Subsidy Treaties:
Global Abatement If Cooperation Is Fragile

Jonas Metzger

November 2024

Abstract

While a coordinated global decarbonization could be mutually beneficial, countries have a unilateral incentive to maintain higher emission levels. The literature has proposed 'emissions treaties' in which deviations from emission targets would be deterred by threats of sanctions or suspension of future cooperation, once detected. However, such treaties rely on a shared belief that this equilibrium will not be disrupted in the future, and pose severe monitoring challenges in practice. In this work, we examine how cooperation could be sustained today if parties anticipate future disruption of cooperation, and how this changes what they should coordinate on. We find that coordination on subsidies for the fixed-costs of renewable technologies, rather than on emissions levels, becomes more efficient and may be the only incentive compatible option if opportunity costs to cooperation are significant: due to the lower variable costs of renewables, short-term fixed-cost subsidies let market forces ensure a displacement of fossil fuels that persists independently of cooperation. Short-term emissions taxes or targets may fail to incentivize the required investments. A separate benefit is that subsidy treaties do not require high-frequency emissions monitoring. They do require ex-ante estimates of their impact on future emissions and national incomes, which we obtain using a structural model of national electricity sectors and damages from warming. We estimate a subsidy treaty among the USA, China, EU and India for PV panels and wind turbines alone could avoid 40% of CO_2 from electricity over the next century, relative to a non-cooperative 'business-as-usual'.

1 Introduction

From a global perspective, it is cheaper to reduce our carbon emissions than to suffer the costs they will cause through global warming. From the perspective of any individual country however, only a small share of the total damage caused by its emissions occurs within its borders. Hence, as long as CO_2 abatement is costly and countries act unilaterally, total emissions remain inefficiently high – a prisoner's dilemma. Yet, the Coase (1960) Theorem asserts a simple contract should make all generations and countries better off, in theory: those who can most cheaply reduce their emissions are reimbursed for doing so, and those who benefit from the avoided damages fund the reimbursements proportionally to their benefit. The geopolitical reality, however, is that no higher power exists that could enforce such a contract. This creates a 'gap' between the immediate opportunity costs an abating country would incur, and the reward it can expect to receive from the actions of other countries in return. The literature has proposed various approximations to this ideal contract, for which this gap could be 'bridged' via economic assumptions.

We contribute a new proposal to this literature, which we hope could offer a palatable alternative for policymakers who rejected previos emissions treaties, depending on their reasons for doing so. Specifically, our proposal aims to be robust to worries that the treaty may fall apart in the future (for possibly exogenous reasons), which we show may suffice to explain why countries reject significant cooperation on emissions today. As a practical benefit, it also avoids the monitoring challenge of assessing whether countries comply with emission targets. Attributing the present cooperation failure to – and seeking a workaround for – these two concerns is largely orthogonal to prior literature, which typically assumed them away. Instead, prior work often deduced the cooperation failure by invoking bargaining solutions which tend to favor defection to varying degrees, which ultimately yield smaller[1] or larger[2] coalitions. The present work does not offer arguments for a plausible coalition size, or ways to increase it despite pessimistic bargaining solutions (which typically requires optimism in other respects[3]). Instead, our insights concern what any group of countries, insisting on each other's participation within a Subgame-Perfect Nash Equilibrium (SPNE), may wish to coordinate on. In our calibrated model, we entertain a mutually beneficial treaty between the US, China, EU and India, which we estimate would together reap 75% of the global net present benefit of reducing global warming.

It is helpful to review the types of emissions treaties proposed to date, and point out how the 'bridge' of arguments required to attain cooperation may be vulnerable to disruption by geopolitical events. The textbook (Barrett (2005)) 'baseline' relies on dynamic 'pledge-and-

[1] The most pessimistic bargaining models would consider the party with the smallest willingness to pay for abatement (India, among our four) refusing participate in the treaty for a while – against its own interest within the SPNE. Might the remaining three eventually accept its refusal and renegotiate to another cooperative equilibrium among each other, rather than carry out the (not individually, but collectively) costly threat of non-cooperation? Who would give in first? Notions of 'renegotiation-proof' or 'collectively rational' equilibria (e.g. Barrett (2002)) or 'stable coalitions' (e.g. Nordhaus (2015)) pessimistically assume the group of countries always gives in. The argument is that they have more to lose, collectively.

[2] Could it not be credible for the remaining countries to delay cooperation until the refusing country joins, given that it should have an incentive to? Why would indivudal countries get to be 'temporarily irrational' within the SPNE to test the credibility of (individually) rational threats, but groups of countries (which require unanimity) don't get to be while enforcing them? E.g. Schmidt and Streekstra (2022) model this negotiation process dynamically and find that a coalitions may 'collectively' prefer to delay cooperation to insist on designated non-members to join, possibly supporting even a grand coalition.

[3] E.g. that policymakers are willing and able to (further) escalate sanctions in some 'more important' game (which may be impossible during global conflict), or some form of increasing returns to scale of cooperation (e.g. manufacturing costs will fall faster than capacity factors even as the share of renewables becomes large).

review' strategies in an iterated prisoner's dilemma: countries negotiate a set of near-term emissions targets. Once the 'near-term' has passed, countries review whether the targets were met. This is the de-facto approach of current international agreements around carbon emissions. The main incentive to comply is created by threatening to reduce future cooperation if targets are missed. Such a threat only has enough bite if the promise of indefinite cooperation following compliance is credible to begin with, and verification is possible at a high enough frequency to minimize free-riding incentives within each iteration. The current unwillingness of countries to engage in significant abatement may thus be in anticipation that cooperation may break down either way, e.g. in periods of conflict, or because monitoring emissions at high frequencies is difficult.

To increase the bite of threats, researchers proposed to 'link' them to other interactions between countries. One proposal are 'carbon linked bonds', in which countries' external debt obligations increase if emissions targets are missed. However, it is difficult to rule out circumstances in which countries choose to inflate away or default on their debt. Another proposal is to link emission targets to trade: Nordhaus (2015, 2021) proposes 'Climate Clubs' in which parties agree to reduce their emissions and penalize missed targets via trade sanctions. The proposal may be less effective if tariffs are costly for sanctioning countries or if sanctioned countries retaliate (Böhringer et al. (2016); Barrett and Dannenberg (2022)), but it may survive some of these concerns (Farrokhi and Lashkaripour (2021)). However, policymakers may view sanctions as a *non-credible threat*, particularly in times of global conflict: sanctioning allies may cause geopolitical damage outweighing the benefit of the treaty, while adversaries are likely already heavily sanctioned.

Motivated by these concerns, we study how the probability of future break-downs of cooperation changes what countries can and should coordinate on today. We find that cooperating on subsidies for expanding renewable energy capacity becomes *more efficient* than cooperating on emissions or carbon taxes, and supports deeper abatement. Even if breakdown probability and opportunity costs are too large to support emissions treaties in a SPNE, subsidy treaties may still be feasible. The key mechanism underlying this theoretical result is the unique *cost structure* of renewable energy generation: a large enough subsidy – financed by those who benefit most from abatement – for the (relatively high) upfront fixed-cost of e.g. solar panels, wind turbines, EVs or grid-scale energy storage can make it profitable for any country to expand capacity more than it would otherwise. Once the fixed costs are sunk, partial displacement of fossil substitutes can be economically guaranteed to persist independently of cooperation, because renewables have *lower variable costs* than their fossil counterparts. No subsequent monitoring required. In contrast, emission targets or carbon taxes which are not anticipated to persist cannot incentivize the same investments – they mainly incentivize temporary consumption reductions which are suspended once cooperation breaks down.

Subsidy treaties also afford the practical advantage of not requiring high-frequency nationwide CO_2 monitoring, the feasibility of which is often assumed implicitly by emissions treaty proposals, but hard to justify. Mere announcements of e.g. a globally optimal carbon tax may not be reliable, as there is an incentive e.g. to covertly refund tax revenues to energy suppliers to keep coal plants running and gasoline prices at nationally optimal levels. Current high-frequency emissions estimates are indirect – they rely on correlations with data sources which may get distorted once there is an incentive to do so ('Goodhart's Law'). Satellites capable of direct CO_2 measurements have very limited spacio-temporal coverage. Thus, establishing a party deviated from an emissions treaty and then suspending cooperation or imposing sanctions seems ambitious even at yearly intervals. The longer this interval, the higher the opportunity cost of cooperation[4]. We do not model this challenge to avoid ruling out emissions treaties trivially, but it is worth contrasting it here with the comparatively transparent transactions repeated in subsidy treaties.

In a subsidy treaty, the monitoring problem reduces to verifying that a certain amount of renewable generation units were either installed or produced in each designated country. The benefit of the former is that installation costs, in addition to unit costs, could be subsidized. Subsequently, other countries either 1) suspend their own future capacity targets if a country is detected to deviate from its targets today, or they 2) subsidize each unit of additional capacity by agreed-upon amounts once verified. A benefit of 2) is that it could be institutionalized as an international body staffed with delegates from each party, which finances the subsidies at the end of a period via contributions it requires at the start of a period. If any party decides to stop cooperating, it would be detected at the contribution stage, when refunds are still possible before investments are made, which eliminates the 'free-riding' incentive but places limited trust in the body. The shorter the period in which renewable generators can be manufactured, installed and verified, the lower the budget the body is entrusted with at a given time.

To study the benefits of various treaties, we model the incentives of policymakers of national economies calibrated to the US, China, EU and India, who get to engage in a conditional transaction each period, facilitated by a (possibly imperfect) international body. Policymakers are assumed to pass policies and participate in treaties if doing so has unanimous approval by their population, taking the strategies of other countries as given. The structural model of national electricity sectors involves endogenous capacity expansion and unit commitment. We take as given variations in demand and supply of different technologies throughout the day and year, and currently preinstalled capacities, as reported in IEA data. We maintain the historically constant share of global income spent on electricity (i.e. assume unit income-elasticity), to account for increased demand as the subsidies reduce electricity prices and countries grow richer. We calibrate national incomes to the RFF socioeconomic projections of Rennert et al. (2022), and adjust them for avoided climate damages under abatement, via a reduced-form relationship between temperature and log(GDP) levels we estimate in the spirit of Burke et al. (2015). We take fixed- and variable-cost estimates of various generation technologies from Jacobson et al. (2022). To capture each government's cost-of-capital, we fix its 10-year bond rate at its respective Aug 2022 value.

Simulating our model, we find that a subsidy treaty between the US, China, EU and India could abate between $45 - 55\%$ of emissions from electricity in each country, relative to the

[4] In iterated prisoners' dilemma models, this would be captured by stronger discounting between periods, increasing the free-riding incentive. We will consider a more abstract notion of opportunity costs of cooperation, which may also stand in for e.g. an international institution misappropriating funds.

nationally optimal 'business-as-usual' policies over the next century. Net of costs, under current bond prices, it would avert a net present value destruction of around $8T in these countries over the next century, comparable to the COVID-19 pandemic. This is achieved by an approximate doubling of renewable installations, which is mutually beneficial even without net-transfers. If the risk of cooperation breaking down is larger than 5% annually (under a risk-neutral measure), subsidy treaties become more efficient than emissions treaties and thus survive higher opportunity costs of cooperation. Interestingly, subsidy treaties remain effective even if cooperation is attempted only every 5-10 years, which becomes optimal under significant breakdown risk. Under a subsidy treaty, average electricity prices in all countries would fall by $10-20\%$. If subsidies are to be financed via national debt, governments would take on additional debt of less than 10% of their GDP early on, and could fully repay it within a century by taxing additional income from avoided damages. Beyond renewable electricity generation, gas would remain a key part of the electricity mix in all countries, while coal would be increasingly phased out.

While we are not aware of a subsidy treaty having been explicitly discussed in previous literature, similar mechanisms to ours have been examined before. Harstad et al. (2019, 2022) modeled a stylized notion of investment in 'abatement technology R&D' that could change abatement costs in a repeated prisoners dilemma, and found that for a certain range of discount factors, the optimal treaty would mandate both emissions targets and R&D investment levels for each country. Behmer (2023) analyzed an economy similar to ours at the national level, finding that a 'green party' worried about its prospects for reelection would adopt subsidies in addition to taxes, due to a similar mechanism.

The paper is organized as follows. Section 2 presents our model: Section 2.1 outlines the general equilibrium model of each national economy, and 2.2 defines the strategic equilibrium between policymakers and our notion of treaties. Section 3 presents our main theoretical results. Next, we calibrate our model in Section 4, in order to make quantitative predictions about the impact of various treaties in Section 5. Section 6 concludes. Proofs of our theoretical results can be found in Appendix A.

2 Model

We will present our model in two steps. First, 2.1 outlines our model of the national economies which, given national policy and anticipated foreign emissions, attain a general equilibrium between heterogeneous electricity consumers and generation technologies who choose optimal quantities given prices. In the second step, 2.2 will define the game between governments, in which they adopt national policies and make compliance decisions with a particular treaty. Governments make Pareto-optimal choices for their population, changing behvior only with unanimous approval by their population (thus having incomplete preferences), taking the choices of other countries as given, in a Subgame-Perfect Nash Equilibrium. A treaty will be a sequence of voluntary conditional transactions between governments, facilitated by an international body that may, with some probability, embezzle funds under its control.

2.1 Structural Model Of National Economies

We will describe the economy of country $i \in \{1, ..., n\}$ starting in year $\tau \geq 0$ and extending until $T \leq \infty$.

Consumers. There is some measure $\mu_i(\theta)$ of each type $\theta \in \Theta$ of individual whose lifespan extends across any subset of future years $t = 0, ..., T < \infty$. Consumers have three types of endowments: they anticipate an exogenous income stream $(I_{it}^{\theta})_{t\geq 0}$, from which externalities $d_{it}^{\theta}(e_t)$ are subtracted that depend on global stock of carbon e_t in the atmosphere. Individuals may receive lumps-sum transfers τ_{it}^{θ} from their government, and own shares in the future profits $(\pi_{ijt}^{\theta})_{t\geq 0}$ of an electricity producer with generation technology $j \in J$. Consumers choose a vector of electricity consumption $\vec{Q}_{it}^{\theta} = (Q_{itk}^{\theta})_{k=1}^{K}$ in each subperiod k of the year, as well as their consumption of the non-electricity good X_t^{θ} and may issue debt b_{it}^{θ}, to maximize some (time-consistent continuation) utility U_i^{θ} from future consumption:

$$
\begin{aligned}
\forall \theta \in \Theta : \max_{(\vec{Q}_{it}^{\theta}, X_{it}^{\theta}, b_{it}^{\theta})_{t\geq 0}} \; & U_i^{\theta}\left((\vec{Q}_{it}^{\theta}, X_{it}^{\theta})_{t\geq 0}\right) \\
\text{s.t. } & \vec{p}_{it}^{\,Q} \times \vec{Q}_{it}^{\theta} + X_{it}^{\theta} \leq I_{it}^{\theta} - d_{it}^{\theta}(e_t) + \tau_{it}^{\theta} + b_{it}^{\theta} + \sum_{j\in J} \pi_{ijt}^{\theta}, \quad t = 0, ..., T \leq \infty \\
& \sum_{t\geq 0} \beta_{it} \cdot b_{it}^{\theta} = 0
\end{aligned}
$$

subject to repaying bonds at market prices β_{it}, which imply interest rates $r_{it} := \beta_{i(t+1)}/\beta_{it} - 1$.

Electricity Producers. There are $|J|$ different types of representative electricity producers, each of which specializes in particular generation technology $j \in J$. They are endowed with a preinstalled capacity K_{i0j} at time $t = 0$, and their output is taxed at a rate of τ_{itj}^{Q}. They choose whether to expand their future capacity by demanding new generation units $(\Delta K_{itj})_{t\geq 0}$, at a price of p_{itj}^{K}. Their generation profile is bounded as $\vec{Q}_{itj} \leq \vec{\gamma}_{ij} K_{itj}$ by j's vector of capacity factors $\vec{\gamma}_{ij} = (\gamma_{ijk})_{k\in K}$ in each subperiod k. Firms maximize their value to shareholders:

$$
\begin{aligned}
\forall j \in J : \max_{(\vec{Q}_{itj}, \Delta K_{itj}, K_{i(t+1)j}, \pi_{itj})_{t\geq 0}} \; & \sum_{t\geq 0} \beta_{it} \cdot \pi_{itj} \\
\text{s.t. } & \pi_{itj} = (\vec{p}_{it}^{\,Q} - \vec{\tau}_{itj}^{\,Q} - \mathbf{1} c_{ij}^{Q}) \times \vec{Q}_{itj} - c_{ij}^{O} K_{itj} - p_{itj}^{K} \Delta K_{itj} \\
& \vec{Q}_{itj} \leq \vec{\gamma}_{ij} K_{itj} \\
& K_{i(t+1)j} \leq K_{itj} e^{-1/\lambda_j} + \Delta K_{itj}, \quad \Delta K_{itj} \geq 0 \\
& K_{i(t+1)j} \leq B_{ij}, \quad \Delta K_{itj} \leq G_{ij} K_{itj}
\end{aligned}
$$

where B_{ij} parameterizes a bound on the total capacity of technology j that can be installed in a given country, and G_{ij} bounds how fast installed capacity can grow each year. Aggregate emissions are the sum of national emissions $e_t = \sum_i e_{it}$, which are the sum of historical emissions $e_{it} := \sum_i \sum_{\tau<t} \sum_j e_j^Q \mathbf{1} \times \vec{Q}_{i\tau j}$ from each technology.

Manufacturers. Generation capacity is supplied by a profit maximizing representative manufacturer incurring a tax τ_{itj}^K and constant marginal costs c_{it}^K:

$$\max_{\Delta K_{itj}^S} (p_{itj}^K - \tau_{itj}^K - c_{it}^K) \Delta K_{itj}^S$$

Market Equilibrium with externalities. Given the anticipated national policy $\vec{P}_i = (P_{it})_{t \geq 0}$, $P_{it} = (\tau_{it}^\theta, \tau_{it}^Q, \tau_{it}^K, b_{it}^G)$, where b_{it}^G is the government's budged deficit in year t, and the anticipated emissions of other countries $(e_{\imath t})_{\imath \neq i, \tau > t}$, a market equilibrium $E_{i\tau}$ in country i starting in period τ is a path of prices and quantities:

$$\vec{E}_i = (E_{it})_{t \geq 0}, \quad E_{it} := (\vec{p}_{it}^{\,Q},\ p_{itj}^K,\ \beta_{it}, \vec{Q}_{it}^\theta, \vec{Q}_{itj}, K_{itj}, \Delta K_{itj}, b_{itj}^\theta, e_{it})$$

solving the optimization problems above, while clearing all markets $\forall t \geq 0$:

$$\int_\Theta \vec{Q}_{it}^\theta \mathrm{d}\mu_i(\theta) = \sum_j \vec{Q}_{itj}, \quad \int_\Theta \pi_{itj}^\theta \mathrm{d}\mu_i(\theta) = \pi_{itj}, \quad b_{it}^G + \int_\Theta b_{itj}^\theta \mathrm{d}\mu_i(\theta) = 0, \quad \Delta K_{itj} = \Delta K_{itj}^S$$

with global emissions satisfying $e_t = \sum_i e_{it}$. Note that there is no market for cross-country debt, which will allow us to calibrate differences in the real costs of capital between countries without modeling national default risks. Another market incompleteness is that multilateral 'Coasean' trades are infeasible both within and between countries, creating a role for policy.

Histories of External Events While our notation left it implicit so far, our model of the national economies may incorporate uncertainty about future exogenous events, which will be important for the next section. In this case, all variables are understood as mappings from possible histories $h_t = (s_t, h_{t-1}) \in \mathcal{H}_t$ of exogenous events $s_t \in S$ into the respective Euclidean domain. This includes all choice variables $\vec{Q}_{itj}[h_t], \vec{Q}_{it}^\theta[h_t], \pi_{itj}^\theta[h_t], \pi_{itj}[h_t], ...,$ the endowments and externality functions $I_{it}^\theta[h_t], d_{it}^\theta[h_t](\cdot)$, the prices $\vec{p}_{it}^{\,Q}[h_t], \vec{p}_{it}^{\,K}[h_t], \beta_{it}[h_t]$ and the policy variables $\vec{\tau}^\theta[h_t], \vec{\tau}^Q[h_t], \vec{\tau}^K[h_t], ...$ themselves. The same holds for the parameters of the production technologies. All (in)equalities, and operations (addition, $\times$-product across subperiods k, multiplication by juxtaposition) are applied 'pointwise' at the respective histories. The only exception is the $\cdot$ product, which denotes the inner product over possible histories $\beta_{it} \cdot b_{it}^\theta := \sum_{h_t \in \mathcal{H}_t} \beta_{it}[h_t] b_{it}^\theta[h_t]$ and $\beta_{it} \cdot \pi_{itj} := \sum_{h_t \in \mathcal{H}_t} \beta_{it}[h_t] \pi_{itj}[h_t]$ at a given time t. Preferences are defined over these mappings, and implicitly define subjective probabilities and risk-preferences over the history tree $(\mathcal{H}_t)_t$. In this interpretation, $\beta_{it}[h_t]$ becomes the price of an Arrow-Debreu security issued at $t = 0$ that repays \$1 if and only if history h_t occurs.

2.2 Treaties and Strategic Equilibrium Between Policymakers

This subsection will define and discuss the notion of treaties between countries we will employ, and formalize the strategic equilibrium between policymakers of each country.

Various models for international climate agreements have been considered in the literature, but their differences should not be understood as an ongoing debate about the 'right' way to model international agreements. Instead, they parsimoneously emphasize different challenges and opportunities for incentivizing cooperative abatement. They may also capture different suggestions for how a given treaty may be implemented in practice. For example, Barrett (2005) presents the simplest model in which the cooperation failure can be understood: a static prisoner's dilemma between countries. One prospect for cooperation arises if this game is infinitely repeated, as considered in e.g. Barrett (2005); Harstad et al. (2019), if parties are 'patient enough' (=defection is detected fast enough) for the threat of reversion to noncooperation to bite. This is suggestive of non-binding 'pledge-and-review' agreements, the de-facto current approach to climate negotiations. Alternatively, cooperation can be sustained even in a static model as in e.g. Nordhaus (2015), if sanctions on other countries may depend on their policy in the same period. Here, the practical suggestion is for groups of cooperating countries to use economic threats to discourage members from reverting to non-cooperation and encourage non-members to join.

We entertain an alternative conception of an international treaty $\Gamma = (\Gamma_t)_t$, as a sequence of voluntary conditional transactions between governments. These are facilitated by some international body (e.g. the UN or WTO) which generally sticks to its commitments, but may be corrupted and embezzle funds with some probability. We adopt this model, which is suggestive of the institutionalized implementation proposal discussed in the introduction, mainly for simplicity: it can model cooperation in a non-static, finite horizon economy, while limiting number of counterfactuals we must numerically solve in Section 5. In principle however, our theoretical results would have analogues in an iterated prisoner's dilemma and no international body. We will restrict our attention to transactions of the form $\Gamma_t(C_{t-1}, \vec{Q}_t, \Delta K_t) = R_t$, where $C_{t-1} = (C_{it-1})_i$ are voluntary deposits that each government makes just before h_t is realized, and $R_t = (R_{it})_i$ are refunds that each government receives at the end of period t, which may depend on the national equilibria realized in that year, specifically on $\vec{Q}_t = (\vec{Q}_{it})_i$ and $\Delta K_t = (\Delta K_{it})_i$. The body must balance its budget in any event $\sum_i C_{it-1} \geq \sum_i R_{it}$. The refunds $R_t[h_t]$ may depend on the external events s_t realized in that year. We let history evolve according to

$$h_t = (s_t, h_{t-1}), \quad s_t \in \{\text{normal}, \text{breakdown}\}$$

such that we can model an exogenous breakdown of cooperation by restricting our attention to Nash Equilibria in which $s_t = \text{breakdown}, C_t > 0 \implies C_\tau[h_\tau] = 0 \; \forall \tau > t$. I.e. we focus on equilibria in which failed attempts to cooperate cause a permanent reversion to non-cooperation. We consider the following two types of treaties, implemented by a body that may embezzle funds in the event of corruption:

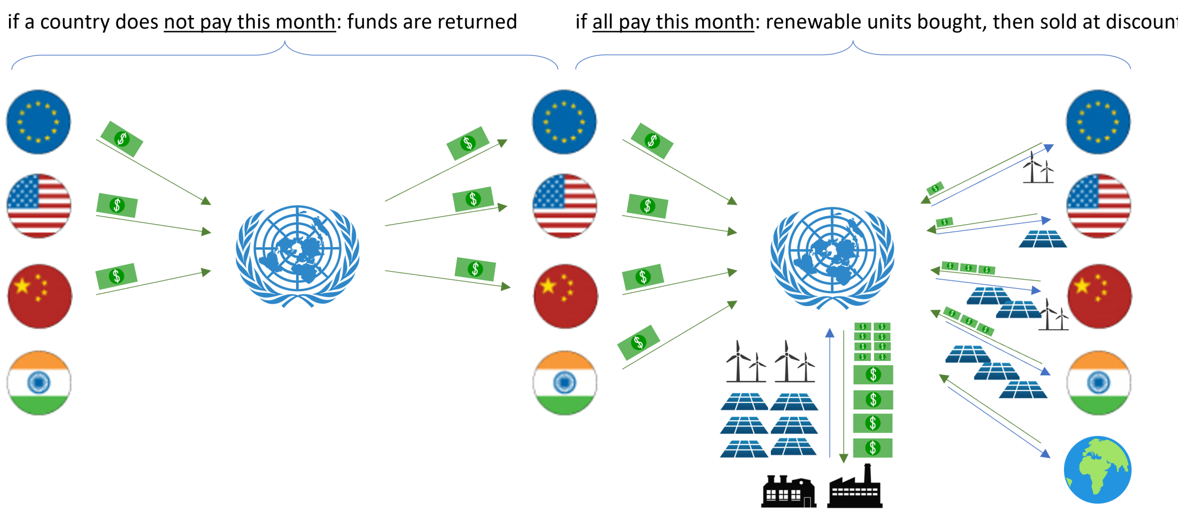

Figure 1: Implementation sketch of a subsidy treaty between CHN, US, EU, and IND, in which non-parties might also be allowed to purchase subsidized renewable infrastructure.

Emissions Treaty The international body implements an emissions treaty if

$$\Gamma_t(C_{t-1}, \vec{Q}_t, \Delta K_t)[h_t] = \begin{cases} C_{t-1}[h_{t-1}] - V_t[h_t] & \text{if } \exists \tau \leq t : s_\tau = \text{breakdown} \\ \overline{R}_{it}[h_t] - \sum_j \overline{\tau}^Q_{jt}[h_t] \vec{Q}_{itj}[h_t] & \text{else if } C_{t-1} = \underline{C}_{t-1} \\ C_{t-1}[h_{t-1}] & \text{else} \end{cases}$$

i.e. refund some fixed payment $\overline{R}_{it}[h_t]$ minus a carbon tax if and only if everyone contributed the prescribed amount $\underline{C}_t$, else refund the contributions unconditionally. In the event of a breakdown, the body may destroy some 'value-at-risk' $V_t[h_t] \geq 0$ of the contributions, which is a placeholder for any opportunity costs of cooperation, e.g. a limited trust in the body itself.

Subsidy Treaty The international body implements a subsidy treaty if

$$\Gamma_t(C_{t-1}, \vec{Q}_t, \Delta K_t)[h_t] = \begin{cases} C_{t-1}[h_{t-1}] - V_t[h_t] & \text{if } \exists \tau \leq t : s_\tau = \text{breakdown} \\ \sum_j -\overline{\tau}^K_{jt}[h_t]\Delta\vec{K}_{itj}[h_t] & \text{else if } C_{t-1} = \underline{C}_{t-1} \\ C_{t-1}[h_{t-1}] & \text{else} \end{cases}$$

i.e. providing a subsidy $-\overline{\tau}^K_{jt} \geq 0$ for the fixed costs of added (renewable) generation units in the next period if and only if everyone contributed the prescribed amount, else refund the contributions unconditionally. This transaction is equivalent to a body purchasing renewable generators if it received sufficient deposits, 'marking' them to ensure the same unit is not subsidized twice, and reselling them at a discount. This is visualized in Figure 1.

Nash Equilibrium. Policymakers in each country choose their national policy and compliance decisions with the treaty in a way that is Pareto-optimal for their population, given the decisions of other countries. Formally, a strategic equilibrium is a tuple $(\mathcal{P}^*, \mathcal{E}^*, C^*)$, where $C^*_i = (C_{it})_t$ prescribes a contribution $C_{it}[h_t]$ for each history, and the mappings $\mathcal{E}^*_i\,(h_t, (C_\tau[h_\tau])_{\tau<t}) = E_{it}[h_t]$ and $\mathcal{P}^*_i\,(h_t, (C_\tau[h_\tau])_{\tau<t}) = P_{it}[h_t]$ prescribe a national equilibrium and a policy for each possible contribution path up to a given history. Given a treaty Γ and the strategies of other countries, the strategy $(\mathcal{P}^*_i, \mathcal{E}^*_i, C^*_i)$ of country i must satisfy:

- **Budget-Balance:** For any contribution history $(C_t)_t \geq 0$, the national equilibria $(E_{it})_{t>0}$ prescribed by $\mathcal{E}^*$ satisfy the market equilibrium conditions given $\mathcal{P}^*$ and the budget constraint of the policymaker:

$$R_{it} - C_{it-1} + \int_\Theta \tau^\theta_{it}\mathrm{d}\mu_i(\theta) + \sum_j \vec{\tau}^Q_{ijt} \times \vec{Q}_{itj} + \tau^K_{ijt}\Delta K_{itj} + b^G_{it} = 0$$

$$\sum_{t\geq 0} \beta_{it} \cdot b^G_{it} = 0, \quad R_t = \Gamma_t(C_{t-1}, \vec{Q}_t, \Delta K_t)$$

- **Pareto-Optimality:** No country i can Pareto-improve its population $U^\theta_i, \theta \in \Theta$ after any possible history h_t or contribution path $(C_t[h_t])_{t<\tau}$ by unilaterally deviating to a strategy $(\mathcal{P}'_i, \mathcal{E}'_i, C'_i)$ differing in future periods, without violating Budget-Balance.

Note that we ruled out dynamic strategies in which contributions may depend on historical contribution paths, or in which national policies may dynamically depend on historical policies by other countries. These cases are not necessary for our core argument and ruling them out simplifies both the theoretical and computational analysis of our model.

3 Theoretical Results

In this section, we will present our theoretical results. First, we establish the unilateral 'business-as-usual' (BAU) scenario and its inefficiency. Next, we present our core argument: that subsidy treaties can be more efficient than emissions treaties in equilibria that are 'pessimistic' about future cooperation, and survive a higher risk of corruption of the international body facilitating the treaty.

Proposition 1 (Unilateral Equilibrium: BAU)**.** *Let $T < \infty$ and assume the preferences $U_{i\tau}^{\theta}$ have the typical convexity and non-saturation properties as in Debreu (1959). Further, let $d_{it}(\cdot)$ be strictly convex and continuuous, and consider the case without a treaty $\Gamma_t = C_t = 0$. Then in all Nash equilibria, governments levy a nationally optimal carbon tax:*

$$\tau_{ij\tau}^{Q} = e_j^Q \beta_{i\tau}^{-1} \sum_{t>\tau} \beta_{it} \cdot \nabla_{e_t} d_{it}(e_t), \quad d_{it}(e_t) := \int_{\Theta} d_{it}^{\theta}(e_t) \, \mathrm{d}\mu_i(\theta)$$

equal to the marginal future damages from emissions in country i, discounted by the prices $\beta_{it}[h_t]$ of Arrow-Debreu securities.

Proposition 1 is reminiscent of the Coase (1960) Theorem, valid under more general (not necessarily quasi-linear) preferences, but with a slightly weaker conclusion: it is the qualitative characterization of the optimal policy – a carbon tax equal to the NPV of damages, together with lump-sum reimbursements – that is independent of initial endowments, while the resulting level of emissions may depend on endowments. It is noteworthy that this result implies that under complete markets, unless it is already in place, every individual in this economy could *benefit* if a nationally optimal carbon tax is introduced – whether they are consumers, shareholders of renewable or fossil energy producers, older or younger or future generations (corresponding to θ who only care about consumption in particular years), borrowers or lenders, whether they live in a low-income or high-income country, and even if they have *subjective probabilities* about the likelihood of damages from global warming. This contrasts with the fact that special interest groups or disagreements about probabilities are often viewed as the main obstacle to climate change mitigation, and even climate advocates commonly argue that consumers or shareholders of fossil energy producers would have to give up some standard of living for the greater good. The proof of Proposition 1 in the appendix characterizes the specific transfer scheme required to ensure carbon taxes are a Pareto-improvement: consumers are compensated for higher energy prices using the increased revenues of energy producers, fossil shareholders are compensated for higher carbon taxes using government debt that is repaid over time by taxing the additional future income from avoided climate damages, and borrowers are reimbursed for higher interest rates using the additional income enjoyed by lenders. However, this static equilibrium is still inefficient, as captured by the next result.

Proposition 2 (Optimal Global Policy)**.** *Regardless of which policies $\underline{P}_{\tau} = (\underline{P}_{i\tau})_i$ are in place, every individual would prefer if all governments jointly adopted the global policy $P_{\tau} =$*

$(P_{i\tau})_i$ instead, in which each country levies the globally optimal carbon tax $\bar{\tau}^Q_{j\tau}$:

$$\bar{\tau}^Q_{j\tau} = \sum_i \tau^Q_{ij\tau},$$

with $\tau^Q_{ij\tau}$ characterized as in Proposition 1. The within-country transfers τ^θ_{it} that are necessary to ensure this policy change constitutes a Pareto-improvement are compatible with national budgets after cross-country net transfers:

$$R_{it} - C_{it-1} = \sum_j \left(\tau^Q_{\imath tj} \sum_\imath \mathbf{1} \times (\vec{Q}_{\imath tj} - \underline{\vec{Q}}_{\imath tj}) - \bar{\tau}^Q_{jt} \mathbf{1} \times (\vec{Q}_{itj} - \underline{\vec{Q}}_{itj}) \right)$$

where quantities before and after the policy change are denoted with and without an underline respectively.

Proposition 2 shows that the populations of all countries could Pareto-improve upon the nationally optimal policies of Proposition 1 by jointly raising the carbon tax to a uniform globally optimal level. In this case, the reimbursements needed to ensure a Pareto-improvement now require cross-country net transfers, in which countries that are able to decarbonize more than others are reimbursed using funds from countries with higher net present benefits than others. While it seems plausible that emissions treaties are therefore the only treaty worth considering, the next result shows that this intuition is misleading.

Proposition 3 (Advantages Of Subsidy Treaties). *To make probabilities explicit, consider agents with Von-Neumann-Morgenstern preferences and let s_t be iid. Consider only 'fragile' equilibria in which $s_t = breakdown \implies C_\tau[h_\tau] = 0 \ \forall \tau > t$. Then, given appropriate choices for the structural parameters, if $P(s_t = breakdown)$ is high, each emissions treaty is Pareto-dominated by a subsidy treaty. If the value-at-risk $V_t[h_t]$ is high, countries will refuse to participate in any emissions treaty, while a subsidy treaty may still be sustained in equilibrium and Pareto-improve upon BAU.*

Appendix A.3 proves the result for a particularly simple economy, choosing the structural parameters carefully: the proof exploits that renewables have higher fixed- and lower variable-costs than (and are gross substitutes for) fossil fuels. This can make it unattractive to invest signficantly into renewable capacity during BAU, and even under the prospect of temporary cooperation on a carbon tax. At the same time, short-term cooperation via subsidy treaties can offset the initial fixed-costs and make expansion profitable. Once the fixed costs are sunk, using renewables is cheaper than the alternative: abatement has *negative* variable opportunity costs. Thus subsidy treaties can economically guarantee abatement past the breakdown of cooperation, without the need for monitoring. Clearly, whether similar effects occur under a realistic calibration of the national economies is an empirical question, which we answer to the positive in the remainder of this paper.

Beyond the PV panels or wind turbines considered in this paper, e.g. electric vehicles and grid-scale batteries share similar properties and could also be incentivized via a subsidy treaty.

Figure 2 labels investments with this property 'self-enforcing abatement', because they can be used to compel another party to persistently reduce their emissions, after a voluntary one-time transaction. Abatement actions which do not have this cost-structure include e.g. land-use or consumption changes, or energy-intensive carbon capture. These cannot benefit from a subsidy-treaty, and thus require ongoing monitoring and are vulnerable to disruptions of cooperation. We hope future work can draw inspiration from this framework for tackling other cooperation problems, both at international and at smaller scales[5].

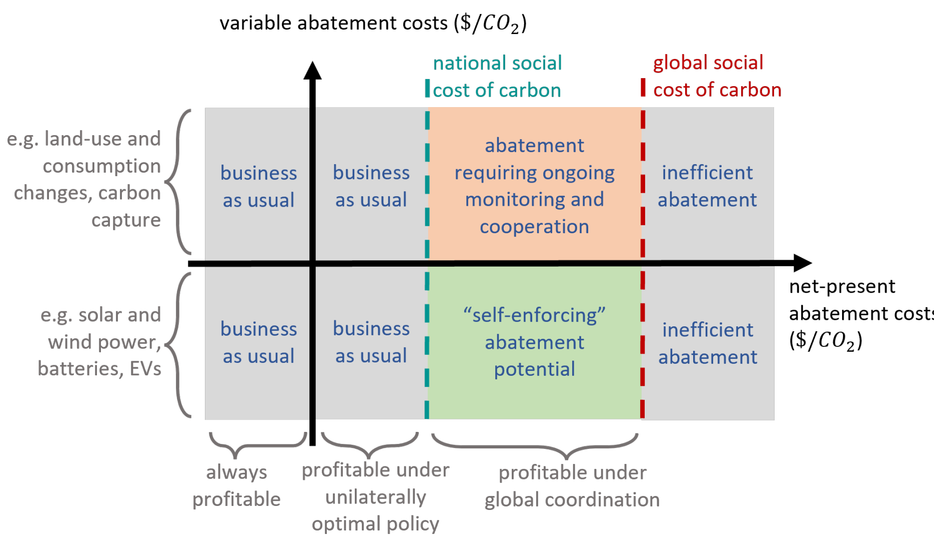

Figure 2: 'Self-enforcing' abatement has negative *variable* opportunity costs: its usage is *cheaper* than the higher-carbon alternative, once fixed-costs are sunk.

4 Empirical Calibration

This section will outline how we calibrate the preferences, technologies and damages from warming in each country, in order to make empirical predictions about the impact of different types of treaties on each country, and to derive mutually beneficial contribution targets. We will conduct a sensitivity analysis to the most crucial choices in Section 5.1.

4.1 Preferences

Clearly, renewable electricity sources are generally imperfect substitutes for fossil sources, because they don't produce electricity around the clock. Their substitutability depends on

[5]Arguably, the Strategic Arms Reduction Treaty (START I) at times used a similar mechanism: destroying weapons is a one-time fixed-cost which can be quickly verified: the US displayed hundreds of chopped-up B-52 nuclear bombers at the Davis-Monthan Air Force Base to aid Soviet verification. A 'destruction treaty' (a 'negative subsidy treaty'), can have persistent effects if a technology became less attractive, and is in use only because its fixed costs are sunk. It could be used for coal plants, once sufficient renewables are deployed.

the existing market shares of each technology: the first solar panel added to a grid can be a perfect substitute for fossil generation as its rated capacity is exactly the amount of 'useful electricity' it can contribute to the grid. Once there are enough solar cells in the grid to serve the full demand during some period of the year however, additional solar cells become less useful and thus solar becomes an imperfect substitute, or may even become complementary.

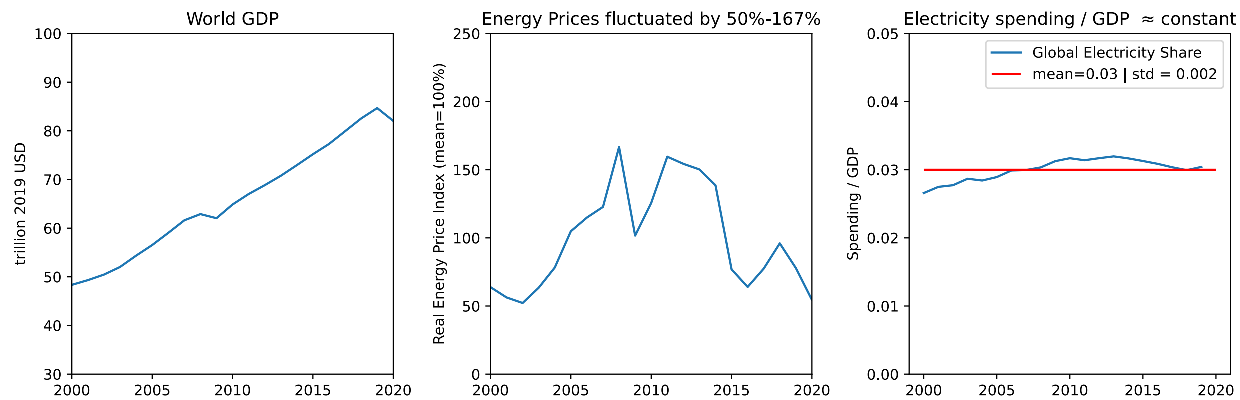

Figure 3: At the global level, electricity spending made up a remarkably constant share of GDP, despite large variations in prices and income levels, justifying a Cobb-Douglas demand aggregation of electricity and non-electricity goods.

We therefore model electricity demand within a given year as a CES aggeration of the electricity consumption Q_{itk} within each of 1000 hours $k \in \{1, ..., 1000\}$ randomly subsampled from the year, and conservatively assume an intertemporal elasticity of substitution of $\sigma = 0.01$. We further assume that the total expenditure share on electricity Q_{it} vs. the non-electricy numeraire X_{it} remains at its historically relatively constant (see Fig. 3) value α_i of 2.5% in US, EU and 4.2% in China and India. This implies a unit-price elasticity of aggregate demand, so our model will produce increases in demand as subsidies lower electricity prices. Within a given year, this yields the demand system:

$$\max_{\{Q_{itk}\}, X_{it}} \quad Q_{it}^{\alpha} \cdot X_{it}^{1-\alpha} \quad \text{s.t.} \quad Q_{it} := \left(\sum_k \omega_k^{\frac{1}{\sigma}} Q_{itk}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \quad \sum_k p_{itk} Q_{itk} + X_{it} \leq I_{it}$$

where the arg max has a closed form solution[6] $Q_k^d(\vec{p}_{it}|I_{it}) = p_{itk}^{-\sigma} \frac{\omega_k \alpha I_{it}}{\sum_k \omega_k p_{itk}^{1-\sigma}}$ for electricity demand[7]. We fix ω_k to be proportional to the respective hourly demand observed within each country during a 'representative year' available in the IEA Real-Time Electricity Tracker. For national incomes, we take the median national income projections from the RFF-SPs (Rennert et al. (2022)).

The representative consumers of different years are treated as different individuals, such that

[6]FOCs imply: $X = (1-\alpha)I$, $Q = \frac{\alpha I}{p}$, $p := \left(\sum_k \omega_k p_k^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$, $Q_k = \omega_k \left(\frac{p_k}{p}\right)^{-\sigma} Q \implies \frac{Q_k}{Q_\kappa} = \frac{\omega_k}{\omega_\kappa} \left(\frac{p_k}{p_\kappa}\right)^{-\sigma}$

[7]The FOCs[6] yield the inverse demand $p_k(\{Q_k\}) = \frac{\alpha I}{Q} \left(\frac{\omega_k Q}{Q_k}\right)^{1/\sigma} = \frac{\alpha I}{Q_k \sum_\kappa (\omega_\kappa/\omega_k)^{\frac{1}{\sigma}} (Q_\kappa/Q_k)^{\frac{\sigma-1}{\sigma}}}$

a Pareto-improving policymaker will only participate in treaties that allow it to make each 'generation' better off. To give rise to a market for government debt, we assume the existence of a representative, infinitely-lived 'bond-market investor' whose discount factor $\beta_i = \frac{1}{1+r_i}$ we calibrate to keep a country's bond rate on government debt fixed.

$$\max_{(b_{it})_t} \sum_t \frac{1}{(1+r_i)^t} \cdot b_{it} \text{ s.t. } \sum_t \beta_{it} \cdot b_{it} = 0$$

Because the income and cost projections we use are in real dollar terms, we calibrate r_i by taking each country's August 2024 nominal interest rate on 10y government debt, adding a 2% risk premium to be conservative, and subtracting long-run inflation rates of 2.3% for the US (the 10y US 'breakeven' rate in 2024), and for China, EU and India respectively, where breakeven rates are unavailable, select conservatively low inflation projections of 1%, 2.2%, 4%, to err on the higher end of real cost of capital. This yields $r_i = 3.9\%, 3.3\%, 3.0\%, 5.5\%$ respectively. Our approach effectively amounts to a partial equilibrium analysis, in which the price of government debt does not change between different equilibria. Rather than introducing a new parameter capturing intertemporal elasticities, our sensitivity analysis will consider a range of interest rates centered around these values. We believe this to be a reasonable approximation, as our main results will not involve increasing deficits by more than 0.3-0.5% of GDP in any given year.

4.2 Supply Parameters

Demand is served by a price-taking supply choosing how each generation technology $j \in \{S, W, C, G, N\}$ (solar, wind, coal, gas, nuclear) contributes to the electricity mix $Q_{itk} \equiv \sum_j Q_{itjk}$ given each j's current $\vec{K}_{it} = (..., K_{itj}, ...)$ and added capacity $\Delta\vec{K}_{it} = (..., \Delta K_{itj}, ...)$, to maximize profits as described in Section 3. We assume constant returns to scale in the manufacturing and operation of all technologies, trading off increasing returns to scale in manufacturing with the increasing marginal costs of resource extraction. We view this as conservative given that historically, at least for renewable energy, the former has dominated the latter.

j	S	W	C	G	N	H
c_j^Q (\$/kWh)	0.0404	0.0404	0.0696	0.0808	0.0496	0.0463
c_j^O (\$/kW/yr)	81.92	155.59	63.56	29.51	153.41	35.97
c_j^K (\$/kW)	6850.82	5045.68	10863.00	2916.22	18352.64	∞
e_j^Q (kgCO$_2$ /kWh)	0.00	0.00	1.024	0.442	0.00	0.00
e_j^K (kgCO$_2$ /kW)	0.00	0.00	0.00	0.00	0.00	0.00
λ_j (average years)	45	25	45	40	55	70

Table 1: Cost estimates from the 'high-cost low-benefit' scenario of Jacobson et al. (2022). e^Q is from EIA.gov.

The key supply side parameters are summarized in Table 1, which are taken from Jacobson et al. (2022). Following Jacobson et al. (2022), we multiply the costs c_j^Q, c_j^O, c_j^K with country and technology-specific cost multipliers given in Table 2. We uniformly subsample $k = 1, ..., 1000$ periods from hourly data on Q_{ijkt}, p_{ijkt} obtained form the IEA Real-Time Electricity Tracker, and normalize $\sum_k \pi_{jk} = 365 * 24$ such that K_j is measured in kW and Q_{jk} in kWh. For $j = S, W$, we set $\pi_{jk} \propto Q_{i\tau jk}$ observed in country i during the most recent available year, and keep π_{jk} constant for $j \neq S, W$.

m_{ij}	S	W	C	G	N	H
US	0.84	0.82	1	1	1	0.89
EU	0.71	1.11	1	1	1	0.89
CN	0.74	1.17	1	1	1	0.89
IN	0.77	1.17	1	1	1	0.89

B_{ij}	S	W	C	G	N	H
US	448865	3376	∞	∞	∞	79
EU	82295	1447	∞	∞	∞	103
CN	561088	2020	∞	∞	∞	339
IN	33816	908	∞	∞	∞	46

Table 2: Cost multipliers m_{ij} (left) and capacity bounds B_{ij} (right) estimated by Jacobson et al. (2022). We get $(c_{ij}^Q, c_{ij}^O, c_{ij}^K) = (c_j^Q m_{ij}, c_j^O m_{ij}, c_j^K m_{ij})$. m_{EUj} are GDP-weighted averages.

4.3 The Emissions Damage Function

4.3.1 Climate Model Parameters

In Figure 4, we plot simulation draws from the Coupled Model Intercomparison project 6. Assuming for now that all countries warm equally, we conclude that global warming can be reasonably well approximated by a linear relationship:

$$T_{it}(e_t) = T_{it}^{\text{BAU}} + \gamma_i(e_t - e_{it}^{BAU})$$

where e_{it}^{BAU} will be fixed to the emissions of our model in absence of a treaty, and T_{it}^{BAU} are taken to be the RCP7 temperature projections. We compute OLS estimates for γ_i individually for each country, where $\gamma_i \approx 0.00036$ °C/GtCO2 on average.

4.3.2 Estimating The Effect of Temperature Changes on National Income

To estimate the aggregate damage function $d_{it}(e_t)$ in our model, we employ a variant of the methodology of Burke et al. (2015), who estimate a global relationship $d_{it}(T_{it}) = \bar{d}(I_{it}, T_{it})$. We will use random weather variation to identify the marginal effect of temperature changes on the level of national income in each country. The key idea is to look at historical temperature and national income data within a country, and compare via a regression whether, on average, hot years (relative to trend) caused higher or lower national income (relative to trend), and vice versa for cold years. This tells us the slope of $\nabla_{T_{it}} \bar{d}(I_{it}, T_{it})$ at historical incomes and temperatures in each country. We can then examine how this slope varies

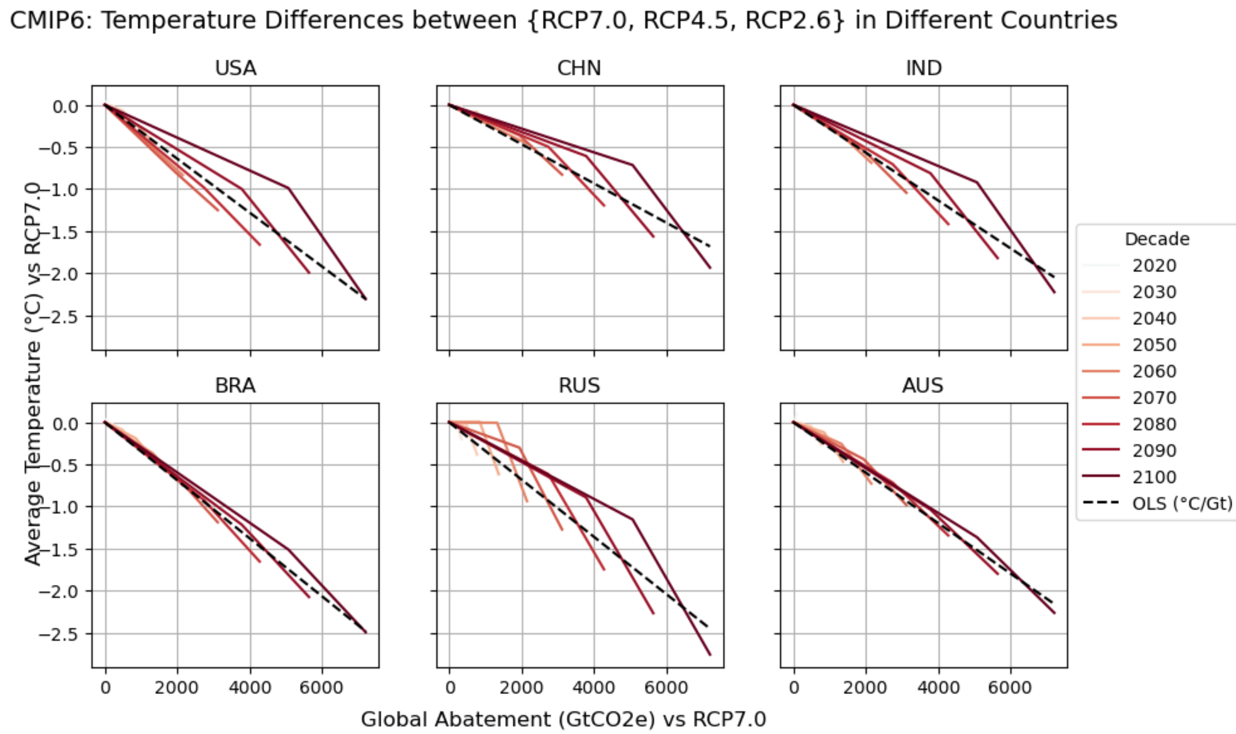

Figure 4: Cumulative emissions and global temperature differences across RCPs in different years are almost linear in each country, across simulation runs from the Coupled Model Intercomparison Project 6.

across countries, to back out an estimate for $d_{it}(T_{it}) = \bar{d}(I_{it}, T_{it})$. Estimating $\bar{d}$ based on data between 1980-2000 and applying it to predict the impact of hot years in various countries between 2000-2020, we will find that this approach almost universally underestimates marginal damages from warming, likely due to an implicit assumption of rapid adaptation, as we will discuss.

Deryugina and Hsiang (2017) provide a theoretical argument why weather can identify the marginal impact of climate: the Envelope theorem asserts that adaptation behaviors have no first-order effect that would create discrepancy between long-run vs short-run impact of marginal temperature changes. Burke and Emerick (2016) provide an empirical argument: finding that adaptation behavior is largely absent for historical climate change, they provide evidence this is driven by an inability to adapt rather than unawareness of its necessity. Other methodologies for estimating the damages from global warming, such as the bottom-up approach, have their own advantages (see Tol (2019) for an overview), but measuring the slope of $d_{it}(T_{it})$ directly, by simply comparing hot and cold years in given country, minimizes the scope for disagreements about contribution targets for each country and makes testable predictions which we will verify below.

Other works have typically estimated the marginal effects of weather-based temperature fluctuations *within a country*, via a regression capturing *"on average, in the past, for each ˚C that* T_{it} *was above (below) trend, i's GDP* I_{it} *was* $\widehat{\beta}_i$*% above (below) trend"*. We extend this approach to also include spillovers from warming in other countries $\imath$, which capture: *for each °C that T_{it} was above (below) trend, I_{it} was $\widehat{\beta}_{i\imath}$% above (below) trend.* Specifically, we estimate:

$$\widehat{\beta} = \arg\min_{\beta} \sum_{i,t} (\Delta \log I_{it} - \sum_{\imath} \beta_{i\imath} \Delta T_{\imath t})^2 + \lambda \sum_{\imath \neq i} \beta_{i\imath}^2$$

where $\Delta \log I_{it} = \log I_{it} - \log Y_{it}^{\text{trend}}$ and $\Delta T_{it} = T_{it} - T_{it}^{\text{trend}}$ are (linearly) de-trended. I.e. we assume weather variations are exogenous with respect to GDP, while climate and GDP *trends* may be historically confounded. We use real GDP from Penn World Tables for I_{it} (147 countries) and yearly average temperature from NASA MERRA-2 data for T_{it} (spanning 1980-2020). Cross-validation favors spillovers and sets $\lambda = 0.5$, which we fix throughout. The %-change of I_{it} from a marginal +1°C of global warming is then estimated by summing the OLS coefficients $\hat{\beta}_{i\imath}$, weighted by relative warming $\bar{\beta}_i := \frac{\partial \log I_{it}}{\partial T_{it}} \approx \sum_{\imath} \hat{\beta}_{i\imath} \frac{\gamma_\imath}{\gamma_i}$.

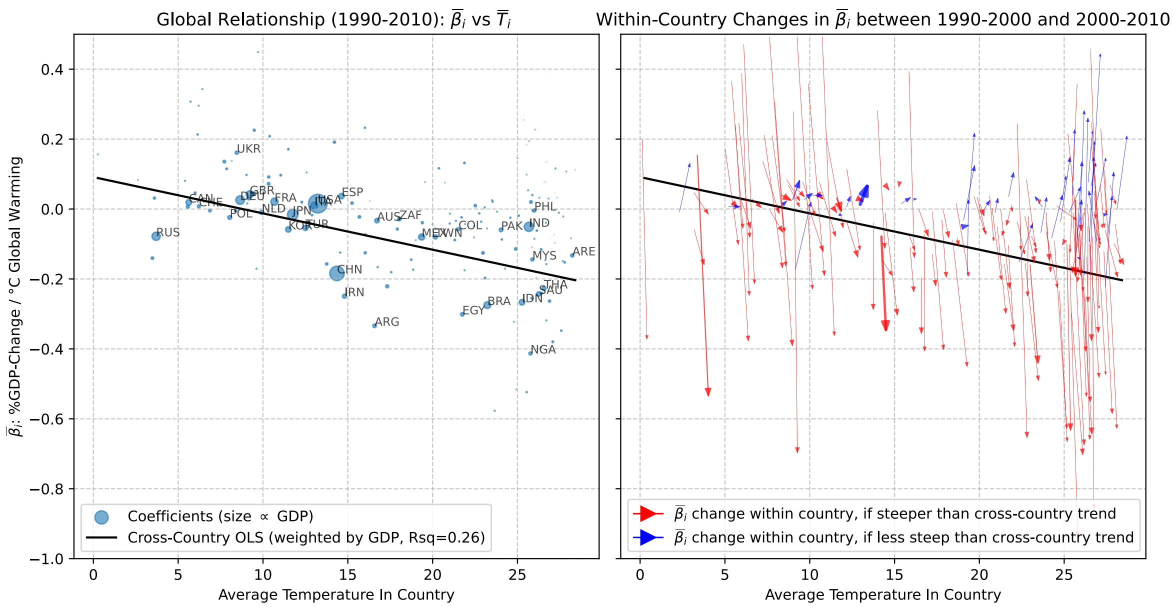

Figure 5: Left: scatterplot of $(\bar{T}_i, \bar{\beta}_i)$ across countries. Dot-size $\propto$ average GDP. Global (Cross-Country) Relationship (GDP-weighted OLS) shown in black. Right: Comparison of within-country change of $\bar{\beta}_i$ in 1980-2000 vs 2000-2020 with the global relationship. Most within-country changes are much steeper than the global relationship, so using the global relationship for Equation 1 is conservative.

Like Burke et al. (2015), we find in the left panel of Figure 5 that susceptibility to global warming is correlated with average local temperature $\bar{T}_i := \frac{1}{40}\sum_{t=1980}^{2020} T_{it}$: $\bar{\beta}_i \approx -0.01(\bar{T}_i - T^{\text{opt}})$, where warming seems to have no marginal impact at local temperatures around $T^{\text{opt}} = 8.8$°C. This suggests that, as countries get warmer, they will become more susceptible to further warming. Assuming the cross-country relationship $\bar{\beta}_i = \frac{\partial \log I_{it}(T_{it})}{\partial T_{it}} = -0.01(T_{it} - T^{\text{opt}})$ suggested by Figure 5 holds within countries, we can solve the differential equation for $I_{it}\,(T_{it})$:

$$I_{it}\left(T_{it}\right) := I_{it}^{\mathrm{BAU}} - d_{it}\left(T_{it}\right) = I_{it}^{\mathrm{BAU}} \exp\left(-0.01 \frac{\left(T_{it} - T^{\mathrm{opt}}\right)^2 - \left(T_{it}^{\mathrm{BAU}} - T^{\mathrm{opt}}\right)^2}{2}\right) \tag{1}$$

if we normalize $I_{it}(T_{it}^{\mathrm{BAU}}) = I_{it}^{\mathrm{BAU}}$ and fix $T^{\mathrm{opt}} = 8.8$°C. The implied $d_{it}\left(T_{it}\right)$ is log-convex[8].

In an effort to downplay their own willingness-to-pay, governments may object that there will be adaptation behavior which prevents marginal damages from warming to rise as quickly in their country as Equation 1 would suggest. Fortunately, we can put this hypothesis directly to the test, as average temperatures have been rising throughout our 40 years of data. Repeating the analysis on the years 1980-2000 and 2000-2020 separately, we can see whether $\bar{\beta}_i$ became more negative – as we'd expect given the average temperature rise within each country. We plot the resulting within-country changes of $\bar{\beta}_i$ in the right panel of Figure 5. We find that marginal damages universally increased much more sharply than we would predict based on the cross-country relationship $\bar{\beta}_i = -0.01(T_{it} - T^{\mathrm{opt}})$ underlying Equation 1. This suggests climate change is outrunning our ability to adapt, which is intuitive: historically, countries had thousands of years to adapt their economies to their local climate, while significant temperature changes now occur within countries over decades. Using the cross-country relationship for Equation 1 thus implicitly assumes countries can adapt infinitely quickly, yielding a lower-bound on each country's willingness-to-pay.

Taking I_{it}^{BAU} and r_i as in Section 4.1, and setting T_{it}^{BAU} to the CMIP6 projections for RCP7.0, we can compute the globally optimal carbon tax from Equation 1, i.e. the social cost of carbon (SCC) as characterized in Proposition 2. We obtain a worldwide SCC of 174\$/tCO2, of which 129\$/tCO2 fall within the USA, China, EU, and India. The fact that 74% of the SCC falls within these four entities motivates our focus on treaties between these countries.

5 Quantitative Results

We will simulate a subsidy and emissions treaties between China, USA, EU and India. We solve for a bond schedule $\{b_{it}\}_t$ that makes each generation within a country equally better off compared to the BAU scenario (in log-income terms). We consider a horizon of 128 years, starting the treaty in 2022, four years after the start of the simulated economy (2018).

5.1 Comparing The Impact Of Treaties

For now, we consider the case without breakdown risk $P(s_t = \text{breakdown}) = 0$ and assume no net transfers – i.e. taxes are refunded and subsidies financed within each country.

[8] Propositions 1 and 2 assumed convexity instead of log-convexity, but our quantitative results would not change meaninfully if we replaced this damage function with a convexified approximation.

To assess the sensitivity of our results to the parameter choices we made above, we multipy log-normal noise of 10% to the cost parameters, 25% to the interest rates, and 20% to the slope of Figure 5 (corresponding to its OLS standard error). We then report the distribution of changes in welfare, average electricity prices and cumulative CO2 emissions under the optimal emissions treaty vs a 100% subsidy treaty.

In Figure 6, we find that the subsidy treaty is a Pareto-improvement upon BAU even without net transfers between countries, creating a total NPV of around \$8T for the parties. In contrast, while the optimal emissions treaty would create around \$17T in NPV, it causes significant electricity price spikes in China, where coal plays a large role. These would have to be reimbursed for the treaty to be individually rational for China. Interestingly, a subsidy treaty leads to similar (ca 70%-80% from BAU) abatement in higher-income countries as an optimal emissions treaty. In lower-income countries, the subsidy treaty abates less than the optimal emissions treaty (ca 40% rather than 80% from BAU), indicating that income-effects play a larger role. In absolute terms however, under either treaty, most abatement would be realized in lower-income countries, which under BAU would pursue a fossil-fueled growth path similar to the one that higher-income countries already experienced.

In Figure 9 in the appendix, we show the distribution of the change in national deficits necessary to ensure each generation is equally better off (in log-income terms) under a subsidy treaty, compared to the BAU scenario. We find that it suffices to increase national deficits by $0.3 - 0.5\%$ of GDP in the first 30 years of the treaty, after which the deficits progressively decrease, ending up at around -2% of GDP to repay the debt by the end of the century.

5.2 Robustness to Pessimism about Sustained Cooperation

We now examine the impact of pessimism about sustained cooperation, by varying the probability of breakdown $P(s_t = \text{breakdown}) > 0$. We now numerically solve for lump-sum transfers between countries which distribute welfare gains equally (in log-income terms). We consider both the case without opportunity costs to cooperation ($V_t = \$0$) and the case where some value is destroyed in the event of breakdown ($V_t = \$2.5T$). Solving the model over multiple counterfactuals simultaneously creates significant computational overhead, so we restrict ourselves to the central parameter values calibrated in Section 4.

The results in Figure 7 show that, for breakdown risks higher than $\approx 5\%$ annually, subsidy treaties become *more efficient* than emissions treaties. In particular, under our assumption that cooperation will not be resumed after a breakdown, we find that attempting to cooperate only about once per decade becomes beneficial, because it decreases the risk of early breakdowns. Subsidy treaties retain much of their efficacy even if cooperation only occurs at these rare intervals, because they exploit the persistent effects of capacity investments. In the presence of opportunity costs to cooperation, we find that subsidy treaties may be the only feasible option for cooperation, if countries are pessimistic about the risk of breakdowns.

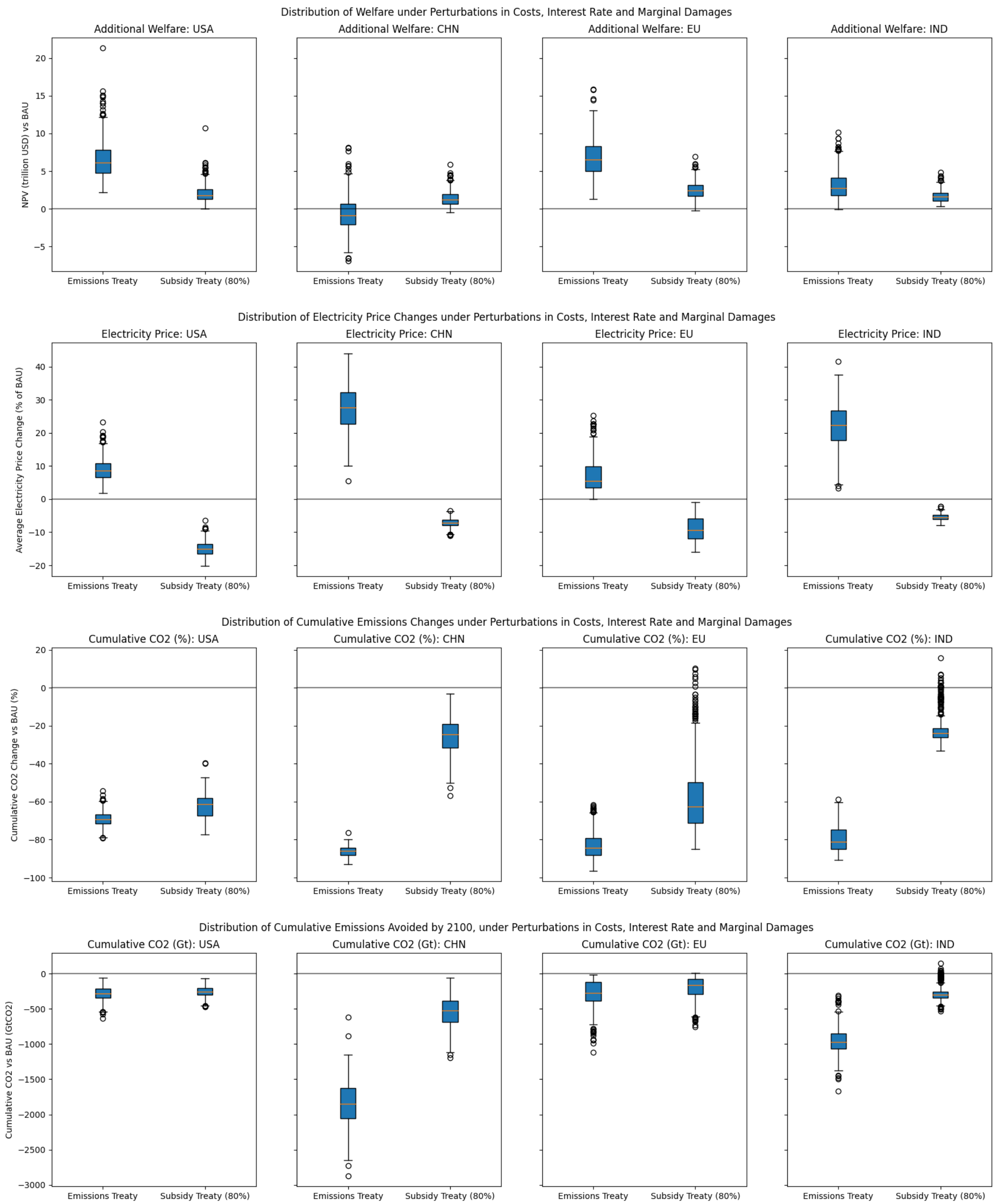

Figure 6: Changes in Welfare, average electricity prices and cumulative CO2 emissions under the optimal emissions treaty vs a 100% subsidy treaty for solar and wind capacity. We show the distribution of changes under perturbations to cost, interest rate and marginal damage parameters.

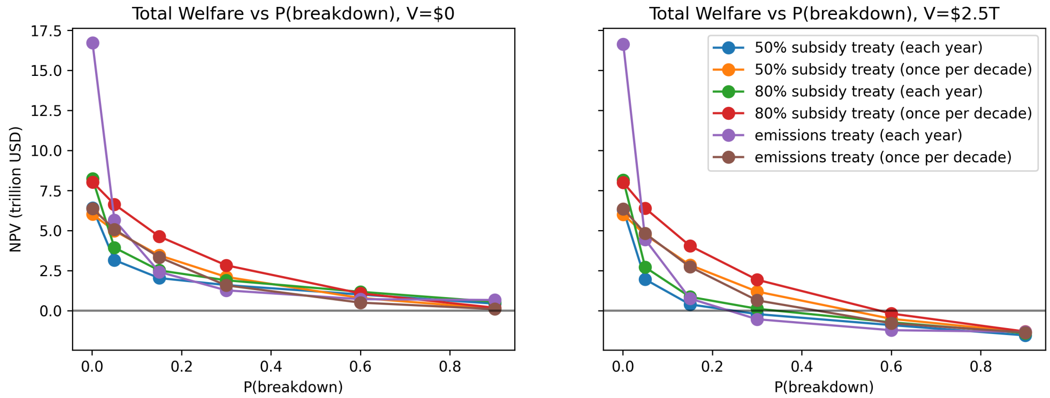

Figure 7: Welfare gains under pessimism about sustained cooperation $P(s_t = \text{breakdown})$, assuming a value at risk of $V_t = \$0$ (left) and $V_t = \$2.5T$ (right). For high breakdown risks, subsidy treaties are more efficient and cooperation should be attempted only once per decade. With opportunity costs to cooperation $V_t > 0$, only subsidy treaties may be beneficial.

6 Conclusion

We proposed *Subsidy Treaties*, which offer a way for countries to cooperate on reducing their inefficiently high carbon emissions over the next century. We shed light on some theoretical and practical advantages they offer relative to the emissions treaties, which have previously received more attention. If pessimism about future cooperation is significant, we find that subsidy treaties become more efficient than emissions treaties. It may also become beneficial to 'risk' significant cooperation only about once per decade, in which case subsidy treaties remain particularly effective. If there are opportunity costs to cooperation, e.g. because international institutions facilitating the transactions are not fully trusted, or because there are monitoring delays, then subsidy treaties may be the only viable option for cooperation. A practical benefit is that subsidy treaties do not require global real-time monitoring of CO2 emissions. The proposal exploits the unique cost structure of renewables: if the upfront fixed costs of renewable generation units are sufficiently subsidized during a period of short-term cooperation, market incentives will drive electricity providers in all countries to buy, operate and maintain them more of them than they would otherwise, because the *variable* costs of renewables are lower than those of fossil fuels. A single iteration of such a treaty can thus directly guarantee lower future emissions, even if cooperation is not anticipated to persist. An emissions treaty cannot incentivize similar investments if there are worries about its persistence. Each party could finance their subsidies via a government debt, and more than repay them with the additional future income from avoided damages. An optimal treaty between China, the US, EU and India would attempt to subsidize the upfront fixed-costs of solar cells and wind turbines for any willing buyer as much as possible, if and only if all parties agree to contribute about as much to the subsidies as they are expected to receive. We estimate the treaty could reduce future worldwide carbon emissions from electricity by $\geq 40\%$ relative to a noncooperative business-as-usual. The welfare gains within the next century would be comparable to avoiding a Covid-19 pandemic: totaling about \$8T in present value for the four parties alone, net of costs. Most abatement would be created in high-growth economies such as India, for whom a fossil-fueled growth path would otherwise be rational.

References

Barrett, S. (2002). Consensus Treaties. *Journal of Institutional and Theoretical Economics (JITE)*, 158(4):529–547.

Barrett, S. (2005). Chapter 28 the theory of international environmental agreements. In Mäler, K.-G. and Vincent, J. R., editors, *Economywide and International Environmental Issues*, volume 3 of *Handbook of Environmental Economics*, pages 1457–1516. Elsevier.

Barrett, S. and Dannenberg, A. (2022). The decision to link trade agreements to the supply of global public goods. *Journal of the Association of Environmental and Resource Economists*, 9(2):273–305.

Behmer, S. (2023). Sticks vs carrots: Climate policy under government turnover.

Burke, M. and Emerick, K. (2016). Adaptation to climate change: Evidence from us agriculture. *American Economic Journal: Economic Policy*, 8(3):106–40.

Burke, M., Hsiang, S., and Miguel, E. (2015). Global non-linear effect of temperature on economic production. *Nature*, 527.

Böhringer, C., Carbone, J. C., and Rutherford, T. F. (2016). The strategic value of carbon tariffs. *American Economic Journal: Economic Policy*, 8(1):28–51.

Coase, R. H. (1960). The problem of social cost. *The Journal of Law and Economics*, 3:1–44.

Debreu, G. (1959). *Theory of value: An axiomatic analysis of economic equilibrium*, volume 17. Yale University Press.

Deryugina, T. and Hsiang, S. M. (2017). The marginal product of climate. *PSN: Global Warming & Climate Change (Topic)*.

Farrokhi, F. and Lashkaripour, A. (2021). Can trade policy mitigate climate change. *Unpublished Working Paper*.

Harstad, B., Lancia, F., and Russo, A. (2019). Compliance Technology and Self-enforcing Agreements. *Journal of the European Economic Association*, 17(1):1–29.

Harstad, B., Lancia, F., and Russo, A. (2022). Prices vs. quantities for self-enforcing agreements. *Journal of Environmental Economics and Management*, 111:102595.

Jacobson, M. Z., von Krauland, A.-K., Coughlin, S. J., Dukas, E., Nelson, A. J. H., Palmer, F. C., and Rasmussen, K. R. (2022). Low-cost solutions to global warming, air pollution, and energy insecurity for 145 countries. *Energy Environ. Sci.*, 15:3343–3359.

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

Nordhaus, W. (2021). Dynamic climate clubs: On the effectiveness of incentives in global climate agreements. *Proceedings of the National Academy of Sciences*, 118(45):e2109988118.

Rennert, K., Prest, B. C., Pizer, W. A., Newell, R. G., Anthoff, D., Kingdon, C., Rennels, L., Cooke, R., Raftery, A. E., Ševčíková, H., et al. (2022). The social cost of carbon: advances in long-term probabilistic projections of population, gdp, emissions, and discount rates. *Brookings Papers on Economic Activity*, 2021(2):223–305.

Schmidt, R. C. and Streekstra, L. (2022). *Coalition Formation In Continuous Time.* Unpublished Manuscript.

Tol, R. S. (2019). A social cost of carbon for (almost) every country. *Energy Economics*, 83:555–566.

A Proofs Of Main Propositions

A.1 Proof Sketch of Proposition 1

Proof. We first prove that the equilibria above exist and are nationally optimal. Consider a modified economy where electricity producers must first purchase the right to emit each ton of carbon, at some price p^e_{it}. Such rights Δe_{it} are created by a single, price-taking producer whose profits are given by:

$$\sum_{t\geq 0} \beta_{it} \cdot \left(p^e_{it} \cdot \Delta e_{it} - d_{it} \left(\sum_{\pi<t} \Delta e_{i\pi} + \sum_{\imath\neq i} e_{\imath t} \right) + d^\theta_{it} \left(\sum_{\imath\neq i} e_{\imath t} \right) \right)$$

that is, the rights are created by a producer who first has to purchase the destroyed future endowments at the capital markets. Given any emissions path of other countries $e_{-i} =: (\sum_{\imath\neq i} e_{\imath t})_t$, we obtain the existence of a market clearing equilibrium in i via a standard application of the Kakutani fixed-point theorem to the correspondence $\varphi_i(p_i, z_i|e_{-i}) := (M(z_i), Z_i(p_i|e_{-i}))$, where $Z_i(p_i|e_{-i})$ denotes the vector of excess demands at some price vector p_i in country i and $M(z_i) := \arg\max_{p_i\in P} p \times z_i$. The resulting fixed point $p^*_i(e_{-i}), z^*_i(e_{-i})$ yields a market clearing price vector. Since there are no externalities in this modified economy, by the first welfare theorem, the resulting market equilibrium is efficient within i. The first order condition of the rights producer yields $p^e_{i\tau} = \sum_{itj} \beta_{it\tau} \nabla_{e_t} d_{it}(e_t)$. Choosing $\tau^Q_{itj} = e^Q_j p^e_{i\tau}$ and reimbursing consumers with $\tau^\theta_{it} = d^\theta_{it}(e_t) - d^\theta_{it}(\sum_{\imath\neq i} e_{\imath t})$ in our original economy therefore yields the same equilibrium. That the policy is feasible (i.e. does not violate the government's budget constraint) follows from revealed preference as choosing $\Delta e_{it} \equiv 0$ would have ensured zero profits to the right producer. We now proved that the policy is feasible and optimal given e_{-i}, and that τ^Q_{itj} must be characterized as in the Theorem by necessesity of the right-producer's first-order condition. What remains to be shown is that there exists an emissions profile $e := (e_{it})_{it},\ e_{it} \equiv \sum_{\pi<t} \Delta e_{i\pi}$ that is jointly consistent with these equilibria. This can be done by applying Kakutani's Fixed Point Theorem to the extended correspondence $\varphi(p, z, e) = (\varphi_i(p_i, z_i|e_{-i}), S_i(p_i, e_{-i}))_i$ where $e_i = (e_{it})_t \in S_i(p_i, e_{-i})$ denotes the (gross) cumulative supply path of emission rights produced in i. Its fixed point p^*, z^*, e^* yields a price and an emissions path consistent with a SPNE in the game between policymakers and simultaneous market clearing in all economies.

While the arguments above suffice to prove the proposition, it is informative to derive the transfers which ensure a Pareto-improvement when switching to $P_{i\tau}$ from some policy $\underline{P}_{i\tau}$. Specifically, these transfers are $\tau^\theta_{it} - \underline{\tau}^\theta_{it} = \tilde{\tau}^\theta_{it} + \sum_j \frac{\pi^\theta_{itj}}{\int_\Theta \pi^\theta_{itj} \mathrm{d}\mu_i(\theta)} \tilde{\tau}_{itj}$, where, denoting equilibrium prices and quantities before and after the policy change with and without an underline respectively,

$$\tilde{\tau}^\theta_{it} = (\vec{p}^{\,Q}_{it} - \underline{\vec{p}}^{\,Q}_{it}) \times \underline{\vec{Q}}^\theta_{it} + d^\theta_{it}(e_t) - d^\theta_{it}(\underline{e}_t) + \frac{\beta_{it\tau} - \underline{\beta}_{it\tau}}{\beta_{it\tau}} \underline{b}^\theta_{it}$$

is a transfer ensuring consumers could afford the same consumption path as before, and

$$\tilde{\tau}_{i\tau j} = (\vec{\tau}^{\,Q}_{i\tau} - \underline{\vec{\tau}}^{\,Q}_{i\tau})\mathbf{1} \times \underline{\vec{Q}}_{ij\tau} - (\vec{p}^{\,Q}_{i\tau} - \underline{\vec{p}}^{\,Q}_{i\tau})\mathbf{1} \times \underline{\vec{Q}}_{ij\tau} + (\vec{\tau}^{\,K}_{i\tau} - \underline{\vec{\tau}}^{\,K}_{i\tau})\underline{\Delta K}_{ij\tau}$$

is a reimbursement to shareholders which fully compensates for any losses in profits. By construction, consumers can afford the same consumption path as before, and shareholders are fully compensated for any losses in profits, the unanimous approval to move from $\underline{P}_{i\tau}$ to $P_{i\tau}$ follows by revealed preference. To see that this policy stricly relaxes the government's inter-temporal budget constraint, note that:

$$\begin{aligned}
&\sum_{t\geq\tau} \beta_{it\tau} \left(\int_\Theta \underline{\tau}^\theta_{it} - \tau^\theta_{it} \mathrm{d}\mu_i(\theta) + \sum_j \tau^Q_{ijt}\mathbf{1} \times \vec{Q}_{ijt} - \underline{\tau}^Q_{ijt}\mathbf{1} \times \underline{\vec{Q}}_{ijt} - \underline{\vec{\tau}}^{\,K}_{i\tau} \underline{\Delta K}_{ij\tau} \right) \\
&= \sum_{t\geq\tau} \beta_{it\tau} \left(d_{it}(\underline{e}_t) - d_{it}(e_t) + \sum_j \tau^Q_{itj}\mathbf{1} \times (\vec{Q}_{itj} - \underline{\vec{Q}}_{itj}) \right) \\
&= \sum_{t\geq\tau} \beta_{it\tau} \left(d_{it}(e_t + \sum_{\tilde{t}=\tau}^{t} \sum_j e_j \mathbf{1} \times (\underline{\vec{Q}}_{i\tilde{t}j} - \vec{Q}_{i\tilde{t}j})) - d_{it}(e_t) + \sum_j \tau^Q_{itj}\mathbf{1} \times (\vec{Q}_{itj} - \underline{\vec{Q}}_{itj}) \right) \\
&= \sum_{t\geq\tau} \beta_{it\tau} \left(\nabla_{\tilde{e}_t} d_{it}(\tilde{e}_t) \sum_{\tilde{t}=\tau}^{t} \sum_j e_j \mathbf{1} \times (\underline{\vec{Q}}_{i\tilde{t}j} - \vec{Q}_{i\tilde{t}j}) + \sum_j e^Q_j \sum_{\tilde{t}\geq t} \beta_{i\tilde{t}t} \nabla_{e_{\tilde{t}}} d_{i\tilde{t}}(e_{\tilde{t}})\mathbf{1} \times (\vec{Q}_{itj} - \underline{\vec{Q}}_{itj}) \right) \\
&= \sum_j e^Q_j \left(\sum_{t\geq\tau} \sum_{\tilde{t}=\tau}^{t} \beta_{it\tau} \nabla_{\tilde{e}_t} d_{it}(\tilde{e}_t)\mathbf{1} \times (\underline{\vec{Q}}_{i\tilde{t}j} - \vec{Q}_{i\tilde{t}j}) + \sum_{t\geq\tau} \sum_{\tilde{t}\geq t} \beta_{i\tilde{t}\tau} \nabla_{e_{\tilde{t}}} d_{i\tilde{t}}(e_{\tilde{t}})\mathbf{1} \times (\vec{Q}_{itj} - \underline{\vec{Q}}_{itj}) \right) \\
&= \sum_{t\geq\tau} \beta_{it\tau} \left(\nabla_{\tilde{e}_t} d_{it}(\tilde{e}_t) - \nabla_{e_t} d_{it}(e_t) \right) (\underline{e}_t - e_t) \geq 0
\end{aligned}$$

where the second line used the market clearing conditions for electricity and bonds, such that the corresponding terms cancel out. The $\underline{\vec{\tau}}_{i\tau}^K \underline{\Delta K}_{ij\tau}$ term cancels out because τ_{it}^θ levies the missing tax revenue from shareholders lump-sum. The fourth line used the mean-value theorem which guarantees the equality for some $\tilde{e}_t = \lambda e_t + (1-\lambda)\underline{e}_t$, $\lambda \in [0,1]$. The penultimate line used $\sum_{t \geq \tau} \sum_{\tilde{t} \geq \tau}^{t} a_t b_{\tilde{t}} = \sum_{u \geq \tau} \sum_{v \geq u} a_v b_u$, which, given absolute convergence, can be seen via term matching: each set of indices $(t, \tilde{t}) = (v, u)$ shows up exactly once on each side: for $\tilde{t} \leq t$ and $v \geq u$ respectively. The last inequality follows from the fact that $\nabla_e d_{it}(e)$ is increasing in e by convexity, whence both factors have the same sign. □

A.2 Proof Sketch of Proposition 2

Proof. This statement can be proved by following the same arguments as in the proof of Proposition 1, but considering an emissions rights producer who has to purchase the endowments that would be destroyed in all economies.

We can again characterize the transfers that ensure adopting this equilibrium is a Pareto-improvement from any other equilibrium. We use the same underlined notation for the equilibrium quantities before the global tax, and denote the quantities after the change without the underline. Given these quantities, we can characterize the transfers τ_{it}^θ using the same expression as before, which guarantee a Pareto-improvement by the same arguments. We only need to ensure the sum over all government budgets remain balanced. Note that the cross-country lump-sum transfers satisfy $0 = \sum_i \bar{\tau}_{it}$. Then i's inter-temporal budget is given by:

$$
\begin{aligned}
&\sum_{t \geq \tau} \beta_{it\tau} \left(\bar{\tau}_{it} + \int_\Theta \underline{\tau}_{it}^\theta - \tau_{it}^\theta \mathrm{d}\mu_i(\theta) + \sum_j \bar{\tau}_{jt}^Q \mathbf{1} \times \vec{Q}_{ijt} - \underline{\tau}_{ijt}^Q \mathbf{1} \times \underline{\vec{Q}}_{ijt} - \underline{\vec{\tau}}_{i\tau}^K \underline{\Delta K}_{ij\tau} \right) \\
&= \sum_{t \geq \tau} \beta_{it\tau} \left(\bar{\tau}_{it} + d_{it}(\underline{e}_t) - d_{it}(e_t) + \sum_j \bar{\tau}_{jt}^Q \mathbf{1} \times (\vec{Q}_{itj} - \underline{\vec{Q}}_{itj}) \right) \\
&= \sum_{t \geq \tau} \beta_{it\tau} \left(\nabla_{\tilde{e}_t} d_{it}(\tilde{e}_t) - \nabla_{e_t} d_{it}(e_t) \right) (\underline{e}_t - e_t) \geq 0
\end{aligned}
$$

where the inequality again follows by convexity of d_{it} and $\exists \lambda \in [0,1] : \tilde{e}_t = \lambda e_t + (1-\lambda)\underline{e}_t$. □

A.3 Proof Sketch of Proposition 3

Proof. To prove the first part of the proposition, it suffices to consider the corner case $P(s_t = \text{breakdown}) = 1$. In this case, cooperation is expected to only occur in the first period. We will consider a toy model with $N = 2$ symmetric countries with singleton population, $T = 3$ periods, two technologies $J = \{F, R\}$ representing fossil and renewable sources respectively, and only one subperiod $k = 1$ in which both technologies can generate $\gamma_{ijk} = 1$. We drop superfluous subscripts. Fossil fuels are characterized by higher variable costs $c_F^Q = 1 > c_R^Q = 0$

and lower upfront fixed costs $c_F^K = 0 < c_R^K = 3$, as well as higher emissions $e_F^Q = 1 > e_R^Q = 0$. We ignore operating costs $c_j^O = 0$ and depreciation $\lambda_j = \infty$. We consider a constant income stream $I = 1$ and assume linear damages $d(e_t) = e_t$. We normalize initial emissions and preinstalled capacity to $e_1 = 0,\ K_R = 0, K_F = 1$. Assume the representative agent in each country maximizes a partial linear expected utility without discounting:

$$U_{i\tau} = \sum_{t \geq \tau}^{3} \mathbb{E}\left[\log(Q_{it}) + X_{it}\right]$$

Assuming for now (and verifying in 1. below) that fossil sources are more profitable, the optimal consumption would be $Q_{it} = \frac{1}{p_{it}^Q} = \frac{1}{1+\tau_{iFt}^Q}$. With similar arguments as in the proof of Proposition 1, we can show that it is optimal for the government to 'forward' the subsidy or tax it is subject to under the respective treaty to the consumer, in addition to the nationally optimal carbon tax. As a result, the optimal emissions treaty charges $\bar{\tau}_{Ft}^Q = (3 - t)$ to incentivize national taxes twice that size. In this toy model, it cannot be optimal to incentivize a higher national tax than the global marginal future damages, as the 1-period tax only affects consumption in the present period and thus future damages, but does not affect investment or future consumption decisions. We compare three scenarios:

1. BAU: Governments levy the nationally optimal carbon taxes $\vec{\tau}_{iF}^Q = (2, 1, 0)$ yielding $\vec{Q}_i = (\frac{1}{3}, \frac{1}{2}, 1)$ and $\vec{e} = (0, \frac{2}{3}, \frac{5}{3})$. The NPV of national incomes net of electricity costs and damages, as well as lump-sum distributions is therefore $\sum_t X_t = 3 - (\frac{1}{3} + \frac{1}{2} + 1) - (0 + \frac{2}{3} + \frac{5}{3}) = -\frac{7}{6}$. This yields a total welfare of $U_{i1} = \log(1/3) + \log(1/2) + \log(1) - \frac{7}{6} \approx -2.96$. Note that as we initially assumed, countries have no incentive to build out renewables, as adding $\Delta K_{R1} = 1/2$ would cost 1.5 and save $1/2 + 1/2 = 1$ total marginal costs in periods 2 and 3, plus 1/2 in period 3 damages, netting out to zero.

2. Optimal Emissions Treaty: incentivizes $\vec{\tau}_{iF1}^Q = (4, 1, 0)$ (given the certain breakdown of cooperation after period 1), resulting in $\vec{Q}_i = (\frac{1}{5}, \frac{1}{2}, 1)$ and $\vec{e} = (0, \frac{2}{5}, \frac{7}{5})$. Hence $\sum_t X_t = 3 - (\frac{1}{5} + \frac{1}{2} + 1) - (0 + \frac{2}{5} + \frac{7}{5}) = -\frac{1}{2}$. This yields a total welfare of $U_{i1} = \log(1/5) + \log(1/2) + \log(1) - \frac{1}{2} \approx -2.8$. There is again no incentive to build out renewables as no treaty is anticipated in future periods, but the coordinated abatement in period 1 improves welfare.

3. Subsidy Treaty: incentivizing $\bar{\tau}_{R1}^K = \tau_{iR1}^K = -1/2$, and $\vec{\tau}_{iF}^Q = (2, 1, 0)$ breaks the indifference to adding renewables in period 1, and makes countries add exactly $K_{iR2} = 1/2$. The resulting consumption remains $\vec{Q}_i = (\frac{1}{3}, \frac{1}{2}, 1)$, but now $\vec{e} = (0, \frac{2}{3}, \frac{2}{3})$ as there are no emissions in period 2. Hence $\sum_t X_t = 3 - 1.5 - (1/3 + 0 + 1/2) - (0 + 2/3 + 2/3) = -\frac{2}{3}$. This yields a total welfare of $U_{i1} = \log(1/3) + \log(1/2) + \log(1) - \frac{2}{3} \approx -2.46$, making it more efficient than an emissions treaty.

Next, we consider the second part of the proposition, setting $P(s_t = \text{breakdown}) = 1 - \epsilon$. Consider an arbitrarily small tax $\bar{\tau}^Q = \delta > 0$ imposed by the treaty, to incentivize

$\tau_{iF1}^Q = 3 + \delta$. To show that no emissions treaty equilibrium exists, it suffices to rule out emissions treaties for small values of δ, where their marginal abatement benefit is most favorable relative to the marginal opportunity costs of risking s_t = corruption. By the envelope theorem applied to the nationally optimal carbon tax, small δ affect national welfare only through its impact on national income (via foreign abatement if the body remains committed and lost funds if it corrupts), but there is no first-order welfare effect through the distortion of domestic electricity consumption. Coordinating on a small δ will require upfront contributions of at least $\underline{C}_1 = \frac{\delta}{3+\delta}$, given the anticipated electricity demand. The first-order impact of foreign abatement on net present national income summed over both future periods is $2 * \nabla_\delta \log(\frac{1}{3+\delta}) = \frac{2\delta}{2+\delta}$, yielding an expected welfare impact of $\frac{\epsilon 2\delta - (1-\epsilon)\delta}{3+\delta} = \frac{(3\epsilon-1)\delta}{3+\delta}$, which is clearly negative for small ϵ. Finally, we show that a subsidy treaty is still beneficial. Note that an arbitrarily small subsidy $\bar{\tau}^K = -\delta > 0$ will incentivize the same equilibrium as before in the absence of corruption, yielding a welfare gain of $-2.46 + 2.96 \approx 1/2$ as the subsidies are only required to break the indifference to renewable investment. This requires upfront contributions of $\underline{C}_1 = \delta/2$ given the anticipated demand for renewables $K_{iR2} = 1/2$. As a result, the expected welfare impact of a subsidy treaty given risk of corruption is $\epsilon/2 - (1-\epsilon)\delta/2$, which is positive for small enough δ. $\square$

B Additional Figures

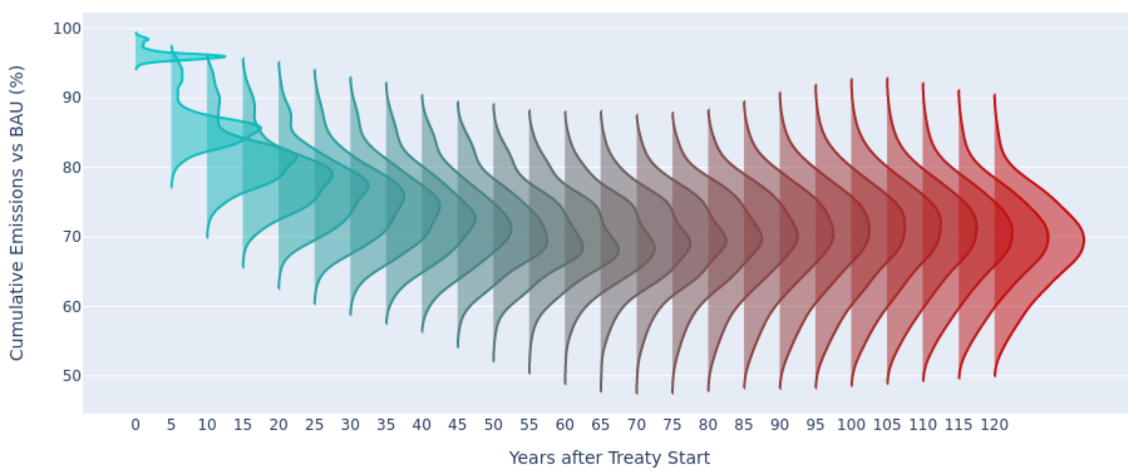

Figure 8: Distribution of global abatement from electricity generation relative to BAU from a permanent subsidy treaty, under different log-normal perturbations to cost, interest rate and marginal damage parameters.

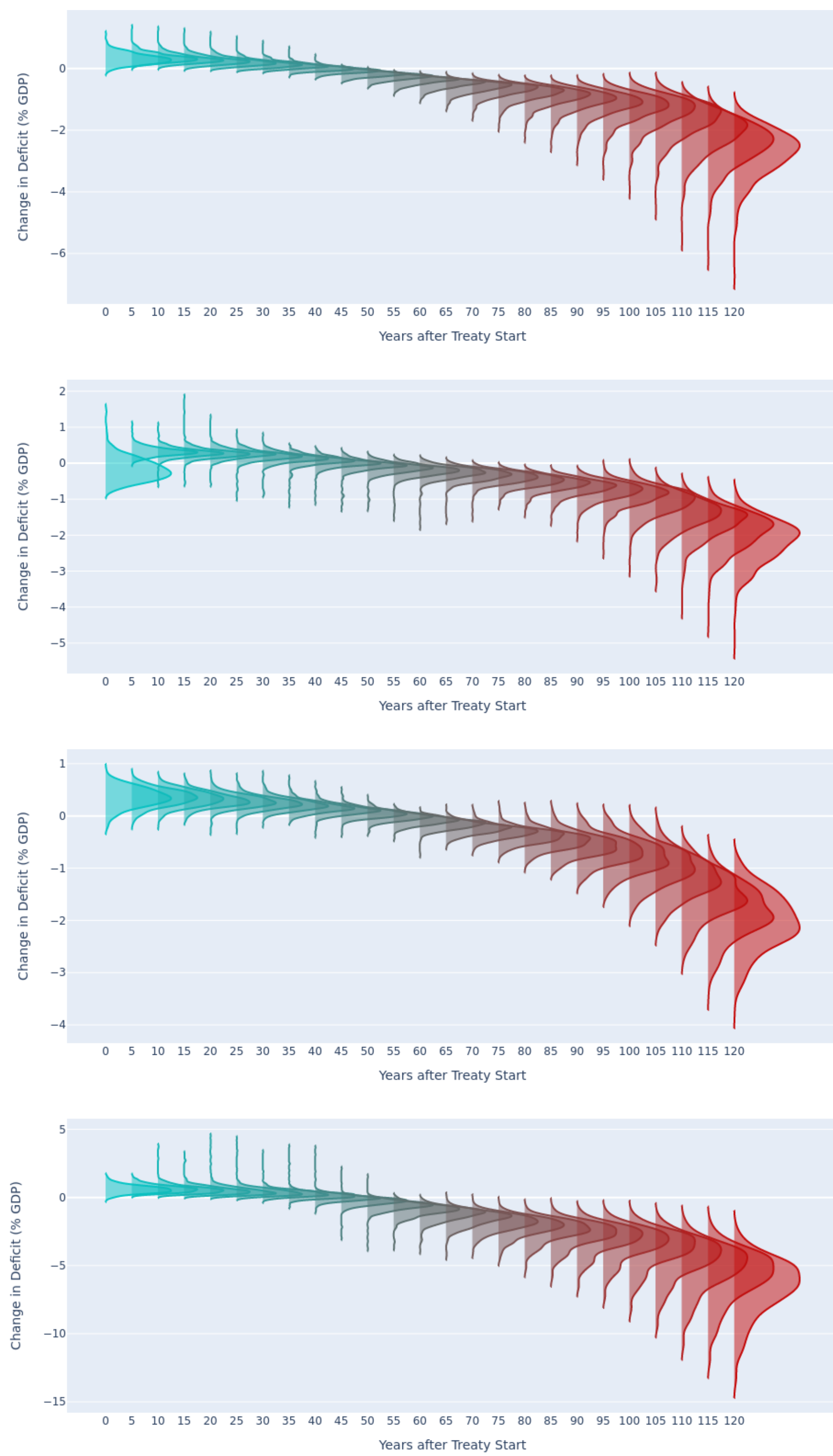

Figure 9: Distribution of additional national deficits under a permanent subsidy treaty, under different log-normal perturbations to cost, interest rate and marginal damage parameters.