WORKING PAPERS

N° 1723

MARCH 2026

"SHOULD WE STOP THE COPs?"

Renaud Bourlès, Jeremy Laurent-Lucchetti
and Jean-Charles Rochet

Toulouse School of Economics

SHOULD WE STOP THE COPs ?

Renaud BOURLÈS* Jeremy LAURENT-LUCCHETTI† Jean-Charles ROCHET‡

March 2026

Abstract

More than a decade after COP21, carbon emission trajectories remain far above the 1.5 ° C threshold, due to lack of international consensus. Departing from cost-benefit approaches, we assess the maximum reduction in carbon emissions that could be accepted by all countries. We characterize the *target-consistent* mechanism that minimizes global emissions subject to the participation constraint of each country. The mechanism can be implemented either via a uniform carbon tax or as a cap–and–trade system. Calibrated to data from 69 countries, including GDP, carbon intensities, and observed tax rates, our model suggests – for our baseline scenario – that the maximum uniform carbon price politically acceptable for all countries is $250 per ton. It could reduce global emissions by 35%, but would require unprecedented international transfers: up to 3% of world GDP, with a large redistribution from high-income, low-emission countries to carbon-intensive emerging economies. Our analysis highlights the structural ambition gap imposed by voluntary cooperation and identifies two levers to overcome it: convergence in green technologies and stronger political support for mitigation. Without progress in these dimensions, international climate policy remains constrained to deliver only modest results.

*Aix-Marseille Univ., CNRS, Centrale Med., AMSE.
†Institute of Economics and Econometrics, GSEM, University of Geneva.
‡Toulouse School of Economics.

1 Introduction

In 2015, the international community committed to the Paris Agreement, aiming to limit global temperature increases to well below 2 degrees Celsius and to continue efforts to constrain the rise to 1.5 degrees Celsius. These targets were collectively endorsed as essential benchmarks for mitigating climate change. However, more than a decade later, emissions trajectories remain well above these thresholds, raising a critical question: What is the most ambitious climate outcome achievable under purely voluntary international cooperation? In other words, what global emissions reduction can the world realistically attain, and by what mechanism? Answering this question is crucial for assessing the "realism gap" between aspiration and outcome in climate negotiations.

Our approach is to focus on climate policies that satisfy voluntary participation constraints by all countries. We develop a theoretical framework in which a supranational principal (interpreted as the Conference of the Parties, COP) designs a mechanism that minimizes global emissions subject to every country being at least as well off as in the status quo. In practice, this means that each country agrees to implement a domestic carbon policy (e.g. a carbon tax or a cap-and-trade system) in exchange for a net transfer such that no country loses from the agreement. This mechanism yields an equilibrium carbon price and a transfer schedule that is budget-balanced and ensures voluntary participation for all participants. The solution can be viewed as a constrained efficient climate treaty. It achieves the lowest feasible emissions level while satisfying each country's own interest. This approach follows the "target-consistent" philosophy often advocated in policy circles, focusing on achieving agreed targets given political constraints, as opposed to the unconstrained global social optimum from a standard cost-benefit analysis. We do not take a position on the normative merit of either approach (see, e.g., Dietz et al. (2018) and Stern and Stiglitz (2022) for discussions); instead, we assume that in practice international negotiations gravitate toward target-consistent outcomes driven by what participants will voluntarily accept.

Our theoretical results reveal important trade-offs. Because countries differ in carbon intensity and abatement cost, an efficient voluntary agreement entails shifting production toward cleaner countries, while high-emission countries scale down their output. Crucially, global welfare gains are redistributed through transfers so that each country's welfare remains at its baseline level (the status quo). In other words, the mechanism achieves emissions reductions without making any country worse off, which is a prerequisite for full participation. A key feature of our framework is the explicit modeling of political constraints: we account for the influence of environmental preferences and producers' lobbies in shaping policy choices. These constraints determine which international agreements are politically viable and help explain observed patterns of carbon pricing. Despite these frictions, the mechanism delivers a simple and powerful result: the optimal carbon price is uniform across countries, with redistribution handled entirely through international transfers. However, this comes at the cost of large international transfers and a preservation of underlying inequalities. Low-emission (often high-income) countries must compensate high-emission (often lower-income) countries, which incur output losses under the climate policy. The outcome does not inherently address inequalities: Every region's welfare stays at the status quo level, meaning that preexisting disparities persist. Moreover, the scale of transfers required for high-emission countries to participate is enormous, underscoring the need for strong commitments to financial support. This stands in stark contrast to the vague and modest transfer promises observed in recent COP negotiations, a gap that calls into question the credibility of achieving ambitious targets without a serious redistribution mechanism.[1]

A central issue is whether such an agreement is credible once implemented. The mechanism itself does not mechanically enforce compliance; rather, its stability rests on a credible reversion strategy chosen by non-deviators. Specifically, we consider a trigger strategy under which countries implement the target-consistent allocation as long as all comply, and revert to the non-cooperative status-quo if any country deviates. Because transfers are calibrated to preserve each country's welfare exactly at its status-quo level, a deviating country cannot obtain a strictly higher payoff than under reversion. The agreement is therefore self-enforcing under voluntary participation. In Section 6.4, we quantify how robust this voluntary frontier is to unilateral and coalition deviations, and interpret such departures as capturing political frictions, including reluctance toward large cross-border transfers.

We then bring the model to the data, quantifying the limits of voluntary cooperation. We calibrate our model using a panel of 69 economies, combining macroeconomic data from the Penn World Table (for output, consumption, and population), emissions and energy intensities from EDGAR, and information on tax systems from the OECD and IMF. In particular, we assemble estimates of effective carbon prices (including explicit carbon taxes, emissions trading system prices, and fuel excise taxes) and broad-based consumption tax rates (VAT/sales taxes) for each country around 2018. We also incorporate country-specific estimates of climate damage: we assume that, in the absence of international cooperation, governments will not set carbon prices above their own marginal benefit from emissions abatement. In calibration, therefore, we bound each country's implied carbon price by its domestic carbon cost estimated by Ricke et al. (2018). This accounts for political realism: absent side payments or extraordinary altruism, no country would in isolation rationally impose a carbon tax exceeding the climate damages it internalizes for itself. We use the observed carbon and consumption tax data to infer the underlying preferences and political constraints of each country. In effect, we estimate two key parameters for every country: an environmental concern parameter (measuring the weight policymakers place on global emissions reductions in their objective) and a political weight on consumers (capturing the weight on consumer welfare relative to fiscal revenue or industrial profit). Intuitively, the latter is identified by the general consumer-tax level a country is willing to impose on its economy, while the former is identified by the stringency of its carbon pricing given that general tax distortion cost. The calibration reduces these parameters by matching the observed consumption tax and the carbon price in each country, subject to the constraint that the carbon price does not exceed the country's own damage to the environment. We find that our model can replicate the observed pattern of production, consumption, and emissions across countries quite well, lending credibility to the structural interpretation of the parameters. In particular, many high-income regions emerge with implied carbon pricing at or near their

[1] In Section 4.7, we extend the mechanism to a dynamic, intertemporal environment, showing that the efficient period-by-period allocation can be decentralized by a time-varying uniform tax path together with period-specific transfers.

domestic damage levels – in other words, they are effectively fully internalizing their own climate externality - while the policies of other countries reflect very low environmental concern. This heterogeneity in environmental preferences helps explain the wide dispersion in carbon prices and underpins the limits of voluntary coordination.

Using the calibrated model, we evaluate the maximum global emission reduction that can be achieved with full voluntary participation. Our central quantitative finding is striking: the most ambitious equilibrium reduces global emissions by roughly one-third relative to the baseline. In our benchmark calibration, the constrained-efficient agreement delivers about a 35% cut in emissions, implemented via a uniform carbon price on the order of \$250 per ton of CO_2. This level of abatement is, in principle, sufficient to meet the 2 ° C target - indeed, it modestly overshoots the aggregate emissions reduction consistent with a 2 ° C pathway (approximately a 28% cut from baseline). However, achieving this outcome relies on enormous international transfers: on the order of 2.5 trillion US dollars annually (roughly 3% of world GDP) flowing from low-emission, high-income economies to high-emission emerging economies. For example, countries such as the United States and the EU would be net contributors, subsidizing the mitigation efforts in countries such as China, India, and other large emerging polluters. These transfers would compensate the latter countries for their output reductions incurred under the high carbon price, ensuring that no country is left worse off, and thereby sustaining universal participation.

Importantly, our results also identify a structural lever that shifts the voluntary frontier: the international diffusion of green technologies. In the model, carbon intensity captures how emissions scale with production. When high-emission countries adopt cleaner technologies, carbon intensity declines, reducing the welfare cost of a given carbon price and relaxing participation constraints. We simulate proportional reductions in carbon intensity for carbon-intensive economies and recompute the participation-constrained optimum, taking into account the implied incentives for low-emitting countries to support technology transfer. In simulations where higher-intensity countries improve carbon intensity (e.g., a 10% reduction), the mechanism can support a higher uniform tax (approx \$324) and deliver substantially larger emissions cuts (roughly 45%). Thus, technology transfer can lower the political and fiscal cost of deeper cooperation—although large-scale diffusion typically unfolds slowly and would require active, coordinated policies (R&D support, technology transfer programs, concessional finance) to be realized at scale. The political reality illuminated by our results is that while substantial emission mitigation is technically achievable under voluntary cooperation, it would require a scale of wealth transfers and international solidarity far beyond what is currently contemplated. Indeed, our calibration suggests that the main reason why current policies fall short of this cooperative ideal is that the required transfers are not being delivered. In effect, the status quo already reflects a partially coordinated outcome – some countries (mostly wealthy ones) are voluntarily bearing more abatement cost than their strictly domestic interest would warrant, thereby providing a global benefit akin to an implicit transfer to others. But this piecemeal altruism and fragmented effort yield only a fraction of the possible gains. Without explicit, large-scale redistribution, further ambition quickly hits a ceiling.

Finally, we contrast the result of our mechanism with the pledged emission cuts under the Paris Agreement. Using data from the Nationally Determined Contribution (NDC) submitted in the wake of

COP21, we estimate that if all current pledges were fully implemented, global emissions would fall by roughly 10% relative to baseline. This is considerably more reduction than has been achieved so far, but it remains well short of the ∼28% required for 2°C, and our results indicate that it is still below the feasible voluntary optimum. Moreover, the pledge-based approach is inefficiently distributed: countries' pledges do not equalize marginal abatement costs, implying a higher overall welfare cost for a given emissions outcome. We calculate that a uniform carbon price of about $70/ton (with appropriate transfers) could achieve the same aggregate 10% reduction at much lower global cost than the patchwork of individual pledges. In other words, there is substantial room to improve the cost-effectiveness of current climate efforts by moving toward a cooperative carbon price mechanism. The comparison underscores a key message: the major constraint in reaching deeper global cuts is not the lack of cost-effective policy tools, but rather the political-economy constraints that prevent their adoption. Our framework quantifies those constraints and thereby delineates the frontier of what voluntary climate cooperation can deliver. The sobering takeaway is that, under the prevailing political forces, this frontier falls short of the aspirations necessary for avoiding dangerous climate change. Strengthening international climate action thus hinges on shifting that frontier – by altering incentive structures, providing greater transfers or technological support, or fundamentally changing how countries value global environmental benefits.

2 Relation with the literature

Our paper relates to three main strands of literature. First, foundational work in climate economics centers on the social cost of carbon (SCC) – the Pigovian price equating marginal climate damages to abatement cost. While Weitzman (1974) famously emphasized the price vs quantity design problem under uncertainty, early models (Nordhaus's DICE and others) advocated an "optimal" carbon tax. More recent debate has focused on whether SCC estimates from integrated assessment models (IAMs) properly reflect climate risk. Stern and Stiglitz (2022) argue that standard IAM-based SCCs (for example, the U.S. interagency estimate of ∼$50–100/ton by 2030) imply warming paths incompatible with the Paris target. They propose a target-consistent approach: calibrating carbon prices to achieve long-run climate goals rather than to median-damage estimates. This "target-consistent" SCC can be several times higher than conventional IAM outputs. Building on this, Aldy et al. (2021) and Stern (2022) incorporate deep uncertainty and ethical considerations—such as intergenerational equity—into SCC valuation. Their work reinforces the need for stronger carbon pricing and critiques traditional IAMs for ignoring distributive consequences and global inequality. These contributions collectively shift the emphasis from purely efficiency-based pricing to a more politically and ethically informed benchmark for climate action. Overall, this literature stresses that plausible climate damages and risk aversion imply much higher carbon prices than those derived from baseline IAM projections. We use the SCC literature to benchmark the ambition gap between Paris pledges and target-consistent carbon price and offer a mechanism allowing to implement such a target-consistent objective.

The second important strand of the literature relates to mechanism design and externalities. In the context of climate change, the main challenge is to design institutions and policies that incentivize agents

– whether countries or firms – to adopt socially optimal behavior, despite having private incentives that may diverge from the common good. A significant body of work explores how mechanism design can help overcome issues of free-riding and enforcement that undermine international climate cooperation. This mechanism-design perspective builds on the classic insight of Barrett (1994) that international environmental agreements must be self-enforcing, with participation and compliance sustained by countries' own incentives rather than by external enforcement. Nordhaus (2015) highlights the difficulties posed by free-rider problems and the failure to enforce climate commitments, proposing mechanisms that could improve cooperation and compliance at the global level. In a complementary vein, Martimort and Sand-Zantman (2016) develop a mechanism to address the challenge of asymmetric information in environmental regulation. They show how the enforcement of climate policies is complicated by the fact that different countries have different preferences, information, or capacities to reduce emissions. Meanwhile, Gersbach et al. (2021) examine the potential of market-based solutions, such as tradable permits and carbon markets, as tools for decentralized environmental policy. This strand of literature emphasizes how political incentives can be structured to address environmental challenges effectively. Adding further nuance, Harstad (2012) introduces a dynamic model of international climate agreements that focuses on the problem of delayed participation, in which countries strategically delay their commitments, expecting others to act first. Harstad argues that this delay exacerbates the free-rider problem and impedes global climate efforts. Battaglini and Harstad (2020) further analyze the design of the treaty in a political-economy context: they show that democratic incumbents tend to negotiate "weak" treaties (which may be barely binding) to differentiate themselves from challengers. In equilibrium, this logic can yield agreements that are either too lax or too technology-focused (avoiding hard sanctions) relative to the global optimum. In sum, the mechanism-design literature clarifies why and how realistic treaties deviate from the textbook social optimum, and it suggests structural solutions (club punishments, finance transfers, innovation incentives) that can push cooperation toward higher ambition. We contribute to this literature by designing a mechanism that implements the most ambitious emission target achievable under voluntary participation. Moreover, we explicitly model political constraints – including the role of lobbying and domestic environmental preferences – that prevent countries from participating in international agreements or adopting stringent climate policies. This allows us to quantify the "realism gap": the divergence between what is environmentally desirable and what is politically and economically feasible given countries' preferences and incentive restrictions.

The last strand examines how trade interacts with environmental policy. Trade can influence environmental outcomes both positively, by spreading cleaner technologies, and negatively, by creating "pollution havens" where environmental regulations are lax. In a recent paper, Copeland et al. (2022) provides a comprehensive review of the literature on the trade-environment nexus, focusing on how trade affects environmental outcomes through channels such as specialization, scale effects, and the diffusion of technology. They emphasize that trade liberalization can either harm or benefit the environment, depending on accompanying environmental policies. Classic models decompose the effect of trade on pollution into scale, technique, and composition effects. Copeland and Taylor (2003) show that, under a wide range of parameters, trade gains in technical efficiency can outweigh scale effects, leading to a reduction in net pollution. Antweiler et al. (2001) famously conclude that, in their empirical estimates, "freer trade appears to be good for the environment" because the effect of technology dominates. On the other hand, asymmetric regulations can create pollution havens: heavily polluting industries may relocate to lax-regulation countries (the pollution haven hypothesis). Baccianti and Schenker (2022) integrate voting mechanisms and citizen preferences into the analysis, showing how political outcomes influence the design and enforcement of environmental policies. Maggi and Staiger (2022) and Kotchen and Maggi (2024) study the role of International Environmental Agreements and global environmental policies in the presence of climate catastrophe and lobbying. Incorporating competition and enforcement issues into the trade–environment literature adds nuance to understanding how international policies are shaped by domestic political and economic considerations. This has motivated a growing structural trade-and-environment literature that studies enforcement and leakage in open-economy settings and how trade instruments can discipline free-riding.

Recent structural general-equilibrium work shows how carbon border adjustments and "climate-club" penalties can be optimally designed and calibrated to the trade and industry structure (Farrokhi and Lashkaripour, 2025). Their focus is on how trade-policy instruments—border adjustments and contingent trade penalties—can mitigate free-riding and raise domestic carbon taxes in a rich quantitative trade environment. We complement this approach by studying a different institutional benchmark: a voluntary frontier, defined as the most ambitious global emissions target that can be implemented while satisfying each country's participation constraint relative to the non-cooperative status quo, together with budget balance. Under this benchmark, redistribution is not a side consideration but a feasibility condition: deeper abatement requires larger compensating transfers to ensure unanimous participation. Our main quantitative objects are therefore the target-consistent carbon price that decentralizes this frontier and the magnitude and direction of international transfers required to sustain it. Because the mechanism must satisfy country-by-country participation constraints expressed in welfare levels, the analysis requires recovering absolute welfare changes and fiscal transfers—objects that cannot be obtained from exact-hat algebra, which characterizes only proportional counterfactuals (see e.g. Dekle et al., 2008). We therefore integrate trade considerations within a parsimonious welfare framework that remains compatible with open-economy linkages while allowing transparent identification of key structural parameters—such as countries' environmental preferences—and a tractable quantification of the "realism gap" between target-consistent ambition and what voluntary participation can sustain. In this sense, the paper speaks to the trade-and-environment literature through a different margin: rather than emphasizing sectoral detail, it focuses on how cross-country differences in carbon intensity, productivity, and political willingness shape the set of climate agreements that can be sustained under voluntary participation.

3 Climate policies with voting and trade

We consider a simple model with two goods: a composite good taken as a numéraire and a consumption good whose production generates carbon emissions. There are R regions $r = 1, 2, \ldots, R$, which use different technologies. The production q_r of the consumption good in each region r leads to aggregate global emissions $E = \sum_r \sigma_r q_r$, where σ_r denotes the carbon intensity of production in region r. These emissions negatively affect the welfare of the citizens in each region r, with a factor α_r that may differ between regions. The vector of outputs is denoted by $q = (q_1, q_2, ..., q_R)$.

Each region is composed of two types of citizens: consumers and firm owners (capitalists), who both vote and consume the numéraire good. Consumers also consume the consumption good, while capitalists produce it. For simplicity, we assume one representative firm. The elected leader of each region strategically sets production and consumption levels (via taxation), constrained by election promises.

In each region, the timing is the same:

1. Two political parties, A and B, compete for votes. They commit to future policies (consumption tax and carbon tax) and are solely motivated by winning elections. Tax revenue is redistributed to firm owners. Citizens vote for the party that maximizes their utilities, which incorporate idiosyncratic preference shocks.

2. Given the policy implemented by the winning party in each region, consumers choose their consumption c_r, and owners of firms choose their production q_r.

3. The equilibrium price p is determined by the equality of supply and demand on the international market. All agents are competitive and take p as given. In region r, the consumption price is $p_r^c = p(1 + t_r)$ where t_r is the consumption tax rate. The producer price is $p_r^p = p - \tau_r \sigma_r$ where τ_r is the carbon tax rate.

3.1 Payoffs and preferences.

A key assumption is that consumers in each region have environmental concerns that influence their voting behavior. These concerns may reflect altruism toward future generations or a general concern about environmental damage. However, we assume that all agents behave competitively and do not internalize emissions when making consumption or production choices. Their environmental preferences are therefore expressed only through voting.

Specifically, each consumer of each region r has net utility

$$u(c_r) - p_r^c c_r - \alpha_r E, \tag{1}$$

where $u(c_r) = h\frac{c^{1-\varepsilon}}{1-\varepsilon}$ is the utility of consuming a quantity c_r of the consumption good, p_r^c is the consumer price (which includes consumption taxes), h is a scaling parameter, and α_r captures the environmental concern of consumers about global emissions. For simplicity, we assume that all consumers within a given region r share the same α_r.

Firm owners in region r earn profits:

$$p_r^p q_r - A_r \times C(q_r/A_r) \tag{2}$$

where p_r^p is the producer price of the consumption good (which includes carbon taxes) in region r, the cost function of region r is $A_r \times C(q_r/A_r)$ with $C(q) = k\frac{q^{1+\eta}}{1+\eta}$, k is a scaling parameter and A_r parameterizes region-specific productivities.

3.2 Political objectives

We adopt a standard probabilistic voting framework in which, in each region, political parties A and B are motivated solely by winning elections (or by vote shares) and can commit to future production policies. As usual in such models, the equilibrium policy corresponds to the maximization of a weighted utilitarian social welfare function (see the appendix for detail).

We normalize the political weight of capitalists to 1, and denote by β_r the political weight of citizens in region r. The policy instruments available to elected leaders are limited to a constant consumption tax rate t_r and a constant carbon tax rate τ_r.

In the unique voting equilibrium, each party in region r proposes the policy that maximizes W_r, a weighted sum of citizens' utility (including perceived cost of carbon emissions E), firms' profit and tax revenue:

$$\mathcal{W}_r = N_r\beta_r[u(c_r) - p(1+t_r)c_r - \alpha_r E] + (p - \tau_r\sigma_r)q_r - A_r C\left(\frac{q_r}{A_r}\right) + pt_rN_rc_r + \tau_r\sigma_r q_r. \tag{3}$$

where N_r denotes the number of consumers in region r.

Consumers and firms optimization implies:

$$p(1+t_r) = u'(c_r), \text{and } p - \tau_r\sigma_r = C'\left(\frac{q_r}{A_r}\right). \tag{4}$$

Thus the political objective in region r can be expressed only as a function of c_r and q_r, the price p being taken as given:

$$\mathcal{W}_r(q_r, c_r) = \beta_r N_r[u(c_r) - \alpha_r E] + (1-\beta_r)N_r u'(c_r)\, c_r - A_r C\left(\frac{q_r}{A_r}\right) + p(q_r - N_r c_r). \tag{5}$$

3.3 Optimal policies and global equilibrium

The first order conditions of the maximization of $\mathcal{W}_r$ with respect to (c_r, q_r) imply that the optimal production and consumption policies in each region satisfy:

$$p = u'(c_r)[1 - (1-\beta_r)\varepsilon] = C'\left(\frac{q_r}{A_r}\right) + N_r\alpha_r\beta_r\sigma_r. \tag{6}$$

These conditions determine optimal consumption policy

$$c_r(p) = u'^{-1}\left(\frac{p}{1-(1-\beta_r)\varepsilon}\right) \tag{7}$$

and optimal production policy

$$q_r(p) = A_r C'^{-1}(p - N_r \alpha_r \beta_r \sigma_r) \tag{8}$$

for each country r. These policies are implemented through consumption taxes

$$t_r = \left(\frac{1}{1 - (1 - \beta_r)\varepsilon} - 1 \right) \tag{9}$$

and carbon taxes

$$\tau_r = N_r \alpha_r \beta_r. \tag{10}$$

These two conditions will be key to the calibration in Section 5.

The equilibrium price p^* clears the global market. It is the unique solution of:

$$\sum_r \left[A_r C'^{-1}(p^* - N_r \alpha_r \beta_r \sigma_r) - N_r u'^{-1} \left(\frac{p^*}{1 - (1 - \beta_r)\varepsilon} \right) \right] = 0 \tag{11}$$

3.4 Comparative statics and leakage

Condition (8) implies that the optimal production of the region r, denoted q_r^*, is decreasing in the relative political weight of its citizens β_r, their environmental concerns α_r, and the carbon intensity of the region σ_r.

In contrast, production in region r increases when the relative political weight of citizens, environmental concerns, or carbon intensity increases in other regions, due to the resulting increase in the international price (equation (11)). This leads to a leakage effect, frequently cited by critics of unilateral climate action: any regional policy that lowers emissions can be partially offset by increased emissions elsewhere.

If the rest of the world is more carbon-intensive, this rebound effect may even lead to higher total emissions. In such a case, an increase in environmental concerns or a decrease in carbon intensity in low-emitting regions may paradoxically raise global emissions. This highlights a key challenge: non-coordinated policies may result in perverse environmental outcomes through trade.

This motivates the need for international coordination. The next section develops a mechanism in which each region commits to a production policy q_r. Since consumer behavior is unchanged, the left-hand side of equation (6) still determines c_r as a function of the price p for each region ($c_r(p)$). The market-clearing condition becomes:

$$\sum_r q_r = \sum_r N_r u'^{-1} \left(\frac{\widehat{p}}{1 - (1 - \beta_r)\varepsilon} \right) \tag{12}$$

which determines the price $\widehat{p}$ as a function of the vector of production levels $q = (q_1, \ldots, q_R)$. Consequently, the political objective for each region (i.e. equation (5)) can be written as $\mathcal{W}_r(q_r, c_r(\widehat{p}(q)))$ that we denote $W_r(q_r, q_{-r})$, with the production of other regions (q_{-r}) impacting both emissions and the international price. Even though consumers and capitalists behave competitively (and take p as given) governments take into account the impact of output decisions of all countries on the world price $\widehat{p}(q)$.

4 Target-consistent mechanisms

4.1 Definitions

We introduce a principal (think of the UNFCCC for example)[2] tasked with designing an international agreement—what we call a *target-consistent mechanism*—that reduces global emissions while respecting voluntary participation by each region. The mechanism specifies, for each region r, the amount it receives m_r (positive or negative) and a production level $\overline{q_r}$. Starting from the non-cooperative status quo q^*, a mechanism is target-consistent if:

1. it satisfies the participation constraint
$$W_r(\overline{q_r}, \overline{q_{-r}}) + m_r \geq W_r(q_r^*, q_{-r}^*), \quad \forall r \in \{1, \ldots, R\}, \tag{13}$$

2. it is budget balanced
$$\sum_r m_r = 0. \tag{14}$$

Among all mechanisms satisfying these constraints, the optimal target-consistent mechanism minimizes aggregate emissions.

Definition 1. *We define $W(q) = \sum_r W_r(q_r, q_{-r})$ as the* global welfare *at production levels $q = (q_1, ..., q_R)$. Because $W(q)$ is strictly concave (see Appendix A.1), it admits a unique maximizer. We refer to this allocation as the* utilitarian solution*.*

In what follows, the target-consistent allocation is our main object of interest, while the utilitarian solution serves as a benchmark.

4.2 Preliminary Results

We first establish two properties of the target-consistent allocation.

Lemma 1. *The production levels implemented by the optimal target-consistent mechanism are weakly below those of the utilitarian solution.*

Proof. Suppose instead that target-consistent quantities exceeded the utilitarian solution. Because global welfare is strictly concave, reducing production would both lower emissions and increase global welfare, contradicting optimality of the mechanism. □

[2]The United Nations Framework Convention on Climate Change (UNFCCC) provides the institutional platform through which international climate negotiations are coordinated, notably via the Conference of the Parties.

Lemma 2. *The optimal target-consistent mechanism satisfies*

$$W(\overline{q}) = W(q^*).$$

Proof. Participation requires that each region obtain at least its status-quo welfare after transfers. Summing the participation constraints across regions and using budget balance ($\sum_r m_r = 0$) implies

$$W(\overline{q}) \geq W(q^*).$$

Hence the mechanism minimizes emissions subject to the single constraint $W(\overline{q}) \geq W(q^*)$. This constraint must bind: otherwise emissions could be reduced further while still satisfying participation. $\square$

4.3 The Main Result

Lemmas 1 and 2 imply that the optimal Target-Consistent scheme solves the following program:

$$\begin{aligned} \min_{\{q\}} \quad & E = \sum \sigma_r q_r \\ \text{s.t.} \quad & W(q) = W(q^*) \end{aligned} \tag{15}$$

which leads to our first main result:

Proposition 1. *The optimal Target-Consistent mechanism $\overline{q}$ is unique and such that*

$$\frac{\sigma_i}{\sigma_j} = \frac{\partial W(\overline{q})/\partial q_i}{\partial W(\overline{q})/\partial q_j} \;\forall i, j \leq R. \tag{16}$$

Proof. Uniqueness follows from the concavity of the Lagrangian, guaranteed by the strict concavity of $W(q)$ as E is linear in q. Let λ denote the Lagrange multiplier associated with the constraint in the above program. The first order conditions are

$$\sigma_r = \lambda \frac{\partial W}{\partial q_r}(\overline{q})$$

for all r, which implies condition (16). $\square$

This result shows that the target-consistent mechanism equalizes the ratio of carbon intensities (σ_r) between regions with the ratio of their marginal contributions to global welfare from a change in production quantities.

This result is illustrated in Figure 4.1 for the two-country case. The axes represent emissions of country i and j. Status-quo emissions $E^* = \sigma_i q_i^* + \sigma_j q_j^*$ are depicted by the black diagonal line. The

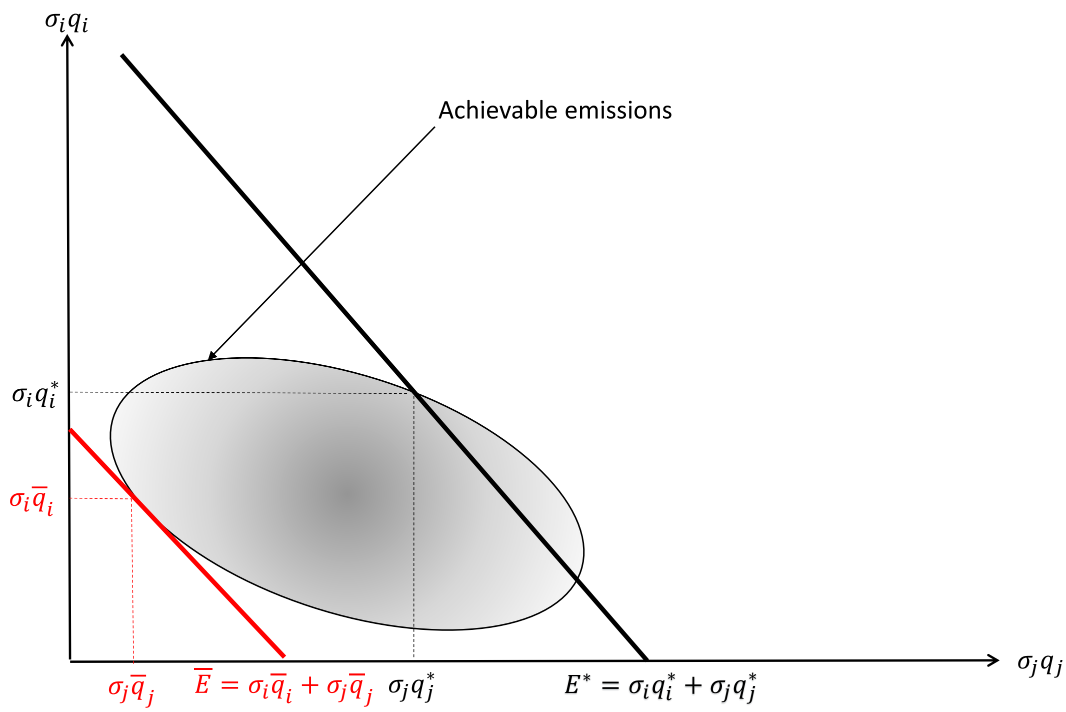

Figure 4.1: Graphical representation of the target-consistent allocation

set of emissions that are achievable via a target-consistent mechanism are represented by the grey area. The concavity of global welfare ensures that this set is convex. We observe that the resulting optimal emission level $\bar{E}$ is the lowest possible along the iso-welfare curve associated with the non-cooperative equilibrium quantities. The optimal target-consistent mechanism is $(\overline{q_i}, \overline{q_j})$.

4.4 Properties of the Optimal Mechanism

An important implication of Proposition 1 is that at the target-consistent optimum, production is higher in regions where carbon intensity is lower.

Proposition 2. *At the target-consistent optimum, the productivity-adjusted level of production (q_r/A_r) is higher in regions with lower carbon intensity.*

Proof. Using the definition of welfare in (5):

$$W = \sum_r W_r = \sum_r \left\{ N_r \left[\beta_r u(c_r) + (1-\beta_r) u'(c_r)\, c_r \right] - A_r C\left(\frac{q_r}{A_r}\right) - N_r \beta_r \alpha_r E \right\}, \tag{17}$$

as by definition $\sum_r q_r = \sum_r N_r c_r$, with c_r determined by optimal consumer behavior at equilibrium price. Therefore, as $cu''(c) = -\epsilon u'(c)$:

$$\frac{\partial W}{\partial q_i} = \sum_r \left\{ N_r u'(c_r)\left(1-(1-\beta_r)\epsilon\right)\frac{\partial c_r}{\partial q_i} \right\} - C'\left(\frac{q_i}{A_i}\right) - \sum_r N_r \alpha_r \beta_r \sigma_i \tag{18}$$

$$= p - C'\left(\frac{q_i}{A_i}\right) - \sum_r N_r \alpha_r \beta_r \sigma_i \tag{19}$$

through consumer optimal behavior (i.e. the left hand side of equation (6)) and as $\sum_r N_r \frac{\partial c_r}{\partial q_i} = 1$ at market equilibrium. Equation (16) then becomes:

$$\frac{\sigma_i}{\sigma_j} = \frac{p - C'\left(\frac{q_i}{A_i}\right) - \left(\sum_r N_r \alpha_r \beta_r\right)\sigma_i}{p - C'\left(\frac{q_j}{A_j}\right) - \left(\sum_r N_r \alpha_r \beta_r\right)\sigma_j} \quad \forall i,j \ \in \{1,...,R\}^2 \tag{20}$$

$$\Leftrightarrow \frac{\sigma_i}{\sigma_j} = \frac{p - C'\left(\frac{q_i}{A_i}\right)}{p - C'\left(\frac{q_j}{A_j}\right)} \quad \forall i,j \ \in \{1,...,R\}^2. \tag{21}$$

The result follows from the convexity of $C(.)$. □

Proposition 2 shows that, in order to meet the environmental target, the mechanism reallocates production to countries with a lower intensity of carbon and / or higher productivity (that is, lower marginal cost). This entails a redirection of trade relative to the noncooperative equilibrium. Moreover, combining this result with optimal firm behavior implies that optimal targets-consistent quantities can be achieved with a carbon tax constant across all regions r.

Proposition 3. *The optimal quantities of the target-consistent mechanism can be implemented through a uniform carbon tax: $T_r = T \ \forall r$, higher than the tax that maximizes global welfare ($T > \sum_r N_r \alpha_r \beta_r$).*

Proof. Using (19), the first-order condition of program (15) with respect to q_i yields:

$$\sigma_i - \lambda\left[p - C'\left(\frac{q_i}{A_i}\right) - \left(\sum_r N_r \alpha_r \beta_r\right)\sigma_i\right] = 0 \tag{22}$$

for all i, where λ is the Lagrange multiplier associated with the global welfare constraint. This can be rearranged as:

$$p = C'\left(\frac{q_i}{A_i}\right) + \left(\frac{1}{\lambda} + \sum_r N_r \alpha_r \beta_r\right)\sigma_i \tag{23}$$

implying that the allocation can be implemented with a carbon tax $T_r = \frac{1}{\lambda} + \sum_r N_r \alpha_r \beta_r = T, \forall i$, common to all regions. Since $\lambda > 0$ (due to the initial inequality constraint), this implies $T > \sum_r N_r \alpha_r \beta_r$ which would be the tax maximizing $W = \sum_r W_r$. □

Proposition 3 highlights that the target-consistent mechanism can be implemented by setting a uniform carbon tax that internalizes the global externality from emissions. The optimal carbon tax is the highest that preserves global welfare at its status quo level. Thus, stronger environmental preferences or greater consumer political power eases participation in the mechanism.

In line with Lemma 1, Proposition 3 also shows that the target-consistent carbon tax exceeds the level that maximizes current global welfare, since by construction the mechanism minimizes emissions subject to maintaining global welfare. In fact, as shown in the next result, it even preserves each regional welfare level:

Proposition 4. *The optimal target-consistent mechanism implements a welfare allocation (with transfers and taxes) that exactly preserves the original welfare distribution:* $W_r(\overline{q_r}, \overline{q_{-r}}) + \ m_r = W_r(q_r^*, q_{-r}^*), \ r \in \{R\}$.

Proof. Suppose, by contradiction, that for some region i: $W_i(\overline{q_i}, \overline{q_{-i}}) + \ m_i > W_i(q_i^*, q_{-i}^*)$. From Lemma 2 and the budget-balance, this implies that for some $j \neq i$: $W_j(\overline{q_j}, \overline{q_{-j}}) + \ m_j < W_j(q_j^*, q_{-j}^*)$ violating the participation constraint for region j. Thus, the only possible allocation must satisfy: $W_r(\overline{q_r}, \overline{q_{-r}}) + \ m_r = W_r(q_r^*, q_{-r}^*), \ \forall r$. □

The fact that the target-consistent mechanism preserves the original welfare inequalities (with tax and transfers) has an important implication: the ex-post distribution of welfare without transfers (i.e. the distribution of $W_i(\overline{q_i}, \overline{q_{-i}})$) is heavily skewed in favor of low-carbon-intensity regions. As shown in Proposition 2, production is reallocated toward regions with lower σ, concentrating economic activity—and thus welfare—there. This shift is essential to minimize global emissions, as production becomes concentrated in low-emitting countries. However, to maintain participation from high-emitting regions, monetary transfers are critical to offset losses in production and changes in trade.

This has important implications for the political economy of international environmental agreements. High-emitting regions, especially developing countries, may resist strict climate targets unless they are provided with assurances of financial support from wealthier, lower-carbon-intensity regions. These transfers are essential for sustaining their welfare while aligning with global emission goals. This persistent challenge in international climate negotiations, as evidenced in recent COP discussions, reflects the need for mechanisms of compensation and redistribution. Although a full political economy analysis is beyond the scope of this paper, our model provides a clear economic rationale for these demands. The calibrated version of the model (next section) will demonstrate that the magnitude of required transfers is currently quite large.

4.5 Self-Enforcement and Stability

A natural concern is that some countries may exhibit political resistance to explicit cross-country transfers—for instance, an intrinsic aversion to transferring resources to geopolitical rivals. In our framework, such considerations can be incorporated in a simple extension by introducing transfer-specific utility costs or by imposing caps on net transfers. Formally, this amounts to tightening the participation constraints. The structure of the mechanism is unchanged: the participation condition continues to discipline the implementable allocation, but the set of emissions targets that can be sustained shrinks. The voluntary frontier should therefore be interpreted as the outer envelope of feasible cooperation under voluntary participation absent transfer frictions; additional political reluctance can only move outcomes inward. In the empirical section 6.4, we quantify the effect of unilateral deviations from the mechanism and interpret them as capturing, among other factors, implicit political costs or reluctance toward international redistribution.

Finally, it is worth noting that a region may have incentives to deviate from the mechanism if all others comply: it may then prefer not receiving m_r and not producing $\overline{q_r}$. This will be the case if for some $q_r \neq \overline{q_r}$:

$$W_r(q_r, \overline{q_{-r}}) \geq W_r(\overline{q_r}, \overline{q_{-r}}) + m_r \tag{24}$$

The literature (see Martimort and Sand-Zantman, 2016; Gersbach et al., 2021) addresses this by designing mechanisms in which regions pay upfront m'_r and receive subsidies $s_r \sigma_r (q^*_r - q_r)$ conditional on emission reduction. This ensures compliance through conditional incentives. While we abstract from these timing issues for expositional simplicity, such an implementation can be easily incorporated into our framework.[3] Moreover, consider the following trigger strategy: countries implement the target-consistent allocation as long as all comply, and revert to the non-cooperative status-quo if any country deviates. Reversion is not mechanically embedded in the mechanism but constitutes a strategic choice of the non-deviators. Assuming that the status-quo allocation constitutes an equilibrium of the underlying policy game, reverting to it is sequentially rational. Given that participation constraints bind with equality (equation 13), a deviating country cannot obtain a strictly higher payoff than under reversion, as the mechanism preserves each country's welfare at its status-quo level. Hence, under this credible reversion strategy, the allocation is self-enforcing. In the calibration analysis, we study the stability of the mechanism by considering unilateral and coalition deviations, evaluating how such departures affect the attainable carbon price, transfers, and overall ambition of the agreement (see section 6.4).

4.6 Equivalence with a Cap–and–Trade implementation.

The uniform carbon price delivered by the mechanism can equivalently be implemented through a cap–and–trade system. Under a frictionless and competitive permit market, setting the aggregate emissions cap equal to $E(\overline{q}) = \sum_r \sigma_r \overline{q}_r$ yields the same carbon price and emissions allocation as the tax regime. In a cap–and–trade system, each country r chooses its production q_r to satisfy $p = C'\left(\frac{q_r}{A_r}\right) - \theta\sigma_r$ (the equivalent of equation 6)), where θ denotes the permit price. Under the tax, the same condition holds with $\theta = T$. Because the aggregate emissions schedule $E(q^*(\theta))$ is continuous and decreasing, market clearing under the cap $E = E(\overline{q})$ implies $\theta = T$. Hence, the cap–and–trade system exactly reproduces the emissions outcomes of the tax mechanism, consistent with the classic price–quantity equivalence result of Weitzman (1974).

Two practical implementations replicate the mechanism's transfers. First, all permits can be auctioned at price $\theta = T$; total auction revenue is $T \cdot E(\overline{q})$. Redistributing this revenue according to the

[3] For example, our framework can be adapted to the mechanism in Martimort and Sand-Zantman (2016), by setting $s_r = T - \tau_r$ and $m'_r = m_r + s_r \sigma_r (q^*_r - q_r)$.

mechanism's transfers m_r (with $\sum_r m_r = 0$) reproduces each country's net monetary position and thus the same welfare and participation outcomes. Second, if permits are freely allocated in quantities ϕ_r (with $\sum_r \phi_r = E(\overline{q})$), the net monetary flow paid–or received– by country r from permit trading is $T(\sigma_r \overline{q}_r - \phi_r)$. To reproduce the mechanism's transfer vector m_r (defined as positive for net contributor), the corresponding initial allocation satisfies

$$m_r = T\big(\sigma_r \overline{q}_r - \phi_r\big) \quad \Longleftrightarrow \quad \phi_r = \sigma_r \overline{q}_r - \frac{m_r}{T}. \tag{25}$$

In the calibration that follows, we reconstruct the initial permit allocation consistent with the transfers implied by the tax implementation of our mechanism (see section 6.3).

4.7 Extension to a dynamic setting

The mechanism extends naturally to a dynamic environment in which regions make decisions over an infinite horizon and the principal minimizes cumulated global emissions. The full dynamic program and its solution are developed in Appendix A.3. Here we summarize the structure and main implications.

Let $W_{r,t}$ denote region r's welfare in period t and ρ the discount rate. In this intertemporal setting, the principal minimizes cumulated global emissions $E = \sum_{t,r} q_{t,r}\sigma_{t,r}$. Assuming that it can credibly propose a take-it or leave-it mechanism, the optimal target-consistent one solves:

$$\min \sum_t \sum_r q_{rt}\, \sigma_{rt}, \tag{26}$$

subject to each region's discounted participation constraint

$$\sum_t \frac{1}{(1+\rho)^t} \left[W_{rt}(q_{rt}, q_{-rt}) + m_{rt}\right] \geq \sum_t \frac{1}{(1+\rho)^t} W_{rt}(q^*_{rt}, q^*_{-rt}) \;\forall r \tag{27}$$

and the discounted budget balance condition. Eliminating transfers as in the static case, the Lagrangian first-order condition for emissions in period t yields (Appendix A.3):

$$\sigma_{r,t} = \lambda(1+\rho)^{-t} \frac{\partial W_t}{\partial q_{rt}}(q_t), \tag{28}$$

Equation (28) shows that the optimal allocation in each period solves a problem formally identical to the static one, with a time-varying multiplier $\lambda(1+\rho)^{-t}$ reflecting discounting. As in the static environment, this condition is equivalent to implementing a *period-specific uniform carbon tax*:

$$T_t \;=\; \frac{(1+\rho)^t}{\lambda} + \sum_r N_{rt}\alpha_{rt}\beta_{rt}, \tag{29}$$

with the corresponding transfers $(m_{r,t})$ chosen to satisfy the discounted participation constraints. Thus, the dynamic mechanism delivers a sequence $\{T_t\}_{t\geq 0}$ that decentralizes the efficient allocation period-by-period while preserving each region's discounted welfare exactly.

In essence, the dynamic extension mirrors the static mechanism period-by-period. At each date, production reallocates toward regions with lower marginal pollution intensities or costs, exactly as in the static model. The sequence of transfers is constructed so that each region's discounted welfare is preserved, ensuring intertemporal participation and making the mechanism welfare-equivalent over time. Finally, discounting and technological change shape the time profile of the optimal tax $\{T_t\}_{t\geq 0}$ and transfers, with higher discounting placing more weight on current welfare and altering the dynamic path of both emissions and payments.

5 Data Sources and Calibration

5.1 Data Sources

Main macroeconomic variables. The main macroeconomic variables used in the model are sourced from the latest version of the Penn World Tables (PWT) for the year 2018. We rely on population data, real GDP on the output side, real household consumption, government consumption and capital stock at purchasing power parities (PPPs) in 2017 US dollars.

Carbon prices and consumption taxes. Carbon price data are drawn from the companion dataset to the *Effective Carbon Rates* (ECR) report published by the OECD. The ECR data set measures carbon pricing on CO_2 emissions from energy use in more than 70 countries, including 45 OECD and G20 nations, accounting for more than 80% of global emissions. The data set incorporates fuel excise taxes, carbon taxes, tradable emission permit prices, and subtracts fuel subsidies, enabling the calculation of the "net carbon price" for each country in 2018. To account for varying coverage, net carbon prices are weighted by the percentage of emissions covered, using data from the OECD's "Country-level Share of GHG Emissions Priced by ECR component". The average net carbon price across countries is \$30.7 in 2018, ranging from a minimum of -\$4.3 to a maximum of \$107. In particular, China's carbon price is \$2.3 and the United States is \$5.3 (including adjustments for fuel subsidies).

For consumption taxes, we use the standard VAT rates per country sourced from the International Monetary Fund (for federal countries with multiple VAT values, like the US, we selected the average VAT rate).

Emissions, carbon intensities and nationally determined contributions. CO_2 emission data are obtained from the EDGAR (Emissions Database for Global Atmospheric Research) Community GHG Database (`https://edgar.jrc.ec.europa.eu/report_2024`). Emissions are reported in megatons of CO_2. Carbon intensities (σ) are calculated by dividing CO_2 emissions by GDP from the PWT (in 2017 US dollars) for each country. Nationally Determined Contributions (NDC) are the carbon target voluntarily pledged by countries following the COP 21 in Paris. We extracted these pledged targets (usually for 2030) from the UNFCCC "Nationally Determined Contributions Registry", retaining only the targets that clearly established a percentage reduction from a baseline emission-year.

Social Costs of Carbon. To calibrate national environmental preferences, we rely on country-specific SCC estimates from Ricke et al. (2018), who develop an empirical framework to allocate global SCC between countries based on projected local climate damage. Their model combines probabilistic damage functions with global climate simulations and economic projections to estimate the marginal economic cost of an additional ton of CO_2 emissions, attributed to each emitting country.

We retain the socioeconomic pathway 2 (SSP2) —- a "middle of the road" scenario that reflects moderate population growth and economic development and its closest corresponding climate scenario (RCP6.0), and with the central specification of Burke - Hsiang - Miguel (BHM) damage function (average short-term / long-term, without income differentiation) as well as a growth adjusted discount rate.[4] Under this scenario, the SCC values vary substantially between countries, reflecting differences in exposure, vulnerability, and adaptive capacity to climate change (see A.1). We use these SCC estimates as upper bounds for the calibration of environmental concern parameters (see below) under the assumption that no country would be willing to implement non-cooperatively a domestic carbon tax exceeding its own projected damages. The sum of SCCs in this scenario for our 69 countries is $277, close to the average estimates in the existing climate literature.

5.2 Calibration and Stylized Facts

Our final data set includes 69 countries for 2018, representing 80% of total global emissions and 83% of world GDP. We calibrate the model at baseline (the "Status quo" scenario in our setup) before computing the tax rates proposed by our mechanism and analyzing the resulting emissions, production, and wealth transfers.

Elasticities and scaling parameters. We use values from the literature to set ϵ and η, the elasticities of the utility and cost functions. Specifically, we select $\epsilon = 0.9$ and $\eta = 1$, consistent with previous studies (e.g. Desmet et al. (2018)). We use the two scaling parameters h and k (in utility and cost functions) to fit the observed GDP and emissions, obtaining at baseline $h = 0.7$ and $k = 1260$.

Environmental concerns A key challenge in calibrating our model is that observed carbon taxes in many countries exceed what would be predicted by non-cooperative behavior under purely domestic environmental preferences. This suggests that these taxes may partly reflect cooperative motives or broader international alignment, rather than genuine internalization of national damages. To address this, we rely on country-specific estimates of the Social Cost of Carbon (SCC) as upper bounds on environmental preferences. Specifically, for countries where observed carbon taxes exceed their estimated SCC, we cap the "imputed" carbon tax at the SCC value, interpreting these higher taxes as emerging from international commitments or coordination rather than reflecting underlying domestic preferences. In our sample, 51 countries have an observed carbon tax above their SCC, and 18 countries have a carbon tax below their SCC (Figure 5.1 plots the difference between observed carbon tax and the SCC

[4]In a sensitivity analysis we run the simulation with the BHM cost function calibrated on the short-run damages (next 5 years, very myopic government) and long-run damages.

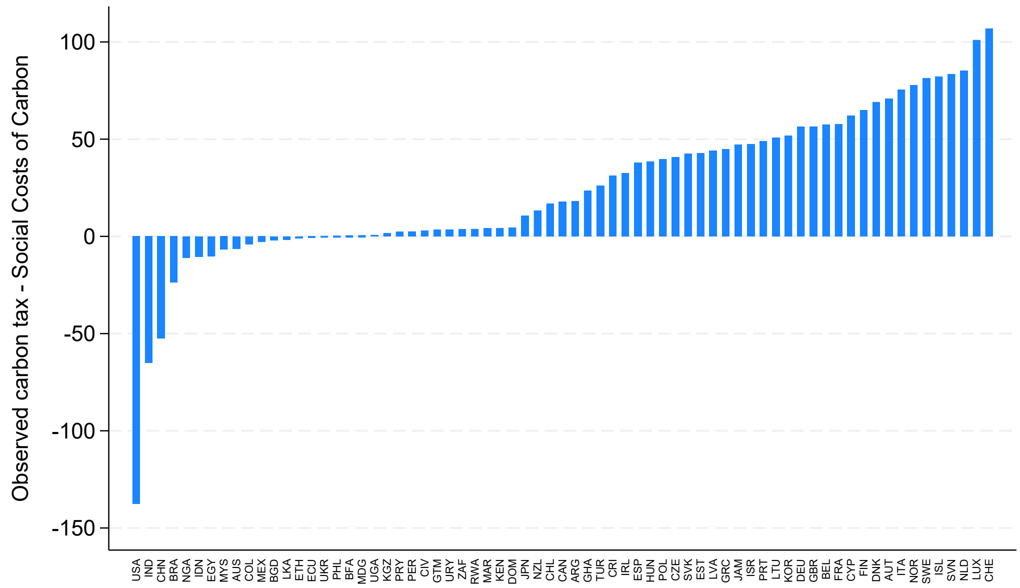

Figure 5.1: Observed carbon tax - Social Costs of Carbon

for each country). We then calibrate environmental concern parameters α_r using the non-cooperative formula directly produced by our model (see equation (10)): $\alpha_r = \frac{\tau_r^{\text{imputed}}}{N_r \beta_r}$ where τ_r^{imputed} is the imputed tax, equal to the observed tax if below the SCC of the country, and equal to the SCC otherwise. This approach yields more plausible values of α between countries. However, it relies on a strong identifying assumption: we impose that any observed carbon tax above a country-specific SCC must result from external pressures, such as participation in cooperative mechanisms (e.g., the EU ETS), thus ruling out the possibility of purely domestic "international altruism." Conversely, by calibrating α to match the SCC when the observed tax exceeds it, we also implicitly assume that environmental preferences in the absence of cooperation cannot be lower than the SCC-implied value. As a result, this method can underestimate or overestimate the true α, depending on the nature of the political motivation of each country. We view this as a pragmatic compromise: it allows us to discipline the calibration using credible bounds while maintaining consistency with observed tax behavior. This approach avoids attributing cooperative behavior to intrinsic domestic environmental preferences and ensures that participation constraints are evaluated relative to a politically credible baseline.

The distribution of estimated αs is given in Figure A.2 in the appendix. The median value of α is 0.076 with a minimum of -0.048 (Egypt) and a maximum of 0.769 (Israel). China and the United States have below median values of 0.001 and 0.016 while most EU countries have above median αs.

To validate our calibration of α, we use the Environmental Performance Index (2018) which ranks countries on 24 performance indicators.[5] Interestingly, figure 5.2 shows a strong positive correlation

[5]This index ranks 180 countries on 24 performance indicators in the following 10 issue categories: air quality, water and sanitation, heavy metals, biodiversity and habitat, forests, fisheries, climate and energy, air pollution, water resources, and agriculture. It is the result of a collaboration between the Yale Center for Environmental Law and Policy (YCELP),

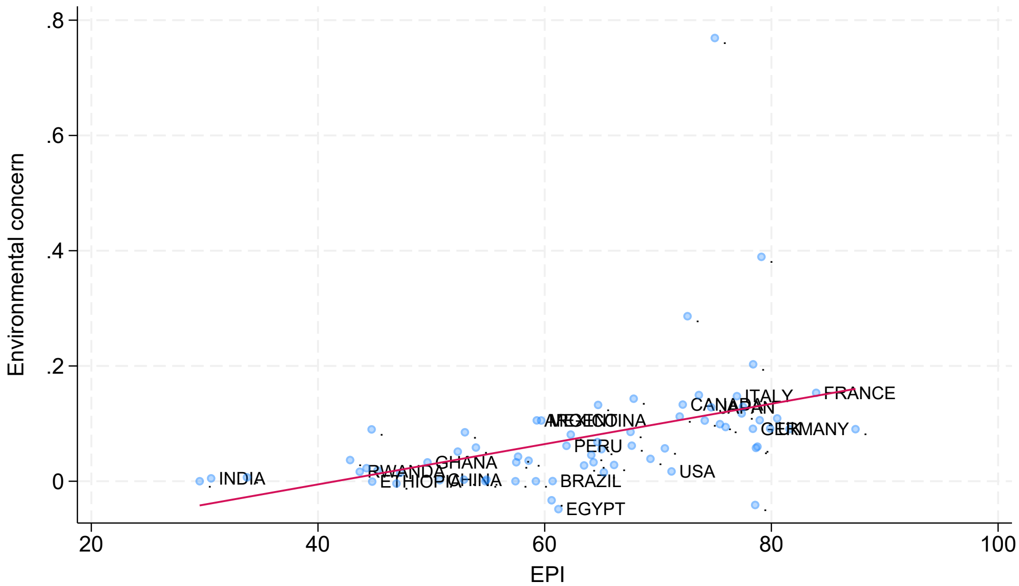

Figure 5.2: Correlation between estimated environmental concerns and the Environmental Performance Index (2018)

between our indicator and the EPI. We also use the number of participants sent to COP meetings as a proxy for environmental engagement (data from Blinova et al. (2024)). Figure A.4 shows a positive correlation between α values and per capita COP participation, indicating alignment between environmental concern and international climate negotiation engagement. Finally, we use our data on carbon targets pledged by each country following the COP 21. Figure A.5 shows a positive association between our measure of environmental concerns and the intensity of emissions reduction pledged for 2030.

Consumer weights. The political weights of consumers β are derived directly from the consumption taxes t_r implementing the status-quo quantities in equation (9). In particular, we have $\beta_r = 1 - \frac{t_r}{\epsilon(t_r+1)}$, where t_r is the standard VAT in country r. The distribution of estimated βs is given in figures A.3 in the appendix.

The set of $\{\alpha_r, \beta_r\}_r$ provides a revealed-political-preferences portrait of the world: it encapsulates how much each government is willing to tax for general revenue and how much it cares about climate mitigation, in a way that is disciplined by real data. We emphasize that α_r in our calibration is a political weight on climate rather than a geophysical measure of damage — it is scaled by political willingness. The use of country-specific SCC estimates ensures that this weight is interpreted within the limits of self-interested behavior: if α_r is high, it means the country is actively addressing climate damages up to (or close to) its own loss from carbon emissions; if α_r is low or zero, the country is largely ignoring even its domestic climate harm in policy. These calibrated weights will be central to our quantitative

Yale University, Columbia University Center for International Earth Science Information Network (CIESIN), and the World Economic Forum (WEF).

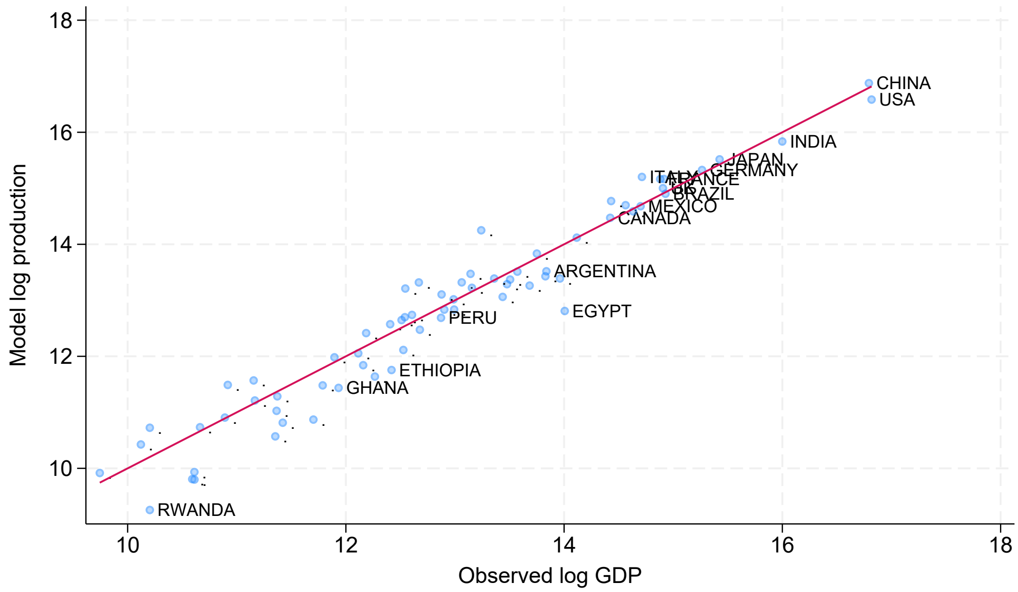

Figure 5.3: Correlation between observed GDP and calibrated production

exercises, as they determine the binding participation constraints in the global mechanism.

Production and consumption. We compute values for q_r and c_r by computing regional demand and supply at equilibrium prices obtained from equation (11).[6] Figures 5.3 and 5.4 demonstrate strong correlations between the constructed values of q_r and c_r (in logs) and their empirical counterparts — GDP and consumption in logs— as reported in the 2018 Penn World Tables. Importantly, consumption was not a targeted moment during calibration.

CO_2 emissions. CO_2 emissions for each country are computed using $E_r = q_r \cdot \sigma_r$. Figure 5.5 shows a strong correlation between our constructed emissions and observed 2018 emissions (in log). Total predicted emissions for the 69 countries in the model are at 30278 megatons, close to the observed emissions of 29865 megatons.

Baseline welfare. Welfare is computed using equation (5) and displayed in figure A.6. As expected from our theoretical framework, Figures A.7 and A.8 illustrate strong positive associations between welfare and observed emissions and GDP.

Summary of data inputs and calibration. As summarized in table 1 in appendix, our quantitative analysis combines a small set of country-level observables for 2018: population, GDP, consumption and capital stock from the Penn World Table; CO_2 emissions from EDGAR (used to construct carbon

[6]The parameter "A_r" is proxied by the capital stock (in 2017 million US dollars) for each country as measured in the Penn World Table in 2018.

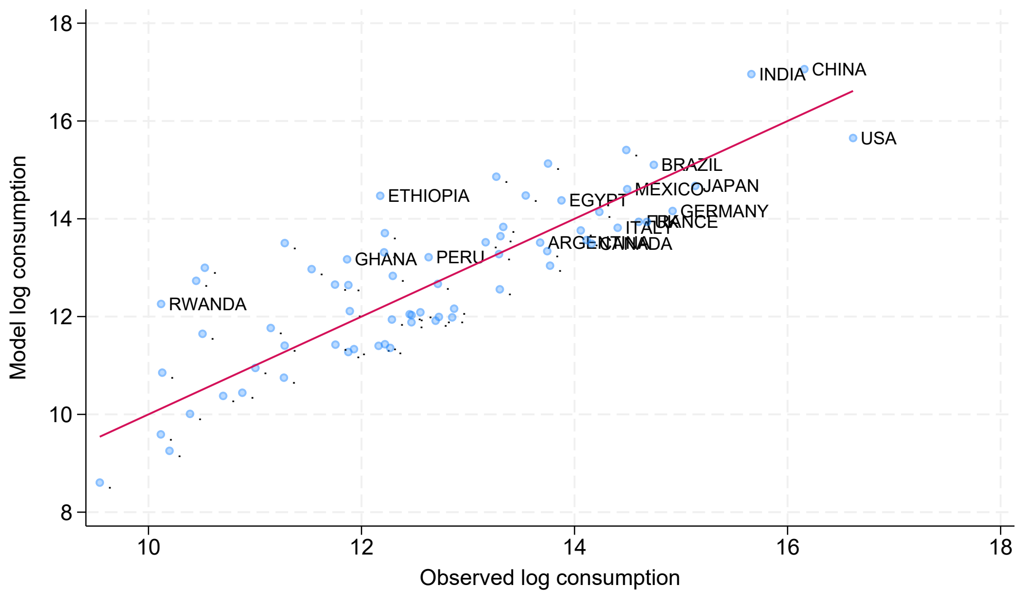

Figure 5.4: Correlation between observed and calibrated consumption

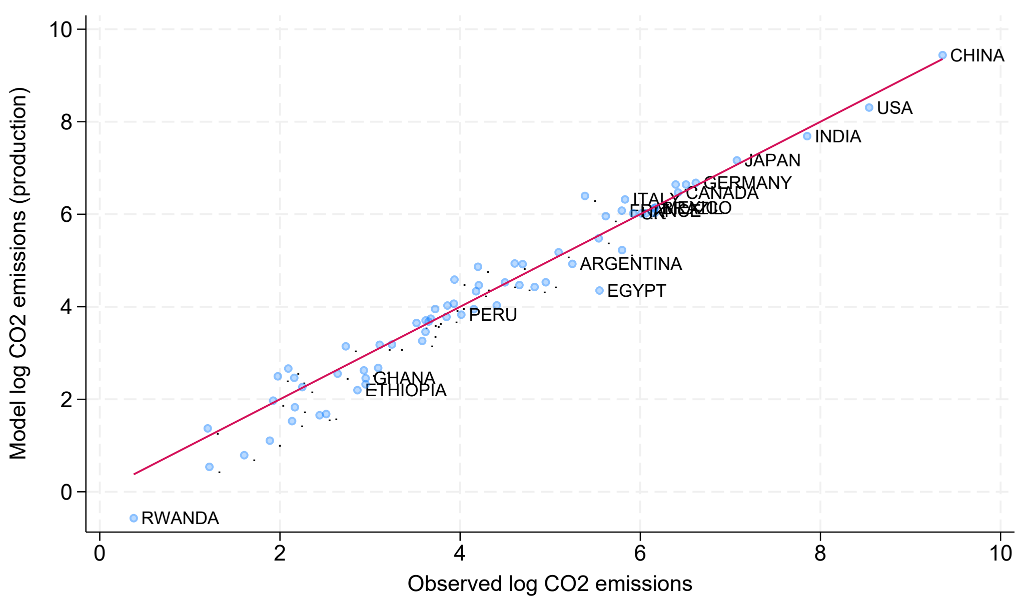

Figure 5.5: Correlation between observed and calibrated CO2 emissions

intensity σ_r); effective carbon prices from the OECD Effective Carbon Rates dataset; and standard VAT rates from the IMF. We additionally use country-specific social costs of carbon from Ricke et al. (2018) as discipline on non-cooperative climate policy, capping the imputed domestic carbon price when observed effective rates exceed SCC. The calibration is deliberately parsimonious. Two global elasticities (ε, η) are taken from the literature, and two scale parameters (h, k) are chosen to match baseline magnitudes. All remaining heterogeneity is identified from policy wedges: the consumer political weight β_r is backed out from VAT rates using equation 9, while environmental concern α_r is recovered from the (capped) carbon price using equation (10). Given these inputs, equilibrium quantities and prices $(p^*, \{q_r, c_r\})$ are recovered from the model's first-order conditions and the market-clearing condition (7)–(11), and counterfactual emissions, welfare, and transfers follow mechanically from the mechanism. Interpreted this way, the calibration recovers a country-level map of politically relevant heterogeneity for climate cooperation: differences in carbon intensity, productivity, tax distortions, and effective climate concern. These are the key objects for the question studied here, namely how far voluntary cooperation can go once participation must be preserved country by country.

6 Main results of the Calibration: Carbon Tax, Welfare and Transfers

6.1 Carbon Tax

We now calculate the carbon tax required to implement the production and emissions reductions under the mechanism, focusing on the most ambitious outcome consistent with voluntary participation by all countries. Following Proposition 3, our baseline calibration indicates that the uniform tax required to reach this participation-constrained optimum is \$252 per ton of CO_2 – very close to the global SCC in 2018 equal to our calculation of \$277. This tax results in a drastic 35% reduction in global emissions, from 30,278 to 19,564 megatons of CO_2. This is above the estimated 28% cut required by 2030–2035 to align with the target of 2 ° C, as described in the most recent UNEP evaluations (UNEP, 2024). This reduction in emissions results from both a decrease in global consumption (and production) of 20 % and a reallocation of production to countries with a lower intensity of carbon, in line with Proposition 2. The reduction combines a scale effect and a composition effect: global output declines while production shifts toward cleaner locations. Framed this way, the mechanism operates through a familiar trade-and-environment logic, here at the cross-country level.

To limit the tax value and the implied emission reductions, we run the same simulation using different SCC values underlying the environmental preference parameters α. In particular, we use the SCC values of Ricke et al. (2018) calibrated with a damage function focusing on short-term damage (damages occurring in the next 5 years, representing a myopic government) and accounting for long-term damages (forward-looking citizens). The tax corresponding to low SCC (short-term damages) is \$53, associated with a reduction in emissions of only 7.5% from baseline. The tax corresponding to the optimistic scenario of forward-looking governments accounting for all long-term discounted costs of climate change is quite high (\$563) associated with a drastic reduction in emissions, close to 59% from baseline.

The map 6.1 displays the large reallocation of emissions throughout the globe implied by the carbon

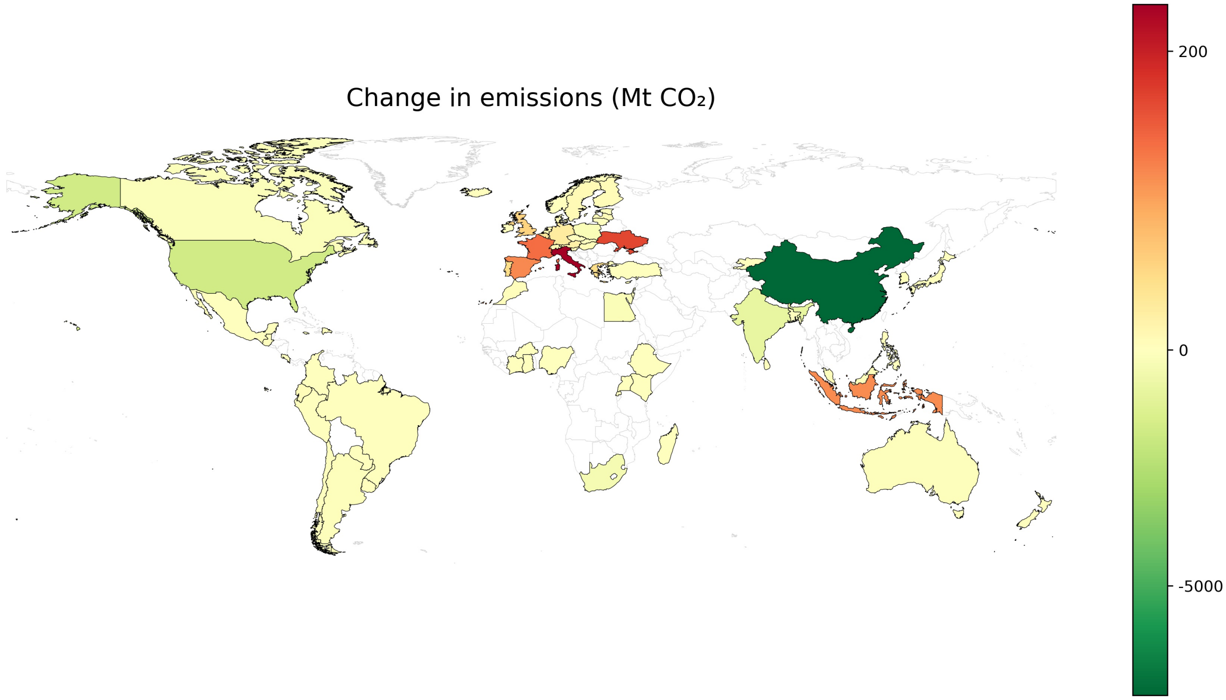

Figure 6.1: Reallocation of emissions implied by the target-consistent carbon tax

tax mechanism in our baseline scenario. In line with Proposition 2, the map shows that emissions—and therefore production— are mostly reallocated toward countries with lower carbon intensity (Figure A.9 in the appendix displays the correlation between carbon intensity and the change in production implied by the mechanism). This spatial reallocation of production has important welfare implications that we will study in more detail.

6.2 Welfare, inequalities and transfers.

The imposition of a large, target-consistent carbon tax not only reduces global emissions, it also induces a pronounced reallocation of production and emissions across countries, with important economic and political consequences. Under our mechanism, the combination of higher carbon costs and comparative-advantage adjustments shifts output away from large, emission-intensive producers, with low baseline carbon tax (notably China and India) toward high-income, low-emission or low-σ_r countries with higher productivity. This reallocation is visible in the welfare distribution before compensatory transfers are applied: the ranked numbers in Figure 6.2 show the magnitude of the divergence in pretransfer welfare (before transfers and after the tax) for each country. Notice that the sum of all these welfare differences amounts to 0 by construction and corresponds to the opposite of transfers received through the mechanism. The map 6.3 reports the spatial distribution of the welfare effects per capita (before transfer). High-income countries – particularly the United States and many EU members — tend to gain because they both receive greater benefits from reduced global emissions (e.g., EU countries, as they have large

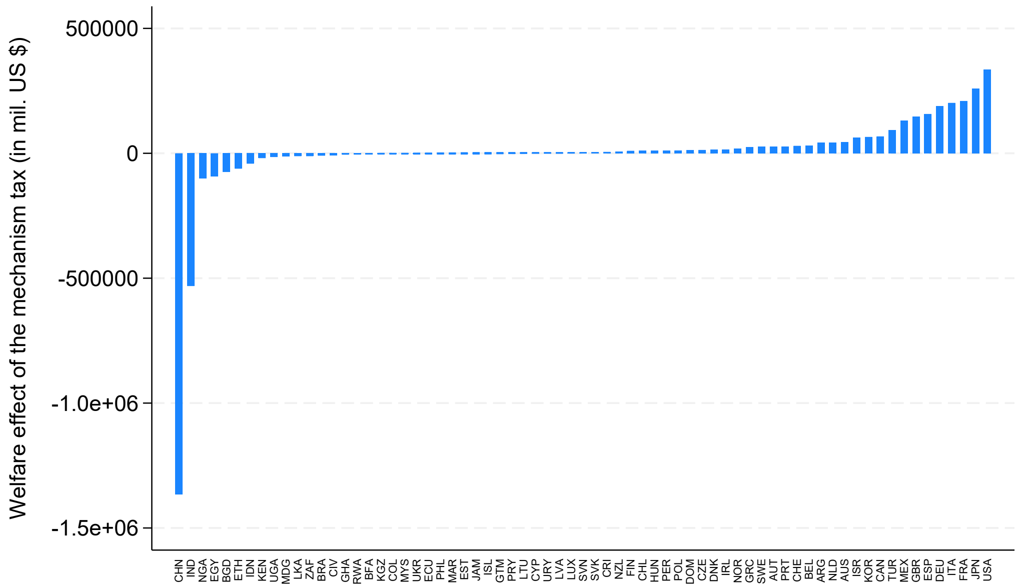

Figure 6.2: Welfare effect of the target-consistent carbon tax

α_r) and capture trade gains as production relocates to low-σ_r, high-productivity locations (e.g., USA).[7] In contrast, large developing economies such as China and India suffer the largest welfare losses, driven primarily by declines in domestic production, adverse terms-of-trade movements, and the relocation of emission-intensive industries.

These patterns also have important distributional and geopolitical implications. Large reallocations of industrial activity and carbon footprints may exacerbate tensions between groups of countries: affected producers may view the policy as a competitive shock to employment and export revenues, while beneficiaries may face domestic pressure to protect newly gained industries or resist large compensatory transfers. Such dynamics raise the risk of trade frictions, demands for compensation, or retaliatory measures (e.g., tariffs or border carbon adjustments), complicating international cooperation on climate policy. While the carbon tax mechanism can make ambitious mitigation technically feasible, its politically salient redistributive impacts reinforce the need for carefully designed compensatory transfers. These pressures are particularly strong in economies where carbon-intensive production is closely tied to export revenues or public finances.

To satisfy every country's participation constraint, huge transfers are required to compensate losers (the opposite of the welfare amounts shown in Figure 6.2). In our calibration, the total volume of monetary transfers needed to implement the 35% emission reduction is on the order of $2.4 trillion (about 3% of world GDP). For instance, the U.S. would need to transfer $335 billion—roughly 5 times the value of its total 2023 foreign aid budget, 1.6% of its GDP.[8] Figure 6.4 shows that for some donor

[7] Although US production (and therefore emissions) decreases under the mechanism, US welfare rises due to terms-of-trade effects: global production also declines, allowing the US to benefit from its relatively low carbon intensity and high production efficiency.

[8] In 2023, the United States had a $68.2 billion foreign aid budget, among which $14.5 billion were spent on humanitarian

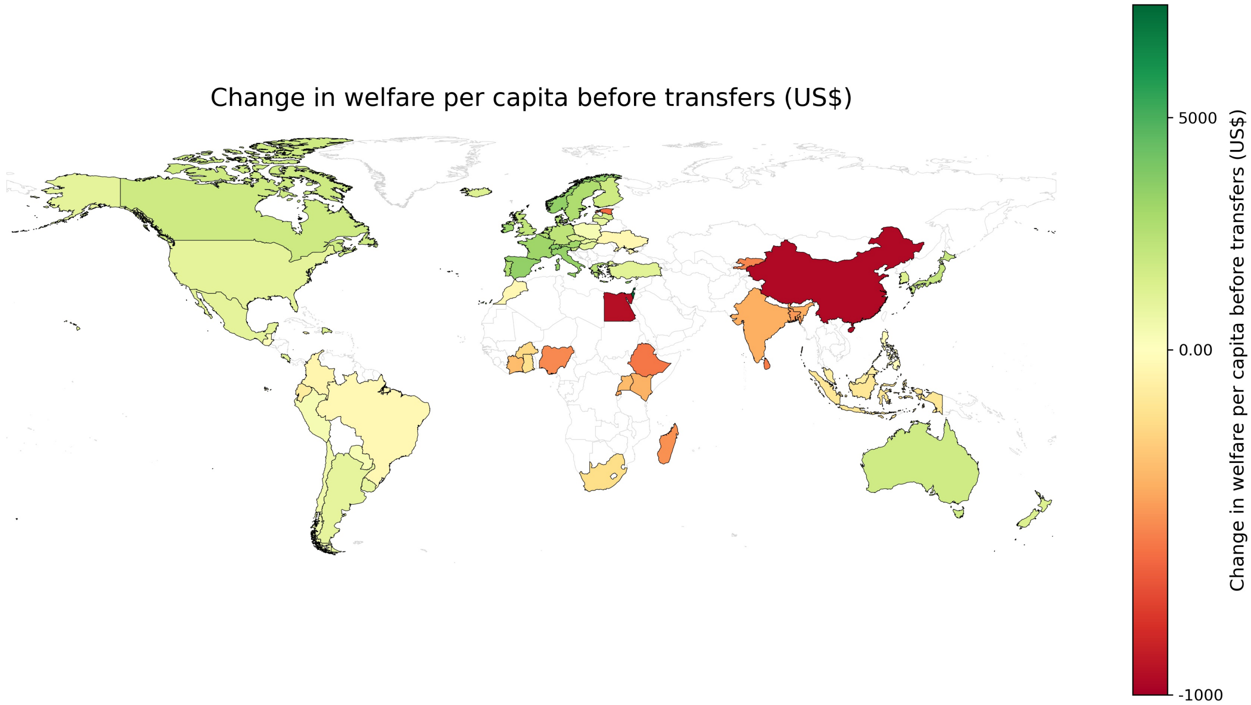

Figure 6.3: Per-capita welfare effect of the target-consistent carbon tax

countries (e.g. Italy or France) the required payments approach 5–8% of GDP, and are even higher for a few smaller high-income economies. On the recipient side, many lower-income countries would receive transfers amounting to a very large share of their GDP (exceeding 15% for Rwanda, for example). The implied redistribution is massive and far exceeds existing international fiscal cooperation. These magnitudes illustrate why securing a voluntary agreement at the optimal target would be so difficult: the necessary side-payments dwarf existing international funding efforts and would likely face serious fairness, reciprocity, and credibility concerns in practice.

6.3 Cap-and-Trade implementation

Using the calibrated outcomes of the tax implementation, we recover the equivalent cap–and–trade allocation implied by our mechanism. Recall from section 4.6 that under the assumption of a frictionless and competitive permit market, the equilibrium permit price θ equals the optimal tax T, ensuring that both systems yield identical incentives and equilibrium emissions. For each country r, the initial allocation of permits ϕ_r that reproduces the transfers m_r under the tax mechanism is given by

$$\phi_r = \sigma_r \overline{q}_r - \frac{m_r}{T}, \tag{30}$$

where $\overline{q}_r$ denotes country r's equilibrium production under the tax mechanism, m_r are the transfers

aid. Source: `https://usafacts.org`.

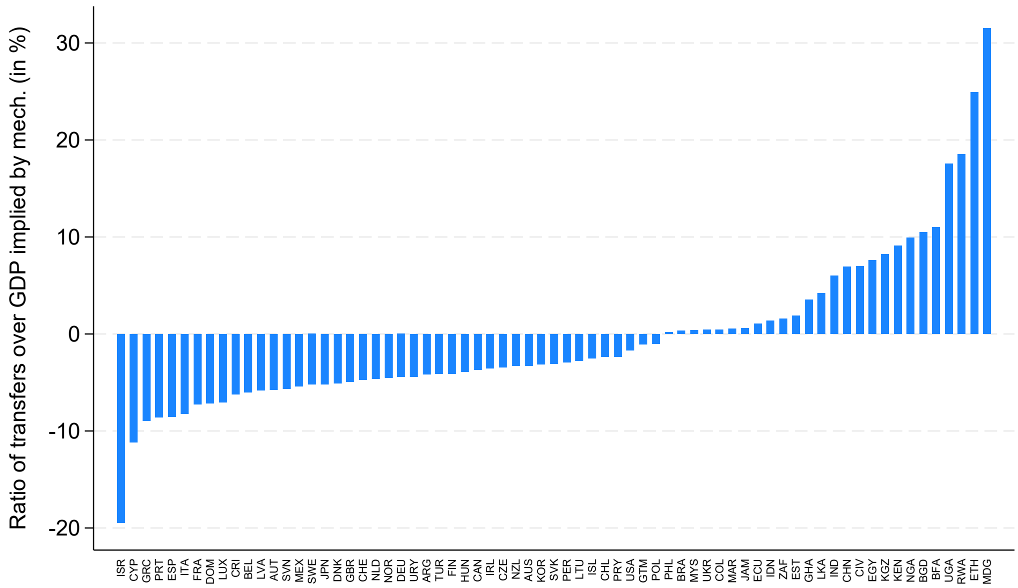

Figure 6.4: Transfers implied by the mechanism over GDP per country

that we identified in the previous section (the opposite of welfare differences in figure 6.2) and T is the optimal tax of \$252. By construction, this allocation yields the same net monetary position for each country and thus preserves the welfare distribution achieved under the tax regime. The two implementations are therefore welfare-equivalent: they generate identical emissions, prices, and welfare outcomes, differing only in the institutional form of implementation.

We apply this formula to our calibrated results to derive the implied initial allocation of permits across countries, denoted by $\{\phi_r\}$. Figure 6.5 reports these allocations, illustrating how the mechanism's transfer structure translates into permit holdings under an equivalent, welfare-preserving cap–and–trade system. By construction, the sum of all ϕ_r equals 19,564 $\mathrm{MtCO_2}$, corresponding to the total emissions cap. A salient observation is that China would receive almost half of the total cap to ensure its participation in the scheme, followed by India and the United States. Together, these three countries account for approximately 82% of the initial quotas, while several others—mostly European countries—start with negative allocations. This quantity-based representation offers a stark visualization of the requirements imposed by the voluntary participation constraints.

6.4 Effect of Non-Compliance.

Although the mechanism is self-enforcing under a credible reversion strategy (see Section 4.5), its implementation may be constrained by additional political frictions. Governments may face domestic resistance to large cross-border transfers or geopolitical reluctance to commit to sizable redistribution. We interpret unilateral or coalition deviations from the voluntary frontier as capturing such implicit political costs of participation. Formally, these frictions can be represented as tighter participation con-

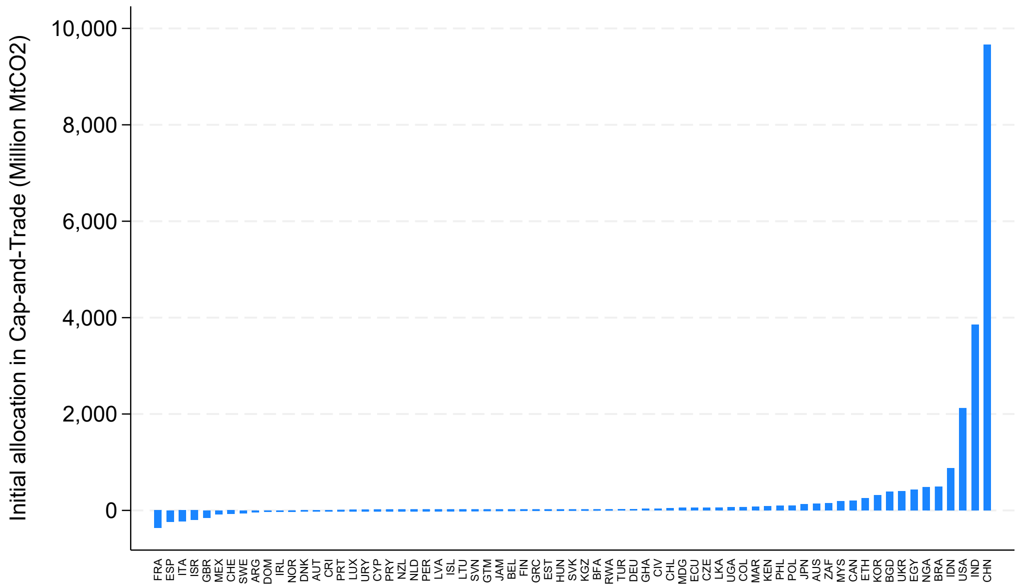

Figure 6.5: Cap-and-trade initial allocation (in MtCo2)

straints or caps on net transfers, shrinking the set of feasible allocations. The voluntary frontier should therefore be understood as the outer envelope of self-enforcing cooperation absent transfer frictions; additional political constraints move outcomes inward and reduce the maximum implementable carbon price.

As discussed in Section 4, incentive compatibility relies on a credible reversion to the non-cooperative status quo following deviation. Reversion is sequentially rational when the non-cooperative allocation is itself an equilibrium of the underlying policy game. If countries anticipate that such reversion would not be fully enforced in practice, those bearing high participation costs may have a stronger incentive to defect. Deviations from the benchmark allocation can thus be interpreted as reflecting either weakened enforcement expectations or the political constraints discussed above.

Figure 6.6 reports the unilateral incentive to deviate for each country. Countries predicted to gain most from free-riding on others' abatement efforts are typically high-productivity and/or low-α_r economies—such as the United States and certain EU members—which would prefer others to reduce emissions while facing lower domestic abatement costs themselves. By contrast, countries such as China and India are strictly better off under the mechanism than outside it, as transfers compensate for the reduction in production induced by the uniform tax. Figure A.10 shows the impact of unilateral deviations on aggregate emissions relative to the mechanism outcome, highlighting the asymmetric global consequences of deviations by large economies such as the United States or the EU.

Considering group deviations, we find that when China and the USA deviate together, the mechanism tax falls to \$133 and total emissions rise to 29,207 $MtCO_2$—only a 3.5% decrease relative to status-quo emissions. When the BRICs countries in our dataset (Brazil, India, China, South Africa, Egypt, Ethiopia, and Indonesia) deviate jointly, the mechanism tax is slightly lower at \$129 and the emissions reduction falls further (29,343 $MtCO_2$, a 3.1% decrease from the status quo). In contrast, when a group of countries that either lacked clear pledges at COP21 or were likely to backtrack deviate—including the USA, Canada, Australia, Argentina, Brazil, Peru, Rwanda, and South Africa—the mechanism imposes a much higher tax of $218 on cooperating countries, and total emissions fall to 21,983 $MtCO_2$ (a 27.3% reduction relative to baseline), much closer to the full-participation outcome.[9]

These simulations highlight the importance of keeping China in the participating coalition, as its participation substantially strengthens the mechanism's ability to redistribute emissions and welfare across countries. As noted above, unilateral deviation by China would not be in its best interest, due to the absence of compensating transfers and the costs imposed by high taxes on its exports to other countries. The credibility of transfers is therefore crucial for sustaining participation and the effectiveness of the mechanism. More broadly, the results highlight the interaction between participation incentives, transfers, and the mechanism's ability to redistribute emissions and welfare across countries.

This finding underscores a critical fragility: absent credible punishment for non-compliance—such as a reversal to non-cooperative policies—even a relatively modest climate agreement could unravel as key players opt to free-ride. In sum, the ambition of any voluntary climate coalition is constrained by the twin requirements of compensation (to ensure that no participant is worse off) and enforcement (to deter defection). These constraints help explain why real-world agreements fall far short of the efficient outcome despite the clear gains from cooperation.

These considerations also help interpret the relatively high carbon price implied by the voluntary frontier. Our target-consistent carbon price is substantially higher than the tax levels typically sustained in climate-club environments—for instance the carbon price of about $89 identified by Farrokhi and Lashkaripour (2025). A first reason is the objective: we characterize the maximum emissions reduction compatible with voluntary participation, rather than the welfare-maximizing tax within a given coalition. The institutional environment also differs. In our framework, ambitious mitigation can be sustained because transfers redistribute the gains from cooperation and allow participation constraints to bind tightly. Climate-club environments instead rely primarily on trade-policy instruments and endogenous membership to discipline free-riding, which provides a more limited and indirect mechanism for redistributing mitigation costs. Differences in sectoral structure, outside options, and welfare accounting across models may also contribute to the gap. Taken together, these considerations suggest that the voluntary frontier should be interpreted as a transfer-supported benchmark, while climate-club outcomes capture what can be sustained when redistribution is more constrained.

6.5 One key policy lever: the diffusion of green technology

The widespread adoption of cleaner, low-carbon technologies would reduce the output cost of abatement across the board, especially for today's high-emitting economies. In our model, a lower intensity of carbon σ_r means that a country can produce the same output with fewer emissions, so cutting emissions entails

[9] This group of non-committing or potentially backtracking countries was selected based on a combination of missing or weak COP21 pledges and political indications of withdrawal risk (e.g., the USA under the Trump administration). The full list is: ARG (Argentina), AUS (Australia), BRA (Brazil), CAN (Canada), EGY (Egypt), ETH (Ethiopia), IDN (Indonesia), MYS (Malaysia), PER (Peru), RWA (Rwanda), ZAF (South Africa), USA (United States).

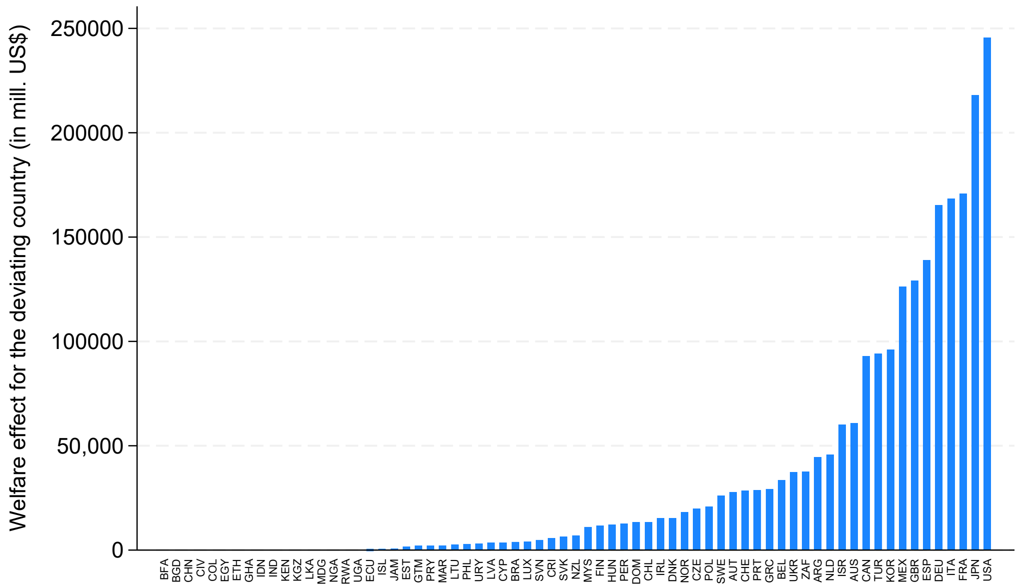

Figure 6.6: Welfare effect of unilateral deviation from the mechanism

a smaller sacrifice of consumption. If all countries experienced a significant technological improvement in carbon efficiency (for example, through cheap renewable energy), the required tax of the mechanism to achieve a given emission target would fall and, critically, the economic losses of major emitters such as China and India under a stringent climate policy would be much smaller. This directly relaxes the participation constraint: when abatement is less economically painful, fewer compensation transfers are needed to keep every country on board.

We model green technology diffusion as a proportional reduction in carbon intensity σ_r for high-emission countries, holding fixed political preference parameters α_r and β_r. A lower σ_r reduces the emissions associated with production in region r, altering comparative advantage and the distribution of abatement costs across countries. For each counterfactual reduction in σ_r, we recompute the participation-constrained allocation and associated transfer vector. To capture incentives for regions to diffuse cleaner technology, we embed them in the participation constraints by comparing welfare under the mechanism with counterfactually lower carbon intensity to status-quo welfare without such reductions. This isolates the technological channel: cleaner production reduces welfare losses in carbon-intensive economies under a uniform tax, relaxes participation constraints, and expands the feasible voluntary frontier. The resulting increase in the implementable carbon price quantifies the role of technology convergence in sustaining deeper cooperation.

Quantitatively, we simulate a scenario in which all countries with carbon intensities above the median (0.00017) adopt cleaner technologies, thus reducing their σ_r. We adopt a baseline scenario where σ_r is reduced by 10% for all affected countries and discuss a reduction of 20% and 50%.[10]

[10] A 10% reduction in carbon intensity is broadly consistent with recent empirical evidence on the effect international green technology transfer, see for example Liu et al. (2024).

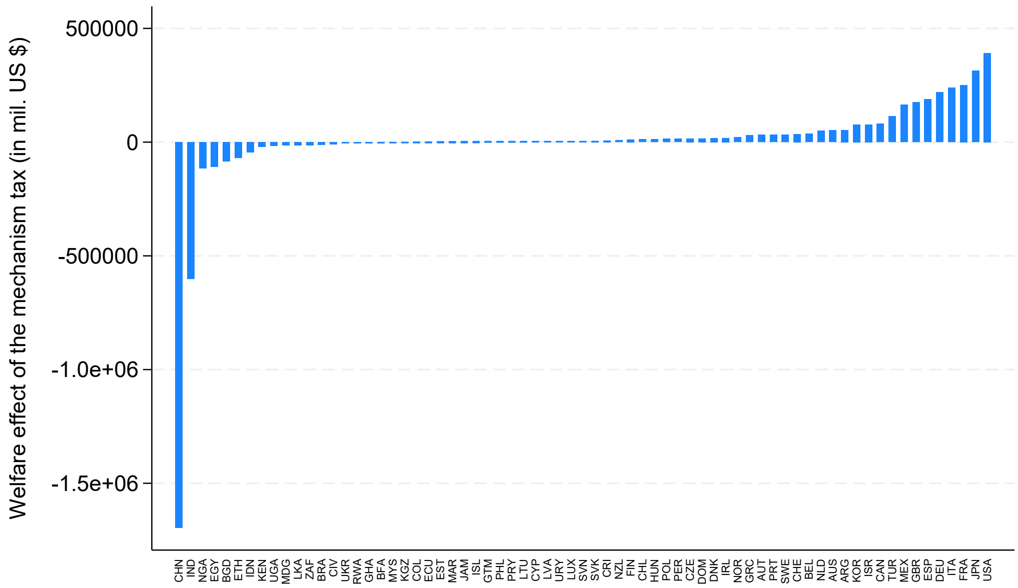

Figure 6.7: Welfare effect of the target-consistent carbon tax after technological transfer (10% efficiency)

To ensure that no country loses from this technological improvement — as highlighted in the theory section, some low-emission countries may experience welfare losses if other countries improve their carbon intensities — we incorporate the technology transfer directly into the mechanism. Specifically, we adjust carbon intensities only when computing the target-consistent allocation, while preserving the original values of σ in the baseline welfare benchmark. Formally, we modify the participation constraint (13) to $W_r(\overline{q_r}, \overline{q_{-r}}; \sigma') + \; m_r \geq \mathcal{W}_r(q_r^*, q_{-r}^*; \sigma)$, where σ denotes the observed carbon intensities and σ' reflects the post-transfer intensities. In this case, the optimal target-consistent emissions fall by an important 45%, with a tax of $324 (attainable thanks to a set of low σ). This would place the global economy well above the requirements for the 2 ° C target, effectively closing much of the current "ambition gap". The green technology transfer allows most high-emitting countries to achieve a more ambitious target at a lower welfare cost, effectively reducing the participation constraint. If the technology transfer is more efficient and allows a 20% (50%) reduction in carbon intensity for all affected countries the effect is even larger as the emissions fall by 51% (64%) thanks to a higher achievable tax of $384 ($844).

Figure 6.7 shows that (pre-transfer) welfare losses are slightly larger than in the baseline. This reflects the fact that, with green technology transfer, the maximum attainable emissions reduction is higher, so bringing each country back to its pre-transfer welfare level requires a greater adjustment. Put differently, green technology transfer allows the baseline target of a 35% emissions reduction to be achieved with lower overall compensatory transfers. This underscores the potentially pivotal role of technology diffusion in narrowing the ambition gap under voluntary cooperation. That said, large-scale diffusion of clean technologies is typically a slow-moving process, unfolding over decades, and would require active policy support—such as R&D subsidies, international technology transfer programs, and institutional frameworks–to be accelerated.

A complementary long-term avenue for deeper climate cooperation is fostering stronger domestic environmental preferences. When voters and governments place greater intrinsic value on emissions reductions, ambitious targets become more politically acceptable, and the need for external incentives or compensatory transfers diminishes. Quantitative simulations (in the appendix A.5) suggest that raising lower-preference countries to the median level could reduce baseline emissions by 8% and enable tax levels consistent with a 72% reduction from baseline (see Appendix, FigureA.12 for welfare effects). Such preference shifts are slow-moving, shaped by education, experience with climate impacts, and institutional quality.

In summary, absent progress along transfer of cleaner technology (or greener preferences) the voluntary cooperation architecture is likely to remain stuck with modest ambition. Our analysis highlights that the stringent limits on climate action are not due to a lack of policy mechanisms or international agreements per se, but rather stem from deeper structural factors: the distribution of abatement costs, preferences, and resources across countries. Recognizing this constraint is crucial for designing realistic climate policy. It suggests that meaningful improvements in global climate outcomes will require either innovations that lower abatement costs, a shift in political will toward valuing the climate, or new forms of burden-sharing beyond the current nation-state-centric approach. Without such changes, any climate agreement that relies on voluntary participation is doomed to remain far less aggressive than what scientific targets would demand.

6.6 Welfare Effects of Implementing COP21 Pledges and the 2 Degree Objective

We assess the welfare implications of the climate pledges submitted after COP21 by comparing their implied emissions reduction to the outcome obtained under our target-consistent mechanism delivering the same aggregate reduction. The goal is to evaluate both the effectiveness and cost-efficiency of current international commitments relative to an incentive-compatible policy.

We compile nationally determined contributions (NDCs) from the UNFCCC "Nationally Determined Contributions Registry," retaining pledges that specify a targeted percentage reduction in emissions by 2030 relative to a clear baseline year.[11] This yields 43 countries with explicit targets and 26 countries without quantified pledges.

Assuming full implementation of these pledges and constant emissions in non-pledging countries, the model predicts a 10% reduction in global emissions relative to the baseline—from 30,175 to 27,160 megatons of CO_2. This is broadly consistent with current projections from international monitoring bodies but remains insufficient to place the world on a trajectory compatible with limiting warming to 2°C, as emphasized in recent UNEP and IPCC assessments (see e.g. UNEP (2024)).

We then compute how the same aggregate reduction could be achieved through our mechanism. A uniform global carbon tax of $70 per ton delivers an almost identical emissions outcome (27,158 megatons of CO_2) but at substantially lower welfare cost. Aggregate welfare losses under the pledge system are estimated to be $1.41 trillion higher than under the $70 tax-and-transfer mechanism. The efficiency

[11] When multiple targets are stated, we retain the most ambitious. For countries without a clear or quantifiable pledge—such as the United States—we assume emissions remain at baseline levels.

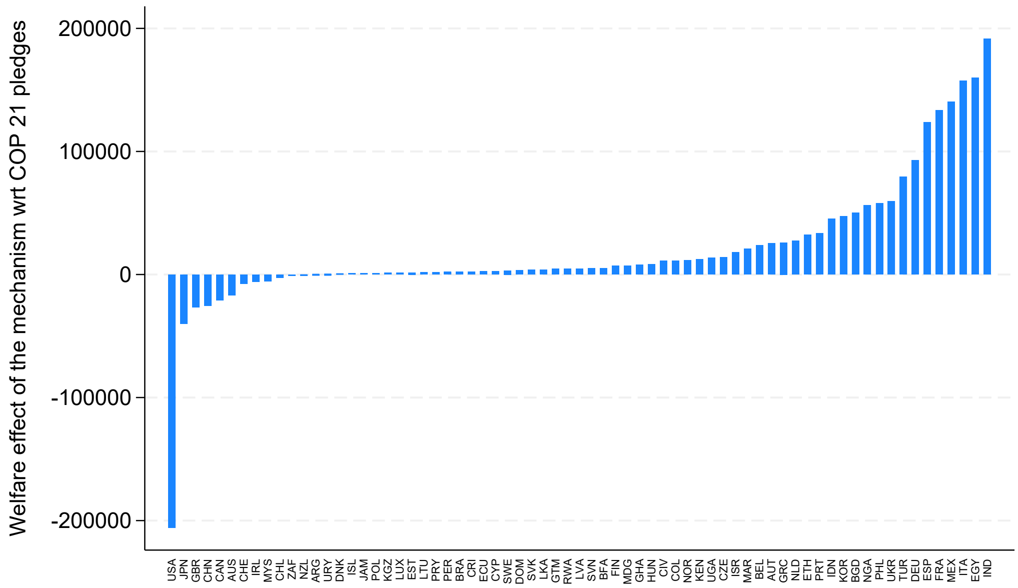

Figure 6.8: Welfare effect of the mechanism wrt COP 21 pledges

gain arises because the mechanism reallocates abatement toward countries with lower marginal costs, whereas the pledge structure imposes reductions in a more rigid and uneven manner.

Figure 6.8 reports the country-level welfare differences between the two approaches. While the mechanism improves global welfare, some countries perform better under the pledge system. Notably, the United States—whose NDC is effectively flat in our dataset—faces higher abatement costs under the mechanism and would therefore require transfers to satisfy its participation constraint.

Implementing the \$70 carbon tax while maintaining voluntary participation requires total transfers of \$271 billion. This is substantially smaller than the \$1,800 billion required to achieve the 28% emissions reduction consistent with the 2°C target, or the \$2,400 billion required for the highest feasible target. Thus, achieving the COP21 emissions outcome through our mechanism is feasible at lower cost and with relatively modest transfers, although it would still involve politically sensitive redistribution from non-pledging to pledging countries.

Implementing the 2 degree target. Finally, we compare our mechanism with what it would take to meet the 2°C target directly. A carbon tax of \$200 per ton reduces emissions by 28%, from 30,272 to 21,685 megatons. This target is in line with both recent UNEP assessments and the revised U.S. EPA Social Cost of Carbon (e.g. see UNEP (2024)).

Figure A.11 reports the welfare changes from this ambitious policy. We compute that achieving voluntary participation would require \$1.8 trillion in global transfers—an order of magnitude larger than current global climate finance flows. This underscores the key result of the paper: without technological convergence or stronger environmental preferences, meeting ambitious climate goals under voluntary cooperation requires financial transfers of unprecedented scale—even if they remain far below the range of expected climate damages[12].

7 Concluding remarks

This paper develops a framework to assess the most ambitious climate target achievable under voluntary international cooperation. We characterize the *voluntary frontier*: the maximum emissions reduction that can be sustained when each country's participation constraint and global budget balance are satisfied. The resulting mechanism combines a uniform carbon price with international transfers that redistribute the gains from cooperation.

Our analysis highlights a fundamental tension between climate ambition and political feasibility. Achieving deeper emissions reductions requires progressively larger transfers to compensate countries facing higher mitigation costs. In our calibration, the carbon price consistent with the voluntary frontier is substantially higher than currently observed policy levels, but sustaining such ambition would require large cross-country transfers. The quantitative results underscore the central role of participation incentives. When major emitters deviate from the agreement, the feasible carbon price falls sharply and emissions reductions decline significantly. In particular, maintaining the participation of large economies such as China proves critical for the effectiveness of the mechanism.

More broadly, our findings help explain why existing climate agreements remain far below the levels implied by global efficiency. Even when cooperation generates large aggregate gains, the distribution of those gains across countries imposes strong political constraints. Transfers therefore emerge as a key instrument for sustaining ambitious international climate policy. The voluntary frontier should be interpreted as a benchmark for what is achievable under cooperative but self-enforcing agreements. Climate-club arrangements and trade-based enforcement mechanisms represent alternative institutional approaches to sustaining cooperation when transfers are limited. Understanding how these different mechanisms interact remains an important direction for future research.

Overall, the results suggest that ambitious climate mitigation is technically feasible but politically constrained. Designing credible mechanisms to redistribute the gains from cooperation may therefore be central to narrowing the gap between climate targets and achievable international agreements.

[12] In a recent study, Bilal and Känzig (2024) find that 1 degree warming could reduce world GDP by 12%

References

Aldy, J. E., Kotchen, M. J., Stavins, R. N., and Stock, J. H. (2021). Keep climate policy focused on the social cost of carbon. *Science*, 373(6557):850–852.

Antweiler, W., Copeland, B. R., and Taylor, M. S. (2001). Is free trade good for the environment? *American Economic Review*, 91(4):877–908.

Baccianti, C. and Schenker, O. (2022). Cournot, pigou, and ricardo walk into a bar: Unilateral environmental policy and leakage with market power and firm heterogeneity. *Journal of the Association of Environmental and Resource Economists*, 9(6):1161–1195.

Barrett, S. (1994). Self-enforcing international environmental agreements. *Oxford Economic Papers*, 46(Supplement 1):878–894.

Battaglini, M. and Harstad, B. (2020). The political economy of weak treaties. *Journal of Political Economy*, 128(2):544–590.

Bilal, A. and Känzig, D. R. (2024). The macroeconomic impact of climate change: Global vs. local temperature. *NBER Working Paper No. 32450.*

Blinova, D., Emuru, R., and Bagozzi, B. (2024). Individual attendance data for over 30 years of international climate change talks. *Nature, Scientific Data*, 11.

Copeland, B. R., Shapiro, J. S., and Scott Taylor, M. (2022). Chapter 2 - globalization and the environment. In Gopinath, G., Helpman, E., and Rogoff, K., editors, *Handbook of International Economics: International Trade, Volume 5*, volume 5 of *Handbook of International Economics*, pages 61–146. Elsevier.

Copeland, B. R. and Taylor, M. S. (2003). *Trade and the Environment: Theory and Evidence*. Princeton Series in International Economics. Princeton University Press.

Dekle, R., Eaton, J., and Kortu, S. (2008). Global rebalancing with gravity: Measuring the burden of adjustment. *IMF Staff Papers*, 55(3):511–540.

Desmet, K., Nagy, D. K., and Rossi-Hansberg, E. (2018). The geography of development. *Journal of Political Economy*, 126(3):903–983.

Dietz, S., Fruitiere, C., Garcia-Manas, C., Irwin, W., Rauis, B., and Sullivan, R. (2018). An assessment of climate action by high-carbon global corporations. *Nature Climate Change*, 8:1072–1075.

Farrokhi, F. and Lashkaripour, A. (2025). Can trade policy mitigate climate change? *Econometrica*, 93(5):1561–1599.

Gersbach, H., Hummel, N., and Winkler, R. (2021). Long-term climate treaties with a refunding clubjo - environmental and resource economics. *Environmental and Resource Economics*, 80:511 – 552.

Harstad, B. (2012). Climate contracts: A game of emissions, investments, negotiations, and renegotiations. *Review of Economic Studies*, 79(4):1527–1557.

Kotchen, M. J. and Maggi, G. (2024). Climate agreements and the role of lobbying. Working paper, Yale University.

Liu, Y., Lei, P., Shen, B., and He, D. (2024). Green technology advancement, energy input share and carbon emission trend studies. *Scientific Report*, (14).

Maggi, G. and Staiger, R. (2022). International climate agreements and the scream of greta. Working paper, Yale University.

Martimort, D. and Sand-Zantman, W. (2016). A mechanism design approach to climate-change agreements. *Journal of the European Economic Association*, 14(3):669–718.

Nordhaus, W. (2015). Climate clubs: Overcoming free-riding in international climate policy. *American Economic Review*, 105(4):1339–70.

Ricke, K., Drouet, L., Caldeira, K., and Tavoni, M. (2018). Country-level social cost of carbon. *Nature Climate Change*, 8(10):895–900.

Stern, N. (2022). A Time for Action on Climate Change and a Time for Change in Economics. *The Economic Journal*, 132(644):1259–1289.

Stern, N. and Stiglitz, J. (2022). The economics of immense risk, urgent action and radical change: towards new approaches to the economics of climate change. *Journal of Economic Methodology*, 29(3):181–216.

UNEP (2024). Emissions gap report 2024: No more hot air . . . please! Technical report, UNEP (United Nations Environment Programme), Nairobi, Kenya.

Weitzman, M. L. (1974). Prices vs. quantities. *The Review of Economic Studies*, 41(4):477–491.

A Appendix

A.1 Concavity of the global welfare

By definition

$$W(q) = \sum_r W_r(q_r, q_{-r}) = \sum_r \left\{ N_r \left[\beta_r u(c_r) + (1-\beta_r) u'(c_r)\, c_r \right] - A_r C\left(\frac{q_r}{A_r}\right) - N_r \beta_r \alpha_r E \right\}, \qquad (31)$$

as $\sum_r q_r = \sum_r N_r c_r$, with c_r determined by optimal consumer behavior at equilibrium price. Therefore, as $cu''(c) = -\epsilon u'(c)$:

$$\frac{\partial W}{\partial q_i}(q) = \sum_r \left\{ N_r u'(c_r)\left(1-(1-\beta_r)\epsilon\right)\frac{\partial c_r}{\partial q_i} \right\} - C'\left(\frac{q_i}{A_i}\right) - \sum_r N_r \alpha_r \beta_r \sigma_i \qquad (32)$$

$$= p(q) - C'\left(\frac{q_i}{A_i}\right) - \sum_r N_r \alpha_r \beta_r \sigma_i \qquad (33)$$

through consumer optimal behavior (i.e. the left hand side of equation (6)) and as $\sum_r N_r \frac{\partial c_r}{\partial q_i} = 1$ at market equilibrium.

For any given vector $q = (q_1, \dots, q_n) \in \mathbb{R}^n_{++}$, the Hessian matrix $H(q)$ of W has entries

$$\frac{\partial^2 W}{\partial q_i^2} = p'(q) - \frac{C_i''(q_i)}{A_i}, \qquad \frac{\partial^2 W}{\partial q_i \partial q_j} = p'(q) \ \ (i \neq j), \qquad (34)$$

with $p'(q) < 0$ (by equation (12)) and each $\frac{C_i''(q_i)}{A_i} > 0$ for all $q_i > 0$.

Consider any nonzero vector $v = (v_1, \dots, v_n) \in \mathbb{R}^n$. The quadratic form associated with $H(q)$ is

$$v^\top H(q) v = p'(q)\Big(\sum_{i=1}^n v_i\Big)^2 - \sum_{i=1}^n \frac{C_i''(q_i)}{A_i}\, v_i^2. \qquad (35)$$

Since $p'(q) < 0$, the first term is non-positive. Since each $\frac{C_i''(q_i)}{A_i} > 0$ and $v \neq \mathbf{0}$ (so at least one $v_i \neq 0$), the second term is strictly positive. Hence

$$v^\top H(q) v < 0 \qquad \text{for all } v \neq \mathbf{0}. \qquad (36)$$

Thus $H(q)$ is negative definite at every vector $q > 0$. Consequently, W is strictly concave on the domain $\mathbb{R}^n_{++}$.

A.2 Political economy framework

Voting game. Each region has two parties, A and B, motivated by winning elections and able to commit to future production policies (as explained in the main text tax decisions determine production policies). The fraction of citizens voting for party A is $\gamma_A = \sum_g \lambda^g \gamma_A^g$, where γ_A^g is the fraction of group

g ($g \in \{c, f\}$, c for consumers, f for firm owners) voting for A and λ^g is the proportion of group g in the population of the region.

The indirect utility function of each citizen is

$$\tilde{w}_i^g(q_r, q_{-r}) = w^g(q_r, q_{-r}) + \mu_i^g(B) \tag{37}$$

where w^g is the utility function of citizens under the proposed production policy (q_r, q_{-r}) (i.e. utility net of environmental concerns for consumers, and profit for producers) and $\mu_i^g(B)$ is the relative preference of individual i for party B (drawn from a known group-specific distribution F^g).

Individual i vote for A if $w^g(q_r^A, q_{-r}) - w^g(q_r^B, q_{-r}) > \mu_i^g(B)$. Therefore, objective of party A is

$$\max_{q_r^A} \sum_g \lambda^g F^g \left(w^g(q_r^A, q_{-r}) - w^g(q_r^B, q_{-r})\right) \tag{38}$$

leading to

$$\sum_g \lambda^g f^g \left(w^g(q_r^A, q_{-r}) - w^g(q_r^B, q_{-r})\right) \frac{\partial w^g(q_r^A, q_{-r})}{\partial q_r^A} = 0 \tag{39}$$

Consequently, as usual in probabilistic voting games, at the Nash Equilibrium of the game $q_r^A = q_r^B = q_r^*$, implying that $w^g(q_r^A, q_{-r}) = w^g(q_r^B, q_{-r})$.

At the voting equilibrium each party proposes the same production policy q_r^* given by

$$\sum_g \beta^g \frac{\partial w^g(q_r^*, q_{-r})}{\partial q_r} = 0 \tag{40}$$

where $\beta^g = \lambda^g f^g(0)$ represents the political weight of each group. It therefore corresponds to the policy maximizing the weighted sum of voters preferences. In our model, to ensure comparability between regions, we further normalize the weight of firm owners to 1 (the weight on consumers is then β^c/β^f) and account for population size in our comparison of consumer weight. β is equation (3) then equals $\beta^c/(\beta^f.N_r)$. We also assume for simplicity that tax revenue is redistributed to firm owners.

A.3 Extension to a dynamic mechanism

We consider in this extension, a setting in which starting from current date $t = 0$ each region can fully anticipate the future path of the economy, including carbon intensities $(\sigma_{rt})_{r,t}$, technology $(A_{rt})_{r,t}$, political preferences $(\alpha_{rt})_{r,t}$, $(\beta_{rt})_{r,t}$, region size $(N_{rt})_{r,t}$ and the corresponding non-cooperative production levels $(q^*_{rt})_{r,t}$. We also assume that citizens care only about contemporaneous emissions E_t, i.e., the emissions they are directly responsible for. In this intertemporal setting, the objective of the Principal is to minimize intertemporal emissions : $E = \sum_{t,r} q_{t,r}\sigma_{t,r}$ and assuming that it can credibly propose a take-it or leave-it mechanism, the optimal target-consistent one solves:

$$\begin{aligned}
&\min_{(q_{rt})_{r,t},(m_{rt})_{r,t}} \sum_t \sum_r q_{rt}\sigma_{rt} \\
&\qquad s.t \sum_t \frac{1}{(1+\rho)^t}\left[W_{rt}(q_{rt},q_{-rt})+m_{rt}\right] \geq \sum_t \frac{1}{(1+\rho)^t} W_{rt}(q^*_{rt},q^*_{-rt}) \ \forall r \\
&\qquad \sum_t \frac{1}{(1+\rho)^t}\sum_r m_{rt} = 0
\end{aligned} \tag{41}$$

with $W_{rt}(q_{rt},q_{-rt})$ the welfare of region r at time t and ρ is the interest rate (supposed constant through time for simplicity).

As in the static case, the participation constraint for every region binds at equilibrium and using the budget constraint, the problem corresponds to:

$$\begin{aligned}
&\min_{(q_{rt})_{r,t}} \sum_t \sum_r q_{rt}\sigma_{rt} \\
&s.t \sum_r \sum_t \frac{1}{(1+\rho)^t} W_r(q_{rt},q_{-rt}) = \sum_r \sum_t \frac{1}{(1+\rho)^t} W_r(q^*_{rt},q^*_{-rt})
\end{aligned} \tag{42}$$

We therefore end up, at each period with a first-order condition similar to the static case:

$$\sigma_{rt} = \lambda \frac{1}{(1+\rho)^t}\frac{\partial W_t}{\partial q_{rt}}(\overline{q_t}) \tag{43}$$

where $W_t(\overline{q}_t) = \sum_r W_r(\overline{q}_{rt},\overline{q}_{-rt})$ and λ the Langrage multiplier associated with the constraint. Denoting $\lambda_t = \lambda\frac{1}{(1+\rho)^t}$, we thus obtain at equilibrium for every region i at every period t:

$$p_t = C'\left(\frac{\overline{q_{it}}}{A_{it}}\right) + \left(\frac{1}{\lambda_t} + \sum_r N_{rt}\alpha_{rt}\beta_{rt}\right)\sigma_{it} \tag{44}$$

meaning that the optimal quantity can be implemented by a uniform carbon tax $T_t = \frac{(1+\rho)^t}{\lambda} + \sum_r N_{rt}\alpha_{rt}\beta_{rt}$ at each period. Moreover, for every couple of regions i, j, at every period:

$$\frac{\sigma_{it}}{\sigma_{jt}} = \frac{p_t - C'\left(\frac{\overline{q_{it}}}{A_{it}}\right)}{p_t - C'\left(\frac{\overline{q_{jt}}}{A_{jt}}\right)} \tag{45}$$

and production reallocation follows the same pattern as in Proposition 1.

Transfers $(m_{r,t})_{r,t}$ are then defined by:

$$\sum_t \frac{1}{(1+\rho)^t}\left[W_r(\overline{q}_{rt},\overline{q}_{-rt},\sigma_{rt},\sigma_{-rt}) - W_r(q^*_{rt},q^*_{-rt},\sigma_{rt},\sigma_{-rt})\right] \geq -\sum_t \frac{1}{(1+\rho)^t} m_{rt} \ \forall r \tag{46}$$

with the sum on r of both side being equal to 0.

Following the same reasoning as in Lemma 2, it has to be that the above equation holds with equality for every region:

$$\sum_t \frac{1}{(1+\rho)^t}\left[W_r(\overline{q}_{rt}, \overline{q}_{-rt}, \sigma_{rt}, \sigma_{-rt}) + m_{r,t} - W_r(q^*_{rt}, q^*_{-rt}, \sigma_{rt}, \sigma_{-rt})\right] = 0 \ \forall r \tag{47}$$

and a solution for $m_{r,t}$ (that doesn't entail inter-generational redistribution) is:

$$m_{r,t} = W_r(q^*_{rt}, q^*_{-rt}, \sigma_{rt}, \sigma_{-rt}) - W_r(\overline{q}_{rt}, \overline{q}_{-rt}, \sigma_{rt}, \sigma_{-rt}) \ \forall r, t \tag{48}$$

A.4 Calibrated model

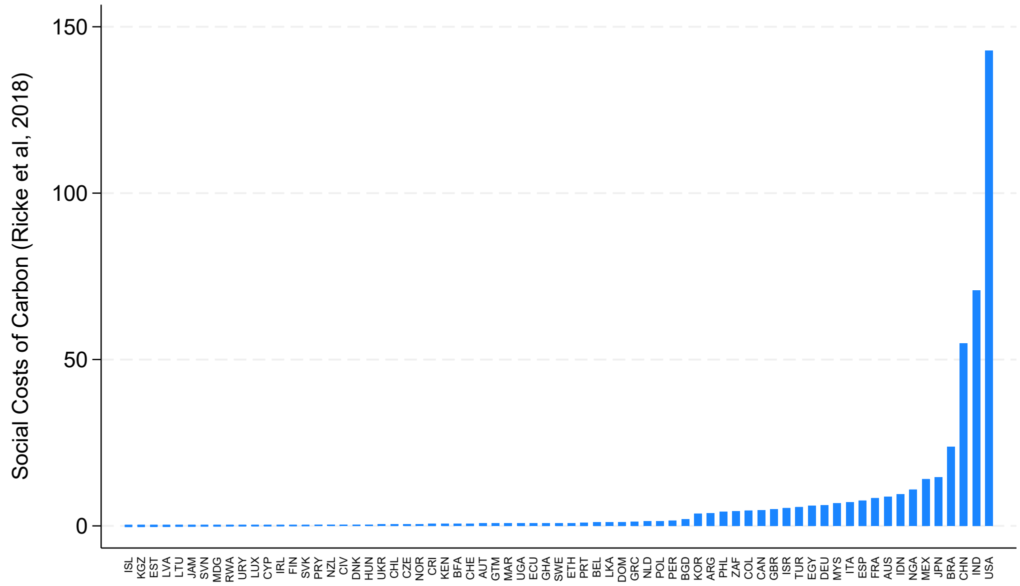

Figure A.1: Social Costs of Carbon (from Ricke et al, 2018)

A.4.1 Calibrated α and β.

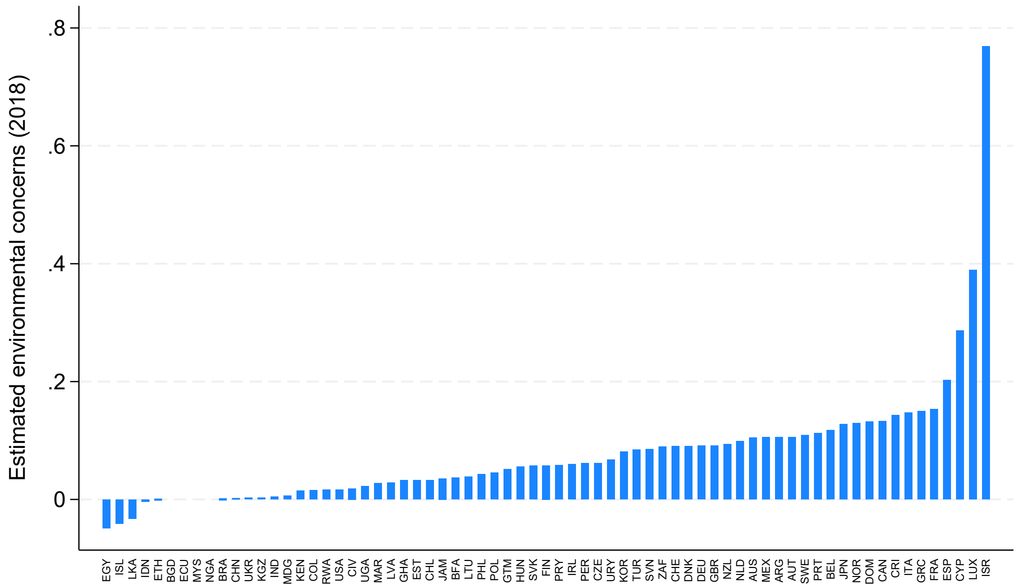

Figure A.2: Estimated environmental concerns over the world

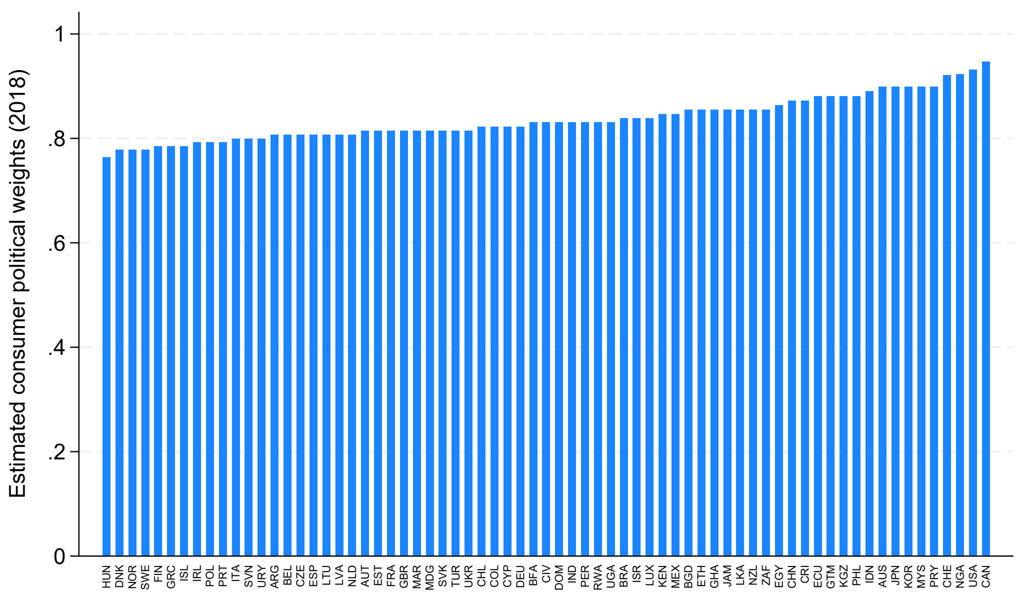

Figure A.3: Estimated consumer political weight over the world

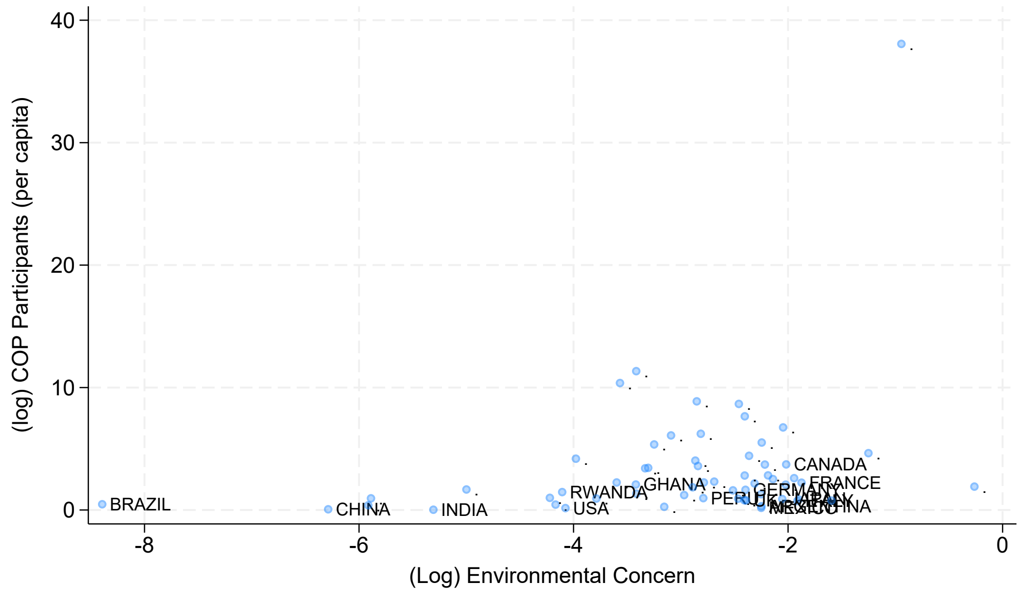

Figure A.4: Correlation between estimated environmental concerns and the number of COP participants (per capita)

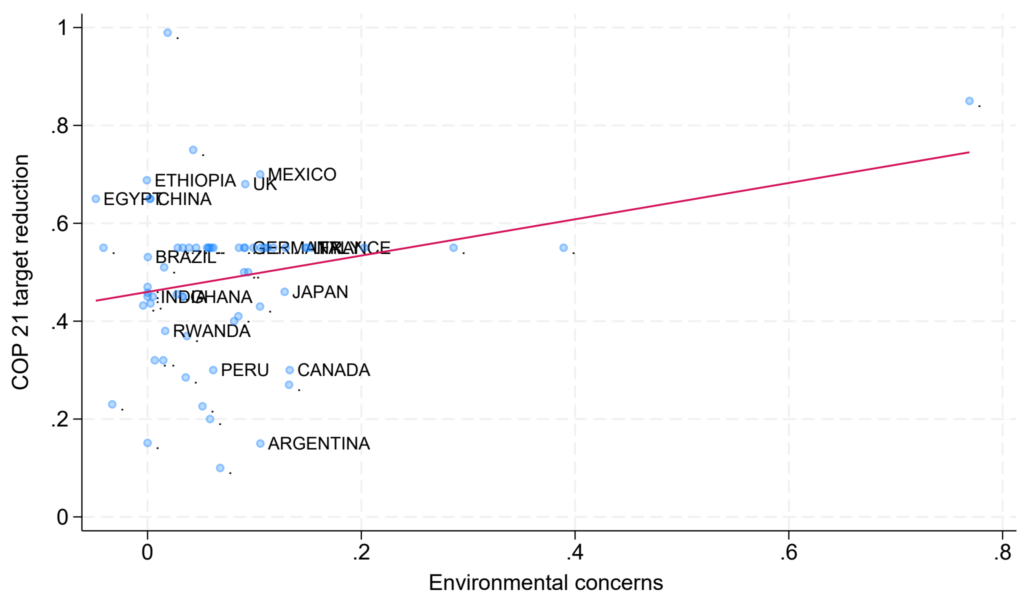

Figure A.5: Correlation between estimated environmental concerns and COP 21 pledges

A.4.2 Summary of Data sources & calibration

Table 1: Data sources, constructed variables, and calibration/identification targets (baseline: 2018)

Object	Definition / role in model	How obtained	Primary source (year)
A. Core data inputs (country-level, r)			
N_r (population)	Number of consumers	Directly observed	Penn World Table (PWT), population, 2018
Y_r (GDP, output-side)	Empirical GDP used for descriptive fit / baseline comparison (Figure 5.3)	Directly observed	PWT, real GDP (output-side, PPP, 2017 USD), 2018
C_r^{data} (HH consumption)	Empirical household consumption used for descriptive comparison (Figure 5.4); not targeted moment	Directly observed	PWT, real household consumption (PPP, 2017 USD), 2018
A_r (productivity shifter)	Proxied by the empirical capital stock; proxy for productivity shifter A_r in production cost	Directly observed (used as proxy)	PWT, capital stock (PPP, 2017 USD), 2018
E_r^{data} (CO_2 emissions)	CO_2 emissions used to compute intensity σ_r and validate implied emissions	Directly observed	EDGAR Community GHG Database, 2018
σ_r (carbon intensity)	Emissions intensity of production;	Constructed: $\sigma_r = E_r^{data}/Y_r$	EDGAR (emissions) + PWT (GDP), 2018
τ_r^{obs} (effective carbon price)	Observed (net) effective carbon rate used to infer/cap climate preference via (10)	Directly observed (coverage-weighted)	OECD Effective Carbon Rates companion dataset (ECR), 2018
t_r^{obs} (consumption tax)	Standard VAT/sales tax used to infer consumer political weight β_r via (9)	Directly observed (avg. for federal states)	IMF VAT rates
SCC_r	Country-specific social cost of carbon used as upper bound for "imputed" carbon price	Directly observed from literature estimates	Ricke et al. (2018), SSP2/RCP6.0, BHM damage function (central spec.)
NDC_r (pledged target)	Pledged emissions reductions (used in pledge comparison exercises)	Directly observed (filtered to comparable % targets)	UNFCCC Nationally Determined Contributions Registry (post-COP21)
B. Parameters taken from the literature (common across countries)			
ε	CRRA / demand elasticity in utility $u(c) = h\frac{c^{1-\varepsilon}}{1-\varepsilon}$; enters (6),(7),(9),(11)	Set externally	Chosen from literature: $\varepsilon = 0.9$ (e.g. Desmet et al. 2018)
η	Convexity / supply elasticity parameter in cost $C(\cdot) = k\frac{(\cdot)^{1+\eta}}{1+\eta}$; enters (8),(11)	Set externally	Chosen from literature: $\eta = 1$ (e.g. Desmet et al. 2018)
C. Calibrated / identified parameters			
h	Utility scale parameter (common); disciplines overall levels (jointly with k)	Calibrated to match baseline magnitudes (GDP/emissions fit)	Internally calibrated: $h = 0.7$ (baseline)
k	Cost scale parameter (common); disciplines production levels (jointly with h)	Calibrated to match baseline magnitudes (GDP/emissions fit)	Internally calibrated: $k = 1260$ (baseline)
β_r	Political weight on consumers (relative to capitalists); identified from VAT via (9)	Recovered from data using model mapping (eq. 9)	IMF VAT rates + model equation (9)
τ_r^{imp} ("imputed" carbon price)	Carbon price used to identify α_r; equals observed if below SCC, else capped at SCC	Constructed rule described in text	OECD ECR + Ricke et al. (2018) SCC
α_r	Environmental concern (political weight on global emissions); identified from (10)	Recovered using eq (10)	OECD ECR + IMF VAT + PWT population + SCC cap

A.4.3 Baseline welfare.

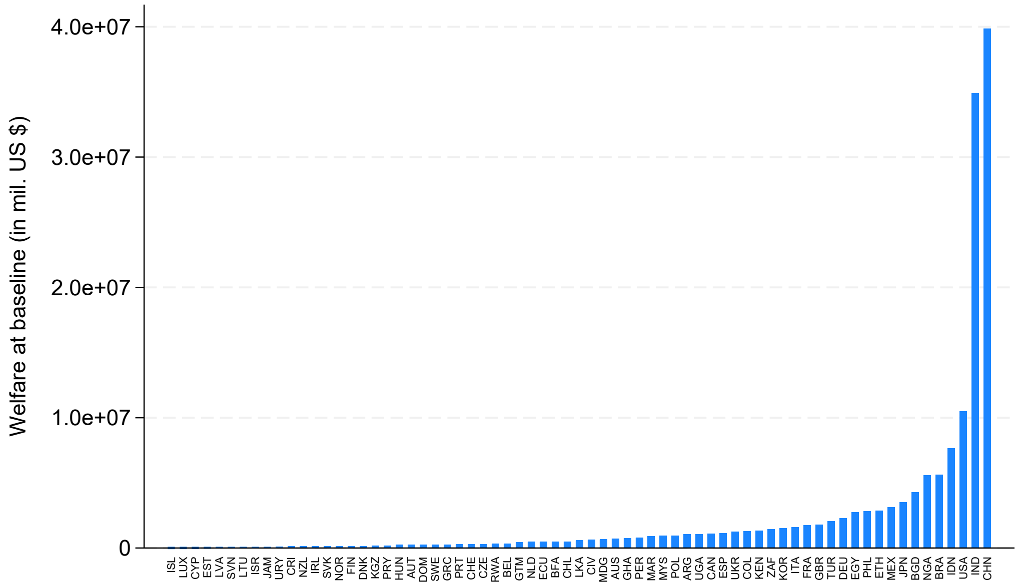

Figure A.6: Baseline welfare (in mil. US$) per country

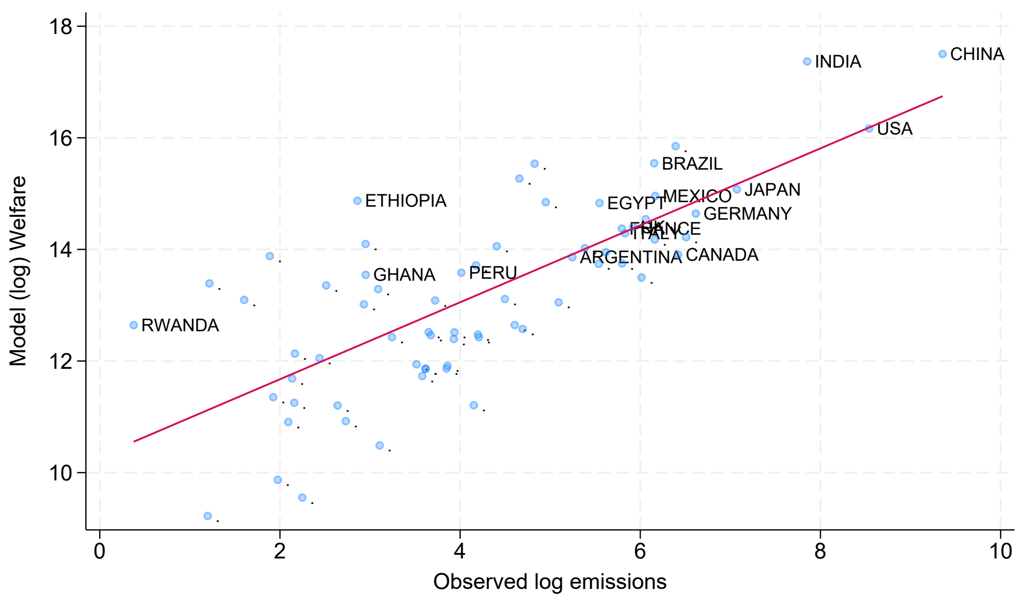

Figure A.7: Correlation between calibrated welfare at baseline and observed CO2 emissions

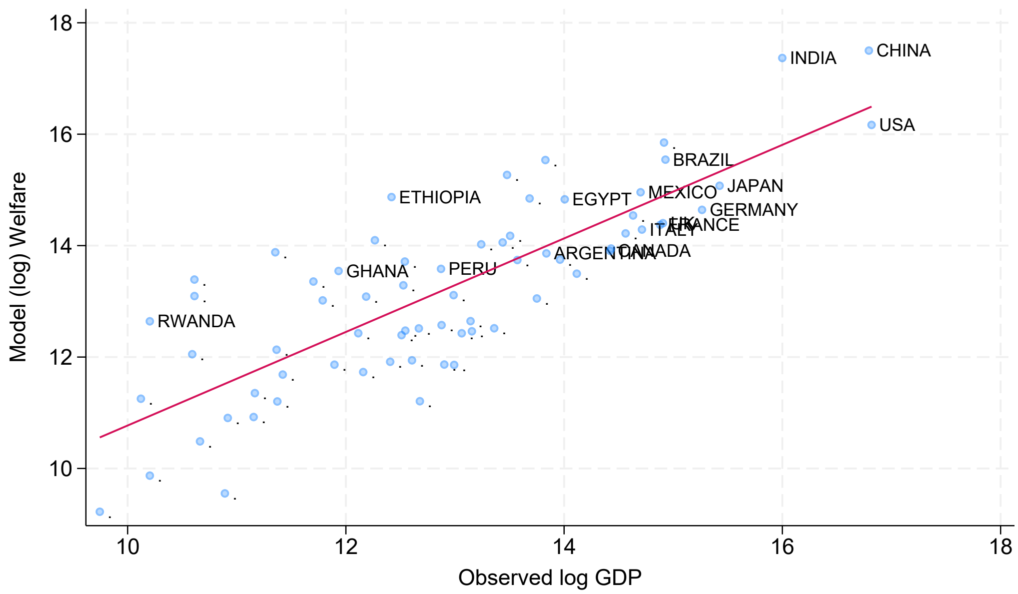

Figure A.8: Correlation between calibrated welfare at baseline and observed GDP (in log)

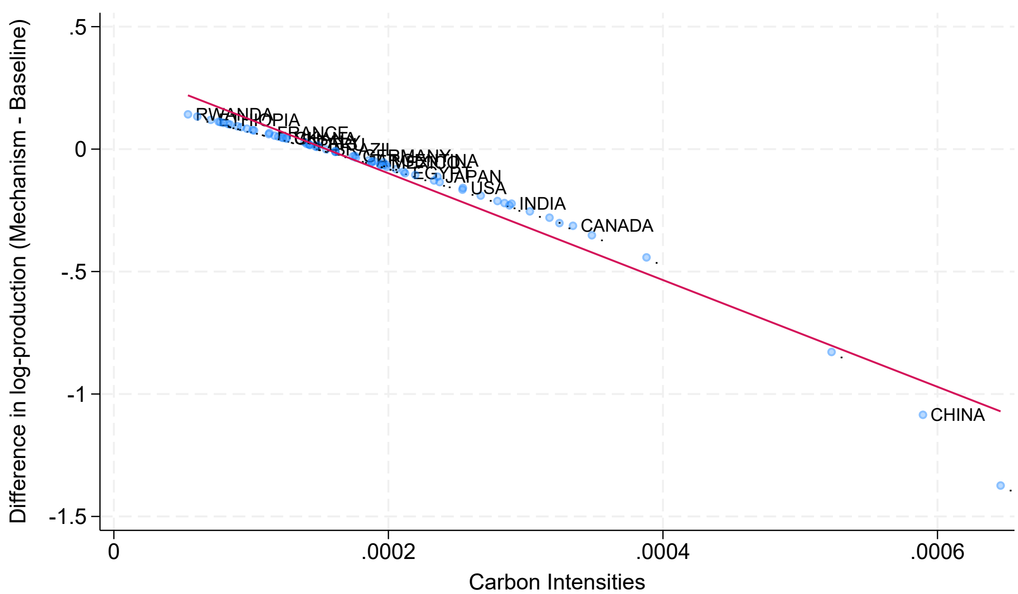

Figure A.9: Production reallocation and carbon intensity

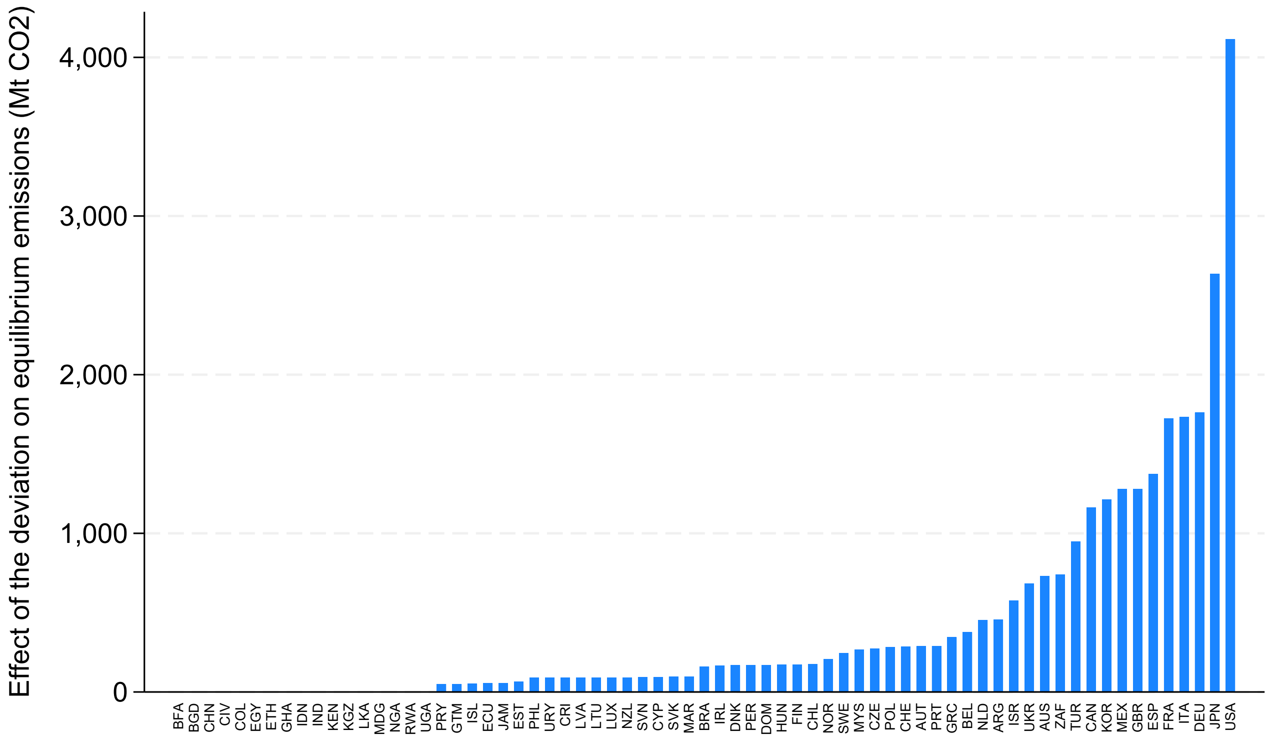

Figure A.10: Effect of the deviation on equilibrium emissions

A.5 A long-term policy: fostering environmental preferences.

A second avenue is a shift in domestic environmental preferences. In our framework, this corresponds to higher weight α_r that voters and governments place on environmental quality relative to consumption. Stronger public concern for climate change – especially in large, high-emission countries – would increase willingness to incur costs for mitigation. If, for instance, the United States, China, and India valued emissions reductions as highly as many European countries currently do, they would be more prepared to accept output losses or to pay transfers as part of a climate agreement. Greater intrinsic motivation to cut emissions effectively raises the bar for voluntary action: the participation constraint would bind at a more ambitious emissions target because fewer countries would require external incentives to participate. Put differently, when the political benefits of abatement (in terms of domestic support or avoided local damages) rise, the need for side-payments to induce cooperation diminishes. We note that such preference shifts may result from changing public awareness, education, or experience with climate impacts, and they can be influenced by policy (e.g. information campaigns or incorporating climate costs into national planning).

Quantitatively, we consider a scenario where all countries with α_r below the median (0.076) adopt the median value. In this case, the baseline emissions, without any intervention, drops to 27,731 megatons of CO2 (8% reduction). Furthermore, the acceptable tax increases drastically to \$1041 which achieve

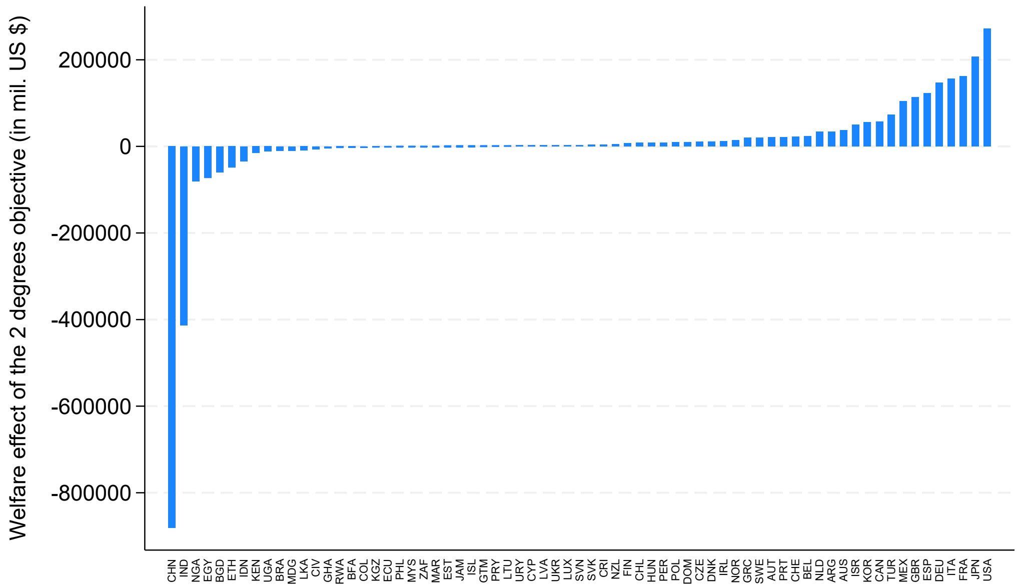

Figure A.11: Welfare effect of the carbon tax needed to reach the 2°C target

a 72% emissions reduction from baseline. Furthermore, increased environmental concern also raises the willingness to pay for abatement, leading to larger compensatory transfers. Figure A.12 illustrates the welfare impacts under this adjustment. In short, improvement in environmental concern makes ambitious cooperation more politically acceptable by mitigating the trade-off between emissions reduction and output. However, such shifts in environmental preferences are typically slow-moving and shaped by long-term factors such as education, institutional quality, and generational value changes. This suggests that while preference alignment can support deeper cooperation, it is unlikely to be a short-run lever for overcoming current coordination barriers.

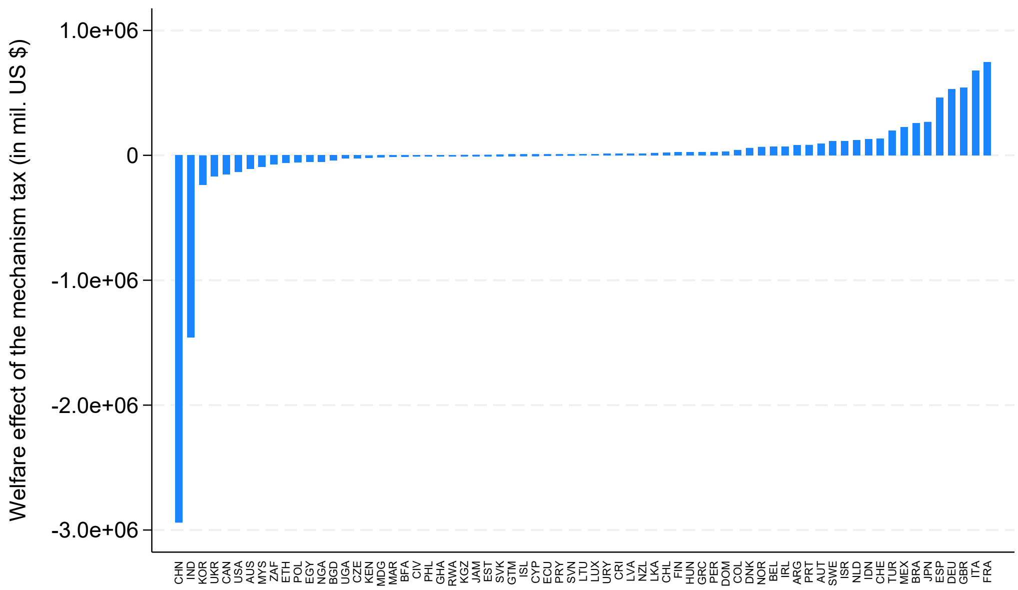

Figure A.12: Welfare effect of the target-consistent carbon tax after an increase in environmental concerns