowment and capital shares have been used in previous versions of the RICE model (Nordhaus and Boyer, 2000; Nordhaus, 2010). The RICE-2010 model uses capital shares but both approaches are numerically close, according to Nordhaus and Boyer (2000).

Table 2: Utilitarian-Negishi uniform carbon price ratio ($\check{\tau}/\tilde{\tau}$). Dynamic model.

	$\eta = 1$		$\eta = 1.5$	
d'_{S2}/d'_{N2}:	1	2	1	2
A) Population growth				
$\bar{g}^L_S = 0\%$, $\bar{g}^L_N = 0\%$	1.00	1.16	1.00	1.22
$\bar{g}^L_S = 1\%$, $\bar{g}^L_N = 0\%$	1.12	1.26	1.16	1.34
$\bar{g}^L_S = 2\%$, $\bar{g}^L_N = 0\%$	1.22	1.33	1.29	1.44
B) Economic growth				
$\bar{g}^w_S = 2\%$, $\bar{g}^w_N = 2\%$	1.22	1.33	1.29	1.44
$\bar{g}^w_S = 3\%$, $\bar{g}^w_N = 2\%$	1.22	1.27	1.23	1.30
$\bar{g}^w_S = 4\%$, $\bar{g}^w_N = 2\%$	1.22	1.23	1.16	1.18

Notes: The carbon price ratios are approximations based on Equation (21). Variable values that are not shown are set as follows: In both panels, $w_{N1}/w_{S1} = 3.2$, $L_{S1}/L_{N1} = 3.7$, $c''_{N1}/c''_{S1} = 1$. In panel A, $\bar{g}^w_S/g^w_N = 1$. In panel B, $\bar{g}^L_S = 2\%$, $\bar{g}^L_N = 0\%$.

(1) the utility function is more concave than logarithmic utility ($\eta > 1$) and $d'_{S2} \geq d'_{N2}$, or (2) the South has disproportionately high climate damages ($d'_{S2} > d'_{N2}$) and $\eta \geq 1$. Since climate damages are expected to be disproportionately large in the South, this last case is the most relevant in practice. Hence, faster economic growth in the South can be expected to reduce the carbon price ratio.

3.3 Welfare-maximizing carbon prices with mitigation finance

The previous section explored how optimal carbon prices depend on welfare weights in the absence of transfers. This section relaxes the constraint of no international transfers and examines the effect of these transfers on welfare-maximizing (utilitarian) carbon prices. I focus on conditional transfers to fund emission reductions, referred to as mitigation finance, which constitutes the main type of international climate finance under the Paris Agreement (OECD, 2024).

I define the conditional transfer for mitigation from the North to the South as the difference between the total abatement cost in the South (funded from domestic and foreign sources) and the abatement cost that is borne by the South,

$$T = C_S(A_{SD} + A_{SF}) - C_S(A_{SD}), \tag{22}$$

where A_{SD} and A_{SF} denotes the abatement in the South that is funded from domestic and foreign sources, respectively[30]. Aggregate abatement is now given by $A = A_{SD} + A_{SF} + A_{ND}$.

The transfer quantity may be capped at T^{max}, and the abatement costs that are borne by the South may be required not to fall below C_S^{min}. This reflects (potential) real-world constraints, imposed by donor countries, that the provision of mitigation finance is limited and that these funds may need to be used for *additional* abatement relative to the abatement level that would have taken place in the absence of transfers. To focus the discussion, I consider both the presence and absence of an "additionality" condition; in the latter case, abatement costs borne by the South are constrained to be non-negative. Thus, $C_S^{min} \in \{0, C_S^0\}$, where C_S^0 denotes South's optimal abatement cost in the absence of transfers. From the social planner's perspective, these constraints reduce the feasible set of allocations, thus (weakly) reducing global welfare.

In the presence of mitigation finance, the utilitarian planner solves the following problem if carbon price differentiation is feasible:

$$\max_{X_N, X_S, A_{ND}, A_{SD}, A_{SF}} \sum_i L_i u\left(\frac{X_i}{L_i}\right) \tag{23}$$

$$\begin{aligned} \text{subject to: } & X_N = W_N - C_N(A_{ND}) - D_N(A) \overbrace{-C_S(A_{SD} + A_{SF}) + C_S(A_{SD})}^{-T} \\ & X_S = W_S - C_S(A_{SD}) - D_S(A) \\ & C_S(A_{SD}) \geq C_S^{min} \\ & \underbrace{C_S(A_{SD} + A_{SF}) - C_S(A_{SD})}_{T} \leq T^{max}. \end{aligned} \tag{24}$$

In the case in which carbon prices are constrained to be uniform, the constraint of equal marginal abatement costs, $C'_S(A_{SD} + A_{SF}) = C'_N(A_{ND})$, is added to Equations (24).

The optimal carbon prices of these problems depend on whether the constraints on transfer and domestic abatement are binding. Tables 3 and 4 summarize the expressions for the optimal carbon prices under the differentiated and uniform carbon price problems, respectively (see Appendix A.10 for derivations and additional details). As the transfer constraint is gradually relaxed, different phases will unfold: the domestic abatement constraint may begin to bind and the transfer constraint ceases to be binding. An example of these phases is sketched in Figure 3.

For the differentiated carbon price solution, a sufficiently small amount of mitigation finance results in a binding transfer constraint and a slack domestic abatement constraint

[30]Note that I assume, for notational simplicity, that the foreign-funded abatement comes in on top of the domestically-funded abatement.

Table 3: Utilitarian differentiated carbon prices in the presence of mitigation finance.

	Region	$C_S(A_{SD}) > C_S^{min}$	$C_S(A_{SD}) = C_S^{min}$
$T = T^{max}$	South	$-\frac{1}{u'_S}\sum_i u'_i D'_i$	$-\frac{1}{u'_S - \nu}\sum_i u'_i D'_i$
	North	$-\frac{1}{u'_N}\sum_i u'_i D'_i$	$-\frac{1}{u'_N}\sum_i u'_i D'_i$
$T < T^{max}$	Both	$-\sum_i D'_i$	$-\frac{1}{u'_N}\sum_i u'_i D'_i$

Notes: The carbon prices in the North and the South are given by the marginal abatement costs $C'_N(A_{ND})$ and $C'_S(A_{SD}+A_{SF})$, respectively.

if $C_S^{min} = 0$ (if $C_S^{min} = C_S^0$, the domestic abatement constraint binds immediately and this phase is skipped by construction). In this phase, the conditions for the optimal carbon prices in the North and the South remain the same as in the case without transfers[31]. The utilitarian planner uses the transfer to reduce abatement costs in the South, as this yields the largest social welfare gain; domestic abatement is thus crowded out by foreign abatement and the transfer has the same effect as a non-conditional transfer would have. However, this shift in the cost burden from the South to the North changes consumption levels and marginal utilities of the two regions. Specifically, $\frac{u'_S}{u'_N}$ decreases[32], thereby increasing (decreasing) the optimal carbon price in the South (North).

Table 4: Utilitarian uniform carbon prices in the presence of mitigation finance.

	$C_S(A_{SD}) > C_S^{min}$	$C_S(A_{SD}) = C_S^{min}$
$T = T^{max}$	$-\sum_i u'_i D'_i \frac{C''_S + C''_N}{u'_S C''_N + u'_N C''_S}$	$-\sum_i u'_i D'_i \frac{C''_N + C''_S}{(u'_S - \nu) C''_N + u'_N C''_S}$
$T < T^{max}$	$-\sum_i D'_i$	$-\frac{1}{u'_N}\sum_i u'_i D'_i$

Notes: The uniform carbon prices are given by the uniform marginal abatement costs $C'_N(A_{ND}) = C'_S(A_{SD} + A_{SF})$.

As the transfer is increased further, one of two things may occur. First, the transfer constraint may cease to be binding while the domestic abatement constraint remains slack. This happens if and only if the transfer is sufficient to equalize consumption across regions. Consequently, the carbon price is equalized and the Samuelson condition is obtained again. However, this case is unlikely to occur, given the large inequality in the real world. Instead, the more realistic case is that the transfer constraint remains binding and the domestic

[31] Also note that the optimality condition for the North remains unchanged independently of the transfer quantity.

[32] To see this, we can appeal to the envelope theorem and note that the direct effects of the transfer are $\frac{\partial X_N}{\partial T} = -1$ and $\frac{\partial X_S}{\partial T} = 1$.

abatement constraint starts to bind. Once this happens, every additional dollar of mitigation finance is directly used for additional abatement in the South. The resulting marginal abatement cost in the South lies between North's and South's domestic marginal abatement costs since $u'_S - \nu \in (u'_N, u'_S)$, where $\nu \geq 0$ is the Lagrange multiplier on the domestic abatement constraint (i.e., it is the social value of marginally relaxing the domestic abatement constraint). Since $u'_S - \nu > u'_N$, the marginal welfare cost of this foreign-funded abatement, $u'_N C'_S(A_{SD}+A_{SF})$, remains lower than the marginal welfare benefit of abatement, $-\sum_i u'_i D'_i$. Thus, relaxing the transfer constraint further increases social welfare.

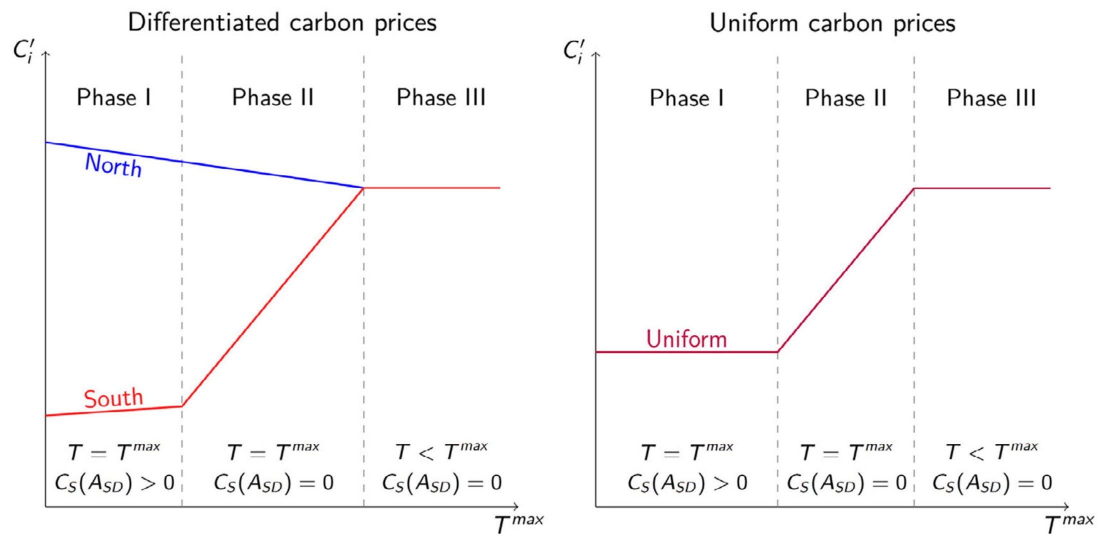

Figure 3: Sketch of carbon price development as mitigation finance is increased.

Notes: In phase I, the transfer constraint binds and the domestic abatement constraint is slack. In phase II, the domestic abatement constraint starts to bind. In phase III, the transfer constraint becomes slack. The graphs illustrate the case of $C_S^{min} = 0$. In the case of $C_S^{min} = C_S^0$, phase I is skipped. The carbon prices in the North and the South are given by the marginal abatement costs $C'_N(A_{ND})$ and $C'_S(A_{SD}+A_{SF})$, respectively. Note that this is a simplified sketch and details (such as the curvature of lines) are omitted. Also note that the line in phase I under uniform carbon prices may increase or decrease (see Footnote 35).

Eventually, the transfer constraint stops binding when the socially optimal level of mitigation finance is reached. The marginal abatement costs in the South and the North are then equalized again, however not at the first-best efficient level given by the Samuelson condition, but at the level that equalizes the marginal *welfare* costs and benefits of domestic and foreign abatement; that is, $u'_N C'_N(A_{ND}) = u'_N C'_S(A_{SD} + A_{SF}) = -\sum_i u'_i D'_i$. This implies that the marginal *monetary* costs of abatement exceed the marginal *monetary* benefits of abatement, $C'_N(A_{ND}) = C'_S(A_{SD} + A_{SF}) > -\sum_i D'_i$. Yet, in a constrained setting without non-conditional transfers, or with restricted exogenous non-conditional transfers[33], this

[33] For the latter case, interpret the endowments in Equation (24) as the endowments including the transfer.

solution maximizes global welfare. It is worth emphasizing, however, that additional welfare gains (and indeed Pareto improvements) could be achieved by increasing non-conditional transfers and decreasing mitigation finance, if this is feasible[34].

The logic of the different phases under the uniform carbon price solution is largely analogous. Initially, mitigation finance replaces domestic abatement in the South[35] (unless this is prevented by $C_S^{min} = C_S^0$). If inequality across regions persists, the uniform carbon price increases once the domestic abatement constraint binds and mitigation finance is increased further. Finally, as in the differentiated carbon price solution, the socially optimal level of mitigation finance is reached when the marginal *welfare* costs and benefits of domestic and foreign abatement are equalized.

3.4 Jointly optimal carbon prices and transfers

The previous two sections examined the welfare-maximizing carbon prices in the absence of transfers and with a fixed, exogenous transfer quantity for mitigation. This section considers a third setting where transfers are endogenous to climate policy. Specifically, it explores the jointly welfare-maximizing policy package of carbon prices and transfers, contingent on North's willingness to increase transfers to the South when it has more resources available, either through reduced domestic abatement costs or decreased climate damages.

The logic that motivates this examination is as follows. In the absence of transfers, differentiated carbon prices maximize utilitarian welfare by accounting for inequality in climate policy. However, differentiated carbon prices do not yield cost-effective emission reductions. Hence, additional global welfare gains would be possible if rich countries would make transfers to poor countries instead of accounting for inequality in climate policy through differentiated carbon prices. Importantly, this not a description of the first best setting in which the social planner has access to unrestricted transfers. Instead, this argument merely relies on the feasibility of increased transfers as rich countries benefit from changes in climate policy.

Indeed, policymakers may trade off investing in domestic emission reductions, supporting international mitigation efforts, or directing resources toward Loss and Damage payments or foreign aid. Moreover, developed countries may be willing to increase transfer payments if developing countries raise their abatement efforts, as this leads to reduced climate damages. However, political support may differ across these expenditures. For example, domestic

[34]This may also be achieved in an international emissions trading scheme through the initial permit allocation. However, this paper focuses on a setting without an international emissions trading scheme.

[35] In this phase, it is ambiguous whether the uniform carbon price increases or decreases as the transfer is increased. If the utilitarian uniform carbon price is above the efficient carbon price given by the Samuelson condition, then increasing mitigation finance, and thereby decreasing inequality, reduces the utilitarian carbon price; and vice versa for the opposite case.

constituencies may have a preference for investments that reduce carbon emissions domestically over similar investments abroad (Buntaine and Prather, 2018), which may, in turn, be preferred over international transfers for Loss and damage and foreign aid (Fabre et al., 2024). As a result, reduced domestic abatement expenditures or climate damages might free up some resources for increasing international transfers, but the relationship may not be one-to-one, reflecting differing levels of political and public support. From a conceptual perspective, North's willingness to increase transfers as it benefits from changes in climate policy affects the feasible set of allocations on which the Pareto frontier is defined.

I capture the possibility of reallocating domestic resources to international transfers in a simplified manner in the stylized two-region model. Suppose that for every dollar the North saves, relative to a reference allocation, it is willing to increase its transfers to the South by a fraction of a dollar, $\phi \in [0, 1]$. The parameter ϕ thus determines the extent to which it is feasible to increase transfers from the North to the South as the North benefits from changes in climate policy. I refer to ϕ as the transfer feasibility parameter. Formally, the endogenous transfer is given by

$$T = \phi \left(Y_N - Y_N^{ref} \right), \tag{25}$$

where $Y_N = W_N - C_N(A_N) - D_N(A)$ is the endowment net of abatement costs and climate damages (for short, net GDP) and $Y_N^{ref} = W_N - C_N(A_N^{ref}) - D_N(A^{ref})$ is the net GDP at the reference allocation. Conceptually, the reference allocation is the feasible allocation that maximizes the utilitarian SWF if transfers were not endogenous to climate policy (i.e., $\phi = 0$). It is characterized by Y_N^{ref} and a restricted transfer T^{ref}, which can be understood as a baseline quantity of foreign aid independent of climate policy. While this reference allocation affects the distribution in the final allocation, it does not affect the optimality conditions, which I am focusing on here. I therefore opt to keep the exposition general and do not further specify the reference allocation.

In this section, I focus on the utilitarian solution that allows for differentiated carbon prices to identify the conditions under which uniform carbon prices prove to be optimal from a global welfare perspective. With this goal in mind, I impose no constraints on the use of the transfer, allowing it to be used in the most advantageous manner. The transfer is thus non-conditional and can be interpreted as payments for Loss and Damage or foreign aid. Table A1 in the appendix shows the results for the case of conditional transfers for mitigation finance, combining the insights from this section and the previous one.

In this setting, the utilitarian planner solves the following problem:

$$\max_{X_N, X_S, A_N, A_S} \sum_i L_i u\left(\frac{X_i}{L_i}\right) \tag{26}$$

$$\begin{aligned}\text{subject to: } X_N &= W_N - C_N(A_N) - D_N(A) - T^{ref} - \overbrace{\phi\left(W_N - C_N(A_N) - D_N(A) - Y_N^{ref}\right)}^{T}, \\ X_S &= W_S - C_S(A_S) - D_S(A) + T^{ref} + \phi\left(W_N - C_N(A_N) - D_N(A) - Y_N^{ref}\right).\end{aligned} \tag{27}$$

The jointly optimal policy is characterized by (1) region-specific carbon prices,

$$\begin{aligned}\tau_N = C_N'(A_N) &= -\underbrace{\frac{u_S'}{u_N' + \phi(u_S' - u_N')}}_{\in\left[1, \frac{u_S'}{u_N'}\right]} D_S'(A) - D_N'(A), \\ \tau_S = C_S'(A_S) &= -D_S'(A) - \underbrace{\frac{u_N' + \phi(u_S' - u_N')}{u_S'}}_{\in\left[\frac{u_N'}{u_S'}, 1\right]} D_N'(A),\end{aligned} \tag{28}$$

and (2) an endogenous transfer,

$$T = \phi\left(W_N - C_N(A_N) - D_N(A) - Y_N^{ref}\right), \tag{29}$$

where all functions are evaluated at the optimal solution.

The main question is: How do the utilitarian carbon prices depend on the transfer feasibility parameter ϕ? It is instructive to first look at the edge cases. If $\phi = 0$, the optimality conditions collapse to the utilitarian differentiated carbon price solution of Equation (10), $C_i' = \frac{1}{u_i'} \sum_{j \in \mathcal{I}} u_j' D_j'$. In this case, it is infeasible to increase transfers as regional abatement efforts are changed. Hence, carbon prices are fully differentiated to equalize the welfare cost of abatement, resulting in welfare-cost-effective emission reductions. Conversely, if $\phi = 1$, the optimality conditions collapse to the Samuelson condition, $C_i' = \sum_i u_i' D_i'$, which is also satisfied at the Negishi solution. Notably, this solution results in uniform carbon prices and cost-effective emission reductions. Therefore, if it is feasible to fully transfer all net GDP gains in the North as a result of changed carbon prices, it becomes welfare-maximizing to adopt uniform carbon prices and address distributional issues entirely through transfers.

For intermediate values of $\phi \in (0, 1)$, carbon prices are differentiated, but the degree of differentiation diminishes as ϕ increases, along with the transfer quantity. Intuitively, as North's willingness to redistribute gains increases, transfers become an increasingly effective way to achieve global welfare improvements compared to carbon price differentiation. In brief, ϕ pins down the transfer quantity and the welfare-maximizing degree of carbon price differentiation. Crucially, uniform carbon prices are welfare-maximizing if and only if $\phi = 1$, reflecting that the North is willing to transfer all its net GDP gains from changes in carbon prices to the South.

3.5 Synthesis: Uniform or differentiated carbon prices

This section summarizes under which conditions optimal carbon prices are uniform or differentiated across regions. This is shown in the form of a decision tree in Figure 4.

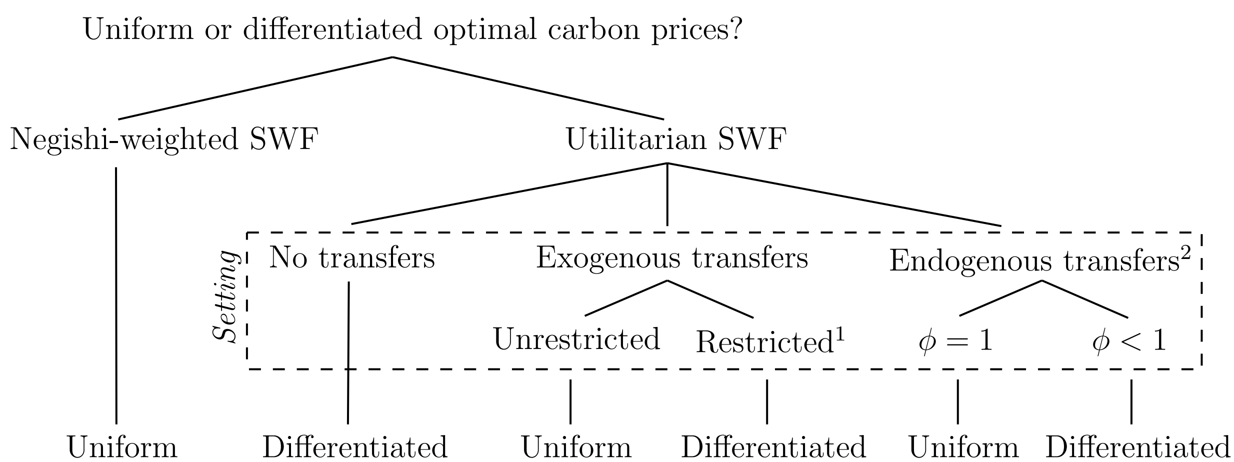

Figure 4: Conditions for uniform or differentiated optimal carbon prices.

Notes: [1]"Restricted" means restricted below the optimal level. [2]I assume restricted transfers here to focus on an additional rationale for uniform carbon prices even if transfers are restricted. If endogenous transfers are not restricted below the optimal level, uniform carbon prices are optimal.

Firstly, uniform carbon prices are optimal if inequality across regions is ignored (or considered optimal) through the use of Negishi weights. Furthermore, when inequality is accounted for, uniform optimal carbon prices remain optimal if distributional issues are better addressed through transfers. Under exogenous transfers, this is the case if transfers are sufficient to equalize consumption in the case of non-conditional transfers, or if there is enough mitigation finance to equalize marginal abatement costs[36]. Additionally, uniform carbon prices become optimal under endogenous transfers, even when the transfer quantity is restricted, if the North is willing to transfer all its net GDP gains from changes in carbon prices to the South.

Importantly, however, the transfer conditions that result in uniform optimal carbon prices are quite demanding. In practice, political constraints may prevent these conditions from being met. In such cases, differentiated carbon prices maximize global welfare.

[36]Note that this could also be achieved in international emissions trading schemes.

4 Simulations

This section presents the simulation results. Section 4.1 introduces the RICE model and methodology. Sections 4.2 and 4.3 discuss how optimal carbon prices are affected by the choice of welfare weights and international climate finance, respectively.

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, 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. Importantly, the functions that determine climate damages and abatement costs are region-specific (see Appendix C.2 for additional information).

4.1.2 Optimizations

For this analysis, I make two important modifications to the Mimi-RICE-2010 model: (1) I implement three different optimization problems, and (2) I incorporate interregional transfers. The final model that includes these modifications is referred to as *Mimi-RICE-plus*.

Optimization problems

The following three optimization problems are implemented:

1. *Negishi solution*: Maximization of the discounted Negishi-weighted SWF with no constraints on the marginal abatement costs and the interregional transfers[37].

[37] 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.

2. *Utilitarian differentiated carbon price solution*: 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. *Utilitarian uniform carbon price solution*: 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.

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[38]: The emissions control rate (which determines 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*

$$\mathcal{W}^N = \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{I}} L_{it} \beta^t \tilde{\alpha}_{it} u\left(x_{it}\right) \tag{30}$$

where $\mathcal{I}$ denotes the set of the 12 RICE regions, and $\mathcal{T} = \{0, 1, 2, ..., 590\}$ is the time horizon of the RICE model[39], 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 $\tilde{\alpha}_{it}$ are the time-variant Negishi welfare weights. The utility function is given by

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

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

The time-variant Negishi weights are given by

$$\tilde{\alpha}_{it} = \frac{1}{u'\left(\tilde{x}_{it}\right)} v_t, \tag{31}$$

[38]Note that I do not optimize the saving rates, as 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.

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

where $\tilde{x}_{it}$ is the consumption at the Negishi solution[40], v_t is the wealth-based component of the social discount factor. 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.3 for more details).

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

$$\mathcal{W}^U = \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{I}} L_{it} \beta^t u\left(x_{it}\right). \tag{32}$$

Carbon prices

In optimization problems (1) and (2), carbon prices are allowed to be differentiated across regions. However, recall that in the Negishi solution, uniform carbon prices are optimal by the construction of the Negishi weights. In the third optimization problem, a constraint of equal marginal abatement costs across regions is added[41].

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, the primary form of international climate finance agreed upon in global climate change negotiations (UNFCCC, 2015; OECD, 2024). 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.5).

I implement the conditional transfer for mitigation as follows. The richest four regions of the RICE model (US, Other High Income countries, Japan, and EU) are the donor regions, with each region contributing in proportion to its net output[42]. The total (potential) transfer quantity is \$100 billion per year in 2025 (in 2025 dollars) and increases over time with the aggregate net output in the donor regions[43]. While highly stylized, this implementation reflects the developed countries' goal of jointly mobilizing \$100 billion per year by 2020, after which a new goal shall be set of at least \$100 billion per year (UNFCCC, 2009; UNFCCC, 2015).

[40]The Negishi weights are obtained by solving the optimization multiple times (in the presence of an implicit no transfer constraint, since regions in RICE are autarkic) and iteratively updating the weights until convergence.

[41]The source code for the implementation of this constraint was adopted from the Mimi-NICE model (Dennig et al., 2017).

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

[43]Note that the transfer potential is not exhausted if it exceeds what is needed to fully abate emissions in recipient regions.

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 (for additional details, see Appendix C.4).

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*[44]. 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, 2020). It may thus be considered closest to the actual political constraints imposed on the transfer provision.

Optimization algorithms

The optimization problems are solved with the numerical 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). Parts of the source code were adopted from the mimi-NICE model (Dennig et al., 2017) and the RICEupdate model (Dennig et al., 2019).

4.2 The role of welfare weights

This section investigates how optimal carbon prices depend on the choice of welfare weights in the absence of international transfers. As in the theory section, I distinguish between two utilitarian solutions contingent on whether carbon prices are constrained to be uniform. I begin by presenting the main finding: an increased optimal climate policy stringency under both utilitarian approaches compared to the Negishi solution. Leveraging the theoretical insights, the remainder of the section explores the reasons behind this result.

[44] 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.

4.2.1 The effect on optimal carbon prices

It is useful to first examine the overall stringency of the optimal climate policy paths. To this end, Figure 5 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 commonly used positive and normative calibrations of the utility discount rates by Nordhaus (1.5%) and Stern (0.1%), respectively (Nordhaus, 2011; Stern et al., 2006)[45].

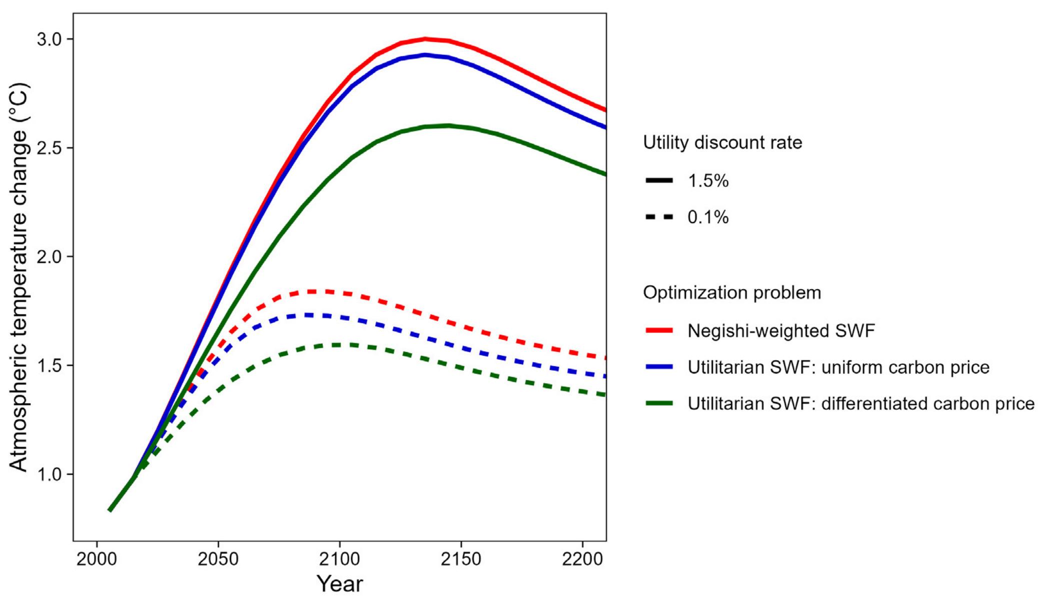

Figure 5: Optimal atmospheric temperature trajectories conditional on the optimization problem and the utility discount rate.

Notes: 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 first main result is that accounting for global inequality increases the optimal climate policy stringency in the RICE model; the utilitarian solutions with uniform and differentiated carbon prices yield lower optimal temperature trajectories than the Negishi solution. Allowing for differentiated carbon prices in the utilitarian optimization results in the lowest warming by accounting for inequality in determining both the carbon price level and differentiation. Figure 5 also shows the well-known large sensitivity of optimal climate policy to the

[45]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.

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[46] carbon dioxide emissions for the entire model horizon from 2005-2595 are shown in Table A2 in the appendix. 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 impacts of future damages are given comparatively greater weight. Specifically, relative to the Negishi solution, 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.

The optimal carbon prices[47] in 2025 are shown in Table 5 (the full carbon price trajectories are shown in Figure A2 in the appendix, along with the corresponding emissions). Welfare-maximizing uniform carbon prices exceed the Negishi-weighted carbon prices for both utility discount rates; specifically, by 15% (21%) under the 1.5% (0.1%) utility discount rate.

Furthermore, to reduce the welfare cost of abatement, there are large differences in welfare-maximizing carbon prices between regions when the constraint of equalized carbon prices is not imposed. Consistent with the theoretical results in Section 3.2.2, this yields high carbon prices in rich regions – exceeding $\sim$\$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 already reach their backstop price in 2025, resulting in zero carbon emissions. Notably, the utilitarian differentiated carbon prices exceed the Negishi-weighted carbon prices for all regions but the poorest three (or four) regions. This also results in large regional changes in cumulative emissions compared to the Negishi solution, featuring substantial emission reductions in rich regions, smaller reductions in middle-income regions, and emission increases in the poorest three regions (see Figure A3 in the appendix). Importantly, the emission reductions in rich and middle-income regions outweigh the emission increases in the poorest regions, resulting in lower global emissions compared to the Negishi solution. The differentiated carbon price optimum is discussed in more detail in Appendix C.6.

[46] 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. Cumulative emissions from land use change are 29 GtCO_2 globally over the entire model horizon from 2005-2595.

[47] Note that all dollar values are 2022 USD. I convert the 2005 USD values of the RICE model to 2022 USD values using the World Bank GDP deflator (World Bank, 2023b).

Table 5: 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

Notes: 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.

4.2.2 Regional heterogeneities and distributional effects

Why do the utilitarian maximizations lead to greater climate policy ambition compared to the Negishi solution? This section addresses this question by analyzing the distributional impacts of the different climate policy pathways and the regional heterogeneities that drive these outcomes.

I begin by examining which regions are better off and which are worse off under the utilitarian solutions compared to the Negishi solution. To provide a simple summary statistic that shows which regions gain or lose overall over the entire model horizon, I focus on regions' aggregate intertemporal welfare changes, expressed as the net present value (NPV) of consumption changes over time. These regional NPV consumption changes are shown in

Figure 6 along with the global welfare gains of the utilitarian solutions (the temporal trajectories in consumption changes are shown in Figure A6 in the appendix). 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. Global utilitarian welfare changes are expressed in the welfare-equivalent consumption change in 2005 if consumption were distributed equally. Details of this calculation are provided in Appendix A.11. For the remainder of the numerical results, I focus on the 1.5% utility discount rate to streamline the discussion. Additional results for the 0.1% utility discount rate are shown in the appendix.

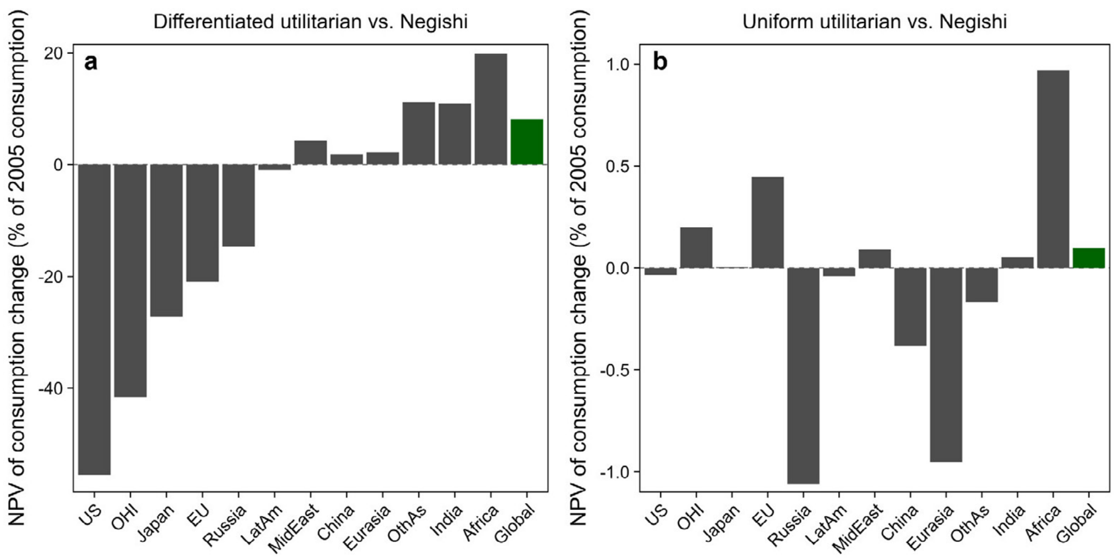

Figure 6: Net present value of consumption changes.

Notes: The values show the welfare-equivalent consumption change in 2005, as a percentage of the consumption in 2005. The “Global” value expresses the global utilitarian welfare change in the welfare-equivalent consumption change in 2005 if consumption were distributed equally (for details, see Appendix A.11). The figure shows the results for the utility discount rate of 1.5%.

The comparison of the utilitarian differentiated carbon price solution and the Negishi solution is straightforward: rich regions are better off in the Negishi solution and poor regions are better off in the utilitarian solution (see Figure 6a). Most importantly, global (unweighted) welfare is higher in the utilitarian solution. This is of course unsurprising since the utilitarian SWF measures global (unweighted) welfare and Negishi weights upweight the welfare of rich regions and downweight the welfare of poor regions. The poorest four regions are better off in all periods in the utilitarian solution, due to both lower abatement costs in their regions and lower global emissions. In contrast, rich regions experience NPV consumption losses as a result of higher abatement costs associated with increased carbon prices.

Crucially, however, all regions enjoy consumption gains after 2150 because of reduced climate damages due to reduced global emissions (see Figure A6a). 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 affected substantially, especially compared to the magnitude of the inequality across regions[48]. Thus, while the utilitarian solutions result in greater global welfare by accounting for global inequality in setting the carbon prices, they do not solve the inequality issue.

The distributional consequences are more complicated for the uniform carbon price solutions. The region that benefits the most from higher carbon prices in the utilitarian solution relative to the Negishi solution is the poorest region, Africa (see Figure 6b). Indeed, the intertemporal welfare gain in utility terms is by far the largest in Africa (see Figure A9b in the appendix). This indicates that the lower carbon prices in the Negishi solution are primarily driven by the down-weighting of Africa in the Negishi-weighted SWF. I have confirmed this through a model run in which Africa is removed from the utilitarian social welfare function[49]. Further analysis reveals that the difference in optimal policy stringency between the utilitarian and Negishi solutions is driven by disproportionately large climate damages in Africa[50]. While all regions benefit again in the long-term, Russia and Eurasia experience the greatest consumption losses in NPV terms due to increased abatement costs.

To understand the underlying drivers of the distributional effects for the uniform carbon price solutions we can leverage the results from the theory model. Proposition 4 and Equation (20) show that relatively higher marginal damages and a more convex abatement cost function in the South, compared to the North, contribute to a higher uniform carbon price in the utilitarian solution than in the Negishi solution. Figure 7 examines whether this is the case in the RICE model, showing the ratios of the regional marginal damages and convexities of the abatement cost functions (both as a percentage of GDP in 2025) relative to the US. Marginal damages are estimated as the present value (in 2025) of the stream of damages associated with a marginal pulse of emissions in 2025 (using region-specific discount rates)[51].

[48] 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 A8 in the appendix.

[49] To be more specific, setting Africa's welfare weights to zero in the utilitarian social welfare function yields an optimal carbon price trajectory nearly identical to that under Negishi weights, resulting in a peak warming of 3.00°C, as in the Negishi solution.

[50] Specifically, I find that if Africa had the same temperature damage function as the US, the utilitarian uniform carbon price solution again results in peak warming of 3.00°C, the same as the Negishi solution.

[51] This is effectively the regional social cost of carbon, which is calculated as the welfare-equivalent regional consumption change in 2025.

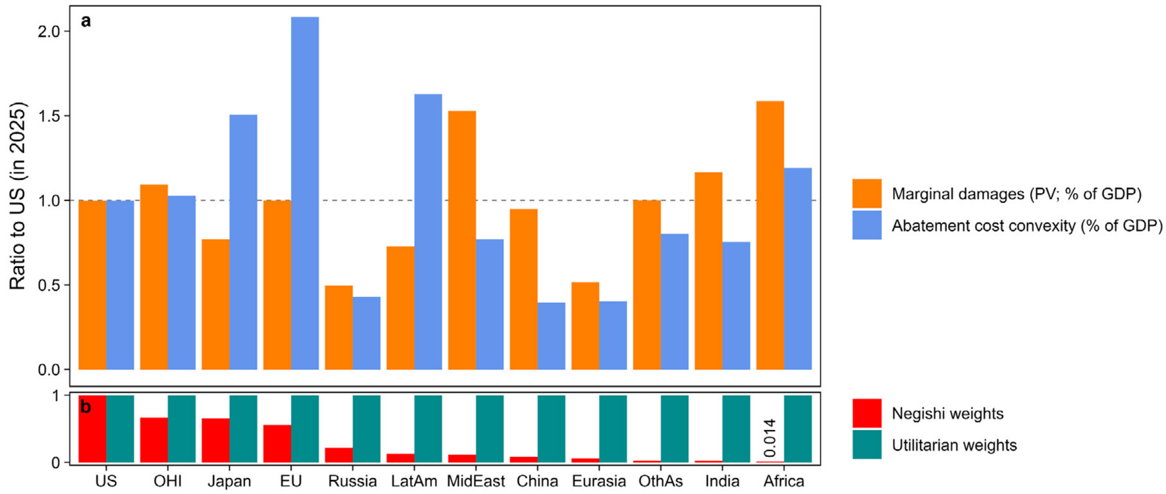

Figure 7: Relative regional marginal damages and abatement cost convexities.

Notes: The ratio of the 2025 present values (PV) of the stream of regional marginal damages as a percentage of the regional GDP in 2025, $\mathbf{d}'_i$, is given by $PV(\mathbf{d}'_i)/PV(\mathbf{d}'_{US})$. The ratio of the convexities in the abatement cost functions is $c''_{i,t}/c''_{US,t}$ (evaluated at uniform carbon prices), where t is the year 2025. Panel (b) shows regions' Negishi and utilitarian welfare weights in 2025 relative to the weights in the US. The figure shows the results for the utility discount rate of 1.5%.

Consistent with the theoretical results, Figure 7 shows that the regions that enjoy NPV consumption gains from the higher utilitarian carbon prices tend to have relatively high marginal damages and/or more convex abatement cost functions. Most notably, Africa is the region with the largest marginal damages relative to the size of its economy. Additionally, Africa has a fairly convex abatement cost function, which reduces its abatement cost burden as carbon prices are increased. Together, these two attributes explain why Africa experiences the largest gains in NPV consumption. The strongly convex abatement cost function in the EU is also noteworthy, resulting in NPV gains in the EU from the higher utilitarian carbon prices. Conversely, Russia and Eurasia experience the largest NPV losses due to relatively low marginal damages and flatter marginal abatement cost curves.

As the poorest region, Africa receives the lowest Negishi weight, which is roughly 70 times smaller than that of the richest region, the US, in 2025 (see Figure7b). Intuitively, the heavy down-weighting of welfare in the region most impacted by climate damages is a key factor behind the lower carbon prices in the Negishi solution.

Figure A12 in the appendix provides a more detailed breakdown of the regional heterogeneities that give rise to differential climate impacts. Rather than computing the present values of the stream of marginal damages, which are affected by population growth and economic growth, it shows the undiscounted marginal damages as a percentage of GDP in a given year alongside population and economic growth. Building on the theoretical insights from the dynamic model, this figure shows that Africa is particularly strongly affected by climate change, not only due to its high climate damages as a percentage of GDP but also because it has the fastest population growth, amplifying the aggregate damages by increasing the number of people affected. However, counterbalancing this to some degree is Africa's fast economic growth, which causes Africa's climate damages to be more heavily discounted under the utilitarian SWF.

4.2.3 Regions' preferred uniform carbon prices

This section presents the preferred uniform carbon prices for different regions, offering a complementary perspective on why utilitarian welfare weights lead to higher uniform carbon prices than Negishi weights. Drawing on the theoretical two-region model from Section 3.2.4, we know that this occurs if and only if the poorer region prefers a higher uniform carbon price than the richer region. By examining these preferences within the RICE model, we can gain further intuition for this result. Additionally, understanding regions' preferences regarding uniform carbon prices is valuable in its own right, as it helps to identify which regions might advocate for more or less stringent global climate policies in international negotiations[52].

Table 6 shows each region's preferred uniform carbon prices and the resulting peak temperature increase (the full carbon price and temperature trajectories are shown in Figure A13 in the appendix). The preferred uniform carbon prices vary widely across regions. In 2025, they differ by nearly an order of magnitude, from \$7 per ton of CO_2 in Russia and Eurasia to over \$60 per ton in Africa and the EU. This also leads to significant differences in peak temperatures, with Africa's preferred policy limiting warming to 2.4°C, while Russia's preferred policy allows for nearly 4°C.

The large differences in regions' preferred uniform carbon prices highlight the importance of how these preferences are weighted in the SWF. This perspective provides additional intuition for why optimal carbon prices are lower in the Negishi solution, which downweights the preferences of poorer regions. Notably, Africa, the poorest region that is most heavily downweighted in the Negishi-weighted SWF, has the highest preferred carbon prices, mainly due to its disproportionately large climate damages. In contrast, the US, the region with the largest Negishi weights, prefers comparatively low carbon prices, particularly after 2050. Specifically, Africa's preferred uniform carbon prices are more than twice as high as those of the US. While the overall effect depends on all regions, the downweighting of Africa's preferences is the primary reason for lower carbon prices in the Negishi solution compared

[52]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.

Table 6: Regions preferred uniform carbon prices (in 2022 $\$/tCO_2$) and resulting peak warming. For comparison, the uniform utilitarian and Negishi-weighted carbon prices are also shown.

	Carbon prices			
Welfare weights	2025	2055	2085	Peak warming (°C)
Negishi	25	60*	116**	3.00
Utilitarian	29	68	128	2.93
US	30	64	111	2.99
OHI	34	77	146	2.86
Japan	35	90	174	2.72
EU	63	133	224	2.42
Russia	7	17	35	3.98
LatAm	38	79	138	2.91
MidEast	37	72	118	3.00
China	13	40	94	3.00
Eurasia	7	20	43	3.65
OthAs	27	62	120	3.00
India	29	67	129	2.94
Africa	64	134	225	2.40

Notes: The table shows the results for the utility discount rate of 1.5%. Temperature changes are relative to 1900. Mimi-RICE-plus only yields an approximately equalized carbon price for the Negishi solution. Specifically, it varied across regions between (*) 59 and 63 $\$/tCO_2$ in 2055, and (**) 113 and 121 $\$/tCO_2$ in 2085.

to the utilitarian solution, which assigns equal weight to all regions' welfare, as discussed above.

4.3 The role of international climate finance

This section evaluates how conditional transfers for mitigation influence welfare-maximizing (utilitarian) carbon prices and explores the optimal allocation of mitigation finance.

4.3.1 The effect on optimal carbon prices

I begin by examining how these transfers affect the overall policy stringency. Figure 8 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. 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 9).

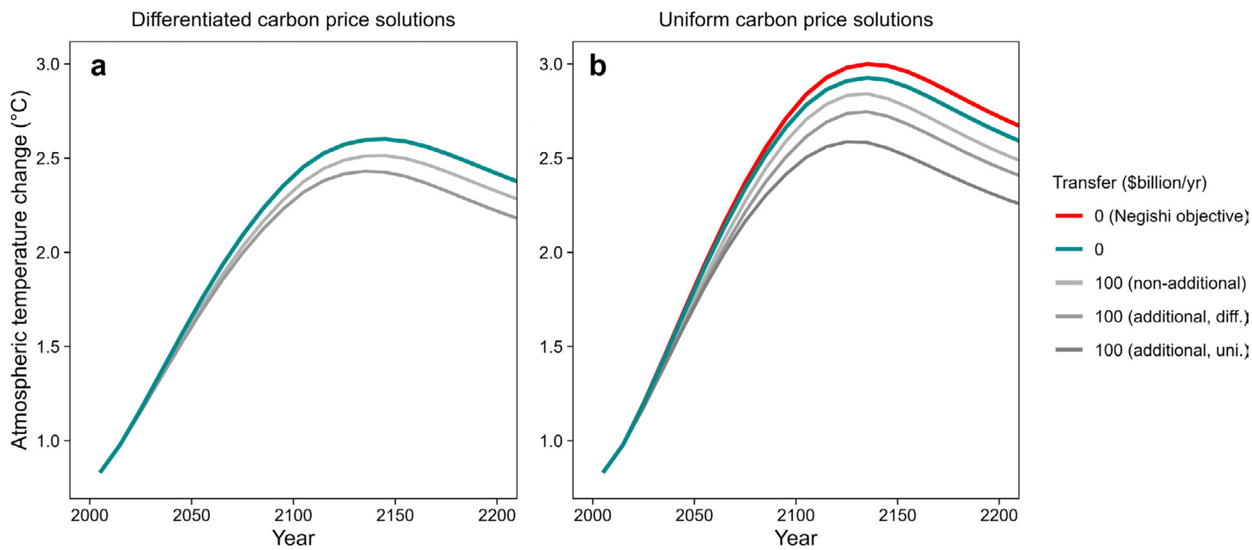

Figure 8: Optimal atmospheric temperature trajectories conditional on the optimization problem and the transfer scenario.

Notes: The Negishi-weighted solution (red) is compared to the utilitarian solutions without (teal) and with foreign abatement (gray). The different shades of gray 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. The figure shows the results for the utility discount rate of 1.5%.

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 9), resulting in a reduction of peak warming by 0.57°C, from 3.00°C to 2.43°C (see Figure 8).

Transfers for mitigation also substantially affect carbon prices. Figure 10 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-

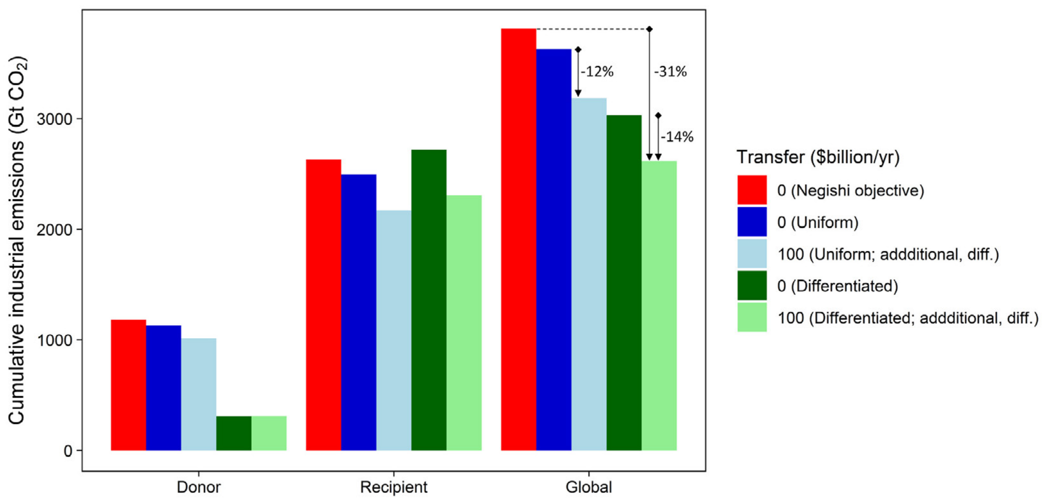

Figure 9: Optimal cumulative industrial emissions depending on the optimization problem and the transfer scenario.

Notes: This figure shows the results for foreign abatement that is additional relative to the utilitarian no-transfer differentiated carbon price solution. The figure shows 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[53]. In other words, the utilitarian 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 10. 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) [54]. 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.

[53]Note that, in some scenarios, the transfer constraint stops binding in the second half of the 22^{nd} century when not all of the available transfer is needed to reach zero emissions. This is indicated by the declining total transfer quantity in Figure 11.

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

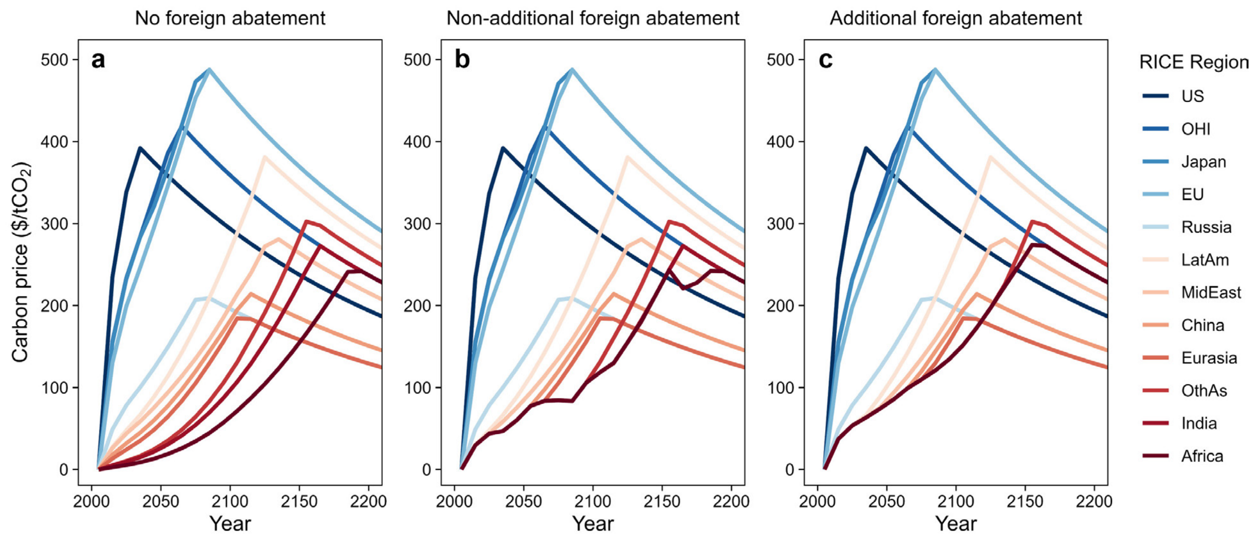

Figure 10: Optimal differentiated carbon price trajectories conditional on the optimization problem and the transfer scenario.

Notes: The figure shows the utilitarian differentiated carbon prices under (a) no foreign abatement, (b) non-additional foreign abatement, (c) additional foreign abatement. Once the carbon price reaches the region-specific backstop price, it decreases with the backstop price. Regions are arranged on the color scale from rich (blue) to poor (red). Results are for the utility discount rate of 1.5%.

Foreign abatement also has a large effect on optimal uniform carbon prices (see Table 7). Importantly, the welfare-maximizing uniform carbon price in 2025 roughly doubles from $29/t$CO_2$ without transfers to $54/t$CO_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 $25/t$CO_2$. The important policy implication is that the availability of international climate finance for mitigation considerably increases the carbon prices that maximize global welfare.

4.3.2 Optimal transfer allocation

The optimal allocation of international mitigation finance is shown in Figure 11, 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 abatement opportunities due to its large economy. The logic is that abatement in China absorbs most of the transfer as the marginal abatement costs are

Table 7: 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

Notes: 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.

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.

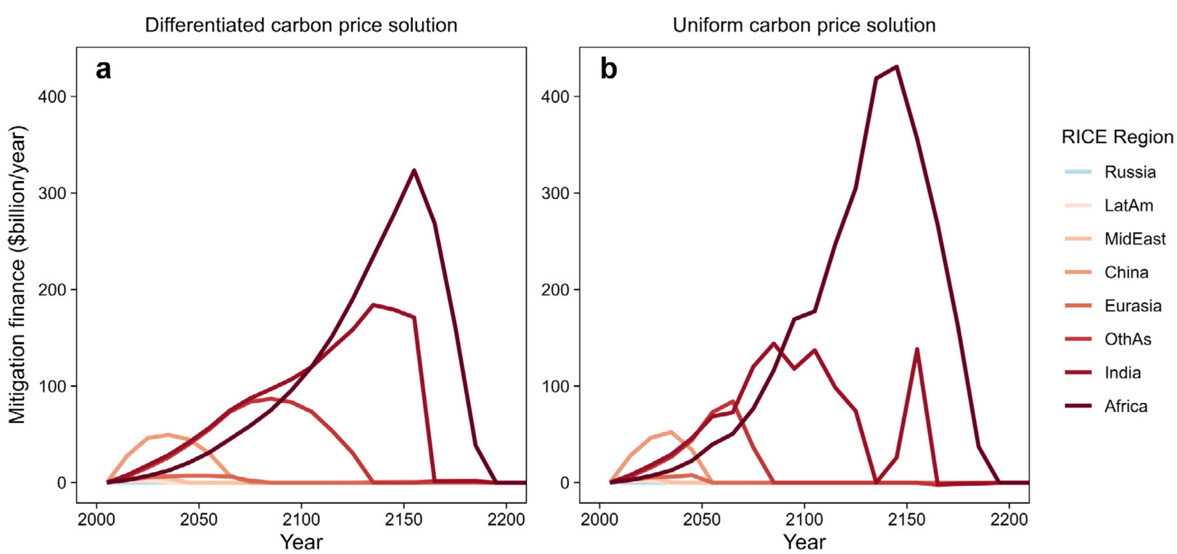

Figure 11: Optimal allocation of international mitigation finance over time.

Notes: 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 investigates the influence of accounting for inequality and international climate finance on optimal carbon prices, two central issues in international climate policy. Specifically, it compares the optimal carbon prices under two 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, employing 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 regional differences in marginal climate damages and the burden of abatement costs, and disparities in population and economic growth. Intuitively, global welfare maximization warrants more stringent climate policy if poor countries are more vulnerable to future climate change—due to higher marginal damages, faster population growth and slow economic catch-up—and if the cost burden of abatement predominately falls on rich countries. This highlights the importance of accurately accounting for regional heterogeneities in climate economic model. 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, I show how international climate finance affects welfare-maximizing carbon prices. 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 uniform and differentiated carbon price solutions. Notably, the welfare-maximizing uniform carbon price in 2025 almost doubles, from \$29/tCO2 to \$54/tCO2, under the default discounting parameters in RICE. Furthermore, compared to the Negishi solution, global cumulative emissions are 31% lower in the utilitarian solution with differentiated carbon prices and international mitigation finance.

To summarize, the key 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. Furthermore, it would be valuable to strengthen the empirical evidence on regional heterogeneities highlighted by this paper's theoretical results, including differences in damage and abatement cost functions, as well as in economic and population growth.

Third, this study relies on deterministic models. Given the substantial uncertainties in both human and physical systems and the associated economic effects of climate change, extending the analysis to a probabilistic framework would be valuable. Building on the findings of this paper, it could be particularly insightful to explore potential disparities in the uncertainties that different regions face.

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.

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. and J. Emmerling (2018). "Inequality and the Social Cost of Carbon". *Journal of the Association of Environmental and Resource Economists* 6.2, pp. 29–59.

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).

Anthoff, D., C. Hepburn, and R. S. J. Tol (2009). "Equity weighting and the marginal damage costs of climate change". *Ecological Economics* 68.3, pp. 836–849.

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 (2020). “Optimal global climate policy and regional carbon prices”. *Handbook on the Economics of Climate Change*. Edward Elgar Publishing, pp. 224–238. ISBN: 978-0-85793-906-7. (Visited on 09/16/2024).

Budolfson, M., F. Dennig, M. Fleurbaey, N. Scovronick, A. Siebert, D. Spears, and F. Wagner (2018). “Optimal Climate Policy and the Future of World Economic Development”. *The World Bank Economic Review*.

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.

Buntaine, M. T. and L. Prather (2018). “Preferences for Domestic Action Over International Transfers in Global Climate Policy”. *Journal of Experimental Political Science* 5.2, pp. 73–87.

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

Carleton, T., A. Jina, M. Delgado, M. Greenstone, T. Houser, S. Hsiang, A. Hultgren, R. E. Kopp, K. E. McCusker, I. Nath, J. Rising, A. Rode, H. K. Seo, A. Viaene, J. Yuan, and A. T. Zhang (2022). “Valuing the Global Mortality Consequences of Climate Change

Accounting for Adaptation Costs and Benefits*". *The Quarterly Journal of Economics*, qjac020.

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

Chawla, K. (2023). "Normative Challenges in Climate Change Economics". *Handbook of the Philosophy of Climate Change*. Ed. by Pellegrino, G. and Di Paola, M. Cham: Springer International Publishing, pp. 249–266. ISBN: 978-3-031-07002-0. DOI: 10.1007/978-3-031-07002-0_17. URL: https://doi.org/10.1007/978-3-031-07002-0_17 (visited on 11/02/2024).

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.

d'Autume, A., K. Schubert, and C. Withagen (2016). "Should the Carbon Price Be the Same in All Countries?" *Journal of Public Economic Theory* 18.5, pp. 709–725.

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.

Douenne, T., A. J. Hummel, and M. Pedroni (2023). *Optimal Fiscal Policy in a Climate-Economy Model with Heterogeneous Households*. Rochester, NY. DOI: 10.2139/ssrn.4018468. URL: https://papers.ssrn.com/abstract=4018468 (visited on 11/15/2024).

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

Eyckmans, J., S. Proost, and E. Schokkaert (1993). "Efficiency and distribution in greenhouse negotiations". *The Economics of International Environmental Agreements*, p. 157.

Fabre, A., T. Douenne, and L. Mattauch (2024). *International Attitudes Toward Global Policies*. Rochester, NY. DOI: 10.2139/ssrn.4448523. URL: https://papers.ssrn.com/abstract=4448523 (visited on 06/21/2024).

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).

Fleurbaey, M. and F. Maniquet (2008). "Utilitarianism versus fairness in welfare economics". *Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls*, pp. 263–280.

— (2018). "Optimal Income Taxation Theory and Principles of Fairness". *Journal of Economic Literature* 56.3, pp. 1029–1079.

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.

Gollier, C. (2012). "The Debate on Discounting: Reconciling Positivists and Ethicists". *Chicago Journal of International Law* 2, pp. 549–564.

— (2015). "Discounting, inequality and economic convergence". *Journal of Environmental Economics and Management* 69, pp. 53–61.

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

Harsanyi, J. C. (1955). "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility". *Journal of Political Economy* 63.4, pp. 309–321.

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.

Kelleher, J. P. (2019). "Reflections on the 2018 Nobel Memorial Prize Awarded to William Nordhaus". *Erasmus Journal for Philosophy and Economics* 12(1), pp. 93–107.

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.

Kotlikoff, L., F. Kubler, A. Polbin, and S. Scheidegger (2024). "Can today's and tomorrow's world uniformly gain from carbon taxation?" *European Economic Review* 168, p. 104819.

Landis, F. and T. Bernauer (2012). "Transfer payments in global climate policy". *Nature Climate Change* 2.8, pp. 628–633.

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.

Méjean, A., P. Collins-Sowah, C. Guivarch, F. Piontek, B. Soergel, and N. Taconet (2024). "Climate change impacts increase economic inequality: evidence from a systematic literature review". *Environmental Research Letters* 19.4, p. 043003.

Mirrlees, J. A. (1971). "An Exploration in the Theory of Optimum Income Taxation". *The Review of Economic Studies* 38.2, pp. 175–208.

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).

— (2024). *Climate Finance Provided and Mobilised by Developed Countries in 2013-2022, ,* . Paris: OECD Publishing. URL: https://doi.org/10.1787/19150727-en.

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

Piketty, T. and E. Saez (2013). "Optimal Labor Income Taxation". *Handbook of Public Economics*. Ed. by Auerbach, A. J., Chetty, R., Feldstein, M., and Saez, E. Vol. 5. handbook of public economics, vol. 5. Elsevier, pp. 391–474. DOI: 10.1016/B978-0-444-53759-1.00007-8. URL: https://www.sciencedirect.com/science/article/pii/B9780444537591000078 (visited on 08/16/2023).

Prest, B. C., L. Rennels, F. Errickson, and D. Anthoff (2024). "Equity weighting increases the social cost of carbon". *Science* 385.6710, pp. 715–717.

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.

Robbins, L. C. R. (1935). *An essay on the nature & significance of economic science.*

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. (1970). “Interpersonal Aggregation and Partial Comparability”. *Econometrica* 38.3, pp. 393–409.

— (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.

Stern, N. H. (2007). *The Economics of Climate Change: The Stern Review.* Cambridge, UK: Cambridge University Press.

Stiglitz, J. E. (1987). “Pareto efficient and optimal taxation and the new new welfare economics”. *Handbook of Public Economics.* Vol. 2. Elsevier, pp. 991–1042. DOI: 10.1016/S1573-4420(87)80010-1. URL: https://www.sciencedirect.com/science/article/pii/S1573442087800101 (visited on 10/19/2024).

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

UNFCCC (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.*

US Office of Management and Budget (2023). *Circular No. A-4.*

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).

World Economics (2024). *Global South.* World Economics. URL: https://www.worldeconomics.com/Regions/Global-South/default.aspx (visited on 08/10/2024).

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 (A19) 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

$$\begin{aligned}\mathcal{L} = L_i \alpha_i u\left(\frac{X_i}{L_i}\right) &- \sum_i \lambda_i \left(X_i - W_i + C_i(A_i) + D_i(A)\right) \\ &- \mu\left(C_N'(A_N) - C_S'(A_S)\right),\end{aligned}$$

where λ_i and μ are Lagrange multipliers.

The first-order conditions (FOCs) are:

$$\begin{aligned}&[X_i] : \alpha_i u'(x_i) = \lambda_i, \quad \forall i \\ &[A_N] : \lambda_N C_N'(A_N) = -\sum_i \lambda_i D_i'(A) - \mu C_N'' \\ &[A_S] : \lambda_S C_S'(A_S) = -\sum_i \lambda_i D_i'(A) + \mu C_S'' \\ &[\lambda_i] : X_i = W_i - C_i(A_i) - D_i(A), \quad \forall i \\ &[\mu] : C_N'(A_N) = C_S'(A_S).\end{aligned}$$

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

$$\alpha_N u'(x_N) C_N'(A_N) = -\sum_i \alpha_i u'(x_i) D_i'(A) - \mu C_N'', \tag{A1}$$

$$\alpha_S u'(x_S) C_S'(A_S) = -\sum_i \alpha_i u'(x_i) D_i'(A) + \mu C_S''. \tag{A2}$$

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 (A1) by Equation (A2), and simple manipulations yield the optimal uniform carbon price (Equation (A19) in the main text),

$$\tau^{uni} = C_i'^*(A_i^*) = -\sum_i \alpha_i u'(x_i^*) D_i'(A^*) \frac{C_S'' + C_N''}{\alpha_N u'(x_N^*) C_S'' + \alpha_S u'(x_S^*) C_N''}.$$

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). Hence, 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 (6),

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

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

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

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

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

A.2.2 Utilitarian uniform carbon price solution

We start with Equation (8),

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

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

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

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

$$u'(\check{x}_N)\frac{d\check{A}}{d\check{A}_S d\check{\tau}}\frac{dC_S(\check{A}_S)dC_N(\check{A}_N)}{dC_S(\check{A}_S)+dC_N(\check{A}_N)} + u'(\check{x}_S)\frac{d\check{A}}{d\check{A}_N d\check{\tau}}\frac{dC_S(\check{A}_S)dC_N(\check{A}_N)}{dC_S(\check{A}_S)+dC_N(\check{A}_N)}$$
$$= -\sum_i u'(\check{x}_i)\frac{dD_i(\check{A})}{d\check{\tau}}.$$

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

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

which simplifies to

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

A.2.3 Regions' preferred uniform carbon price

We start with Equation (13),

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

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

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

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

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

A.3 Proof of Proposition 1

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

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

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 $\tilde{D}'_i < \check{D}'_i$ for all i (note that marginal damages of abatement are negative).

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

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

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

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

Multiplying out and collecting terms, we have

$$C''_N\Big(\check{u}'_N\check{D}'_N + \check{u}'_S\check{D}'_S - \check{u}'_S\tilde{D}'_N - \check{u}'_S\tilde{D}'_S\Big) < C''_S\Big(\check{u}'_N\tilde{D}'_N + \check{u}'_N\tilde{D}'_S - \check{u}'_N\check{D}'_N - \check{u}'_S\check{D}'_S\Big). \tag{A3}$$

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

$$\Big(\check{u}'_N\tilde{D}'_N + \check{u}'_N\tilde{D}'_S - \check{u}'_N\check{D}'_N - \check{u}'_S\check{D}'_S\Big) > 0 \tag{A4}$$

for the inequality in Equation (A3) to hold, since $C''_i > 0$ for all i.

Cross-division, collecting common terms, and rearranging yields

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

Note that

$$\underbrace{\tilde{D}'_N - \check{D}'_N}_{<0} + \tilde{D}'_S < \tilde{D}'_S$$

and

$$\underbrace{\tilde{D}'_S - \check{D}'_S}_{<0} + \tilde{D}'_N < \tilde{D}'_N$$

Thus, for the numerator we have

$$\begin{aligned}
\overbrace{\check{u}'_N \underbrace{(\tilde{D}'_N - \check{D}'_N + \tilde{D}'_S)}_{<\tilde{D}'_S} - \check{u}'_S \check{D}'_S}^{>0 \text{ by Equation (A4)}} &< \check{u}'_N \tilde{D}'_S - \check{u}'_S \check{D}'_S \\
&< \check{u}'_N \check{D}'_S - \check{u}'_S \check{D}'_S \\
&> 0,
\end{aligned}$$

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

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

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 (A4)) numerator.

Putting this together we have

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

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

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

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

We begin 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 $\tilde{D}'_i \geq \check{D}'_i$ for all i (note that marginal damages of abatement are negative).

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

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

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

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

Multiplying this out and collecting common terms gives

$$C''_N \Big(\check{u}'_N \check{D}'_N + \check{u}'_S \check{D}'_S - \check{u}'_S \tilde{D}'_N - \check{u}'_S \tilde{D}'_S \Big) \geq C''_S \Big(\check{u}'_N \tilde{D}'_N + \check{u}'_N \tilde{D}'_S - \check{u}'_N \check{D}'_N - \check{u}'_S \check{D}'_S \Big). \tag{A5}$$

We know that

$$\Big(\check{u}'_N \tilde{D}'_N + \check{u}'_N \tilde{D}'_S - \check{u}'_N \check{D}'_N - \check{u}'_S \check{D}'_S \Big) > 0 \tag{A6}$$

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 (A5) and (A6) imply

$$\Big(\check{u}'_N \check{D}'_N + \check{u}'_S \check{D}'_S - \check{u}'_S \tilde{D}'_N - \check{u}'_S \tilde{D}'_S \Big) > 0 \tag{A7}$$

since $C''_i > 0$ 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

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

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

$$\tilde{C}' \geq \check{C}' \iff \frac{C''_N}{C''_S} \geq \frac{\check{u}'_N(\tilde{D}'_N - \check{D}'_N + \tilde{D}'_S) - \check{u}'_S\check{D}'_S}{\check{u}'_N\check{D}'_N - \check{u}'_S(\tilde{D}'_S - \check{D}'_S + \tilde{D}'_N)}. \tag{A8}$$

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

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

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 (A7)[55]. The last inequality follows from the implication of $\tilde{C}' \geq \check{C}'$ documented in Equation (A8).

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

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

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

□

[55]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:

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

A.4 Proof of Corollary 1

Proposition 1 establishes that $\check{\tau} > \tilde{\tau}$, if and only if $\frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S}$. We can rewrite this condition as

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

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

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

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 3 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 3 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 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 (14) 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,

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

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[56]. 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

[56]This can be seen as follows:

$$\frac{d\tau}{dA(\tau)} = \frac{d\tau}{dA_N(\tau) + dA_S(\tau)} = \frac{d\tau}{\frac{1}{C''_N}d\tau + \frac{1}{C''_S}d\tau} = \frac{1}{\frac{1}{C''_N} + \frac{1}{C''_S}},$$

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

know that this inequality holds from the strict convexity of the damage function and since $\breve{\tau} > \mathring{\tau}^N$ implies $\breve{A}_i(\breve{\tau}) > \mathring{A}_i(\mathring{\tau}^N)$ for all i for strictly convex abatement cost functions.

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

Together, this establishes the following inequalities for the North:

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

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

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

Together, this establishes the following inequalities for the South:

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

We have thus shown that

$$\frac{-\frac{d\breve{D}_S}{d\breve{\tau}}}{\frac{d\breve{C}_S}{d\breve{\tau}}} > 1 > \frac{-\frac{d\breve{D}_N}{d\breve{\tau}}}{\frac{d\breve{C}_N}{d\breve{\tau}}}.$$

A.5 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

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

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

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

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.6 Proof of Proposition 2

Proof. We first need to obtain expressions for the abatement as a function of the marginal abatement cost. Since I assume that the abatement cost functions are smooth, strictly increasing, strictly convex, and the third derivative is equal to zero for all $A_i > 0$, they have the following form[57]: $C_i(A_i) = k_i A_i^2 + m_i A_i + n_i$, with $k_i > 0$. The marginal abatement cost is thus $C'_i(A_i) = 2k_i A_i + m_i$ and the second derivative is $C''_i = 2k_i$. Therefore, $k_i = \frac{C''_i}{2}$. We invert the marginal abatement cost function to obtain an expression for the abatement:

$$A_i = \frac{1}{2}(C'_i - m_i)k_i^{-1}.$$

We split the proof into the forward and backward implications.

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

We begin 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

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

Plugging in the expressions for the marginal abatement costs detailed in Definitions 6

[57]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 C.2).

and 10, and rewriting, we get

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

Next, note that $\tilde{A} < \hat{A}$ implies $\tilde{D}'_i < \hat{D}'_i$, for all i. Therefore, the previous inequality implies[58]

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

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

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

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

$$\frac{\hat{u}'_N}{\hat{u}'_S}\frac{\hat{D}'_S}{\hat{D}'_N} > \frac{C''_N}{C''_S}.$$

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

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

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

Plugging in the expressions for the marginal abatement costs detailed in Definitions 6 and 10, and rewriting, we get

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

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

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

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

$$\left(\hat{D}_N' - \tilde{D}_N' + \frac{\hat{u}_S'}{\hat{u}_N'}\hat{D}_S' - \tilde{D}_S'\right)\frac{1}{C_N''} \geq \left(\tilde{D}_S' - \hat{D}_S' + \tilde{D}_N' - \frac{\hat{u}_N'}{\hat{u}_S'}\hat{D}_N'\right)\frac{1}{C_S''}. \tag{A9}$$

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

$$\hat{D}_N' - \tilde{D}_N' + \frac{\hat{u}_S'}{\hat{u}_N'}\hat{D}_S' - \tilde{D}_S' < 0. \tag{A10}$$

Moreover, Equations (A9) and (A10) imply

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

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

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

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

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

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

So far, we have established that

$$\tilde{A} \geq \hat{A} \iff \frac{\hat{u}_N'}{\hat{u}_S'}\frac{C_N''}{C_S''} \geq \frac{\hat{u}_N'\left(\hat{D}_N' - \tilde{D}_N' - \tilde{D}_S'\right) + \hat{u}_S'\hat{D}_S'}{\hat{u}_S'\left(\tilde{D}_S' - \hat{D}_S' + \tilde{D}_N'\right) - \hat{u}_N'\hat{D}_N'}. \tag{A11}$$

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:

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

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 (A10) and (A.6) [59]. The last inequality follows from the implication of $\tilde{A} \geq \hat{A}$ documented in Equation (A11).

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

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

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

$\square$

A.7 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.

[59] 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:

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

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

I begin by establishing the conditions under which $\check{\tau} > \mathring{\tau}^N$, or equivalently, $\check{C}' > \mathring{C}^{N\prime}$[60]. First note that, for strictly convex abatement cost functions, $\check{C}' > \mathring{C}^{N\prime}$ implies $\check{A}_i > \mathring{A}_i^N$ for all i, and thus $\check{A} > \mathring{A}^N$. For strictly convex damage functions, this implies $\check{D}'_i > \mathring{D}_i^{N\prime}$ for all i (note that marginal damages of abatement are negative).

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

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

which can be rewritten as

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

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

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

Rewriting yields

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

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

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

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 begin 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 $\check{A}_i \leq \mathring{A}_i^N$ for all i, and thus $\check{A} \leq \mathring{A}^N$. For strictly convex damage functions, this implies $\check{D}'_i \leq \mathring{D}_i^{N\prime}$ for all i (note that marginal damages of abatement are negative).

[60] 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.

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

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

which can be rewritten as

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

Let us temporarily define $\delta_N \equiv \mathring{D}^{N\prime}_N - \check{D}'_N$[61]. 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

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

Rewriting yields

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

So far, we have established that

$$\check{C}' \leq \mathring{C}^{N\prime} \iff \frac{\check{D}'_S}{\mathring{D}^{N\prime}_N} \leq \frac{C''_N}{C''_S} - \underbrace{\frac{\delta_N}{-\mathring{D}^{N\prime}_N} \frac{\check{u}'_N}{\check{u}'_S}}_{\geq 0}. \tag{A12}$$

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:

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

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 (A12).

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 $\frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S} \implies \check{C}' > \mathring{C}^{N\prime}$.

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

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

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

$\square$

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 begin 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 $\check{A}_i < \mathring{A}_i^S$ for all i, and thus $\check{A} < \mathring{A}^S$. For strictly convex damage functions, this implies $\check{D}'_i < \mathring{D}_i^{S\prime}$ for all i (note that marginal damages of abatement are negative).

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

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

which can be rewritten as

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

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

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

Rewriting yields

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

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 begin 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 $\check{A}_i \geq \mathring{A}_i^S$ for all i, and thus $\check{A} \geq \mathring{A}^S$. For strictly convex damage functions, this implies $\check{D}'_i \geq \mathring{D}_i^{S\prime}$ for all i (note that marginal damages of abatement are negative).

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

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

which can be rewritten as

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

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

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

Simplifying and rearranging yields

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

So far, we have established that

$$\check{C}' \geq \mathring{C}^{S\prime} \iff \frac{\check{D}'_N}{\mathring{D}_S^{S\prime}} \geq \frac{C''_S}{C''_N} + \underbrace{\frac{\delta_S}{-\mathring{D}_S^{S\prime}} \frac{\check{u}'_S}{\check{u}'_N}}_{\geq 0}. \tag{A13}$$

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:

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

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 (A13).

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

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

□

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 begin 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 $\tilde{D}'_i > \mathring{D}^{N\prime}_i$ for all i (note that marginal damages of abatement are negative).

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

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

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

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

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

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

which simplifies to

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

We can now establish the following inequalities:

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

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

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

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

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 begin 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 $\tilde{A}_i \leq \mathring{A}_i^N$ for all i, and thus $\tilde{A} \leq \mathring{A}^N$. For strictly convex damage functions, this implies $\tilde{D}_i' \leq \mathring{D}_i^{N\prime}$ for all i (note that marginal damages of abatement are negative).

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

$$-\tilde{D}_N' - \tilde{D}_S' \leq -\mathring{D}_N^{N\prime} \frac{C_S'' + C_N''}{C_S''},$$

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

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

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

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

which simplifies to

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

So far, we have established that

$$\tilde{C}' \leq \mathring{C}^{N\prime} \iff \frac{C''_N}{C''_S} \geq \frac{\tilde{D}'_S}{\mathring{D}^{N\prime}_N} + \underbrace{\frac{-\delta_N}{\mathring{D}^{N\prime}_N}}_{\geq 0}. \tag{A14}$$

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.

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

where the second and third inequalities follow from $\tilde{D}'_i \leq \mathring{D}^{N\prime}_i$ for all i, and the last inequality follows from the implication of $\tilde{C}' \leq \mathring{C}^{N\prime}$ documented in Equation (A14).

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

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

□

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 begin 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 $\tilde{A}_i < \mathring{A}^S_i$ for all i, and thus $\tilde{A} < \mathring{A}^S$. For strictly convex damage functions, this implies $\tilde{D}'_i < \mathring{D}^{S\prime}_i$ for all i (note that marginal damages of abatement are negative).

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

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

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

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

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

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

which simplifies to

$$\frac{C_S''}{C_N''} > \frac{\tilde{D}_N'}{\mathring{D}_S^{S\prime}} + \frac{-\delta_S}{\mathring{D}_S^{S\prime}}.$$

We can now establish the following inequalities:

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

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

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

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

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 $\tilde{D}_i' \geq \mathring{D}_i^{S\prime}$ for all i (note that marginal damages of abatement are negative).

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

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

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

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

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

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

which simplifies to

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

So far, we have established that

$$\tilde{C}' \geq \mathring{C}^{S\prime} \iff \frac{C''_S}{C''_N} \leq \frac{\tilde{D}'_N}{\mathring{D}_S^{S\prime}} + \underbrace{\frac{-\delta_S}{\mathring{D}_S^{S\prime}}}_{\leq 0}. \tag{A15}$$

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:

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

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 (A15).

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

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

□

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 the preferred uniform carbon prices of the Global North ($\mathring{\tau}^N$) and the*

Global South ($\mathring{\tau}^S$), 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 $\mathring{\tau}^S > \check{\tau} > \mathring{\tau}^N$ if and only if $\frac{\check{D}'_S}{\check{D}'_N} > \frac{C''_N}{C''_S}$. Similarly, it can be shown that $\mathring{\tau}^S < \check{\tau} < \mathring{\tau}^N$ if and only if $\frac{\check{D}'_S}{\check{D}'_N} < \frac{C''_N}{C''_S}$. Hence, we have $\mathring{\tau}^S = \check{\tau} = \mathring{\tau}^N$ if and only if $\frac{\check{D}'_S}{\check{D}'_N} = \frac{C''_N}{C''_S}$, 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 $\frac{\tilde{D}'_S}{\tilde{D}'_N} > \frac{C''_N}{C''_S}$. Similarly, it can be shown that $\mathring{\tau}^S < \tilde{\tau} < \mathring{\tau}^N$ if and only if $\frac{\tilde{D}'_S}{\tilde{D}'_N} < \frac{C''_N}{C''_S}$. Hence, we have $\mathring{\tau}^S = \tilde{\tau} = \mathring{\tau}^N$ 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.

It follows that $\mathring{\tau}^S = \check{\tau} = \tilde{\tau} = \mathring{\tau}^N$ if and only if $\frac{D'_S}{D'_N} = \frac{C''_N}{C''_S}$, where $D'_i = \check{D}'_i = \tilde{D}'_i$.

□

A.8 Proof of Proposition 3

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

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

Simplifying and rearranging yields

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

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

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

We thus have

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

We have thus shown that

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

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

Proposition 1 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 $\mathring{\tau}^S > \check{\tau} > \mathring{\tau}^N$. Therefore,

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

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

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

□

A.9 Proof of Proposition 4

Proof. Proof of forward implication: $\tilde{\tau} < \check{\tau} \implies \frac{\check{u}'_{N2}\check{D}'_{N2}+\check{u}'_{S2}\check{D}'_{S2}}{\tilde{D}'_{N2}+\tilde{D}'_{S2}} > \nu \frac{\check{u}'_{N1}C''_{S1}+\check{u}'_{S1}C''_{N1}}{C''_{S1}+C''_{N1}}$.

We ask under which conditions $\tilde{\tau} < \check{\tau}$, or equivalently, $\tilde{C}'_1 < \check{C}'_1$. For strictly convex abatement cost functions, $\tilde{C}'_1 < \check{C}'_1$ implies $\tilde{C}_{i1} < \check{C}_{i1}$ and $\tilde{A}_{i1} < \check{A}_{i1}$ for all i, and thus $\tilde{A}_1 < \check{A}_1$. This implies $\tilde{D}_{i2} > \check{D}_{i2}$ and $\tilde{D}'_{i2} < \check{D}'_{i2}$ for all i (note that marginal damages of abatement are negative). Using the budget constraints, this implies $\check{X}_{i1} < \tilde{X}_{i1}$ and $\check{X}_{i2} > \tilde{X}_{i2}$ for all i. For an exogenous population, we thus have $\check{x}_{i1} < \tilde{x}_{i1}$ and $\check{x}_{i2} > \tilde{x}_{i2}$ for all i. Therefore, $\check{u}'_{i1} > \tilde{u}'_{i1}$ and $\check{u}'_{i2} < \tilde{x}'_{i2}$ for all i.

We have $\tilde{C}'_1 < \check{C}'_1$ if and only if

$$-\beta\nu\left(\tilde{D}'_{N2} + \tilde{D}'_{S2}\right) < -\beta\left(\check{u}'_{N2}\check{D}'_{N2} + \check{u}'_{S2}\check{D}'_{S2}\right)\frac{C''_{S1} + C''_{N1}}{\check{u}'_{N1}C''_{S1} + \check{u}'_{S1}C''_{N1}}.$$

Cancelling β, multiplying both sides by the denominator on the RHS (which is positive), and rearranging, we get

$$\left(\check{u}'_{N2}\check{D}'_{N2} + \check{u}'_{S2}\check{D}'_{S2}\right)\left(C''_{S1} + C''_{N1}\right) < \nu\left(\tilde{D}'_{N2} + \tilde{D}'_{S2}\right)\left(\check{u}'_{N1}C''_{S1} + \check{u}'_{S1}C''_{N1}\right).$$

Since $\tilde{D}'_{i2} < \check{D}'_{i2}$ for all i, the previous inequality implies

$$\left(\check{u}'_{N2}\check{D}'_{N2} + \check{u}'_{S2}\check{D}'_{S2}\right)\left(C''_{S1} + C''_{N1}\right) < \nu\left(\check{D}'_{N2} + \check{D}'_{S2}\right)\left(\check{u}'_{N1}C''_{S1} + \check{u}'_{S1}C''_{N1}\right).$$

Noting that $(\check{D}'_{N2} + \check{D}'_{S2})$ is negative so the sign flips upon division, rearranging yields

$$\frac{\check{u}'_{N2}\check{D}'_{N2} + \check{u}'_{S2}\check{D}'_{S2}}{\check{D}'_{N2} + \check{D}'_{S2}} > \nu\frac{\check{u}'_{N1}C''_{S1} + \check{u}'_{S1}C''_{N1}}{C''_{S1} + C''_{N1}}.$$

<u>Proof of backward implication</u>: $\frac{\check{u}'_{N2}\check{D}'_{N2} + \check{u}'_{S2}\check{D}'_{S2}}{\check{D}'_{N2} + \check{D}'_{S2}} > \nu\frac{\check{u}'_{N1}C''_{S1} + \check{u}'_{S1}C''_{N1}}{C''_{S1} + C''_{N1}} \implies \tilde{\tau} < \check{\tau}$.
We will prove the contrapositive of the stated result. That is, we will prove

$$\tilde{\tau} \geq \check{\tau} \implies \frac{\check{u}'_{N2}\check{D}'_{N2} + \check{u}'_{S2}\check{D}'_{S2}}{\check{D}'_{N2} + \check{D}'_{S2}} \leq \nu\frac{\check{u}'_{N1}C''_{S1} + \check{u}'_{S1}C''_{N1}}{C''_{S1} + C''_{N1}}.$$

We start by establishing the implications of $\tilde{\tau} \geq \check{\tau}$, or equivalently, $\tilde{C}'_1 \geq \check{C}'_1$.

For strictly convex abatement cost functions, $\tilde{C}'_1 \geq \check{C}'_1$ implies $\tilde{C}_{i1} \geq \check{C}_{i1}$ and $\tilde{A}_{i1} \geq \check{A}_{i1}$ for all i, and thus $\tilde{A}_1 \geq \check{A}_1$. This implies $\tilde{D}_{i2} \leq \check{D}_{i2}$ and $\tilde{D}'_{i2} \geq \check{D}'_{i2}$ for all i (note that marginal damages of abatement are negative). Using the budget constraints, this implies $\check{X}_{i1} \geq \tilde{X}_{i1}$ and $\check{X}_{i2} \leq \tilde{X}_{i2}$ for all i. For an exogenous population, we thus have $\check{x}_{i1} \geq \tilde{x}_{i1}$ and $\check{x}_{i2} \leq \tilde{x}_{i2}$ for all i. Therefore, $\check{u}'_{i1} \leq \tilde{u}'_{i1}$ and $\check{u}'_{i2} \geq \tilde{x}'_{i2}$ for all i.

We have $\tilde{C}'_1 \geq \check{C}'_1$ if and only if

$$-\beta\nu\left(\tilde{D}'_{N2} + \tilde{D}'_{S2}\right) \geq -\beta\left(\check{u}'_{N2}\check{D}'_{N2} + \check{u}'_{S2}\check{D}'_{S2}\right)\frac{C''_{S1} + C''_{N1}}{\check{u}'_{N1}C''_{S1} + \check{u}'_{S1}C''_{N1}}.$$

Cancelling β, multiplying both sides by the denominator on the RHS (which is positive), and rearranging, we get

$$\left(\check{u}'_{N2}\check{D}'_{N2} + \check{u}'_{S2}\check{D}'_{S2}\right)\left(C''_{S1} + C''_{N1}\right) \geq \nu\left(\tilde{D}'_{N2} + \tilde{D}'_{S2}\right)\left(\check{u}'_{N1}C''_{S1} + \check{u}'_{S1}C''_{N1}\right).$$

Since $\tilde{D}'_{i2} \geq \check{D}'_{i2}$ for all i, the previous inequality implies

$$\left(\check{u}'_{N2}\check{D}'_{N2} + \check{u}'_{S2}\check{D}'_{S2}\right)\left(C''_{S1} + C''_{N1}\right) \geq \nu\left(\check{D}'_{N2} + \check{D}'_{S2}\right)\left(\check{u}'_{N1}C''_{S1} + \check{u}'_{S1}C''_{N1}\right).$$

Noting that $(\check{D}'_{N2} + \check{D}'_{S2})$ is negative so the sign flips upon division, rearranging yields

$$\frac{\check{u}'_{N2}\check{D}'_{N2} + \check{u}'_{S2}\check{D}'_{S2}}{\check{D}'_{N2} + \check{D}'_{S2}} \leq \nu\frac{\check{u}'_{N1}C''_{S1} + \check{u}'_{S1}C''_{N1}}{C''_{S1} + C''_{N1}}.$$

We have thus shown that

$$\tilde{\tau} \geq \check{\tau} \implies \frac{\check{u}'_{N2}\check{D}'_{N2} + \check{u}'_{S2}\check{D}'_{S2}}{\check{D}'_{N2} + \check{D}'_{S2}} \leq \nu \frac{\check{u}'_{N1}C''_{S1} + \check{u}'_{S1}C''_{N1}}{C''_{S1} + C''_{N1}}.$$

By contraposition, we therefore have

$$\frac{\check{u}'_{N2}\check{D}'_{N2} + \check{u}'_{S2}\check{D}'_{S2}}{\check{D}'_{N2} + \check{D}'_{S2}} > \nu \frac{\check{u}'_{N1}C''_{S1} + \check{u}'_{S1}C''_{N1}}{C''_{S1} + C''_{N1}} \implies \tilde{\tau} < \check{\tau}.$$

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

$$\tilde{\tau} < \check{\tau} \iff \frac{\check{u}'_{N2}\check{D}'_{N2} + \check{u}'_{S2}\check{D}'_{S2}}{\check{D}'_{N2} + \check{D}'_{S2}} > \nu \frac{\check{u}'_{N1}C''_{S1} + \check{u}'_{S1}C''_{N1}}{C''_{S1} + C''_{N1}}.$$

□

A.10 Derivation of optimal carbon prices with mitigation finance

This section shows the derivation of the utilitarian uniform carbon prices in the presence of mitigation finance (Table 4 in the main text). The derivation of the differentiated carbon prices is largely analogous. For the case of endogenous mitigation finance, $T^{max} = \phi\big(W_N - C_N(A_{ND}) - D_N(A) - Y_N^{ref}\big)$ (see Table A1 for results and Section 3.4 for notation and setup).

The Lagrangian of this problem is:

$$\begin{aligned}
\mathcal{L} = \sum_i L_i u\left(\frac{X_i}{L_i}\right) &- \lambda_N\big(X_N - W_N + C_N(A_{ND}) + D_N(A) + \overbrace{C_S(A_{SD} + A_{SF}) - C_S(A_{SD})}^{T}\big) \\
&- \lambda_S\big(X_S - W_S + C_S(A_{SD}) + D_S(A)\big) \\
&- \gamma\big(C'_N(A_{ND}) - C'_S(A_{SD} + A_{SF})\big) \\
&+ \nu(C_S(A_{SD}) - C_S^{min}) \\
&+ \xi\big(T^{max} \underbrace{- C_S(A_{SD} + A_{SF}) + C_S(A_{SD})}_{-T}\big),
\end{aligned}$$

where $\nu \geq 0$ and $\xi \geq 0$ are Lagrange multipliers.

The Karush-Kuhn-Tucker (KKT) conditions are:

$$
\begin{aligned}
&[\text{Stationarity}]: \\
&[X_i]: u'(x_i) = \lambda_i, \quad \forall i \\
&[A_{ND}]: \lambda_N C_N'(A_{ND}) = -\sum_i \lambda_i D_i'(A) - \gamma C_N'' \\
&[A_{SD}]: \lambda_S C_S'(A_{SD}) = -\sum_i \lambda_i D_i'(A) - \lambda_N (C_S'(A_{SD}+A_{SF}) - C_S'(A_{SD})) + \gamma C_S'' \\
&\qquad\qquad + \nu C_S'(A_{SD}) - \xi (C_S'(A_{SD}+A_{SF}) - C_S'(A_{SD})) \\
&[A_{SF}]: \lambda_N C_S'(A_{SD}+A_{SF}) = -\sum_i \lambda_i D_i'(A) + \gamma C_S'' - \xi C_S'(A_{SD}+A_{SF}) \\
&[\text{Primal feasibility}]: \\
&[\lambda_N]: X_N = W_N - C_N(A_{ND}) - D_N(A) \overbrace{- C_S(A_{SD}+A_{SF}) + C_S(A_{SD})}^{-T}) \\
&[\lambda_S]: X_S = W_S - C_S(A_{SD}) - D_S(A) \\
&[\gamma]: C_N'(A_{ND}) = C_S'(A_{SD}+A_{SF}) \\
&[\nu]: C_S(A_{SD}) - C_S^{min} \geq 0 \\
&[\xi]: \overbrace{C_S(A_{SD}+A_{SF}) - C_S(A_{SD})}^{T} \leq T^{max} \\
&[\text{Dual feasibility}]: \\
&\nu \geq 0, \xi \geq 0 \\
&[\text{Complementary slackness}]: \\
&\nu (C_S(A_{SD}) - C_S^{min}) = 0 \\
&\xi (T^{max} \underbrace{- C_S(A_{SD}+A_{SF}) + C_S(A_{SD})}_{-T}) = 0.
\end{aligned}
$$

Plugging the optimality condition for A_{SF} in the one for A_{SD}, we get

$$
\lambda_S C_S'(A_{SD}) = \lambda_N C_S'(A_{SD}) + \nu C_S'(A_{SD}) + \xi C_S'(A_{SD}), \tag{A16}
$$

and thus

$$
u_S' = u_N' + \nu + \xi.
$$

We need to distinguish two cases depending on whether the domestic abatement constraint binds.

(CASE 1) Domestic abatement constraint not binding: $C_S(A_{SD}) > C_S^{min}$

If $C_S(A_{SD}) > C_S^{min}$, then, by complementary slackness, we must have $\nu = 0$.

Again, there are two cases depending on whether the transfer constraint binds.

(CASE 1.1) Transfer constraint not binding: $T < T^{max}$

If $T < T^{max}$, then, by complementary slackness, we must have $\xi = 0$. Therefore, $u'_N = u'_S$ by Equation (A.10). Using this in the optimality conditions for A_{ND} and A_{SF}, we have

$$-\sum_i u'_i D'_i(A) - \gamma C''_N = -\sum_i u'_i D'_i(A) + \gamma C''_S.$$

Since $C''_i > 0$ for all i, this equation only holds for $\gamma = 0$. This has an intuitive interpretation: the uniform carbon price constraint is not binding when marginal utilities are equalized. With $\gamma = 0$, we obtain the Samuelson condition

$$C'_N(A_{ND}) = C'_S(A_{SD} + A_{SF}) = -\sum_i D'_i(A).$$

(CASE 1.2) Transfer constraint binding: $T = T^{max}$

If $\xi > 0$, then, by complementary slackness, we must have $T = T^{max}$. Using this in Equation (A.10), we get $\xi = u'_S - u'_N$. The Lagrange multiplier on the transfer constraint, ξ, which reflects the value of marginally relaxing the transfer constraint, is given by the difference in marginal utilities. This is intuitive. Since $C_S(A_{SD}) > 0$, relaxing the transfer constraint shifts the abatement cost burden from the South to the North. The social value of this cost shift is given by the difference in marginal utilities.

Using this in the optimality condition for A_{SF}, we have the following optimality condition for abatement in the South:

$$u'_S C'_S(A_{SD} + A_{SF}) = -\sum_i u'_i D'_i(A) + \gamma C''_S.$$

The optimality condition for abatement in the North is

$$u'_N C'_N(A_{ND}) = -\sum_i u'_i D'_i(A) - \gamma C''_N.$$

Using the uniform carbon price constraint, $C'_S(A_{SD} + A_{SF}) = C'_N(A_{ND})$, we can solve these two equations for the optimal uniform carbon price,

$$C'_N(A_{ND}) = C'_S(A_{SD} + A_{SF}) = -\sum_i u'_i D'_i \frac{C''_S(A_{SD} + A_{SF}) + C''_N(A_{ND})}{u'_N C''_S(A_{SD} + A_{SF}) + u'_S C''_N(A_{ND})}.$$

(CASE 2) Domestic abatement constraint binding: $C_S(A_{SD}) = C_S^{min}$

If $\nu > 0$, then by complementary slackness, we must have $C_S(A_{SD}) = C_S^{min}$.

As before, there are two cases depending on whether the transfer constraint binds.

(CASE 2.1) Transfer constraint not binding: $T < T^{max}$

If $T < T^{max}$, then, by complementary slackness, we must have $\xi = 0$. By Equation (A.10), we get $\nu = u_S' - u_N'$. The social value of marginally relaxing the domestic abatement constraint is again given by the difference in marginal utilities since relaxing the domestic abatement constraint allows for shifting the abatement cost burden from the South to the North.

Using $\xi = 0$ in the KKT condition for A_{SF}, we have the following optimality condition for abatement in the South:

$$u_N' C_S'(A_{SD} + A_{SF}) = -\sum_i u_i' D_i'(A) + \gamma C_S''.$$

The optimality condition for abatement in the North is again given by (A.10). By the uniform carbon price constraint, we thus have

$$-\sum_i u_i' D_i'(A) + \gamma C_S'' = -\sum_i u_i' D_i'(A) - \gamma C_N''.$$

Since $C_i'' > 0$ for all i, this equation can only be satisfied for $\gamma = 0$, indicating that the uniform carbon price constraint is not binding. However, in contrast to Case 1.1, this is now not because marginal utilities are equalized, but because of the availability of sufficient mitigation finance to equalize carbon prices.

With $\gamma = 0$, we obtain the optimal uniform carbon price

$$C_N'(A_{ND}) = C_S'(A_{SD} + A_{SF}) = -\frac{1}{u_N'} \sum_i u_i' D_i'(A).$$

(CASE 2.2) Transfer constraint binding: $T = T^{max}$

If $\xi > 0$, then, by complementary slackness, we must have $T = T^{max}$. By Equation (A.10), we have $\nu + \xi = u_S' - u_N'$. The social value of marginally relaxing both the transfer constraint and the domestic abatement constraint is the difference in marginal utilities. We can also write the previous equation as $u_N' + \xi = u_S' - \nu$, and since $\nu > 0$ and $\xi > 0$, we have $u_S' > u_S' - \nu > u_N'$.

Rearranging the optimality condition for A_{SF} and plugging in $u_N' + \xi = u_S' - \nu$, we have the following optimality condition for abatement in the South:

$$\underbrace{(u'_S - \nu)}_{\in(u'_N, u'_S)} C'_S(A_{SD} + A_{SF}) = -\sum_i u'_i D'_i(A) + \gamma C''_S.$$

The optimality condition for abatement in the North is again given by (A.10). Using the uniform carbon price constraint, $C'_S(A_{SD}+A_{SF}) = C'_N(A_{ND})$, we can solve these two conditions for the optimal uniform carbon price, yielding

$$C'_S(A_{SD} + A_{SF}) = C'_N(A_{ND}) = -\sum_i u'_i D'_i(A) \frac{C''_N + C''_S}{\underbrace{(u'_S - \nu)}_{\in(u'_N, u'_S)} C''_N + u'_N C''_S}.$$

A.11 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 begin by computing the net present value (NPV) of the utilitarian welfare changes across two solutions for each region (the numerator in Equation (A17)) [62], 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:

$$\frac{\sum_t L_{it}\beta^t u(x_{it}^{Util}) - \sum_t L_{it}\beta^t u(x_{it}^{Neg})}{L_{i0}} = u(x_{i0}^{cf}) - u(x_{i0}^{Neg}), \tag{A17}$$

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, $u(x_{it}) = \frac{x_{it}^{1-\eta}}{1-\eta}$ (for $\eta = 1$, $u(x_{it}) = \log(x_{it})$, where η is the elasticity of the marginal utility of consumption),

[62] 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.

we can solve for the counterfactual per capita consumption in the initial period:

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

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

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

In addition, I express the utilitarian welfare changes as the welfare-equivalent consumption change in 2005 if consumption were distributed equally (specifically, the "Global" values in Figures 6 and A10 and all values in Figures A9 and A11). Let $\bar{x}_0$ be the world average per capita consumption in 2005; that is $\bar{x}_0 = \frac{\sum_i L_{i0} x_{i0}}{\sum_i L_{i0}}$.

I then proceed as above to calculate the counterfactual per capita consumption in the initial period for the world average consumer:

$$\bar{x}_{i0}^{cf} = \left[(1-\eta) \frac{\Delta PV(U)}{\sum_i L_{i0}} + (\bar{x}_{i0})^{1-\eta} \right]^{\frac{1}{1-\eta}},$$

where $\Delta PV(U) = \sum_t L_{it} \beta^t u(x_{it}^{Util}) - \sum_t L_{it} \beta^t u(x_{it}^{Neg})$ for the regional values and $\Delta PV(U) = \sum_t \sum_i L_{it} \beta^t u(x_{it}^{Util}) - \sum_t \sum_i L_{it} \beta^t u(x_{it}^{Neg})$ for the global values.

Finally, the aggregate welfare-equivalent consumption change in 2005, if consumption were distributed equally, is calculated as follows:

$$\Delta \bar{X}_0 = \sum_i L_{i0} \left(\bar{x}_0^{cf} - \bar{x}_0^{Neg} \right).$$

Appendix B: Additional Tables and Figures

B.1 Additional tables

Table A1: Utilitarian differentiated carbon prices in the presence of endogenous mitigation finance.

	Region	$C_S(A_{SD}) > C_S^{min}$	$C_S(A_{SD}) = C_S^{min}$
$T = T^{max}$	South	$-D'_S - \underbrace{\frac{u'_N + \phi(u'_S - u'_N)}{u'_S}}_{\in\left[\frac{u'_N}{u'_S},1\right]} D'_N$	$-\underbrace{\frac{u'_S}{u'_S - \nu}}_{\in\left[1,\frac{u'_S}{u'_N}\right]} D'_S - \underbrace{\frac{u'_N + \phi(u'_S - u'_N - \nu)}{u'_S - \nu}}_{\in\left[\frac{u'_N}{u'_S},1\right]} D'_N$
	North	$-\underbrace{\frac{u'_S}{u'_N + \phi(u'_S - u'_N)}}_{\in\left[1,\frac{u'_S}{u'_N}\right]} D'_S - D'_N$	$-\underbrace{\frac{u'_S}{u'_N + \phi(u'_S - u'_N - \nu)}}_{\in\left[1,\frac{u'_S}{u'_N}\right]} D'_S - D'_N$
$T < T^{max}$	Both	$-\sum_i D'_i$	$-\frac{1}{u'_N} \sum_i u'_i D'_i$

Notes: The carbon prices in the North and the South are given by the marginal abatement costs $C'_N(A_{ND})$ and $C'_S(A_{SD} + A_{SF})$, respectively. Note that if $\phi = 0$, the social value of marginally relaxing the domestic abatement constraint is zero, $\nu = 0$. The results are derived as in Appendix A.10.

Table A2: Cumulative global industrial CO_2 emissions ($GtCO_2$) depending on the optimization problem and the utility discount rate (ρ).

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

B.2 Additional figures

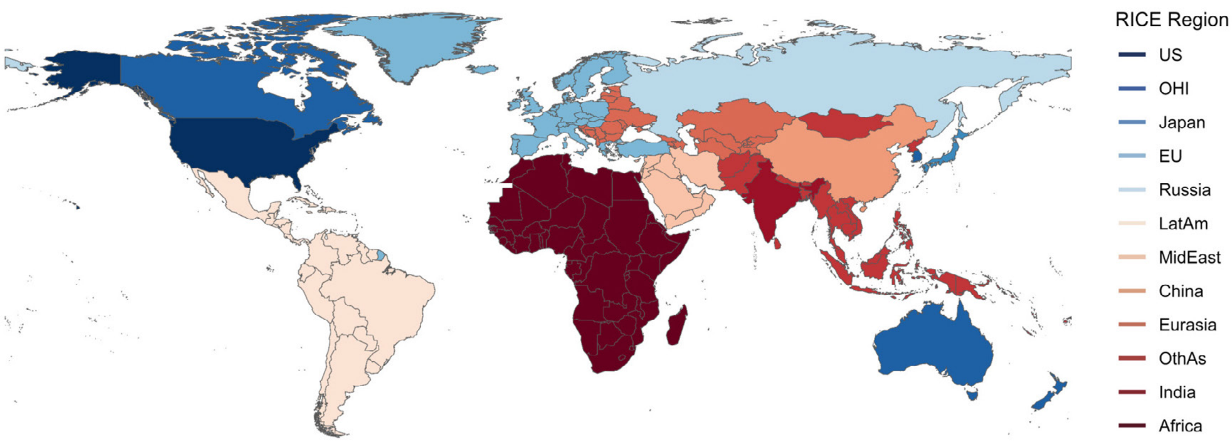

Figure A1: Regions of the RICE model.

Notes: Countries of the same color belong to the same region (OHI = Other High Income countries, OthAs = Other Asia). Regions are arranged on the color scale from rich (blue) to poor (red).

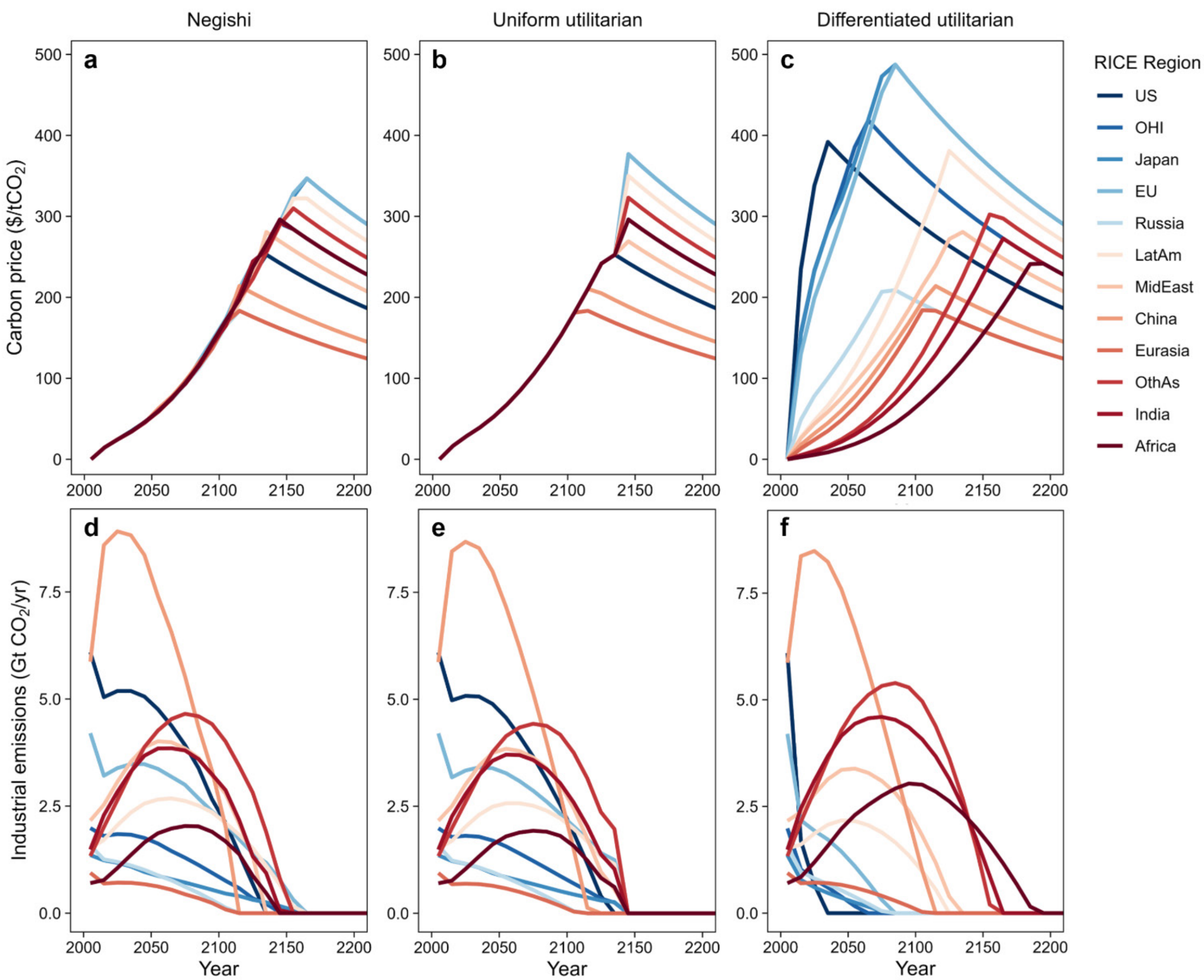

Figure A2: Optimal trajectories for carbon prices and industrial emissions conditional on the optimization problem.

Notes: Results are for the utility discount rate of 1.5%. Panels (a)-(c) show 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. Panels (d)-(f) show 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.

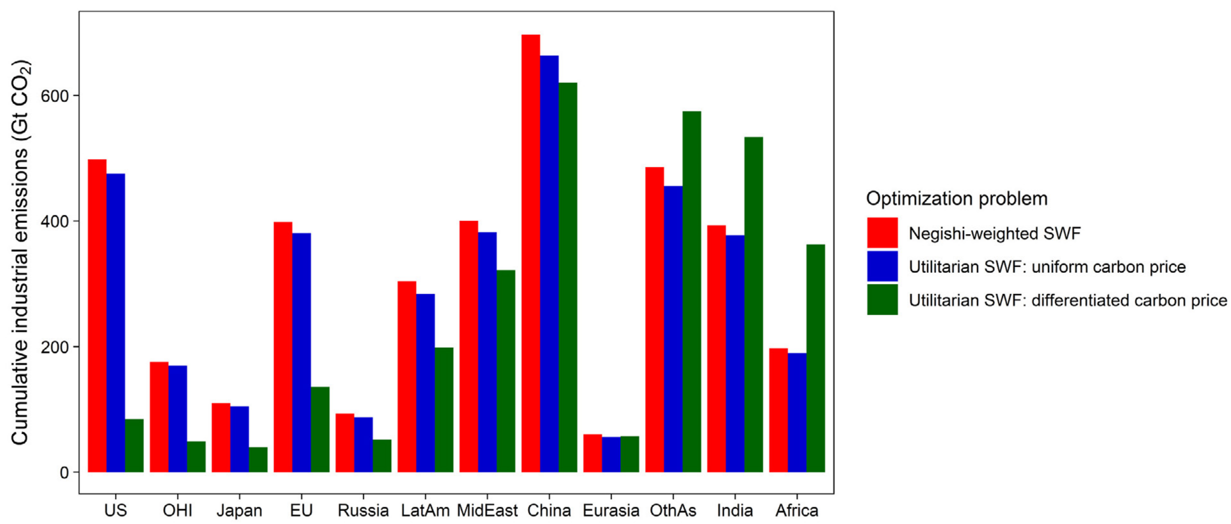

Figure A3: Optimal cumulative industrial emissions depending on the optimization problem.

Notes: The figure shows the results for the utility discount rate of 1.5%.

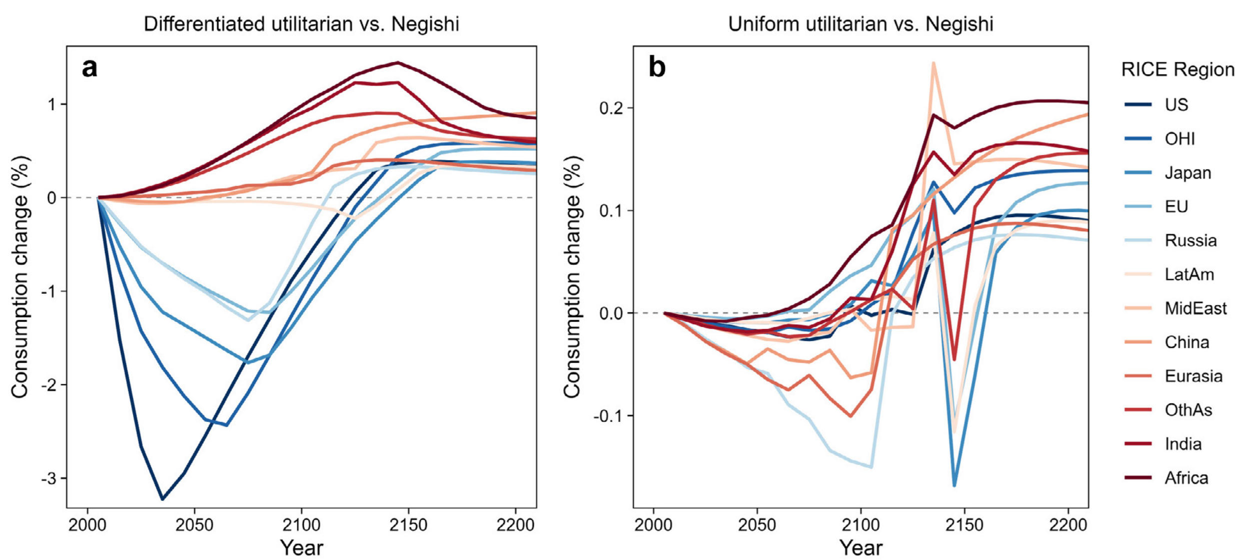

Figure A4: Relative regional consumption changes between the Negishi solution and the utilitarian solutions.

Notes: 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. The figure shows the results for the utility discount rate of 1.5%.

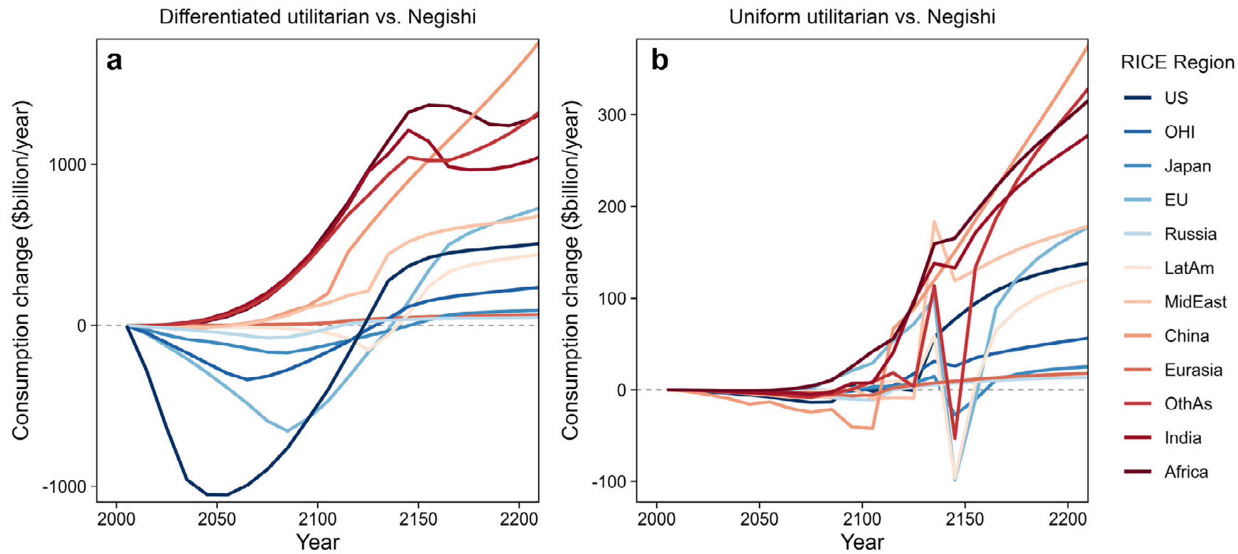

Figure A5: Regional consumption changes between the Negishi solution and the utilitarian solutions.

Notes: Positive values indicate a higher consumption level in the utilitarian solutions. The figure shows the results for the utility discount rate of 1.5%.

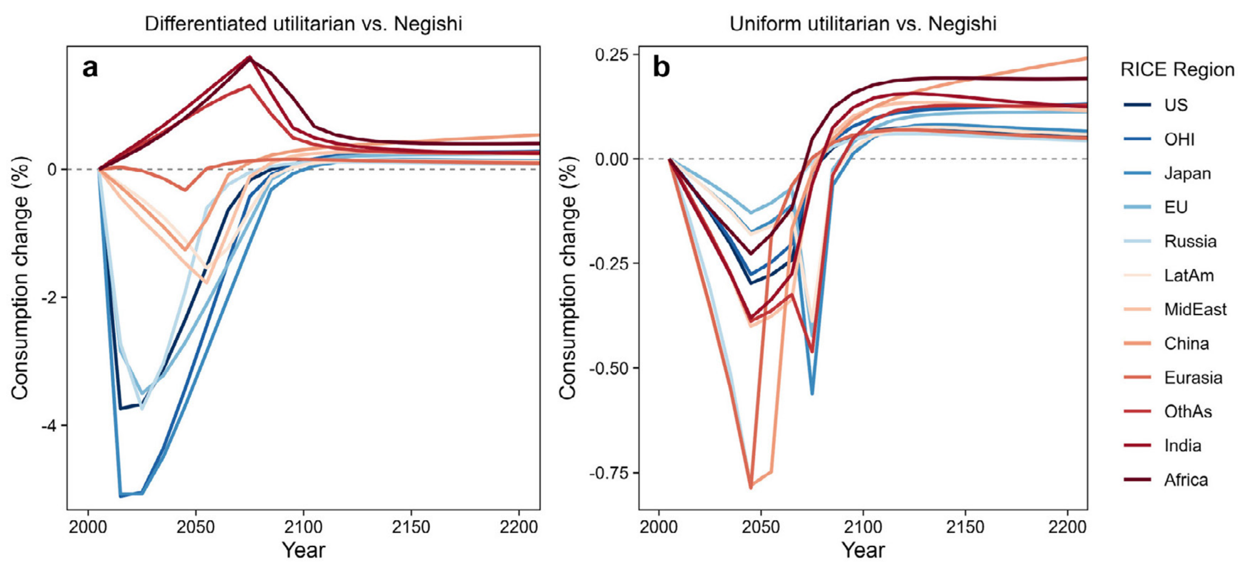

Figure A6: Relative regional consumption changes between the Negishi solution and the utilitarian solutions.

Notes: 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. The figure shows the results for the utility discount rate of 0.1%.

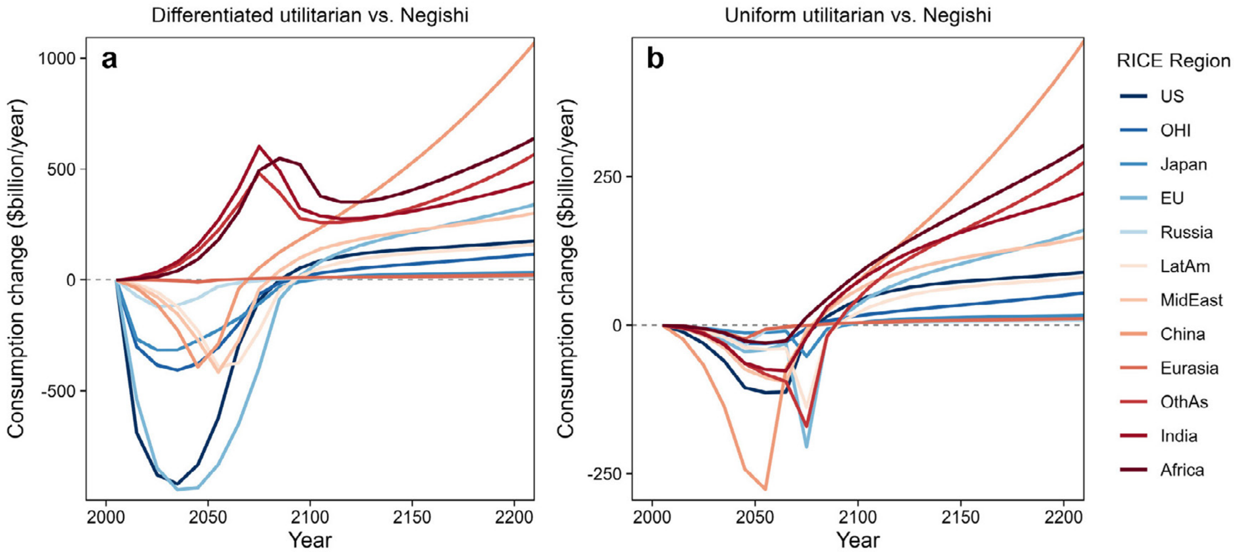

Figure A7: Regional consumption changes between the Negishi solution and the utilitarian solutions.

Notes: Positive values indicate a higher consumption level in the utilitarian solutions. The figure shows the results for the utility discount rate of 0.1%.

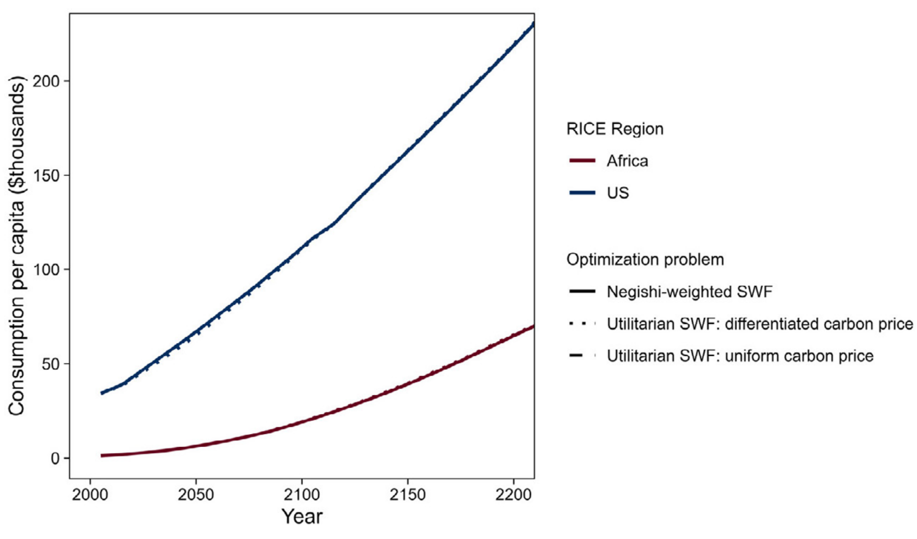

Figure A8: Consumption per capita trajectories for Africa and the US.

Notes: The figure shows the results for the utility discount rate of 1.5%.

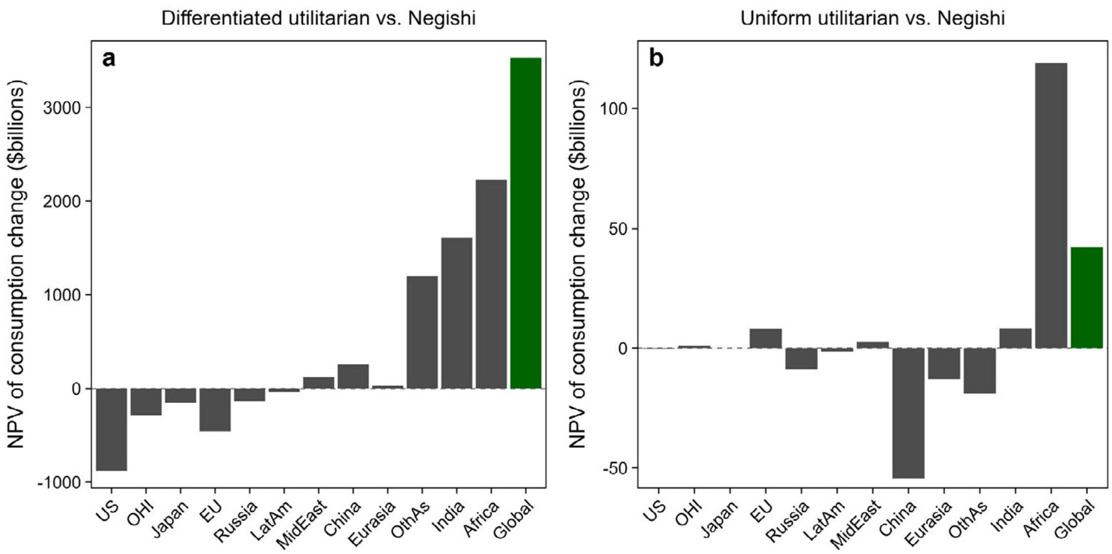

Figure A9: Utilitarian welfare changes.

Notes: The values express the regional and global utilitarian welfare change in the welfare-equivalent consumption change in 2005 if consumption were distributed equally (for details, see Appendix A.11). The figure shows the results for the utility discount rate of 1.5%.

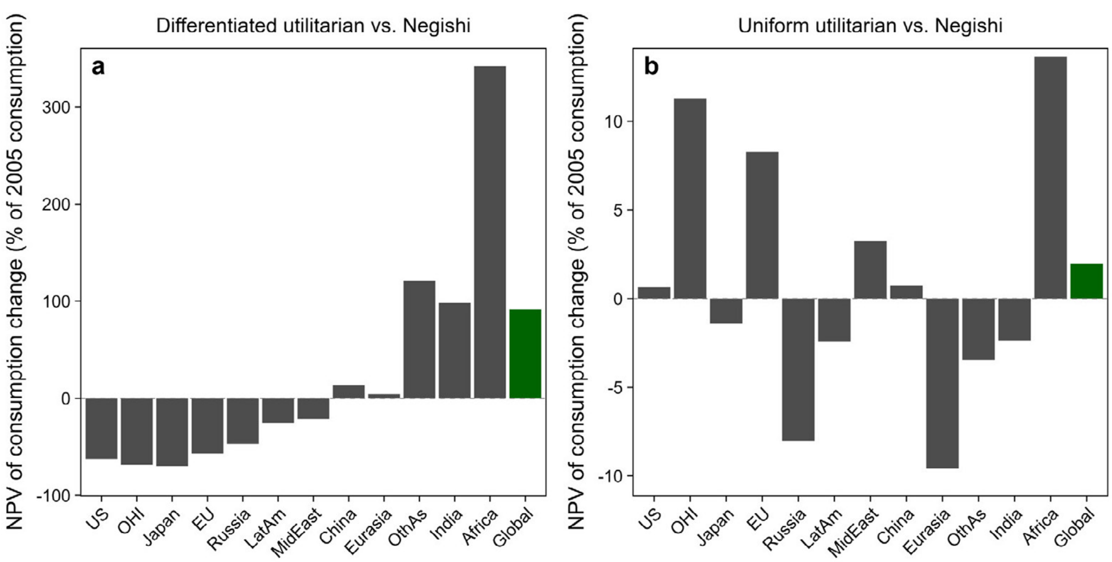

Figure A10: Net present value of consumption changes.

Notes: The values show the welfare-equivalent consumption change in 2005, as a percentage of the consumption in 2005. The "Global" value expresses the global utilitarian welfare change in the welfare-equivalent consumption change in 2005 if consumption were distributed equally (for details, see Appendix A.11). The figure shows the results for the utility discount rate of 0.1%.

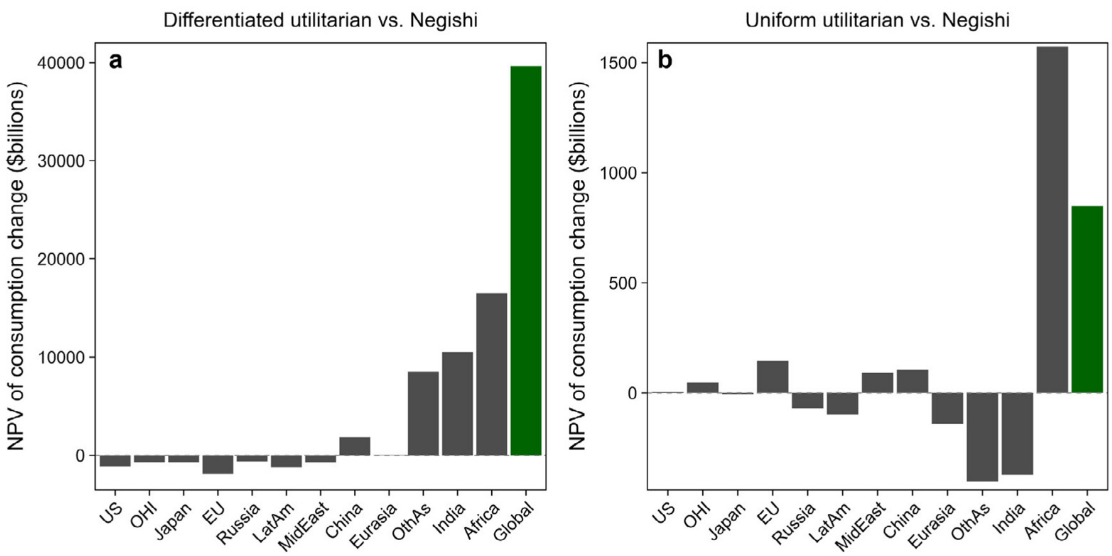

Figure A11: Utilitarian welfare changes.

Notes: The values express the regional and global utilitarian welfare change in the welfare-equivalent consumption change in 2005 if consumption were distributed equally (for details, see Appendix A.11). The figure shows the results for the utility discount rate of 0.1%.

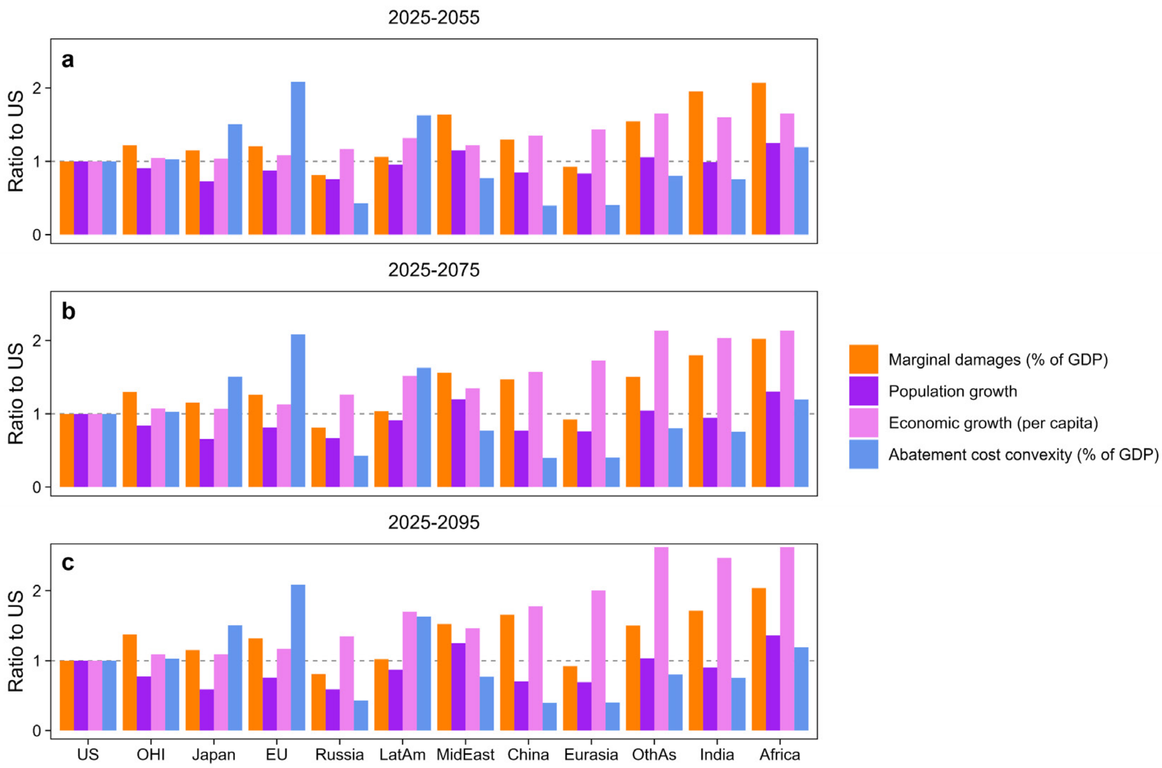

Figure A12: Relative regional marginal damages, abatement cost convexities, population and economic growth.

Notes: The ratio of the convexities in the abatement cost functions is $c''_{i,2025}/c''_{US,2025}$ (evaluated at uniform carbon prices). The ratio of the marginal damages as a percentage of GDP is $d'_{i,t}/d'_{US,t}$. The ratios of population growth and economic growth are given by the relative growth factors $\frac{L_{i,t}/L_{i,2025}}{L_{US,t}/L_{US,2025}}$ and $\frac{y_{i,t}/y_{i,2025}}{y_{US,t}/y_{US,2025}}$. where y is the GDP per capita. The year t corresponds to either 2055, 2075, or 2095 in panels (a), (b), and (c), respectively. The figure shows the results for the utility discount rate of 1.5%.

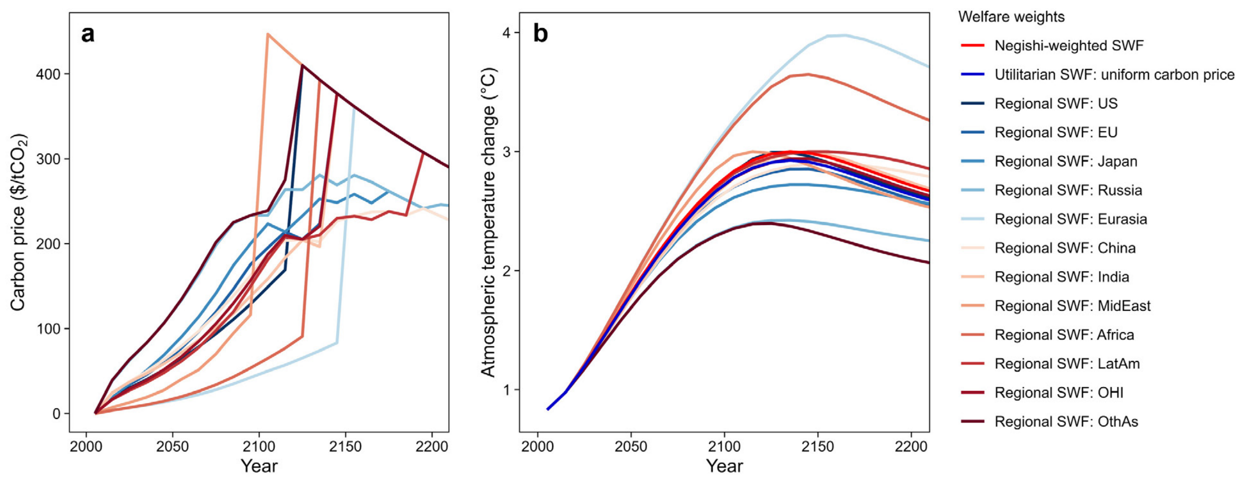

Figure A13: Regions' preferred uniform carbon prices and corresponding temperature trajectories.

Notes: The figure shows the results for the utility discount rate of 1.5%. Temperature changes are relative to 1900. Note that Russia, Eurasia and China have the lowest backstop technology prices, causing the large carbon price increases once it is beneficial for these regions to increase the globally uniform carbon price above the level of their respective backstop prices.

Appendix C: Supplementary Information

C.1 Optimal carbon prices for arbitrary welfare weights

Definition 7. *The **optimal differentiated carbon price** (for arbitrary welfare weights) for region i is implicitly defined by*

$$\tau_i^{diff} = C_i'(A_i^*) = -\frac{1}{\alpha_i u'(x_i^*)} \sum_{j \in \mathcal{I}} \alpha_j u'(x_j^*) D_j'(A^*). \tag{A18}$$

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 Eyckmans et al. (1993) and 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.

It is insightful to rearrange Equation (A18) to

$$\alpha_i u'(x_i^*) C_i'^* = -\sum_{j \in \mathcal{I}} \alpha_j u'(x_j^*) D_j'(A^*).$$

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

$$\tau^{uni} = C_i'(A_i^*) = -\sum_i \alpha_i u'(x_i^*) D_i'(A^*) \frac{C_S'' + C_N''}{\alpha_N u'(x_N^*) C_S'' + \alpha_S u'(x_S^*) C_N''}. \tag{A19}$$

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[63]; that is, $C_i'' = 0$ for one i. Specifically, if

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

the Global North has a linear abatement cost function, then the expression collapses to the differentiated carbon price expression for the Global North[64]; 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.

C.2 Abatement cost and damage functions of the RICE model

In the RICE model, regional climate damages are given by

$$D_{it} = Y_{it} d_{it},$$

where Y_{it} is the GDP gross of damages and abatement costs, and d_{it} denotes the climate damage as a fraction of GDP, which is composed of damages from atmospheric temperature changes and sea level rise (which are ultimately functions of emissions/abatement).

Regional abatement costs are given by

$$C_{it} = Y_{it} \underbrace{\frac{b_{it}\sigma_{it}}{\theta} \left(\frac{A_{it}}{\sigma_{it} Y_{it}} \right)^{\theta}}_{c_{it}},$$

where b_{it} is the price of a backstop technology (i.e., the marginal abatement cost 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 (in RICE, $\theta = 2.8$). Note that the abatement costs per GDP, c_{it}, are a function of $\frac{A_{it}}{Y_{it}}$.

The damage function from atmospheric temperature changes is shown in Figure A14. The trajectories of the regional baseline carbon intensities and backstop technology prices are shown in Figure A15.

[64]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.

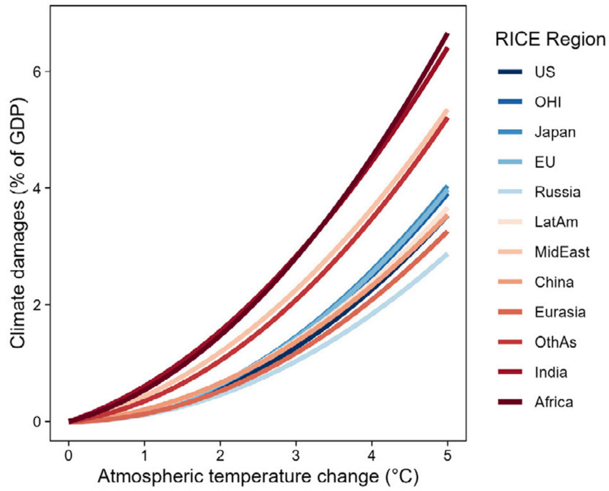

Figure A14: Regional damage functions for atmospheric temperature changes in the RICE model.

Notes: Temperature changes are relative to temperatures in 1900.

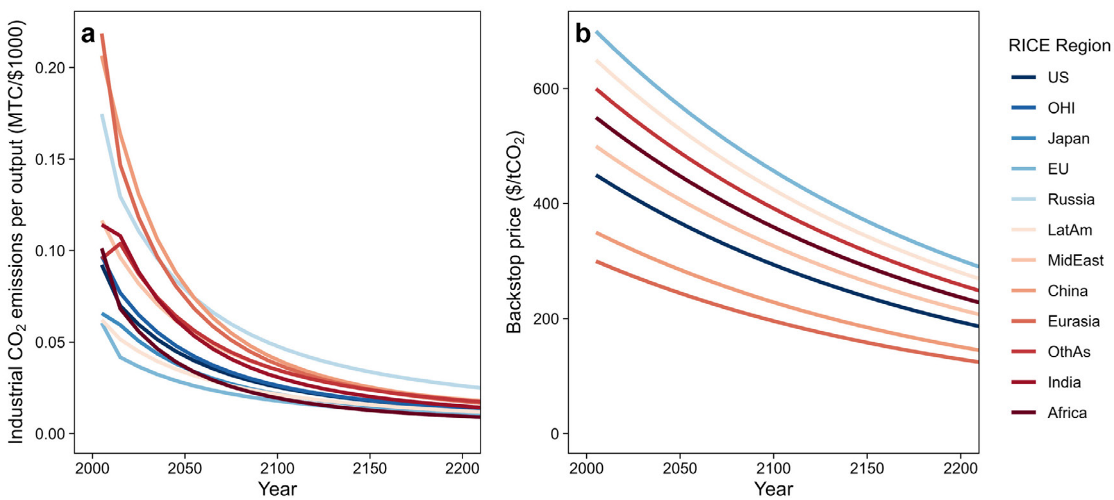

Figure A15: Regional baseline carbon intensities (a) and backstop technology prices (b) in the RICE model.

Notes: 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. The following regions have identical backstop prices: (1) Russia and Eurasia, (2) Other High Income (OHI) countries, Africa, and India, (3) Japan and the EU.

C.3 Time-variant Negishi weights

The time-variant Negishi welfare weights are given by

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

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:

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

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.4 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) [65]. 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 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)[66].

[65] 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).

[66] 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.

More specifically, the total transfer quantity, T_t^{tot}, increases from its baseline value in 2025, T_{2025}^{tot} [67], 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

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

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

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

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:

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

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:

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

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.

C.5 The role of non-conditional transfers

The optimal temperature trajectories in the presence of interregional non-conditional transfers are shown in Figure A16. 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

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

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.

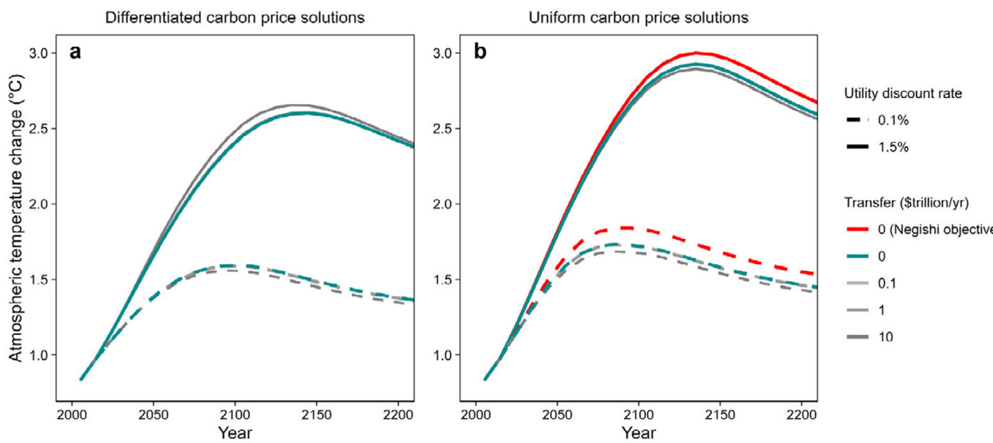

Figure A16: Optimal atmospheric temperature trajectories conditional on the optimization problem, the total non-conditional interregional transfer, and the utility discount rate.

Notes: 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.6 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 A2 and Table 5). 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, 2020). 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 (2020) 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 emissions 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, 2020). If the permit market is fully competitive, this would result in a globally harmonized carbon price. However, as Budolfson and Dennig (2020) 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 emissions 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 A2 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.