The Optimal Design of Climate Agreements

Inequality, Trade, and Incentives for Climate Policy

Job Market Paper

Most recent version accessible here – Updated regularly

Thomas Bourany*

The University of Chicago

October 28, 2024

Abstract

Fighting climate change requires ambitious global policies, which are undermined by free-riding incentives. The heterogeneity in both the impacts of climate change and the costs of carbon taxation exacerbate non-cooperation, which makes the implementation of multilateral climate agreements difficult. This paper studies how to design an optimal climate club – in the spirit of Nordhaus (2015) – to maximize global welfare, incorporating strategic behavior when countries can exit climate agreements. In an Integrated Assessment Model with heterogeneous countries and international trade, I study the choice of countries in the agreement, the optimal level of carbon tax that members set on fossil fuels, and the tariffs they impose on non-members to incentivize participation. The choice balances an intensive margin – a club with few countries and large individual emission reductions – and an extensive margin – accommodating more countries at the cost of lowering the carbon tax. I find that the optimal climate club consists of all countries except Russia, a \$100 tax per ton of CO_2 within the club, and a 50% tariff on goods from non-members. In contrast, the globally optimal carbon tax is \$150 when free-riding is absent. In several extensions, I study additional policy instruments, such as transfers or fossil-fuel-specific tariffs.

*Thomas Bourany, *thomasbourany@uchicago.edu*. I thank my advisors Mikhail Golosov, Esteban Rossi-Hansberg, Lars Hansen and Michael Greenstone for valuable guidance and advice. I also thank Jordan Rosenthal-Kay, Tom Hierons, Aditya Bhandari, Elena Aguilar and other seminar participants at UChicago and Booth for stimulating discussions. All errors are mine.

1 Introduction

Fighting climate change requires ambitious global policies. To avoid severe consequences of global warming, carbon emissions must reach net zero in the next decades, and our economies need to phase out fossil fuels in a concerted effort to keep the world temperature under $2°C$ (IPCC et al. (2022)). However, the world is currently facing climate inaction. One of the main reasons behind this lack of cooperation is the presence of free-riding in climate policy: the benefits of fighting climate change are global, while the costs of reducing emissions using carbon pricing are local. Individual countries have incentives to free-ride on the rest of the world's reduction in emissions without implementing costly carbon abatement themselves.

Taxation of carbon and fossil fuels has strong redistributive effects across countries, determining their willingness to implement climate policy. First, emerging economies may face challenges in reducing the fossil fuel consumption necessary to continue their economic development. Second, carbon taxation has a substantial impact on energy markets, affecting the surplus of fossil fuel exporters and importers. Finally, imposing a carbon tax in one country reallocates economic activity and carbon emissions toward other countries through international trade – or "carbon leakage". All these effects reinforce free-riding incentives and climate inaction.

Multilateral climate agreements have been the traditional answer to address climate inaction, with the United Nations Conference of the Parties (COP) as an example. Unfortunately, they have failed to achieve decisive binding policy agreements. More recently, trade instruments have been the focus of policy discussions as trade policy offers the potential to give incentives to other countries to reduce emissions. In particular, Nordhaus (2015) proposes the idea of a "climate club", a voluntary agreement where members implement common carbon taxation as well as retaliatory tariffs on countries that do not participate in the club. In this context, trade sanctions are necessary to foster participation in the club and reduce free-riding incentives.[1]

What should be the design of a climate agreement that accounts for free-riding incentives as well as redistributive effects? What is the optimal climate club, its composition and level of carbon tax? This paper addresses these questions by examining the conditions necessary to construct a universal climate agreement with globally optimal carbon tax and tariffs. I explore which factors incentivize countries to join such an agreement, and I investigate how carbon and trade policy needs to be implemented to promote participation, maximize welfare, and fight climate change.

I tackle these policy questions in a climate-economy framework – or Integrated Assessment Model (IAM) – augmented with heterogeneous countries and international trade. I build a multi-country IAM, extended with bilateral trade frictions and energy markets. Individual countries differ in their vulnerability to climate change, income levels, their energy mix in oil, gas, coal, and

[1] Another notable example is the European Union's Carbon Border Adjustment Mechanism (CBAM), which is proposed to address the carbon leakage problem. This policy is a "carbon tariff" – i.e. a tariff whose rate increases with the carbon content of imported goods. This also has the potential to generate incentives for trade partners to implement climate policy in order to lower the carbon footprint of their exports.

non-carbon energy, their costs of producing fossil fuels as well as trade costs in trade in goods. This framework allows me to account for the multifaceted redistribution and leakage effects that arise in general equilibrium as a result of climate change and climate policy. The model serves as a laboratory for evaluating the welfare effects of different agreement designs.

With endogenous participation, countries have differing incentives to join a climate agreement. As a result, the decisions on the optimal levels of carbon tax and trade tariffs, as well as the choice of participants in the club, should be made jointly. Indeed, the optimal design reveals a tradeoff between an intensive and an extensive margin. At the intensive margin, an agreement could gather a small set of countries that can individually implement large emissions reductions with high carbon taxes. However, this is not sufficient to reduce global emissions and combat climate change effectively. In contrast, building a more extensive climate club requires accommodating the participation of a larger number of countries, which can only be done at the cost of lowering the carbon tax.

In this context, I address the policy problem where a global social planner maximizes the world's welfare by designing a climate agreement, or "climate club", that consists of three elements: (1) a set of countries included in the agreement – also called "climate coalition" – that are subject to the climate and trade policies, (2) the level of the carbon tax that club members set on their oil, gas, and coal energy consumption and (3) the level of the uniform trade tariffs that members impose on imported goods from non-member countries, while club members benefit from free-trade among themselves.[2] This policy design follows closely Nordhaus (2015)'s climate club setting.

Countries make individual choices to join or leave the agreement, and such strategic participation needs to be accounted for in the design of the agreement. I consider Nash equilibria, where countries make participation decisions either unilaterally or in "coalition deviations", i.e. when a subset of countries decide jointly to deviate and leave the agreement. The policy thus mirrors an optimal taxation and trade policy problem with limited instruments, together with a choice of countries. Given policy instruments, the coalition choice resembles the type of combinatorial discrete choice problem (CDCP) that arises in trade economics, e.g. Arkolakis et al. (2023). I propose different numerical solution methods to tackle this problem in the presence of participation constraints. I consider a restricted set of instruments, a single carbon tax, and a single good tariff. This follows the idea that the Social Planner choice represents the outcome of bargaining between club members. With bargaining frictions and transaction costs, negotiating over a small number of instruments – in particular, a single carbon tax – facilitates the finding of an agreement and undermines free-riding.[3]

[2]In the main club agreement design, I assume that fossil fuels are still freely traded for all countries. Moreover, non-members are *passive* in the sense that they do not retaliate with additional trade tariffs against the clubs. These two assumptions are relaxed in the extensions.

[3]Weitzman (2015) argues that a single carbon price serves as a "focal point" and is superior to binding quantity targets. Indeed, with transaction and bargaining costs, the Coase theorem may fail, preventing the agreement from reaching international cooperation. He attributes the failure of the Kyoto Agreement to the fact that quantity quotas represent a subdivision of efforts over countries and are, hence, more subject to disagreements and free-riding.

I contrast this framework with policy benchmarks absent endogenous participation. First, I consider the optimal carbon policy when the coalition gathers the entire world without participation constraints. I show that the choice of the carbon tax depends crucially on the availability of redistribution instruments, such as lump-sum transfers, in the First-Best allocation. Without such transfer instruments, I show how the choice of the Second-Best carbon tax accounts for distributional motives. Indeed, the carbon tax accounts for income inequality and its effect on demand distortion, supply redistribution through fossil-fuel energy markets, as well as trade leakage. As a result, the optimal carbon tax is \$150 per ton of CO_2 in the Second-Best and is lower than the Social Cost of Carbon, i.e. the marginal cost of climate change, a result that contrasts with the conventional Pigouvian principle.[4].

Second, I also compare the "climate club" framework to the non-cooperative Nash equilibrium, in which each individual countries choose its "unilaterally optimal" carbon taxation and trade tariffs. The unilateral carbon tax policy can become a subsidy if the Local Cost of Carbon – the cost of climate change as internalized by an individual country – is lower than two terms-of-trade manipulation terms, one for the good market and one for the fossil-fuel energy market. Similarly, optimal tariffs are also used for terms-of-trade manipulation of goods, a logic that aligns with conventional results in Ossa (2014) or Farrokhi and Lashkaripour (2024).

In comparison, climate agreements provide an "issue linkage" (Maggi (2016)) by linking the implementation of carbon policy with a reduction in tariffs, as the club promotes free trade among coalition members. The countries' participation choice depends on a balance between two effects: the distortionary cost of carbon taxation against the cost of tariffs – related to gains from trade. To choose whether to exit the club, individual countries consider if the first outweighs the second.

I find that the optimal climate club consists of all countries with the exception of Russia and former Soviet countries, and the agreement imposes a moderate carbon tax of \$100 per ton of CO_2 and a 50% tariff on traded goods of non-participants. The optimal climate agreement cannot achieve the world's optimal policy with complete participation – an agreement with a \$150 carbon tax and all the countries – despite full discretion on the choice of carbon tax and tariffs.

The reason is threefold. First, to increase participation, it is beneficial to reduce the carbon tax. Several Middle Eastern countries and developing economies in South Asia and Africa would not join an agreement with a high carbon tax, regardless of the level of the tariffs, since the gains from trade are bounded. Therefore, it is optimal to lower the tax from \$150 to \$100/tCO_2 to include those countries and share the "burden" of carbon abatement across more countries.

Second, it is beneficial to leave several fossil fuel producers like Russia and former Soviet countries outside of the climate agreement. Indeed, they suffer large welfare costs from carbon taxation, being relatively cold, closed and exporters of oil and gas. They would never join an

[4] The optimal policy problem with limited instruments is treated extensively in Bourany (2024) in a large class of climate-macroeconomic models. In the present paper, I draw a particular emphasis on international trade and leakage effects, a novel channel that needs to be accounted for in optimal carbon taxation.

agreement unless the carbon tax was very small, which is not optimal from a global perspective.

Third, trade policy is a key strategic instrument to undermine free-riding and incentivize countries to join the agreement. All the countries for which the cost of large tariffs outweighs the distortionary cost of carbon taxation are willing to participate in such climate clubs. That is especially the case for countries in Europe, East Asia, including China, and South-East Asia, which trade internationally a large share of goods production and have large gains from trade. Absent tariff retaliation, free-riding prevails over the cost of climate actions, as discussed in Nordhaus (2015). However, if moderate tariffs spur participation for a low carbon tax, this incentive effect vanishes quickly as the carbon tax increases and larger emissions reductions are required. The gains from trade are bounded – and small for some countries like the Middle East and Russia – and therefore, there is a limit to what carbon policy can achieve.

Additional policy instruments – such as transfers with a "loss and damage" fund or fossil-fuel-specific tariffs – improve the climate agreements and increase the carbon tax, bringing it closer to the Second-Best allocation. Indeed, these two instruments address two channels to incentivize participation. First, redistributing part of revenues from the carbon tax to poorer economies like South Asia and Africa – which consume less fossil fuel per capita – improves their welfare much more than the loss incurred by the richer economies of North America, Europe, and East Asia. Second, fossil-fuel-specific tariffs have strong effects on the energy rents of oil-exporting countries. This targets directly the large fossil-fuel producers, like Middle-Eastern Gulf countries and Russia, and increases the retaliatory power of the climate club by enforcing their participation. As a result, this allows the club to increase the attainable carbon tax and global carbon abatement. With these two instruments, we can reach the globally optimal allocation with complete participation and a more ambitious carbon policy since the agreement is more influential in penalizing non-members.

Lastly, I compare how these results on the optimal agreement change depending on the strategic response of countries outside the club. In the baseline result, non-members are passive and do not impose tariffs on the climate club members. If, instead, we consider that non-members cooperate strategically to impose retaliatory tariffs on club members, this has important consequences on the stability of the club. First, it makes the intensive-extensive margin tradeoff more salient, as countries have even more incentive to deviate from the club for high carbon taxes and moderate tariffs. However, retaliation in the face of high tariffs from club members also raises the costs of a trade war for *both* club members and non-members. This implies that if the club is large enough, it can threaten to escalate a trade war to enforce the participation of the non-members and ultimately achieve the optimal policy with a high carbon tax and complete participation.

Literature

This work relates to a large literature on the economics of climate change and bridges a gap with both the international trade policy and the game theoretical literature. First, I contribute to the debate on the formation of Climate Clubs, following the pioneering contribution of Nordhaus

(2015). The implementation of climate policy suffers from a free-riding problem and Nordhaus proposed a simple framework to evaluate the principle of issue linkage, i.e. linking the enforcement of a climate policy with trade tariffs. He shows with the C-DICE model that for different – exogenously set – carbon prices and tariffs rates, we can achieve varying participation to a climate club. With a low carbon price – up to 25$\$/tCO_2$ – and high tariffs – above 10%, the climate club can achieve a club with all the 15 regions he considers.

I depart from Nordhaus' Climate Club framework in three directions. First, I show that when a social planner chooses endogenously and optimally both the carbon tax, the tariffs, and the club members, we observe an intensive margin - extensive margin tradeoff. A lower carbon tax and higher tariffs increase participation. Second, I depart from the C-DICE model that uses ad-hoc functions for the carbon abatement – inspired by the DICE model – and the gain from trade and costs of tariffs – a quadratic approximation of the results of Ossa (2014). I show that modeling the energy market – both with heterogeneity in demand and supply of fossil energy – and trade in goods – accounting for leakage effects and terms-of-trade manipulation – highlights the tradeoff between the cost of carbon taxation and the cost of tariffs. In particular, in this micro-funded setting, gains from trade are bounded, which makes some countries unwilling to join an agreement if the loss from phasing out fossil fuels is too large, and this *for any level of tariffs*. Third, I model the cost of climate on production as endogenous to policy, which makes the optimal carbon tax account for redistributive effects through income inequality, trade leakage, and energy markets. Iverson (2024) take the same C-DICE model as Nordhaus and analyze how a "Tiered Climate Club" can achieve higher carbon abatement, where two different "tiers" imply different levels of carbon tax and tariff retaliations for different sets of countries. Hagen and Schneider (2021) also analyze how retaliation can undermine the stability of the club and results in potentially suboptimal climate clubs where the gains from climate policy are undermined by the costs of a trade war. In comparison to these three models, I provide a quantitative analysis where optimal policy is chosen strategically to enforce participation.

Farrokhi and Lashkaripour (2024) also study how climate policy can be conducted with trade instruments. They solve for the optimal trade policy in a rich multi-industry trade model, inspired by Copeland and Taylor (2004), and show that unilateral policy accounts for carbon leakage when setting tariffs. In this setting, they explore the sequential construction of a climate club, where European Union starts a coalition, implements the unilaterally optimal trade-climate policy, and iteratively grows the participation to the club. In contrast, I show how the club should *design* the trade-climate policy *strategically* to spur participation. My framework also incorporates several redistribution channels absent from their framework. Non-linear damage makes the cost of climate change endogenous to policy, and inequality across countries creates differences between policies maximizing output, reducing emissions, and improving welfare.

This project lies at the intersection of three bodies of literature: one on trade policy, one on the game-theoretical aspects of climate policy cooperation, and one on macroeconomic models of climate change.

First, the interdependence between climate, environmental, and trade policies is explored extensively in Kortum and Weisbach (2021), Barrett (2001), Bohringer et al. (2016), Bohringer et al. (2012) or Hsiao (2022). These articles explore the differences between unilateral policies implemented at the country level and the potential for climate cooperation using trade policies. Other articles in this trade literature explore the underpinnings of optimal trade policies, e.g. Costinot et al. (2015), Ossa (2014), Adao et al. (2023), Antras et al. (2024). More specifically they study the choice of trade tariffs with different objectives, like terms of trade manipulation or supply-chains considerations. I show how these policy instruments can be used for issue linkage and climate policy and study the optimal design of climate agreements in the presence of free-riding.

Moreover, I also borrow from the theoretical literature on climate cooperation, with classical references such as Barrett (1994), Harstad (2012), Barrett (2003), Barrett (2013), Nordhaus (2015) or the older literature collected in Batabyal (2000) and summarized in Maggi (2016). There is also a large literature on dynamic games and coalition formation games, that focus on the building of agreements, either through coordination games or through bargaining procedures, and summarized in Ray and Vohra (2015), or Okada (2023) more recently. Similarly, Nordhaus (2021), Harstad (2023), or Maggi and Staiger (2022) study those questions as well as other dynamic features, such as technical change, the path of climate dynamics or intertemporal decision-making. I draw inspiration from many of these references and provide a quantitative framework with many dimensions of heterogeneity to reveal the different factors that drive the countries' decisions to participate in climate clubs and provide policy recommendations.

Third, I also draw from the macroeconomic literature on the implications of climate change and carbon policy. Indeed, building on my own work Bourany (2024), I show that the optimal carbon policy accounts for several general equilibrium channels, as well as macroeconomic dynamics – in the spirit of Integrated Assessment Models. Starting from a static version of the DICE/RICE models as in Barrage and Nordhaus (2024) and Nordhaus and Yang (1996), I study the optimal fossil fuel taxation, as in Golosov et al. (2014) with heterogeneous countries/regions, as in Krusell and Smith (2022), Cruz and Rossi-Hansberg (2024, 2022), Kotlikoff, Kubler, Polbin, Sachs and Scheidegger (2021), Kotlikoff, Kubler, Polbin and Scheidegger (2021), as well as international trade. Similarly, I study a quantitative model with many dimensions of heterogeneity but keep it static to be able to study strategic interactions between large countries and a climate agreement design, which includes a combinatorial discrete choice joint with an optimal policy choice.

The remainder of this paper is structured as follows. In Section 2, I lay out the Integrated Assessment Model that we study in the policy analysis. The design of the climate agreement is exposed in Section 3. In Section 4, I present how I match the model to the data. In Section 5, I discuss the optimal policy benchmarks with or without cooperation in a setting without free-riding incentives. In Section 6, I present the main result of our analysis on the optimal climate agreement. In Section 7, I develop extensions that relax some of the assumptions made in the baseline results, such as the availability of additional instruments or retaliation.

2 An integrated assessment model with heterogeneous regions and trade

I build an integrated assessment model (IAM) that incorporates various dimensions of heterogeneity influencing individual countries' incentives to join climate agreements. This framework is the simplest model that includes both climate externality, a non-trivial energy market for fossil energy, and a realistic trade structure that reproduces the leakage effects of taxation.

I study a static economy[5] with I countries indexed by $i \in \mathbb{I}$, each with population $\mathcal{P}_i$. All the economic variables are expressed per capita.[6] Each country is composed of five representative agents: (i) a household that consumes the final goods, (ii) a final-good firm producing goods using labor and energy, (iii) a fossil energy firm extracting oil and gas, (iv) a producer of coal energy, and (v) a producer of renewable/non-carbon energy. Moreover, each country has a government that sets taxes and tariffs.

2.1 Household problem

The representative household in country i imports from all countries $j \in \mathbb{I}$ and consumes the aggregate quantity c_i. I consider an Armington structure, c.f. Anderson (1979), Arkolakis, Costinot and Rodriguez-Clare (2012), where each country produces its own variety. The household preferences have constant elasticity of substitution θ over goods from different countries.

$$\mathcal{U}_i = \max_{\{c_{ij}\}} u(\{c_{ij}\}_j) = u\Big(\mathcal{D}_i^u(\mathcal{E})c_i\Big) \, , \qquad c_i = \Big(\sum_{j\in\mathbb{I}} a_{ij}^{\frac{1}{\theta}} c_{ij}^{\frac{\theta-1}{\theta}}\Big)^{\frac{\theta}{\theta-1}} \, , \tag{1}$$

where a_{ij} are the preference shifters for country i on the good purchased from country j, which also include the home-bias a_{ii}.[7] The climate externality affects consumption which a factor $\mathcal{D}_i^u(\mathcal{E})$ which summarizes climate damages, given world emissions $\mathcal{E}$. It is a reduced-form representation of the climate system – and the path of temperatures – and decreases in $\mathcal{E}$ and is country-specific due to differences in the vulnerability and costs of climate change. In the quantification Section 4, we detail how we calibrate this function for each country i. Households earn labor income, energy rent, and transfers, and their budget constraints is given by:

$$\sum_{j\in\mathbb{I}} c_{ij}\,(1+\mathrm{t}_{ij}^b)\tau_{ij}\mathrm{p}_j = w_i\ell_i + \pi_i^f + \mathrm{t}_i^{ls} \, , \tag{2}$$

where w_i is the wage rate, ℓ_i the exogenous labor supply is normalized to 1, π_i^f the profit earned from the ownership of the energy firms, and t_i^{ls} the lump-sum transfer received from the government.

[5]More particularly, the static is a stationary representation of a dynamic model, as I describe in Appendix B.3. This allows keeping the framework simple enough to study the strategic interaction between countries as well as the joint design between a combinatorial discrete choice and optimal policy choice for the carbon tax and tariffs.

[6]For example, y_i or e_i^f are final output and fossil energy use respectively, and $\mathcal{P}_i y_i$ and $\mathcal{P}_i e_i^f$ represent the total quantities produced/consumed in the country. I allow for population growth n and TFP growth $\bar{g}$ in the dynamic model, and we display here the stationary version of the Balance Growth Path.

[7]We assume that preferences $\{a_{ij}\}$ and iceberg trade costs $\{\tau_{ij}\}$ are policy-invariant, in particular, they are not sensitive to price changes and tariffs.

On the expenditure side, the household in i imports quantities c_{ij} from j, purchased at price p_j, and subject to iceberg cost τ_{ij} and to trade-tariffs $1+\mathrm{t}^b_{ij}$. The choice of trade policy will be made explicit below.

The optimal consumption choice of the household yields the following quantities and Armington trade shares given by:

$$\begin{aligned} c_{ij} &= a_{ij} c_i \Big(\frac{(1+\mathrm{t}^b_{ij})\tau_{ij}\mathrm{p}_j}{\mathbb{P}_i}\Big)^{-\theta} , \\ s_{ij} &\equiv \frac{c_{ij} p_{ij}}{c_i \mathbb{P}_i} = a_{ij} \frac{((1+\mathrm{t}_{ij})\tau_{ij}\mathrm{p}_j)^{1-\theta}}{\sum_k a_{ik}((1+\mathrm{t}_{ik})\tau_{ik}\mathrm{p}_k)^{1-\theta}} , \end{aligned} \tag{3}$$

where $p_{ij} = (1+\mathrm{t}^b_{ij})\tau_{ij}\mathrm{p}_j$ is the effective price for a variety from country j sold in country i, and $\mathbb{P}_i$ is the price index of country i:

$$\mathbb{P}_i = \Big(\sum_{k\in\mathbb{I}} a_{ik}((1+\mathrm{t}_{ik})\tau_{ik}\mathrm{p}_k)^{1-\theta}\Big)^{\frac{1}{1-\theta}} .$$

As a result, we summarize the budget constraint as $c_i\mathbb{P}_i = \sum_{j\in\mathbb{I}} c_{ij}\,(1+\mathrm{t}^b_{ij})\tau_{ij}\mathrm{p}_j$, and the per-capita welfare of country i is then summarized by the indirect utility as the utility of income discounted by the price level and climate damages, namely:

$$\mathcal{U}_i = u\Big(\mathcal{D}^u_i(\mathcal{E})c_i\Big) = u\Big(\mathcal{D}^u_i(\mathcal{E})\,\frac{w_i\ell_i+\pi^f_i+\mathrm{t}^{ls}_i}{\mathbb{P}_i}\Big) . \tag{4}$$

2.2 Final good firm problem

The representative final good producer in country i is producing the domestic variety at price p_i. The firm's profit maximization is:

$$\max_{\ell_i, e^f_i, e^c_i, e^r_i} \ \mathrm{p}_i\,\mathcal{D}^y_i(\mathcal{E})\,z_i\,F(\ell_i, e^f_i, e^c_i, e^r_i) - w_i\ell_i - (q^f+\xi^f\mathrm{t}^\varepsilon_i)e^f_i - (q^c_i+\xi^c\mathrm{t}^\varepsilon_i)e^c_i - q^r_i e^r_i$$

where the production function $\bar{y}_i = F(\ell_i, e^f_i, e^c_i, e^r_i)$ is constant returns to scale and concave in all inputs. It uses labor, ℓ_i, at wage w_i, fossil energy, e^f_i, purchased at price, q^f, coal, e^c, at price, q^c_i, and renewable energy, e^r_i, at price, q^r_i. Energy from oil-gas, e^f_i, and coal, e^c_i, differ from renewable in the sense that they emit greenhouse gases, with respective carbon concentration ξ^f and ξ^c, as we will see in Section 2.4. As a result, there is a motive for taxing oil, gas, and coal energy with the carbon tax t^ε_i. We discuss the choice of this tax in the next sections.

The productivity of the domestic good firm, $y_i = \mathcal{D}^y_i(\mathcal{E})\,z_i\,\bar{y}_i$, can be decomposed in two terms. First, the TFP, z_i, represents productivity as well as institutional/efficiency differences between countries. Invariant to prices and policy, this technology wedge accounts for income inequality across countries. These differences in TFP translate into differences in consumption that create redistribution motives for tax policy.

The second difference in productivity comes from the climate externality summarized by the net-of-damage function $\mathcal{D}_i^y(\mathcal{E})$, given world emissions $\mathcal{E}$. This function is also a reduced-form representation of the climate system from future temperatures, decreases in $\mathcal{E}$, and is country-specific due to differences in costs of climate change, as we detail how we detail in the quantification Section 4.

The firm input decisions solve the optimality conditions, where we define the marginal product of an input x as $MPx_i \equiv \mathcal{D}_i^y(\mathcal{E})\, z_i F_x(\ell_i, e_i^f, e_i^c, e_i^r)$ for $x \in \{\ell_i, e_i^f, e_i^c, e_i^r\}$. For example, in the case of oil and gas e_i^f, the first-order condition can be written as:

$$\mathrm{p}_i \mathcal{D}_i^y(\mathcal{E})\, z_i\, F_{e^f}(\ell_i, e_i^f, e_i^c, e_i^r) =: \mathrm{p}_i MPe_i^f = q^f + \xi^f \mathrm{t}_i^{\varepsilon} \ , \tag{5}$$

and similarly for other inputs ℓ_i, e_i^c, e_i^r. Crucially, the private decision of firms do not internalize climate externalities of their own fossil-fuel energy use and only responds to carbon tax $\mathrm{t}_i^{\varepsilon}$.

2.3 Energy markets

The final-good firm is consuming three kinds of energy sources – oil-gas, coal, or renewable (non-carbon) – which are supplied by three representative energy firms in each country. Oil-gas sources are traded internationally, and countries can be exporters or importers. Coal and renewable sources are both traded locally, an empirically relevant assumption given the substantial trade costs in coal shipping or electricity transfers.

2.3.1 Fossil firm

In each country $i \in \mathbb{I}$, a competitive energy producer extracts fossil fuels – oil and gas – e_i^x and sells it to the international market at price q^f. The energy is extracted at convex cost $\mathcal{C}_i^f(e_i^x)$, where the convex costs are paid in the unit of the consumption bundle of the household.[8] The energy firm's profit maximization problem is given by:

$$\mathcal{P}_i \pi_i^f = \max_{e_i^x} q^f e_i^x - \mathcal{C}_i^f(e_i^x) \mathbb{P}_i \ , \tag{6}$$

where $\mathcal{P}_i \pi_i^f$ is the total energy rent of country i. Since the extraction costs are convex, the production function has decreasing return to scales,[9] and hence, even with competitive firms, taking the fossil price as given, a positive energy rent exists. Moreover, for the sake of simplicity, we do not consider that energy firms have market power in the setting of energy prices – for example, in the case of OPEC – even though this framework could easily allow for such an extension. Any sources of misallocation – in the sense of Hsieh and Klenow (2009) – are accounted for in the calibration of the cost function $\mathcal{C}_i^f(\cdot)$ as we will see in the quantification Section 4.

[8]This allows to account for international inputs in goods and services for building capital for resource extraction.

[9]We can also define a fossil production function with inputs x_i^f such that $e^x = g(x_i^f)$ and profit $\pi = q^f g(x) - x\mathbb{P}_i$ instead of $\pi = q^f e^x - \mathcal{C}(e^x)\mathbb{P}_i$, in which case $g(x) = \mathcal{C}^{-1}(x)$

Naturally, the optimal extraction decision follows from the optimality condition:

$$q^f = \mathcal{C}_i^{f\,\prime}(e_i^x)\mathbb{P}_i \ , \tag{7}$$

which yields the implicit function $e_i^{x\star} = e^x(q^f/\mathbb{P}_i) = \mathcal{C}_i^{f\prime\,-1}(q^f/\mathbb{P}_i)$. Finally, the energy rent comes from fossil firms' profits $\pi^f(q^f, \mathbb{P}_i) = q^f e^x(q^f/\mathbb{P}_i) - \mathcal{C}_i^f(e^x(q^f/\mathbb{P}_i))\mathbb{P}_i > 0$ and depends on the marginal costs as well as the elasticity $\nu_i = \frac{\mathcal{C}_i^{f\,\prime\prime}(e^x)}{\mathcal{C}_i^{f\,\prime}(e^x)e^x}$.

As we will see below, the profit $\pi^f(q^f, \mathbb{P}_i)$ and its share in income $\eta_i^{\pi f} = \frac{\pi_i{}^f}{y_i \mathrm{p}_i + \pi_i{}^f}$ are key to determine the exposure of a country to carbon taxation. Indeed, reducing carbon emissions by phasing out of fossil fuels reduces energy demand and its price q^f and hence affects energy profit π_i^f and the welfare of large oil and gas exporters.

2.3.2 International fossil energy markets

I assume that oil and gas are traded frictionlessly in international markets.[10] The market clears such that

$$E^f = \sum_{i\in\mathbb{I}} \mathcal{P}_i e_i^f = \sum_{i\in\mathbb{I}} e_i^x \ . \tag{8}$$

Countries have different exposure to this fossil energy market. As country i consumes fossil fuels in total quantity $\mathcal{P}_i e_i^f$, and produces total quantity e_i^x, its net exports of oil and gas are $e_i^x - \mathcal{P}_i e_i^f \lesseqgtr 0$.

2.3.3 Coal firm

A representative firm produced coal, that is consumed by the final good firm. I differentiate coal from other fossil fuels like oil and gas because coal production typically does not generate large energy rents for producing countries as a share of GDP. Moreover, large coal producers also consume a large fraction of that coal locally, as trade costs for coal transportation are larger. Hence, I make this empirically grounded assumption that coal is not traded.

The production $\bar{e}_i^c$ is constant returns to scale and uses final good inputs. I assume the production function is of the form $\bar{e}_i^c = x_i^c/\mathcal{C}_i^c$, where x_i^c is a CES aggregator of exactly the same form as eq. (1). As a result, the inputs $x_i^c = \mathcal{C}_i^c \bar{e}_i^c$ is paid in the consumption bundle at price $\mathbb{P}_i$ and the profit maximization problem is:

$$\pi_i^c = \max_{e_i^c} q_i^c \bar{e}_i^c - \mathcal{C}_i^c \bar{e}_i^c \mathbb{P}_i \ ,$$

where the marginal cost $\mathcal{C}_i^c$ is a constant. This implies that there is no coal profit[11] in equilibrium,

[10] For the sake of simplicity, I refrain from considering a general Armington structure, combining different fossil varieties with demand $e_i^f = \left(\sum_j (e_{ij}^f)^{\frac{\Theta-1}{\Theta}}\right)^{\frac{\Theta}{\Theta-1}}$. I make the simplifying assumption that fossil fuels produced in different countries are not distinguishable – crude oil or natural gas from Nigeria, Saudi Arabia, or Russia are not differentiated varieties – corresponding to the limiting case $\Theta \to \infty$

[11] This is motivated by evidence that even the largest coal producers do not have coal rents above 1% of GDP.

i.e. $\pi_i^c = 0$. The price for coal and the market clearing condition are given by:

$$q_i^c = \mathcal{C}_i^c \mathbb{P}_i \ , \qquad\qquad \bar{e}_i^c = e_i^c \ . \tag{9}$$

Hence, for a given price index of inputs $\mathbb{P}_i$, this implies a perfectly elastic supply curve for coal energy, something we observe in practice as coal production is easily scalable in response to oil and gas price fluctuations.

2.3.4 Renewable, non-carbon, firm

The final good firm also uses renewable and other low-carbon energy sources, such as solar, wind or nuclear electricity. This provides a way of substituting away from fossil fuel in the production function $F(\cdot)$.

A representative firm produces renewable or non-carbon energy, and this supply, $\bar{e}_i^r$, is not traded. This assumption is verified by the fact that electricity is rarely traded across countries – and when it is, it only is only the result of temporary differences in electricity production due to intermittency, rather than large structural imbalances. The production $\bar{e}_i^r$ is also constant returns to scale, with production $\bar{e}_i^c = x_i^r / \mathcal{C}_i^r$, and x_i^r a CES aggregator of the same form as eq. (1). This input is paid in units of the final good at price $\mathbb{P}_i$. Hence, the renewable firm maximization problem is:

$$\pi_i^r = \max_{\bar{e}_i^r} q_i^r \bar{e}_i^r - \mathcal{C}_i^r \bar{e}_i^r \mathbb{P}_i \ ,$$

where $\mathcal{C}_i^r$ is a constant and resulting in zero profits $\pi_i^r = 0$. As a result, the price of renewable and the market clearing are given by:

$$q_i^r = \mathcal{C}_i^r \mathbb{P}_i \ , \qquad\qquad \bar{e}_i^r = e^r \ . \tag{10}$$

This once again returns a perfectly elastic supply curve, which is a slightly stronger assumption in the context of renewable energy. In the short run, renewable energy requires investments in capacity, implying a fairly inelastic supply curve. This is especially true considering the intermittency problems of wind and solar energy, c.f. Gentile (2024). However, in the long run, technological progress and learning-by-doing create positive externalities, substantially decreasing the cost of clean energy, resulting in a decreasing supply curve, c.f. Arkolakis and Walsh (2023). I take the intermediary conservative assumption that the supply curve is flat and will explore the robustness of this assumption in future extensions.

2.4 The climate system

Carbon emissions released from the burning of fossil fuels create a climate externality as they feed back into the atmosphere, increasing temperatures and affecting damages. Despite the model being static, I incorporate climate system dynamics[12] as in standard Integrated Assessment Models. These future damages, summarized in the stationary equilibrium, affect the Social Cost of Carbon and the Pigouvian level of carbon taxation, as we will see in Section 5.1.

I model the damage functions affecting country i utility $\mathcal{D}_i^u(\mathcal{E})$ and productivity $\mathcal{D}_i^y(\mathcal{E})$ as a reduced-form summary of the impact of climate change. I develop a standard dynamic climate system that can be summarized in a static form in a simple way. It expresses the mapping from (i) emissions $\mathcal{E}$ to a path of atmospheric carbon concentration $\mathcal{S}_t$, (ii) from carbon concentration to a path of global and local temperatures T_{it}, and then from local temperatures to damage $\hat{\mathcal{D}}(T_{it}-T_i^\star)$, and (iv) finally summarizes it in present discounted value to obtain $\mathcal{D}_i^y(\mathcal{E})$ and $\mathcal{D}_i^u(\mathcal{E})$.

First, the static model represents stationary decisions on energy choices taken "once and for all". These yearly emissions from fossil fuels sum up to

$$\mathcal{E} = \sum_{i\in\mathbb{I}} \mathcal{P}_i(\xi^f e_i^f + \xi^c e_i^c) \ ,$$

where ξ^f and ξ^c represent the carbon concentration of oil-gas and coal, respectively. Accounting for population and TFP growth, this leads to a path of emissions given by $\mathcal{E}_t = e^{(\bar{g}+n)t}\mathcal{E}$, $\forall t$.[13] They represent trajectories of emissions given the emissions and policies decisions in the initial equilibrium, e.g. e_i^f and e_i^c.

Second, I consider a dynamic system – in continuous time – for carbon concentration in the atmosphere:

$$\dot{\mathcal{S}}_t = \zeta_t \mathcal{E}_t - \delta_s \mathcal{S}_t \qquad \text{with} \qquad \mathcal{S}_{t_0} = \mathcal{S}_{2024} \ ,$$

where δ_s is the exit rate of carbon out of the atmosphere, which is typically small for standard calibrations. To make the carbon concentration bounded and non-exploding – given the constant path of emissions – I follow Krusell and Smith (2022) by assuming that part of emissions $\mathcal{E}_t$ is abated via carbon capture and storage (CCS) modeled by the exogenous parameter ζ_t. The share of emissions abated grows to 100% in the long-run, implying that $\zeta_t \to_{t\to\infty} 0$. Increasing CCS allows the system to reach net-zero in several centuries, stabilizing cumulative carbon emissions and temperature.

Third, I assume a linear relationship between the cumulative CO_2 emissions $\mathcal{S}_t$ and the global temperature anomaly $\mathcal{T}_t$ compared to preindustrial levels.

$$\mathcal{T}_t = \chi \mathcal{S}_t = \chi\Big(\mathcal{S}_{t_0} + \int_{t_0}^{\infty} e^{-\delta_s(t-t_0)}\zeta_t \mathcal{E}_t dt\Big) \ ,$$

where χ is the climate sensitivity parameter, i.e. how much warming a ton of CO_2 causes, and

[12] For simplicity, I refrain from using a larger scale climate system as in Dietz et al. (2021) or Folini et al. (2024).

[13] In the stationary model, we consider a balanced-growth path with global population and TFP growth and all the path of variables over time are expressed as level per "effective" capita. In future calibration, I will consider country-specific growth rates for TFP $\bar{g}_i$ and population n_i.

where $\mathcal{E}_t$ is measured in carbon units, and $\mathcal{S}_{t_0}$ is the initial stock of carbon before all the policy decisions are made – i.e. in 2024. This specification is rationalized by a large climate-sciences literature, e.g. Dietz et al. (2021), that shows there exists an approximately linear relationship between $\mathcal{S}_t$ and $\mathcal{T}_t$, as is shown in the following Figure 1. It displays the relationship between temperature anomaly and cumulative CO_2 emissions over time, both for historical data in black and a large class of climate models in different Representative Concentration Pathways (RCP).

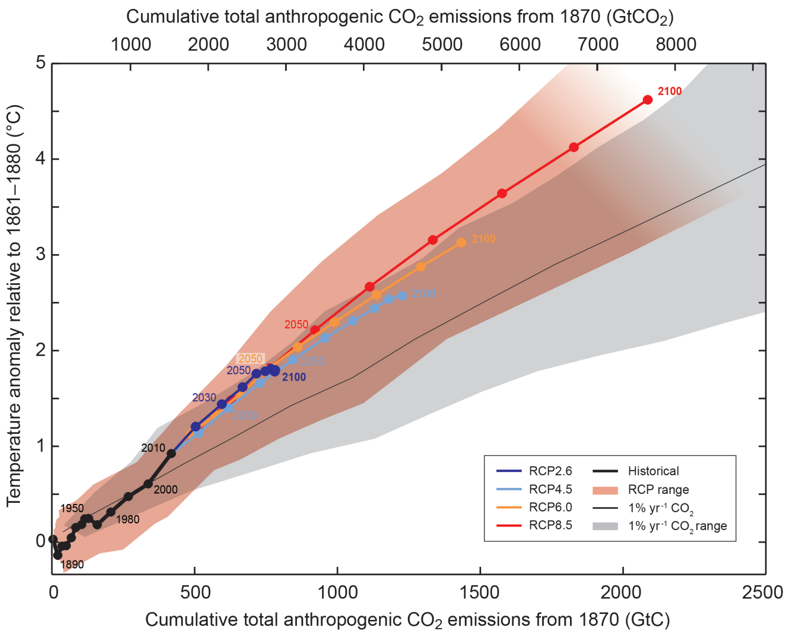

Figure 1: Linearity – Cumulative emissions and temperature, IPCC et al. (2022)

Fourth, I consider linear relationship between global and local temperatures, namely:

$$T_{it} = \bar{T}_{i0} + \Delta_i \mathcal{T}_t = \bar{T}_{i0} + \Delta_i \chi \mathcal{S}_t \ ,$$

where Δ_i is a linear pattern scaling parameter that depends on geographical factors such as albedo or latitude. In the quantification Section 4.5, I explain how I estimate this pattern scaling by regressing local temperatures on global temperature.

Fifth, I consider a period damage function $\hat{\mathcal{D}}(T_{it} - T_i^\star)$ where $T_i^\star$ is the "optimal" temperature for country i. The function $\mathcal{D}(\hat{T})$ is a reduced-form representation of the economic damage to productivity. In the baseline quantification, I assume damages are quadratic, as in standard Integrated Assessment Models. This methodology follows Krusell and Smith (2022), Kotlikoff, Kubler, Polbin, Sachs and Scheidegger (2021) and Burke et al. (2015). Such damage creates winners and losers: countries that are substantially warmer than a target temperature $T_i^\star$ are extremely affected by increases in temperature due to climate change. In contrast, regions with negative $T_{it} - T_i^\star$ benefit – at least in the short-run – from a warmer climate. I consider a slight deviation to the above articles by assuming that the target temperature $T_i^\star$ might be different across countries: an already warm regions have different adaptation costs compared to a country which is historically cold. As a result, the target temperature $T_i^\star = \alpha T^\star + (1-\alpha)\bar{T}_{i0}$ can be more or less tilted toward historical temperature. I discuss this quantification in Section 4.5.

Finally, to obtain a reduced-form static damage function $\mathcal{D}_i(\mathcal{E})$ we summarize the future costs of climate change in present-discounted value

$$\mathcal{D}_i(\mathcal{E}) = \bar{\rho}\int_{t_0}^{\infty} e^{-\bar{\rho}t}\,\hat{\mathcal{D}}(T_{it} - T_i^\star)dt\ ,$$

with $\bar{\rho} = \rho - n + \eta\bar{g}$ the "effective discount factor" and ρ is the household discount factor, n is the global population growth and $\bar{g}$ the global TFP growth. This net-of-damage function $\mathcal{D}_i(\mathcal{E})$ will be internalized by the Social Planner when making optimal climate policy choices.

2.5 Equilibrium

To close the model, we need to determine the final good prices for each country p_i, and we consider the market clearing for each good i

$$\begin{aligned}
\mathcal{P}_i y_i &= \sum_{k\in\mathbb{I}} \mathcal{P}_k \tau_{ki} c_{ki} + \sum_{k\in\mathbb{I}} \mathcal{P}_k \tau_{ki} (x_{ki}^f + x_{ki}^c + x_{ki}^r) \\
\mathcal{P}_i\, \mathrm{p}_i \underbrace{y_i}_{=\mathcal{D}_i^y(\mathcal{E}) z_i F(\cdot)} &= \sum_{k\in\mathbb{I}} \frac{s_{ki}}{1+\mathrm{t}_{ki}^b} \left(\mathcal{P}_k \mathrm{p}_k y_k + q^f(e_k^x - \mathcal{P}_k e_k^f) + \mathcal{P}_k \mathrm{t}_k^{ls}\right)
\end{aligned} \tag{11}$$

where x_{ki}^f, x_{ki}^c and x_{ki}^r are the good inputs used by country k and imported from country i to produce fossil and renewable energy respectively. The second equation is a reformulation of the market clearing where the sales of countries i equals the expenditures from all countries k, coming from their incomes in good sales as well as net-exports of fossil energy.

To summarize, the competitive equilibrium of this economy is defined as follows:

Definition. Competitive equilibrium (C.E.):
For a set of policies $\{\mathrm{t}_i^\varepsilon, \mathrm{t}_{ij}^b, \mathrm{t}_i^{ls}\}_i$ across countries, a C.E. is a set of decisions $\{c_{ij}, e_i^f, e_i^c, e_i^r, e_i^x, \bar{e}_i^c, \bar{e}_i^r\}_{ij}$, and prices $q^f, \{\mathrm{p}_i, w_i, q_i^c, q_i^r\}_i$ such that
(i) Households choose consumption $\{c_{ij}\}_{ij}$ maximizing utility eq. (1) s.t. the budget constraint eq. (2), which yield trade shares eq. (3)
(ii) Final good firms choose inputs $\{\ell_i, e_i^f, e_i^r\}_i$ to maximize profits, resulting in eq. (5)
(iii) Fossil energy firms maximize profits eq. (6) and extract/produce $\{e_i^x\}_i$ given by eq. (7)
(iv) Renewable and coal energy firms maximize profits, and supplies $\{\bar{e}_i^c, \bar{e}_i^r\}$ are given respectively by eq. (9) and eq. (10)
(v) Energy markets clears for fossils as in eq. (8) and for coal and renewable in eq. (9) and eq. (10)
(vi) Good markets clear for final good for each country as in eq. (11), and trade is balanced by Walras Law.

3 The optimal agreement design with endogenous participation

Because of unequal exposure to climate change and carbon policy, countries have different incentives to enforce climate policy, creating a free-riding problem. Therefore, the optimal carbon tax needs to account for endogenous participation. Designing a climate agreement reveals a trade-off between an intensive margin – associated with the choice of the policy instruments – and an extensive margin – related to the extent of participation in the agreement.

3.1 Agreement design and participation constraints

The social planner solves a Ramsey problem, choosing the optimal agreement, which boils down to a carbon tax, retaliatory tariffs on non-participants, and a set of countries participating in the agreement, subject to participation constraints. I first design the set of climate agreements considered and then define the planner's objective.

Definition. Climate Agreements

A climate agreement is a set $\{\mathbb{J}, \mathrm{t}^\varepsilon, \mathrm{t}^b\}$, with a coalition of countries $\mathbb{J} \subseteq \mathbb{I}$, a carbon tax t^ε, and a tariff t^b, such that in the competitive equilibrium,

- Countries $i \in \mathbb{J}$ are subject to a carbon tax t^ε on fossil energy e_i^f.
- If country j exits the agreement, club members $i \in \mathbb{J}$ charge uniform tariffs $\mathrm{t}_{ij}^b = \mathrm{t}^b$ on the final good from j.
- Countries in the club rebate the revenues of the carbon tax and tariffs to the household $\mathrm{t}_i^{ls} = \mathrm{t}^\varepsilon(\xi^f e_i^f + \xi^c e_i^c) + \sum_{j \notin \mathbb{J}} \mathrm{t}^b \tau_{ij} \mathrm{p}_j (c_{ij} + x_{ij}^f + x_{ij}^c + x_{ij}^r)$
- Countries inside the club, $i, j \in \mathbb{J}$, benefit from free-trade $\mathrm{t}_{ij}^b = 0$.
- Countries outside the club, $k \notin \mathbb{J}$, keep a passive trade policy,[14] $\mathrm{t}_{k\ell}^b = 0, \ \forall \ell \in \mathbb{I}$.
- All countries – members as well as non-members – still trade in fossil (oil-gas) energy at international price q^f.[15]

I keep the number of policy instruments $\mathbf{t} = \{\mathrm{t}^\varepsilon, \mathrm{t}^b\}$ considered in the agreement purposefully small for two reasons. First, this is consistent with the idea behind deviations of the Coase theorem: when bargaining over many policy instruments is associated with transaction costs, a negotiation between n parties – $\mathbb{J}$ countries here – "can be prevented from attaining a socially desirable outcome", c.f. Weitzman (2015). For example, quantity targets ε_i and bilateral tariffs t_{ij}^b exacerbate free-riding incentives since an agreement requires all countries to accept the policies of all the other countries.[16] Second, carbon pricing is based on standard principles of Pigouvian taxation, where the optimal carbon tax equals the marginal cost of emitting one additional ton

[14]This assumption will be relaxed in the extension, Section 7.3.

[15]This assumption will be relaxed in Section 7.2.

[16]For this reason, I refrain from studying agreements that bargain over a set of country-specific taxes t_i^ε or emissions quantity targets ε_i, and bilateral tariffs t_{ij}^b, since the bargaining costs increase at least proportionally – if not exponentially – in the number of countries. The case of individual carbon taxes is analyzed in Bourany (2024) absent free-riding. Accounting for strategic interactions between I countries also makes the problem computationally more involved, as in theory all the instruments need to account jointly for the Lagrange multipliers of all the participation constraints, for any possible coalition.

of carbon – the Social Cost of Carbon – common for all countries. This optimal uniform carbon price serves as a "focal point" of an international agreement, and the goal of this policy problem is to compare how it needs to be changed when accounting for free-riding.

Participation constraints

I define indirect utility $\mathcal{U}_i(\mathbb{J}, \mathrm{t}^\varepsilon, \mathrm{t}^b) \equiv u\big(c_{i(\mathbb{J},\mathrm{t}^\varepsilon,\mathrm{t}^b)}\mathcal{D}_i(\mathcal{E}_{(\mathbb{J},\mathrm{t}^\varepsilon,\mathrm{t}^b)})\big)$ as in Section 2.1. Then we can define participation constraints in two ways, depending on the type of deviations we consider.

1. *Unilateral deviation:* country i can choose to exit the agreement unilaterally. This does not affect the composition of the agreement or the decision of the other members. Country i in the agreement will participate if the value of staying is larger than the value of being outside the agreement:

$$\mathcal{U}_i(\mathbb{J}, \mathrm{t}^\varepsilon, \mathrm{t}^b) \geq \mathcal{U}_i(\mathbb{J}\backslash\{i\}, \mathrm{t}^\varepsilon, \mathrm{t}^b) \qquad \forall\, i \subseteq \mathbb{J}\,. \qquad \text{[Unilateral-Nash PC]} \tag{12}$$

2. *Sub-coalition deviation:* country i can choose to exit the agreement in cooperation with other members of a potential sub-coalition $\hat{\mathbb{J}}$. All these members leave the agreement. The decision of all those countries $i \in \hat{\mathbb{J}}$ to leave is made jointly: the value of being outside is above the value of staying for all $i \in \hat{\mathbb{J}}$. This makes the participation constraints more intricate and write as follow:

$$\mathcal{U}_i(\mathbb{J}, \mathrm{t}^\varepsilon, \mathrm{t}^b) \geq \mathcal{U}_i(\mathbb{J}\backslash\hat{\mathbb{J}}, \mathrm{t}^\varepsilon, \mathrm{t}^b) \qquad \forall\, i \in \hat{\mathbb{J}} \;\&\; \forall\, \hat{\mathbb{J}} \subseteq \mathbb{J}\,. \quad \text{[Coalition-Nash PC]} \tag{13}$$

The optimal agreement needs to account for these participation constraints, and be robust to unilateral or sub-coalition deviations.

Welfare criterion and planner's objective

We consider a global social planner maximizing the world welfare:

$$\max_{\mathbb{J},\mathrm{t}^\varepsilon,\mathrm{t}^b} \mathcal{W}(\mathbb{J}, \mathrm{t}^\varepsilon, \mathrm{t}^b) = \max_{\mathbb{J},\mathrm{t}^\varepsilon,\mathrm{t}^b} \sum_{i\in\mathbb{I}} \mathcal{P}_i\omega_i\, \mathcal{U}_i(\mathbb{J}, \mathrm{t}^\varepsilon, \mathrm{t}^b)\ , \tag{14}$$

subject to participation constraint – Unilateral Nash, robust to deviation

$$\mathbb{J} \in \mathbb{S}(\mathrm{t}^\varepsilon, \mathrm{t}^b) = \{\mathcal{J} \mid \mathcal{U}_i(\mathcal{J}, \mathrm{t}^\varepsilon, \mathrm{t}^b) \geq\ \mathcal{U}_i(\mathcal{J}\backslash\{i\}, \mathrm{t}^\varepsilon, \mathrm{t}^b), \quad \forall\, i \in \mathcal{J}\}\ , \tag{15}$$

or Coalitional-Nash, robust to sub-coalition $\hat{\mathbb{J}}$ deviations:

$$\mathbb{J} \in \mathbb{C}(\mathrm{t}^\varepsilon, \mathrm{t}^b) = \{\mathcal{J} \mid \mathcal{U}_i(\mathcal{J}, \mathrm{t}^\varepsilon, \mathrm{t}^b) \geq \mathcal{U}_i(\mathcal{J}\backslash\hat{\mathbb{J}}, \mathrm{t}^\varepsilon, \mathrm{t}^b), \quad \forall\, i \in \mathcal{J},\ \forall\, i \in \hat{\mathbb{J}} \,\&\, \forall\, \hat{\mathbb{J}} \subseteq \mathcal{J}\}\,. \tag{16}$$

where ω_i are the Pareto weights, and $\mathcal{P}_i$ the population size of country i. The social planner maximizes world welfare, in part with the goal of fighting the climate change externality. As a result, the planner maximizes the sum over $\mathbb{I}$ instead of $\mathbb{J}$. In the case where a planner only maximizes the coalition $\mathbb{J}$'s welfare, it would yield the unintended consequences that the optimal agreement could be restricted to a subset of rich, cold, high-value $\mathcal{U}_i$ countries which would manipulate terms-of-trade and potentially subsidize fossil fuels. I give intuitions for such results in Section 5.2 when countries choose their climate and trade policy unilaterally. However, these outcomes are unintended for global climate agreements, which aim at maximizing world welfare.

The set of agreements stable under Coalition-Nash resembles the concept of "core" in general equilibrium theory. Both of these sets $\mathbb{S}(\mathrm{t}^\varepsilon, \mathrm{t}^b)$, $\mathbb{C}(\mathrm{t}^\varepsilon, \mathrm{t}^b)$ could be empty: it is possible that no country finds it beneficial to be part of the agreement for a given policy $\mathrm{t}^\varepsilon, \mathrm{t}^b$.

Extensive margin vs. intensive margin tradeoff

The problem of a world planner determines jointly the policy instruments $\mathbf{t} = \{\mathrm{t}^\varepsilon, \mathrm{t}^b\}$ and the choice of the country subject to participation constraints $\mathbb{J} \in \mathbb{S}(\mathrm{t}^\varepsilon, \mathrm{t}^b)$ or $\mathbb{C}(\mathrm{t}^\varepsilon, \mathrm{t}^b)$. As a result, this reveals an extensive-intensive margin trade-off. For a given set of participants $\mathbb{J}$, higher carbon tax t^ε and lower tariffs t^b increase global welfare. That is, this planner would like to reduce carbon emissions and promote free trade at the intensive margin to maximize welfare. However, this choice of instruments also affects countries' participation: higher taxes t^ε and lower tariffs t^b reduce incentives for countries to participate. If a country deviates by exiting the agreement, it increases its emissions, and international trade is reduced, lowering welfare at the extensive margin. As a result, the planner would like to balance these two countervailing effects. This tradeoff is analyzed in detail in the context of this model in Section 6.3.

3.2 Optimal design and solution method

This design problem combines a choice of instruments and a choice of countries, making it difficult to solve. I provide two methods to handle the joint optimal policy/combinatorial discrete choice problem.

3.2.1 Framework for the optimal design

I formalize the policy problem under the two types of participation constraints – Unilateral deviations vs. Coalition deviations – subject to the allocation being a competitive equilibrium. I consider a general class of policy instruments $\mathbf{t}$ that encompass carbon tax $\mathrm{t}_i^\varepsilon = \mathrm{t}^\varepsilon \mathbb{1}_{\{i \in \mathbb{J}\}}$, uniform tariffs on non-members $\mathrm{t}_{ij}^b = \mathrm{t}^b \mathbb{1}_{\{i \in \mathbb{J}, j \notin \mathbb{J}\}}$ as well as potential additional instruments as analyzed in Section 7. The design problem can be stated as

$$\max_{\mathbb{J},\mathbf{t}} \mathcal{W}(\mathbb{J},\mathbf{t}) = \max_{\mathbb{J},\mathbf{t}} \sum_{i\in\mathbb{I}} \mathcal{P}_i\, \omega_i\, \mathcal{U}_i(\mathbb{J},\mathbf{t}) \ ,$$
$$s.t. \qquad \mathbb{J} \in \mathbb{S}(\mathbf{t})\,, \quad or \quad \mathbb{J} \in \mathbb{C}(\mathbf{t}) \ .$$

Participation constraints $\mathbb{S}(\mathbf{t})$ or $\mathbb{C}(\mathbf{t})$, as defined in eq. (15) and eq. (16), make the problem intricate as they limit the instruments the planner can use for each set of countries in the agreements.

I take the following approach: I split the problem into an inner problem and an outer problem. First, the planner chooses the policy instruments $\mathbf{t}$. Then, given $\mathbf{t}$, participation constraints yield a set of achievable coalitions $\mathbb{J}$. If no coalition is achievable, then the welfare for those instruments is $-\infty$. This choice of country in the inner problem is analogous to a combinatorial discrete choice problem (CDCP).

$$\max_{\substack{\mathbb{J},\mathbf{t} \\ \mathbb{J}\in\mathbb{S}(\mathbf{t})}} \mathcal{W}(\mathbb{J},\mathbf{t}) = \max_{\mathbf{t}} \max_{\mathbb{J}\,|\,\mathbb{J}\in\mathbb{S}(\mathbf{t})} \mathcal{W}(\mathbb{J},\mathbf{t}) \ .$$

I explain in Appendix E why the opposite approach – solving for the coalition as an outer problem and for policy instruments in the inner problem – is intractable.[17].

In the approach presented here, the *outer problem* for the choice of instrument $\mathbf{t}$ is solved with a simple grid search[18] since the indirect welfare is now discontinuous and non-convex in the application:

$$\max_{\mathbf{t}} \widehat{\mathcal{W}}(\mathbf{t}) \ , \qquad \text{where} \qquad \widehat{\mathcal{W}}(\mathbf{t}) = \max_{\mathbb{J}\,|\,\mathbb{J}\in\mathbb{S}(\mathbf{t})} \mathcal{W}(\mathbb{J},\mathbf{t}) \ .$$

3.2.2 Solution methods

I propose two methods to solve the *inner problem* of the optimal choice of countries $\mathbb{J}^\star$, out of all the possible combinations $\mathcal{P}(\mathbb{I})$. This combinatorial discrete choice problem is prohibitive numerically for large numbers of countries $\#\mathbb{I}$. To handle this challenge, I first use an exhaustive search method (brute force method), and then I propose a squeezing procedure inspired by tools from the trade literature, e.g. Arkolakis, Eckert and Shi (2023), as a more efficient alternative in the unilateral-Nash case. I introduce the combinatorial problem before presenting each method in turn. The combinatorial discrete choice problem for a given policy $\mathbf{t}$ is given by:

$$\max_{\mathbb{J}\in\mathcal{P}(\mathbb{I})} \mathcal{W}(\mathbb{J},\mathbf{t}) \ ,$$
$$s.t. \qquad \mathbb{J}\in\mathbb{S}(\mathbf{t}) \ .$$

I express the Lagrangian of the constrained optimization, with multiplier ν_i for country i's participation, we obtain, with a slight abuse of notation[19], as follow:

$$\max_{\mathbb{J}\in\mathcal{P}(\mathbb{I})} \ \mathcal{W}(\mathbb{J},\mathbf{t}) + \sum_{i\in\mathbb{J}} \nu_{i,\mathbb{J}} \Big(\mathcal{U}_i(\mathbb{J},\mathbf{t}) - \mathcal{U}_i(\mathbb{J}\backslash\{i\},\mathbf{t}) \Big) =: \max_{\mathbb{J}\in\mathcal{P}(\mathbb{I})} \widetilde{\mathcal{W}}(\mathbb{J},\mathbf{t}) \tag{17}$$

in the case where the participation constraints only account for unilateral deviations.[20]

[17] In subgame perfect equilibria, it makes the Lagrange multipliers on the participation constraints ν_i depends not only on the coalition considered $\mathbb{J}$, but all the coalition in every subgame, e.g. $\mathbb{J}\backslash\{i\}$ etc. These multipliers affect the policy choice and make the problem unsolvable

[18] For this reason, keeping the number of instruments small $\mathbf{t} = \{\mathrm{t}^\varepsilon, \mathrm{t}^b\}$ is a computational advantage for searching over a low-dimensional $\mathbf{t}$.

[19] The objective function $\mathcal{W}(\mathbb{J},\mathbf{t})$ is not continuous, differentiable, or even convex. The handling of the inequality constraints could not, in theory, rely on the KKT theorem which applies in the $\mathcal{C}^1$ and convex case. However, with $\nu_{i,\mathbb{J}} = 0$ if $\mathbb{J}\in\mathbb{S}(\mathbf{t})$ and $\nu_{i,\mathbb{J}} = \infty$ if $\mathbb{J}\notin\mathbb{S}(\mathbf{t})$, the problem well defined.

[20] A longer list of constraints needs to be included if we consider coalition deviations.

First method: Exhaustive enumeration

First, when the number of countries $I = \#\mathbb{I}$ is small, one obvious yet costly solution is to perform an exhaustive search over $\mathcal{P}(\mathbb{I})$. The idea is to enumerate all the combinations $\mathbb{J} \in \mathcal{P}(\mathbb{I})$, and evaluate welfare $\mathcal{W}(\mathbb{J}, \mathbf{t})$. This has evidently a computational cost proportional to $2^{\#\mathbb{I}}$, i.e. the number of potential combinations.

This solution has, however, the advantage of considering all the participation constraints – including the coalition-robust agreements – "for free". Indeed, we can assess if the coalition is stable both in the case of unilateral-Nash $\mathbb{J} \in \mathbb{S}(\mathbf{t})$ and coalitional-Nash $\mathbb{J} \in \mathbb{C}(\mathbf{t})$, for all sets $\mathbb{J}$. This is feasible because every possible deviation of sub-groups $\hat{\mathbb{J}}$ yields a new agreement $\mathbb{J}' = \mathbb{J} \backslash \hat{\mathbb{J}}$ which is already computed as another coalition $\mathbb{J}' \in \mathcal{P}(\mathbb{I})$. If one of the participation constraints is violated, the set considered $\mathbb{J}'$ is discarded, i.e. $\nu_{i,\mathbb{J}'} = \infty$, $\mathcal{W}(\mathbb{J}') = -\infty$. In practice, several coalitions can be stable for a given policy $\mathbf{t}$, and the exhaustive search selects the one that maximizes welfare. In practice, among all the stable coalitions the one that maximizes welfare is the largest one since for a given policy $\mathbf{t} = \{\mathrm{t}^{\varepsilon}, \mathrm{t}^{b}\}$, the larger the coalition, the higher the gains from trade and the gains from reducing emissions.[21]

Second method: Squeezing procedure for CDCP with Participation Constraints

Second, since full enumeration is costly, I provide an alternative algorithm inspired by methods used in the international trade literature to solve combinatorial discrete choice problems. The additional difficulty that needs to be considered is the presence of participation constraints. In this section, we only consider unilateral deviations. The idea behind this method is greatly inspired by Arkolakis, Eckert and Shi (2023) and Farrokhi and Lashkaripour (2024).

The idea is to build iteratively sets that are lower bound $\underline{\mathcal{J}}$ and upper bound $\overline{\mathcal{J}}$ for the optimal coalition $\mathbb{J}$: subset $\underline{\mathcal{J}}$ *includes* all the countries that are known to be part of the optimal set $\mathbb{J}$ and $\overline{\mathcal{J}}$ is a superset, which it *excludes* the countries that we know are not part of the optimal set. The set $\overline{\mathcal{J}} \backslash \underline{\mathcal{J}}$ is the set of potential countries. The natural starting point is $\underline{\mathcal{J}} = \varnothing$, $\overline{\mathcal{J}} = \mathbb{I}$.

The squeezing step in standard CDCP is a mapping from $\mathcal{J}$ to members that bring a positive marginal value to the objective $\mathcal{W}(\mathbb{J}) := \mathcal{W}(\mathbb{J}, \mathbf{t})$. The modification needed in settings with participation constraints is that the country also needs to gain marginal *individual* value $\mathcal{U}_i(\mathbb{J}) := \mathcal{U}_i(\mathbb{J}, \mathbf{t})$ to be part of the coalition:

$$\Phi(\mathcal{J}, \mathbf{t}) \equiv \{ j \in \mathbb{I} \,|\, \Delta_j \mathcal{W}(\mathcal{J}, \mathbf{t}) > 0 \;\&\; \Delta_j \mathcal{U}_j(\mathcal{J}, \mathbf{t}) > 0 \} \tag{18}$$

where the marginal values for global welfare and individual welfare are

$$\Delta_j \mathcal{W}(\mathcal{J}, \mathbf{t}) \equiv \mathcal{W}(\mathcal{J} \cup \{j\}, \mathbf{t}) - \mathcal{W}(\mathcal{J} \backslash \{j\}, \mathbf{t}) = \sum_{i \in \mathbb{I}} \mathcal{P}_i \omega_i \big(\mathcal{U}_i(\mathcal{J} \cup \{j\}, \mathbf{t}) - \mathcal{U}_i(\mathcal{J} \backslash \{j\}, \mathbf{t}) \big)$$

$$\Delta_j \mathcal{U}_i(\mathcal{J}, \mathbf{t}) \equiv \mathcal{U}_i(\mathcal{J} \cup \{j\}, \mathbf{t}) - \mathcal{U}_i(\mathcal{J} \backslash \{j\}, \mathbf{t})$$

[21] As long as t^{ε} is below the globally optimal carbon tax as we derive it in Section 5.1.2

The iterative procedure builds the lower bound $\underline{\mathcal{J}}$ and upper bound $\overline{\mathcal{J}}$ by successive application of the squeezing step.

$$\underline{\mathcal{J}}^{(k+1)} = \Phi(\underline{\mathcal{J}}^{(k)}, \mathbf{t}) \qquad\qquad \overline{\mathcal{J}}^{(k+1)} = \Phi(\overline{\mathcal{J}}^{(k)}, \mathbf{t}) \tag{19}$$

Under some conditions (complementarity, as defined next section and in the appendix) this sequential procedure yields two sets $\underline{\mathcal{J}}$ and $\overline{\mathcal{J}}$ such that $\underline{\mathcal{J}} \subseteq \mathbb{J} \subseteq \overline{\mathcal{J}}$. In some cases $\underline{\mathcal{J}} = \overline{\mathcal{J}} = \mathbb{J}$, yielding the optimal coalition. If not, with $\overline{\mathcal{J}} \backslash \underline{\mathcal{J}} = \mathcal{J}^{pot}$, we find the optimal coalition by searching exhaustively over all coalitions $\mathcal{J} = \underline{\mathcal{J}} \cup \hat{\mathcal{J}}$, with $\hat{\mathcal{J}} \in \mathcal{P}(\mathcal{J}^{pot})$.

Applicability of the squeezing procedure

From the combinatorial discrete choice literature, Arkolakis, Eckert and Shi (2023), we know that the squeezing procedure applies in cases where the model exhibits "complementarity" or single-crossing differences in choices. We detail how these conditions can be expressed in Appendix E.2.

Indeed, we say that the objective $\mathcal{W}(\mathcal{J})$ satisfies "complementarity", if the marginal gain $\Delta_j \mathcal{W}(\mathcal{J})$ of the objective is monotone in the set $\mathcal{J}$, i.e. $\Delta_j \mathcal{W}(\mathcal{J}) \leq \Delta_j \mathcal{W}(\mathcal{J}')$, for $\mathcal{J} \subseteq \mathcal{J}'$ & $j \in \mathbb{I}$. In the climate agreement setting, participation constraints and stability require to adjust the welfare objective, from $\mathcal{W}(\mathbb{J})$ to $\widetilde{\mathcal{W}}(\mathbb{J})$ as in eq. (17). In this context, the complementarity (or single crossing differences in choice for its weaker form), with participation constraints, takes an intricate form (SCD-C-PC) which we detail in Appendix E.2.

Theorem The SCD-C-PC from below is *sufficient* for the application of modified squeezing algorithm, i.e. successive application of eq. (18), starting from $\{\varnothing, \mathbb{I}\}$ and eq. (19), to yield bounding sets $\underline{\mathcal{J}} \subseteq \mathbb{J} \subseteq \overline{\mathcal{J}}$ in CDCPs with participation constraints.

One of the advantages is that, for a small number of countries $\#\mathbb{I} \approx 10$, we can evaluate numerically if the sufficient conditions mentioned above are satisfied. The fact that the model is rich, with many dimensions of heterogeneity and general equilibrium effects through energy markets and international trade, prevents the simple evaluation of those sufficient conditions analytically.

4 Quantification

The model is calibrated to a panel of ten regions to provide realistic predictions on the impact of optimal carbon policy. I first describe the data used. I then provide details on the quantification, which functional forms are used, and how the parameters are calibrated to match the data. I summarize in Table 1 the dimensions of heterogeneity of the model. Table 2 in the appendix contains the summary table for the calibration described in this section.

4.1 Data

First, I describe briefly the data used to calibrate the model. I use data for the year 2018-2023, taking the average over that period to smooth out the effect of the COVID-19 recession on energy and macroeconomic data.

I use a sample of 10 "regions": (i) US and Canada, (ii) China and Hong Kong, (iii) the European Union, United Kingdom, and other countries of the Schengen Area, (iv) South Asia (India, Pakistan, Bangladesh, Nepal) (v) Sub-saharian Africa, (vi) Middle-East and North Africa, (vii) Russia and CIS, (viii) Japan, Korea, Australia, Taiwan and Singapore, (ix) South-East Asia (Asean), (x) South and Central America.[22]

I use data for GDP per capita, in Purchasing Power Parity (PPP, in 2011 USD) from the World Bank, as collected and processed by the Maddison Project (Bolt and van Zanden (2023)). For the energy variables, I use the comprehensive data collected and processed in the Statistical Review of Energy (Energy Institute (2024)) that includes the production and consumption of various energy sources, including Oil, Gas, and Coal. It also includes proven reserves of those fossil fuels. For energy rent, I use the World Development Indicators that use national accounts to measure the share of GDP coming from energy (oil, gas and coal) and natural resource rents. Finally, for temperature, I use the same time series as Burke et al. (2015), which use the temperature at country level, averaged over the year and weighted by population across locations. For trade variables, I take the trade flows and gravity variables compiled by the CEPII in Conte et al. (2022).

4.2 Welfare and Pareto weights

The welfare function that the climate agreement designer would maximize is the weighted sum of individual utilities in all countries:

$$\mathcal{W}(\mathbb{J}, \mathbf{t}) = \sum_{i \in \mathbb{I}} \mathcal{P}_i \, \omega_i \, \mathcal{U}_i$$

with $\mathcal{P}_i$ the population size per country, ω_i the Pareto weights and $\mathcal{U}_i$ the country indirect utility per capita. Note that the climate agreement designer maximizes the *world* welfare.

Following the discussion in Anthoff et al. (2009), Nordhaus (2011) and Nordhaus and Yang (1996), one would like to choose Pareto weights that eliminate redistributive effects that are orthogonal to climate change and climate policy. To that purpose, I choose the "Negishi" Pareto weights that make the preexisting competitive equilibrium efficient under that welfare metric. This implies that:

$$\begin{aligned} \omega_i = \frac{1}{u'(\bar{c}_i)} \qquad \qquad & \Leftrightarrow \qquad \qquad C.E.(\bar{c}_i) \in \underset{\bar{c}_i}{\operatorname{argmax}} \sum_i \mathcal{P}_i \omega_i u(\bar{c}_i) \\ \omega_i u'(\bar{c}_i) = \omega_j u'(\bar{c}_j) \qquad \forall i, j \ \in \mathbb{I} & \end{aligned}$$

where $\bar{c}_i$ is the consumption level in the present competitive equilibrium – the period 2018-2023 – absent future climate damage. This implies that the climate agreement and the carbon policy do not look for redistributing across countries through goods and energy general equilibrium effects.

[22] In a future iteration of this project, I consider a panel of 25 countries and seven regions, which gather the G20 and additional large countries.

However, global warming, carbon taxation, and tariffs have redistributive effects, as they change the distribution of c_i. These effects are taken into account in the choice of policies, as we see in Section 5.1.2. In Figure 2, I display the weights ω_i, and $\omega_i \widehat{\mathcal{P}}_i$ adjusted for population $\widehat{\mathcal{P}}_i = \mathcal{P}_i / \mathcal{P}$.

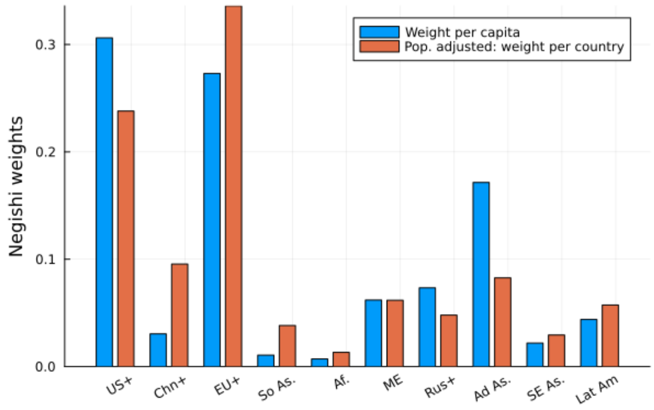

Figure 2: Pareto weights across regions
weights ω_i (blue, left), $\widehat{\mathcal{P}}_i \omega_i$ (red, right)

4.3 Macroeconomy, trade and production

For the macroeconomic part of the framework, I consider standard utility and production functions. First, I consider constant relative risk aversion (CRRA) utility over consumption, as well as climate damage in the utility function. This implies that countries that have a very low production / GDP per capita still suffer potential large losses due to climate damages:

$$\mathcal{U}_i = \frac{\left(c_i \widetilde{\mathcal{D}}_i^u(\mathcal{E})\right)^{1-\eta}}{1-\eta} = \frac{c_i^{1-\eta}}{1-\eta} \mathcal{D}_i^u(\mathcal{E})$$

I calibrate the CRRA/IES parameter to be $\eta = 1.5$, taken from Barrage and Nordhaus (2024).[23] Moreover, the damage $\mathcal{D}_i^u(\mathcal{E})$ is adjusted for the curvature in the utility function.

For production, I use a nested CES framework. The firm combines a Cobb-Douglas bundle of capital k_i and labor ℓ_i[24] with a composite of energy e_i, with elasticity σ^y. Second, the energy e_i aggregates the different energy sources: oil and gas e^f, coal e_i^c, and renewable/non-carbon e_i^r, with elasticity σ^e.

[23]This is slightly lower than the standard value $\eta = 2$, for the reason that higher curvature would imply more unequal weights, ω_i, across different countries.

[24]Labor is inelastically supplied $\ell_i = \bar{\ell}_i$ in each country and normalized to 1 – since the country size $\mathcal{P}_i$ is already taken into account. As a result, all the variables can be seen as input per capita.

$$\text{Output} \qquad y_i = \mathcal{D}_i^y(\mathcal{E})\, z_i\, \bar{y}_i = \mathcal{D}^y(\mathcal{E})\, z_i \Big((1-\varepsilon)^{\frac{1}{\sigma^y}} (e_i)^{\frac{\sigma^y - 1}{\sigma^y}} + \varepsilon^{\frac{1}{\sigma^y}} (k_i^{\alpha} \ell_i^{1-\alpha})^{\frac{\sigma^y-1}{\sigma^y}} \Big)^{\frac{\sigma^y}{\sigma^y - 1}} \,,$$

$$\text{Energy} \qquad e_i = \Big((\omega^f)^{\frac{1}{\sigma^e}} (e_i^f)^{\frac{\sigma^e-1}{\sigma^e}} + (\omega^c)^{\frac{1}{\sigma^e}} (e_i^c)^{\frac{\sigma^e-1}{\sigma^e}} + (\omega^r)^{\frac{1}{\sigma^e}} (e_i^r)^{\frac{\sigma^e-1}{\sigma^e}} \Big)^{\frac{\sigma^e}{\sigma^e-1}} \,.$$

To calibrate these functions, I set the capital-labor ratio to be $\alpha = 0.35$ to match the cost share of capital. For the energy share, I set $\varepsilon = 0.10$ to match the average energy cost share of $\frac{q_i^e e_i}{\mathrm{p}_i y_i} = 6\%$, as measured in Kotlikoff, Kubler, Polbin and Scheidegger (2021) and used in Krusell and Smith (2022). For the elasticity between energy and other inputs, I set $\sigma^y = 0.3$ for all countries, which is in the range of estimates in Papageorgiou et al. (2017), among others.[25] This implies that capital/labor and energy are complementary in production: an increase in the price of energy has a strong impact on output as it is less productive to "substitute away" to other inputs – capital, labor here. This aligns with other empirical and structural evidence on the impact of energy shocks, e.g. Hassler et al. (2019). For each energy source, I calibrate the energy mix for oil-gas, with $\omega^f = 0.56$, coal $\omega^c = 0.27$, and non-carbon $\omega^r = 0.17$, to match the aggregate shares in each of these energy sources in the data. In the next section, I document how to match the individual countries' energy mix using energy prices/costs. Finally, for the elasticity between energy inputs, I use the value $\sigma_e = 2$, following the rest of the literature, i.e. Papageorgiou et al. (2017), Kotlikoff, Kubler, Polbin and Scheidegger (2021), and Hillebrand and Hillebrand (2019), among others.

I calibrate the productivity z_i of the production function $y_i = \mathcal{D}_i^y(\mathcal{E})\, z_i\, \bar{y}_i$ to match exactly the GDP, $y_i \mathrm{p}_i$, across countries. This parameter z_i, represents productivity residuals as well as institutional/efficiency differences across countries. In Figure 3, I show the GDP levels, as they replicated with this model.

Finally, we use trade flow data as seen in Figure 22 to match the pattern of international trade in goods. First, I estimate a gravity regression between trade flow and geographical distance[26]– with fixed effects for importers and exporters – finding an elasticity with distance $\kappa = -1.72$. To rationalize it in the model, I project iceberg trade costs on this geographical distance $\tau_{ij} = d_{ij}^{\beta}$. All the residual differences in trade flows, not rationalized by trade costs, τ_{ij}, prices, p_j, or demand, y_i, are then explained by differences in preferences a_{ij}. We calibrate β to minimize dispersion in a_{ij} over countries j, which implies $\beta = 0.375$ and yields the trade elasticity $\theta = 5.6$. I calibrate those parameters a_{ij} to minimize the distance – mean squared error – between model-generated trade shares $s_{ij} = \frac{c_{ij}\tau_{ij}\mathrm{p}_j}{c_i \mathbb{P}_i} = a_{ij}\frac{(\tau_{ij}\mathrm{p}_j)^{1-\theta}}{\sum_k a_{ik}(\tau_{ik}\mathrm{p}_k)^{1-\theta}}$, and observed shares $\bar{s}_{ij}$ in the data. Since our model imposes a trade balance, those trade shares cannot be matched exactly, only approximately. It does match the fact that some countries are relying more on trade exports and imports – like China, East Asia and South-East Asia – compared to others – Middle East, Africa, and Russia.

[25] It also aligns with my own estimation in Bourany (2022)

[26] The gravity regression is standard: $\log x_{ij} = \kappa \log d_{ij} + \alpha_i + \gamma_j$. In the model, $\tau_{ij} = d_{ij}^{\beta}$, we get $\kappa = (1-\theta)\beta$.

4.4 Energy markets

For the energy market, I match the energy mix of different countries, using the CES framework displayed above, as well as differences in cost of production. For the supply side, we use iso-elastic fossil extraction cost, to replicate the oil-gas supply of fossil producers.

First, in this model, oil and gas are traded on international markets, with demand $\mathcal{P}_i e_i^f$ from the final good firm and supply e_i^x from the fossil energy firm, extracting oil and gas from its own reserves. We use the extraction function $\mathcal{C}_i^f$ to have the following isoelastic form

$$\mathcal{C}_i^f(e_i^x, \mathcal{R}_i)\mathbb{P}_i = \frac{\bar{\nu}_i}{1+\nu}\Big(\frac{e_i^x}{\mathcal{R}_i}\Big)^{1+\nu}\mathcal{R}_i\mathbb{P}_i \ .$$

which is homogeneous of degree one in $(e_i^x, \mathcal{R}_i)$. The inputs are paid in the price of the consumption bundle $\mathbb{P}_i$ since the input $x_i^f = \mathcal{C}_i^f(e_i^x, \mathcal{R}_i)$ takes the same CES form as the consumption demand c_i.[27] This implies the profit function

$$\mathcal{P}_i\pi_i^f = q^f e_i^x - \mathcal{C}_i^f(e_i^x, \mathcal{R}_i) = \frac{\nu\bar{\nu}_i}{1+\nu}\big(\frac{e_i^x}{\mathcal{R}_i}\big)^{1+\nu}\mathcal{R}_i\mathbb{P}_i \ .$$

I calibrate the three parameters $\mathcal{R}_i$, ν_i and $\bar{\nu}_i$ to match two country-level variables e_i^x and π_i^f. The reserve $\mathcal{R}_i$ is taken directly from the data on oil and gas reserves documented by Energy Institute (2024). I calibrate the slope of this cost function $\bar{\nu}_i$ to match exactly the production of oil and gas e_i^x, as informed by that same data source. This is displayed in Figure 4. I then calibrate the curvature of the cost function to match the share $\eta_i^\pi = \frac{\pi_i^f}{y_i p_i + \pi_i^f}$ of fossil energy profit as share of GDP. I choose ν to minimize the distance – mean squared error – between the model share η_i^π and the data, successfully matching the share within $5-10$ percentage points. Differences in oil and gas energy rent across countries are not only determined by differences in cost and technology, but also in differences in trade costs and market power – by the existence of OPEC which control more than 28% of oil supply and around 15% of natural gas supply. This explains why it is difficult to match exactly the value η_i^π. However, to keep the simplicity and tractability of the model, I refrain from adding an additional Armington structure over energy sources, or oligopoly power over oil and gas as discussed in Bornstein et al. (2023) and Hassler et al. (2010).

Second, I match the energy mix of the different countries by relying on the two assumptions made in the model: (i) coal and renewable are only traded at the country level: $\bar{e}_i^r = e_i^r$ and $\bar{e}_i^c = e_i^c$, and (ii) the cost function is linear in goods, i.e. the production is Constant Returns to Scale, implying $q_i^c = \mathcal{C}_i^c\mathbb{P}_i$ and $q_i^r = \mathcal{C}_i^r\mathbb{P}_i$. This allows me to match the energy mix of each country by calibrating the energy costs parameters $\mathcal{C}_i^c$ and $\mathcal{C}_i^r$ for each country to match the data $\frac{e_i^c}{e_i^f + e_i^c + e_i^r}$ and

[27] I express the oil-gas extraction with a cost function $x_i^f = \mathcal{C}_i^f$. We can also express analogously with the following production function:

$$e_i^x = g(x_i^f) = \big(\frac{1+\nu_i}{\bar{\nu}_i}\big)^{\frac{1}{1+\nu}}\mathcal{R}_i^{\frac{\nu_i}{1+\nu_i}}(x_i^f)^{\frac{1}{1+\nu_i}}$$

where the inputs x_i^f are paid in the final good bundle. This production has constant returns to scale in $(x_i^f, \mathcal{R}_i)$.

$\frac{e_i^r}{e_i^f+e_i^c+e_i^r}$. Using the CES framework above, I match exactly the energy shares, successfully identifying countries that are more reliant on coal, vs. oil and gas vs. non-carbon/renewable: for example China and India are very coal dependent and Russia, Middle-East and United-States/Canada are the biggest consumer of oil and gas.

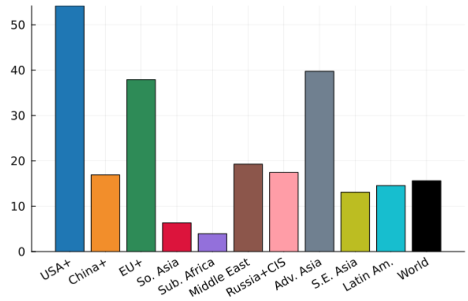

Figure 3: GDP per capita
Thsds 2011-USD PPP, avg. 2018-2023

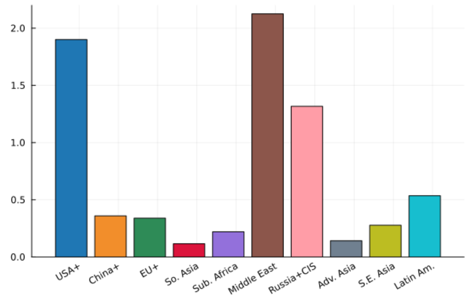

Figure 4: Oil and gas production
GTOE (gigatons oil equiv.), avg. 2018-2023

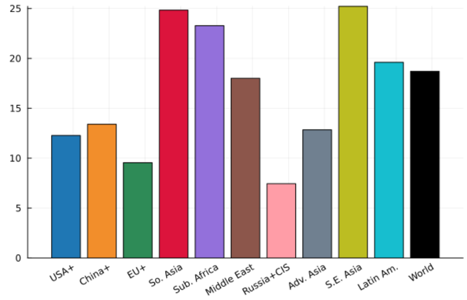

Figure 5: Temperatures
Avg., population-weighted, 2015

4.5 Climate system

Finally, I calibrate the climate model described in Section 2.4 to match important features of the relationship between carbon emissions, temperatures and climate damages.

First, I calibrate two parameters of the global climate system: the climate sensitivity χ, i.e. the reaction of global temperature $\mathcal{T}_t$ to a change in atmospheric concentration of CO_2 $\mathcal{S}_t$, and, the carbon decay rate δ_s representing the exit of carbon of the atmosphere into carbon sinks – oceans, biosphere – and out of the higher atmosphere. To this end, as is standard in the Integrated Assessment models literature, we match the pulse experiment dynamics of larger IAMs – CMIP5 in this case: for a "pulse" of $100GT$ of carbon released – corresponding to 10 years of emissions – the global temperature reaches its peak between $0.20°C$ and $0.25°C$ after 10 years and then decreases slightly to stabilize around $0.17°C$ after 200 years. I follow Dietz et al. (2021), and calibrate $\chi = 0.23$ and $\delta = 0.0004$ to match these two moments, as seen in Figure 23 displayed in appendix.

Moreover, since this climate system is inherently unstable for a given trend of emissions – given once and for all by our static economic model $\mathcal{E}_t = \sum_i \mathcal{P}_i e^{(n_i+\bar{g}_i)t}(e_i^f + e_i^c)$ with n_i the population growth and $\bar{g}_i$ the long-term GDP/TFP growth – I follow Krusell and Smith (2022) by assuming that part of emissions $\mathcal{E}_t$ are captured and stored. I assume the exponential form:

$$\mathcal{T}_t = \chi \mathcal{S}_t \qquad \dot{\mathcal{S}}_t = \zeta_t \mathcal{E}_t - \delta_s \mathcal{S}_t \qquad \zeta_t = e^{-\zeta t}$$

and match the moment suggested by Krusell and Smith (2022): 50% are captured by 2125, and 100% is by 2300 – which is > 99.99% in our model. This implies that in the Business-as-Usual scenario, global temperatures reach $\sim 5°$ by 2100 and are stabilized around $9°$ by 2400[28]. More optimistic scenarios for carbon capture and storage could be imagined, without affecting the main results, since most of the damages are discounted heavily after 2100.

Second, we calibrate the initial temperature T_{it_0} using data from Burke et al. (2015), and we display those difference across regions in Figure 5. Furthermore, we consider the linear pattern scaling $\dot{T}_{it} = \Delta_i \mathcal{T}_t$. I identify the scaling parameter in reduced-form by estimating this linear regression over the period $t = 1950-2015$ for each country and then aggregating by region i[29]. This procedure does not require extensive and granular data such at geographical characteristic, albedo, etc.

Third, to calibrate the damage function, I use the following quadratic function common in many Integrated Assessment Models:

$$\hat{\mathcal{D}}(T_{it}-T_i^\star) = \exp\big(-\gamma^+ \mathbb{1}_{\{T_{it}>T_i^\star\}}(T_{it}-T_i^\star)^2 - \gamma^- \mathbb{1}_{\{T_{it}<T_i^\star\}}(T_{it}-T_i^\star)^2\big)$$

with the damage parameter $\gamma^+ = 0.00340$. This value is intermediary between the value $\gamma^+ = 0.00311$ in Krusell and Smith (2022), calibrated to match Nordhaus' DICE calibration of 6.6% of loss of global GDP when temperature anomaly $\mathcal{T}_t = 5$, and the updated calibration in Barrage and Nordhaus (2024) which calibrate it at $\gamma^+ = 0.003467$. For small values, I consider $\gamma^- = 0.3\gamma^+$, following the quantification done in Rudik et al. (2021) that find that the negative productivity impact of cold temperatures is much weaker than for hot temperature.

Finally, to calibrate $T_i^\star$, I use also an intermediary assumption between the following two cases: (i) the representative agent economy, like Barrage and Nordhaus (2024), would assume $T_i^\star = T_{it_0}$, which implies that $T_{it}-T_i^\star = \Delta_i(\mathcal{T}_t - \mathcal{T}_{t0})$: differences in damages only comes from increases in aggregate temperature. The analysis by Bilal and Känzig (2024) shows that climate damages on GDP comes for a large part from the increase in global temperature, causing extreme events. In contrast, (ii) a different view in heterogeneous countries economies would set $T_i^\star = T^\star$ the same for all regions, at an "ideal" temperature, as in Krusell and Smith (2022) and Kotlikoff, Kubler, Polbin, Sachs and Scheidegger (2021). In this case, differences in climate damages comes essentially from differences in initial temperatures. I take the intermediary step and assume:

$$T_i^\star = \alpha^T T^\star + (1-\alpha^T) T_{it_0}$$

where $\alpha^T = 0.5$ and $T^\star = 14.5$ is the average spring temperature of developed economies – and around the yearly average of places like California or Spain.

[28] Such high temperatures between 2100 and 2400 comes from our static model assumption that the model and emission decisions are made once and for all. In a dynamic model, the damages over time decreases TFP and economic activity leading to an endogenous reduction in the path of emissions and temperature. In Bourany (2024), I simulate the dynamic model over time which align with standard path of future temperatures from IAMs.

[29] To control for the fact that country j has an influence on world temperature $\mathcal{T}_t = \sum_i g_i T_{it}$, I estimate the linear equation with $\mathcal{T}_{t,\neq j} = \sum_{i\neq j} g_i T_{it}$ for each j, i.e. $T_{jt} = \Delta_j \mathcal{T}_{t,\neq j}$

4.6 Heterogeneity

In this section, I summarize the different dimensions of heterogeneity included in the model, and aggregate the parameters of the calibration in appendix Table 2.

Table 1: Heterogeneity across countries

Dimension of heterogeneity	Model parameter	Matched variable from the data	Source of the data
Population	Country size $\mathcal{P}_i$	Population	UN Population Prospect
TFP/technology/institutions	Firm productivity z_i	GDP per capita (2011-PPP)	World Bank/Maddison project
Productivity in energy	Energy-augmenting productivity z_i^e	Energy cost share	SRE Energy Institute (2024)
Cost of coal energy	Cost of coal production $\mathcal{C}_i^c$	Energy mix/coal share e_i^c/e_i	SRE Energy Institute (2024)
Cost of non-carbon energy	Cost of non-carbon production $\mathcal{C}_i^r$	Energy mix/coal share e_i^r/e_i	SRE Energy Institute (2024)
Local temperature	Initial temperature T_{it_0}	Pop-weighted yearly temperature	Burke et al. (2015)
Pattern scaling	Pattern scaling Δ_i	Sensitivity of T_{it} to world $\mathcal{T}_t$	Burke et al. (2015)
Oil-gas reserves	Reserves $\mathcal{R}_i$	Proved Oil-gas reserves	SRE Energy Institute (2024)
Cost of oil-gas extraction	Slope of extraction cost $\bar{\nu}_i$	Oil-gas extracted/produced e_i^x	SRE Energy Institute (2024)
Cost of oil-gas extraction	Curvature of extraction cost ν_i	Profit π_i^f / energy rent	World Bank / WDI
Trade costs	Distance iceberg costs τ_{ij}	Geographical distance $\tau_{ij} = d_{ij}^\beta$	CEPII Conte et al. (2022)
Armington preferences	CES preferences a_{ij}	Trade flows	CEPII Conte et al. (2022)

5 Optimal policy benchmarks without participation constraints

I provide two benchmarks for optimal policy when participation is exogenous. First, I consider a global social planner policy that maximizes aggregate welfare, representing the cooperative allocation. Second, in the non-cooperative Nash-equilibrium, each country implements its unilaterally optimal policy.

5.1 Global Climate Policy with cooperation

The cooperative policy depends on the availability of redistribution instruments. In the First-Best, with unlimited instruments, and in particular lump-sum transfers, the optimal tax is the Social Cost of Carbon (SCC), a measure of the marginal cost of climate change. Without transfers, the optimal tax needs to account for inequality across countries and trade leakage effects. Accounting for inequality and lack of redistribution, the optimal tax is lower than the SCC. The main lessons from this analysis are also described in detail in Bourany (2024) where I develop this argument in a large class of climate-economy, Integrated Assessment Models.

5.1.1 First Best allocation with unlimited instruments

With unlimited instruments, the social planner uses lump-sum transfers to redistribute across countries and offset the negative effects of climate change and carbon taxation. In this context, the optimal tax is the standard Pigouvian tax, which is the Social Cost of Carbon.

Consider a planner that maximizes global welfare by choosing the allocation $\boldsymbol{x} = \{c_{ij}, x_{ij}^{\ell}, e_i^{\ell}\}$

$$\mathcal{W} = \max_{\{c_i, e_i, e_i^x, \dots\}_i} \sum_{i \in \mathbb{I}} \mathcal{P}_i \omega_i \; u(c_j) = \textstyle\sum_{\mathbb{I}} \mathcal{P}_i \omega_i \mathcal{U}_i \tag{20}$$

subject to the market clearing of energy and goods.

The first lesson from this exercise is that the planner would like to equalize marginal utilities through the following condition:

$$\frac{\omega_i u(c_i) \overline{\mathcal{D}}_i^u(\mathcal{E})}{\mathbb{P}_i} = \frac{\omega_j u(c_i) \overline{\mathcal{D}}_j^u(\mathcal{E})}{\mathbb{P}_j} = \overline{\lambda} \qquad\qquad \forall \; i, j \in \mathbb{I}$$

This implies arbitrary large redistribution, using lump-sum transfers, such that:

$$\begin{aligned} c_i &= u'^{-1}\big(\overline{\lambda}\,\mathbb{P}_i / \overline{\mathcal{D}}_i^u(\mathcal{E})\big), \forall i \in \mathbb{I} \\ &= w_i \ell_i + \pi_i^f + \mathrm{t}_i^{ls} \end{aligned}$$

In this case, the transfers t_i^{ls} are designed such that the consumptions are equalized. This implies redistribution, as $\mathrm{t}_i^{ls} < 0$ for some countries and $\mathrm{t}_i^{ls} > 0$ for some other countries.

Second, the Social Cost of Carbon defined as the ratio of marginal value of emissions over marginal utility of consumption can then be simply reformulated with the following multipliers:

$$SCC = -\frac{\frac{\partial \mathcal{W}}{\partial \mathcal{E}}}{\frac{\partial \mathcal{W}}{c_i}} = \frac{\phi^{\mathcal{E}}}{\overline{\lambda}}$$

where I define the multiplier $\phi^{\mathcal{E}}$ as the welfare cost of one additional ton of carbon, which correspond to the constraint $\mathcal{E} = \sum_i \xi^f e_i^f + \xi^c e_i^c$, and $\overline{\lambda}$ the average marginal utility of consumption – or marginal value of wealth.

In that context, the optimal tax is simply the Social Cost of Carbon (SCC)

$$\mathrm{t}^{\varepsilon} = \frac{\phi^{\mathcal{E}}}{\overline{\lambda}} = -\sum_{\mathbb{I}} \mathcal{P}_i \omega_i \Big[\frac{u(c_i)}{u'(c_i)} \mathbb{P}_i \overline{\mathcal{D}}^{u'}(\mathcal{E}) + \mathcal{D}_i^{y\prime}(\mathcal{E}) z_i F(\ell_i, e_i) \mathrm{p}_i\Big] > 0$$

The optimal carbon tax is the Pigouvian level that summarizes the marginal cost of climate change for all countries i. This recovers the result in Golosov et al. (2014) that prevails in representative agents economies: absent redistributive motive, the optimal tax is the Pigouvian level.

In the next section, we see how this result changes when transfers and other instruments, like tariffs or subsidies, are constrained, preventing the planner to perform this redistribution.

5.1.2 Second-Best Ramsey problem without transfers

When the social planner does not have access to transfers, the carbon tax needs to account for redistributive effects. The optimal tax needs to be corrected for (i) the heterogeneous effects of climate change, (ii) the redistributive effects on energy markets, (iii) the distortion of demand,

(iv) the leakage effects of trade.

Consider the Ramsey planner maximizing welfare, as in eq. (20). They choose $\boldsymbol{x} = \{c_{ij}, x^{\ell}_{ij}, e^{\ell}_i\}$, i.e. the traded good for consumption c_{ij}, for energy inputs for the production of in fossil x^f_{ij}, coal x^c_{ij}, non-carbon x^r_{ij} or capital x^k_{ij}, and the energy demand, in fossil e^f_i, coal e^c_i and non-carbon e^r_i, as well as the carbon tax t^{ε} and the prices $\boldsymbol{p} = \{\mathrm{p}_i, q^f, q^c_i, q^r_i\}_i$. However, the allocation and prices are constrained to be a competitive equilibrium: in that case, the planner is restricted to choose controls that respect the individual optimality conditions.

In the Primal approach, these optimality conditions are internalized by the planner through a large array of multipliers for all these constraints which summarize the distortionary effects – through the multipliers on inputs and energy choice of households and firms – the general equilibrium effects – through the multipliers on market clearing and – the redistributive effects – through the multipliers on the household constraints for example. The optimality condition of the planners are numerous and technical and we restrict their complete exposition to Appendix C.2.

The optimal carbon tax t^{ε} corrects climate externality, but also need to account for: (i) redistribution motives, through the multipliers λ_i, (ii) G.E. effects on energy markets, both of the supply – which affect energy rent – and the demand through the distortion of the optimality condition of energy choice, with multipliers υ^f_i for fossils for example and (iii) G.E. effects on good markets, or trade leakage, through multipliers of the market clearing μ_i for good from country i.

The optimal tax formula for carbon – in the case of fossil for example – can be summarized as:

$$\xi^f \mathrm{t}^{\varepsilon} = \xi^f \underbrace{\textstyle\sum_i \widehat{\lambda}_i\, LCC_i}_{=SCC} + \textstyle\sum_i \widehat{\lambda}_i \,\text{Supply Redistrib}^{\circ}_i + \sum_i \widehat{\lambda}_i \,\text{Demand Distort}^{\circ}_i - \sum_i \text{Trade Redistrib}^{\circ}_i$$

$$\text{with} \qquad \widehat{\lambda}_i = \frac{\omega_i \widehat{\mathcal{P}}_i u'(c_i)}{\sum_i \omega_i \widehat{\mathcal{P}}_i u'(c_i)}$$

which can be unpacked as :

$$\xi^f \mathrm{t}^{\varepsilon} = \xi^f \mathcal{P} \sum_i \widehat{\lambda}_i LCC_i + q^f \mathcal{P} \frac{\bar{\nu}}{E^f} \sum_i \widehat{\lambda}_i \Big(e^f_i - \frac{e^x_i}{\mathcal{P}_i}\Big) - (q^f + \xi^f \mathrm{t}^{\varepsilon}) \sum_i \widehat{\mu}_i - \sum_i \widehat{\bar{\upsilon}}^f_i$$

$$\xi^f \mathrm{t}^{\varepsilon} = \xi^f \underbrace{\mathcal{P}\mathbb{E}_i[LCC_i] + \mathcal{P}\mathbb{C}\mathrm{ov}_i(\widehat{\lambda}_i, LCC_i)}_{=\text{Social Cost of Carbon}} + \underbrace{q^f \mathcal{P} \frac{\bar{\nu}}{E^f} \mathbb{C}\mathrm{ov}_i\Big(\widehat{\lambda}_i, e^f_i - \frac{e^x_i}{\mathcal{P}_i}\Big)}_{E^f \text{ supply redistribut}^{\circ}} - \underbrace{\mathbb{E}_i[\widehat{\bar{\upsilon}}^f_i]}_{e^{\ell'}_i \text{ demand distort}^{\circ}} - (q^f + \xi^f \mathrm{t}^{\varepsilon}) \underbrace{\mathbb{E}_i[\widehat{\mu}_i]}_{y_i \text{Trade redistribut}^{\circ}}$$

with the aggregate inverse supply elasticity for fossil $\bar{\nu} = \big(\sum_i \lambda^x_i \nu_i^{-1}\big)^{-1}$, and the social welfare weights", which are the rescaled multipliers for the budget constraint: $\widehat{\lambda}_i = \frac{\omega_i \widehat{\mathcal{P}}_i \lambda_i}{\bar{\lambda}} = \frac{\omega_i \widehat{\mathcal{P}}_i \lambda_i}{\sum_i \omega_i \widehat{\mathcal{P}}_i \lambda_i}$, the multiplier the FOC demand $\widehat{\upsilon}_i = \frac{\omega_i \widehat{\mathcal{P}}_i \upsilon_i}{\bar{\lambda}}$, for the multiplier for market clearing for good: $\widehat{\mu}_i = \frac{\omega_i \widehat{\mathcal{P}}_i \mu_i}{\bar{\lambda}}$. In Appendix C.2, I further develop the distortion terms $\widehat{\upsilon}_i$ to understand how this distortion can be expressed as function of the elasticities of demand σ^y and σ^e for fossil fuels.

This implies that the carbon tax can be different from the Social Cost of Carbon when the planner has redistributive motives. This is described in more details in the companion paper

Bourany (2024). This matters quantitatively for our model as I show in the following figure.

In Figure 6, I display the Social Cost of Carbon and optimal carbon tax in the different equilibria we studied. In the competitive – Business-as-Usual – equilibrium, we see that the endogenous cost of climate change is very large due to climate inaction, i.e. above \$230 per ton of CO_2. However, in the First-Best, thanks to large redistribution, the planner lower the marginal value of wealth $\bar{\lambda}$ for the "average household". This increases the Social Cost of Carbon $SCC = \phi^{\varepsilon}/\bar{\lambda}$ and allows to set a higher carbon tax $\mathrm{t}^{\varepsilon} = SCC$ at $\$220/tCO_2$.

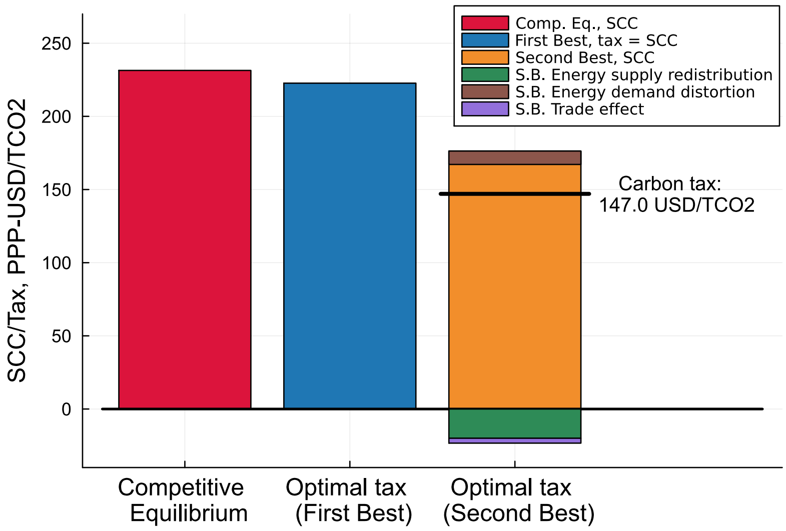

Figure 6: Social Cost of Carbon and optimal carbon taxation

However, in the Second-Best allocation, the social planner cannot redistribute easily and therefore accounts for the redistributive effects in the choice of the carbon tax. The average household is now "poorer", increasing the aggregate weight $\bar{\lambda}$, lowering the Social Cost of Carbon from \$220 to approximately \$170 per ton of CO_2. Moreover, the redistributive effects through the energy supply dries part of the energy rent of large oil and gas producers like the US, Middle East and Russia. This is accounted for in the optimal policy which is now set at \$147 per ton of CO_2.

Winners and losers from cooperative carbon taxation

This optimal second-best Carbon Tax, despite accounting for redistributive effects, still have heterogeneous impact across countries. In Figure 7 I display the welfare change from this policy $\mathcal{U}_i(\mathbb{I}, \mathrm{t}^{\varepsilon})/\mathcal{U}_i(\mathbb{I}, 0)$ in consumption equivalent, in comparison to the competitive equilibrium.

We first observe that most regions gain large benefit from cooperation, with an aggregate effect of 5% for global welfare. However, this hides a large heterogeneity across countries.

The biggest winners are without contest the countries that are the most affected by climate change: South-East Asia, Africa and South Asia, which gains between 6% and 13% of consumption change. However, countries who consume a large share of coal like China and India are not gaining as much because of the large redistributive effect imposed by carbon taxation. Finally, large fossil fuels – oil and gas – exporters like Russia and Middle-East are losers because they see a drying up of their energy rents. In addition, Russia is also a cold countries that do not gain anything for slowing down climate change.

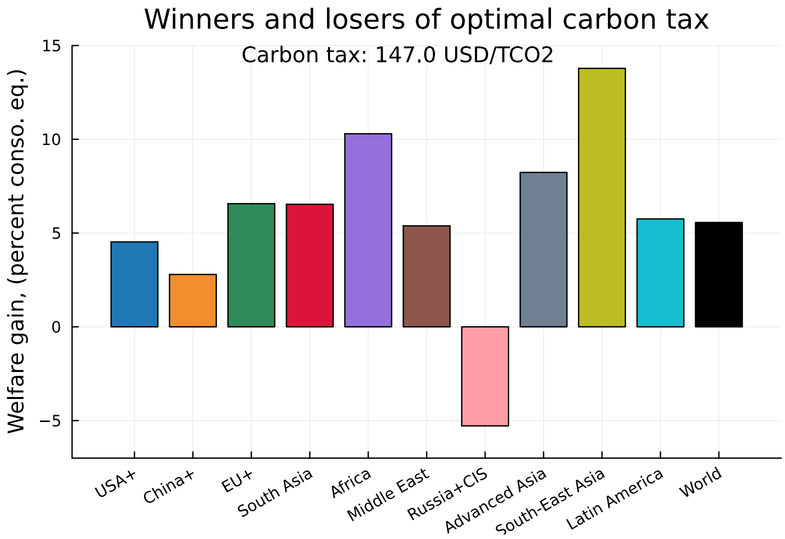

Figure 7: Welfare gains across countries

5.2 Unilateral policies: Nash-equilibrium

In this section, I consider the case where countries act non-cooperatively. In such cases, the optimal carbon policy does not account for the externality imposed on other countries. Moreover, each country might use energy policy and trade tariffs strategically for terms-of-trade manipulation. As a result, for some countries, the carbon tax may become a subsidy. This exercise extends the approach of Ossa (2014), Farrokhi and Lashkaripour (2024) or Kortum and Weisbach (2021) who all solve for optimal unilateral tariffs. In the setting I consider, redistributive effects, general-equilibrium on energy markets and terms-of-trade manipulation yield results that contrast with the above references. I describe the complete setting in Appendix C.3

In this exercise, I consider a Local Social Planner who chooses a local carbon tax $\mathrm{t}_i^{\varepsilon}$ and trade tariffs t_{ij}^{b} to maximize household welfare:

$$\mathcal{V}_i = \max_{\mathrm{t}^{\varepsilon}, \mathrm{t}^{b}} \mathcal{P}_i \mathcal{U}_i$$

subject to the world equilibrium constraints and taking as given the policies of the other countries – as a Nash Equilibrium.

In the spirit of the previous section, using the Primal approach, we see that the planner i chose the policy instruments $\mathrm{t}_i^{\varepsilon}, \mathrm{t}_{ij}^{b}$ internalizing a wide array of market clearing and optimality conditions. More specifically, the planner accounts for the market clearing for good j as household and energy firms in country i imports this variety, e.g. as consumption c_{ij}. I denote this shadow value for good j eq. (11) as $\mu_j^{(i)}$. If $\mu_j^{(i)} > 0$, planner i would like to expand j's supply – relaxing its market clearing and lowering its price. In contrast, we usually obtain $\mu_i^{(i)} < 0$: the planner would like to reduces its own i-supply, to increase its price and manipulate terms of trade $T.o.T_i = \mathrm{p}_i/\mathrm{p}_j$, $\forall j \neq i$.

Similarly, the planner account for the fossil energy market clearing eq. (8), which is global, with multiplier $\mu^{f(i)}$. In addition, the planner internalizes the country level constraints, such as budget with multiplier $\lambda_i = u'(c_i)/\mathbb{P}_i$, or the optimality condition for energy demand, energy supply and good imports. For ease of exposition, I display the result where the only energy input is oil-gas $e_i = e_i^f$. The general case is detailed in Appendix C.3.

One key result in this setting is the absence of local distortions. Several distortionary effects that appeared in the previous Second-Best framework disappear. It is for example the case for the optimality conditions for good imports, c_{ij}, or energy demand e_i which are not distorted by planner's choice. As a result, the planner decision would align with the household and firms of country i, because of the existence of tariff and country-specific tax.

Optimal tariffs

The optimal tariffs are chosen to manipulate terms-of-trade and take the following form:

$$\mathrm{t}_{ij}^b = \frac{\mu_j^{(i)}}{\lambda_i}$$

where $\mu_j^{(i)}$ are the multiplier for good j as accounted by planner i, and λ_i the marginal utility of consumption. Tariffs increase when the planner seeks to decrease demand for good j. This is especially the case when the household in country i is richer: the redistributive effect of tariffs are amplified when $c_i \to \infty$, $\lambda_i \to 0$ and thus $\mathrm{t}_{ij}^b \to \infty$. Unfortunately, with the primal approach, these multipliers $\mu_j^{(i)}$ typically do not have closed forms expressions in the Armington trade model.

Local Social Cost of Carbon

For designing climate policy, we need to evaluate the impact of climate on country i welfare. For that, we can summarize this welfare cost as the Local Cost of Carbon:

$$LCC_i = -\frac{\frac{\partial \mathcal{V}_i}{\partial \mathcal{E}}}{\frac{\partial \mathcal{V}_i}{\partial c_i}} = \frac{\phi_i^{\varepsilon}}{\lambda_i}$$

where ϕ_i^{ε} is the multiplier for carbon dynamics – or shadow cost of increasing emissions by 1 ton of CO_2. In the unilaterally optimal allocation, this cost of climate change summarizes:

$$LCC_i = \mathcal{P}_i\Big[\frac{u(c_i)}{u'(c_i)}\mathbb{P}_i\overline{\mathcal{D}}^{u'}(\mathcal{E}) + \mathcal{D}_i^{y'}(\mathcal{E})z_iF(\ell_i, e_i)\mathrm{p}_i + \sum_j \frac{\mu_j^{(i)}}{\lambda_i}\mathcal{D}_j^{y'}(\mathcal{E})z_jF(\ell_j, e_j)\mathrm{p}_j\Big]$$

we note that in addition to considering the direct damage on household utility and country i output, the planner in i also internalize the damage that climate has on the production of trade partner y_j, through multiplier $\mu_j^{(i)}$ again, and this impact is even larger for rich, low λ_i-countries.

This introduces a novel channel through which local costs of climate change can become correlated due to international trade linkages. Another argument made in Dingel et al. (2019) relates the spatial correlation of the costs of climate change, the gains from trade, and welfare change. Here, I also show how this policy relevant local costs of carbon can be correlated, even in non-cooperative Nash-equilibria, potentially providing a motive for coordinated climate actions.

Unilaterally optimal carbon tax – or subsidy

Finally, I derive the unilaterally optimal carbon tax – here exposed in the case where energy is entirely consumed in fossil-fuel (oil and gas).

$$\xi^f \mathrm{t}_i^\varepsilon = \underbrace{\xi^f LCC_i}_{\substack{\text{Pigouvian}\\ \text{motive}}} \underbrace{- q^f \frac{\mu_i^{(i)}}{\lambda_i}}_{\substack{\text{terms-of-trade}\\ \text{manipulation}}} \underbrace{+ q^f \nu_i \frac{\mathcal{P}_i e_i^f - e_i^x}{e_i^x}}_{\substack{\text{energy supply}\\ \text{redistribution}}}$$

In the Nash-equilibrium of this model, the carbon tax is the sum of three terms: First, the planner i internalizes the effect its emissions have on the welfare of its own country. However, it does not internalize the rest of the world: we can recover a symptom of free-riding as country i does not account for the impact of its emissions on other countries welfare.

Second, it features a terms-of-trade manipulation term: if country i would like to reduce its demand, acting like a monopolist on its variety i, for example in the case where $\mu_i^{(i)} < 0$, then it would tax carbon to lower production. In the opposite logic, if country i looks after expanding its supply and lowering its prices, when $\mu_i^{(i)} > 0$, then the planner would lower taxation of energy, providing a subsidy motive.

Third, taxation of fossil fuel has an impact on international energy market, creating a redistribution term. This is positive for net-importers and negative for exporters. Energy exporters would like to subsidize energy to increase demand, in an attempt to raise the equilibrium price q^f and benefit from better terms-of-trade. This is weighted by the country i production e_i^x and inverse elasticity ν_i: a more inelastic supply, with large ν_i, would amplify this effect. Note that this is the same logic as the global social planner of Section 5.1.2, at a local level.

To sum up, if the terms-of-trade manipulation motive $\mu_i^{(i)} > 0$ is large enough or if the energy-supply redistribution term is negative, for example for oil-gas exporters, and if the local cost of carbon LCC_i is small, the optimal carbon tax can become a *subsidy*

6 Optimal Climate agreement

I now turn to the main result of this paper. The optimal design of climate agreement is a climate club that consists of all the countries at the exception of Russia and CIS countries. Members of the club impose a \$98 carbon tax per ton of CO_2 and a 50% tariff on goods from non-members. The intuition behind this result can be summarized by the tradeoff between the distortionary effect of carbon taxes and the cost of tariffs, which relate to the gains from trade. For some countries, like fossil-fuel producers or developing economies, the first outweighs the second, implying that they would not participate in a climate agreement unless the tax is decreased from \$147 to \$98. This encourages the participation of Middle-East and South Asia, but the optimal agreement does not include the entire world. Indeed, lowering the tax so low to incentivize Russia to participate would compromise climate action and reduce world's welfare.

I first provide details on the trade-off behind country participation, both numerically and theoretically, using first-order decomposition of the model. I then present the main result of the optimal climate club.

6.1 Trade-off: distortionary effects of carbon taxation vs. gains from trade

The following two effects influence the design of the agreement. First, the distortionary effect of the carbon tax differs across countries, and some countries – poor, closed to trade or cold countries and fossil-fuel producers have very large gains from free-riding. Second, the cost of tariffs from trade partners, related to the gains from trade, also differ across countries.

In the Figure 8, I present an experiment where all countries $j \in \mathbb{I}$ implement the optimal level of fossil-fuel taxation $\mathrm{t}^{\varepsilon} = \$147/tCO_2$, except for country i, which deviates from that policy, setting the tax to zero. In this experiment, other countries j do not impose retaliatory tariffs on country i, and continue to implement the optimal policy. For each country i, I plot, in consumption-equivalent units, the welfare gain of such "deviation" compared to the case where the country i stays in this "agreement". This represents the gains from free-riding, while the Rest of the World, or equivalently, is the cost of the distortionary taxation of carbon and fossil fuels.

We first see that such gains from "deviating" range from 1.5% – for Europe, which uses more renewable energy and less coal – to close to 4% for Russia, former soviet countries, and Middle-Eastern countries, whose economies are relying on oil and gas both for good production and energy exports. These distortionary costs are also relatively high for developed economies like South Asia, Sub-Saharan Africa, and Latin America, for the reason that energy, and fossil and coal in particular, are necessary inputs in production and the welfare cost scale with marginal utility of consumption $u'(c_i)$. I provide a welfare decomposition in Section 6.2 to show the sources of these welfare costs and how it differ across regions.

Now, let us compare this cost of carbon taxation to the cost of trade tariffs. In Figure 9, I measure the welfare costs of tariffs in the following experiment: all countries $j \neq i$ impose a very

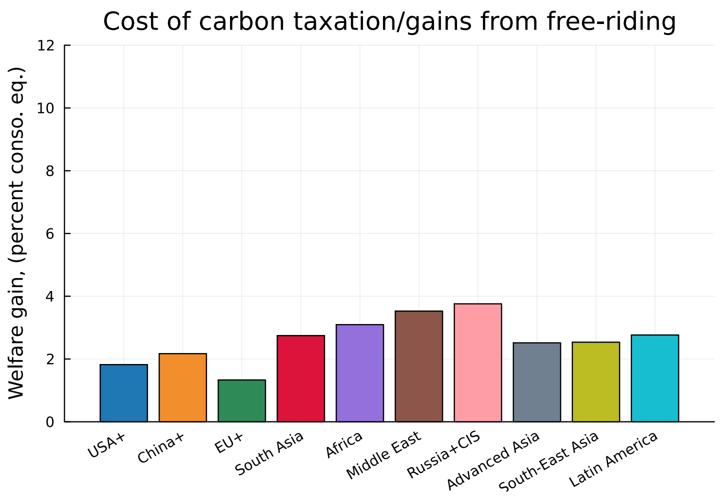

Figure 8: Welfare gain for country i of unilaterally deviating from a world agreement setting a \$147 carbon tax

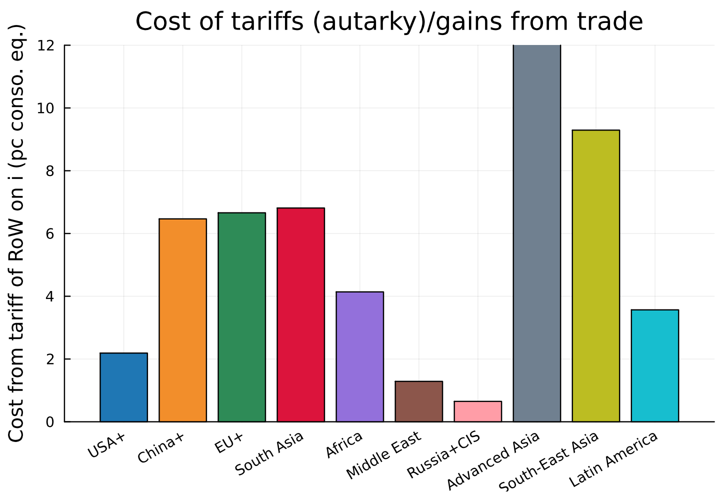

Figure 9: Losses for country i of country $j \neq i$ imposing a 500% tariff on country i

large tariff – 500% – on country i. For each country i, I display the welfare loss, in consumption equivalent percentage changes in the figure. This is a good representation of the upper bound on welfare cost of tariffs – as those welfare costs are virtually identical for higher values of tariffs. This is closely related to the cost of autarky, or gains-from-trade which are bounded and relatively small in standard trade models, c.f. Arkolakis et al. (2012).[30]

Asian countries – advanced economies like Japan and Korea, as well as China and South-East Asian economies – have the most to lose to be in subject to tariffs – respectively around 12%, 8% and 6% welfare equivalent consumption losses. This is in part due to the large trade shares of these countries with each other and with Europe. In comparison, countries like the US, Middle-East and Russia, which are much more closed to international trade, suffer less from tariffs, which change their willingness to join a climate club.

6.2 Welfare decomposition

To understand the mechanisms through which climate change, carbon taxation and tariffs affect welfare, I provide a first-order approximation of welfare to shed light on different mechanisms. This welfare decomposition is described in thorough detail in Appendix D, and it is inspired by Kleinman et al. (2020).

I compute the change in welfare, linearizing the model around the competitive equilibrium where $\mathrm{t}^\varepsilon = \bar{\mathrm{t}}^\varepsilon = 0$ and $\mathrm{t}^b_{ij} = \bar{\mathrm{t}}^b_{ij} = 0$, where policies are identical to the "status-quo". I start from a climate agreement $\mathcal{J}$ of J countries, which are indifferent between being in the club or not, since the policy $(\mathrm{t}^\varepsilon_i, \mathrm{t}^b_{ij}) = (0, 0)$ does not change the equilibrium.

I consider a perturbation where those policy instruments are increased by a small amount,

[30] In this experiment, autarky would be the case when both countries j impose large tariffs on i and i imposes tariffs on countries j as well. I consider the experiment of one-side tariffs, which is closer to the policy implemented in our climate club

denoting $d\ln z_i = \frac{dz_i}{z_i}$:

$$\mathbf{J}d\mathbf{t}^{\varepsilon} = \big\{\mathbb{1}_{\{i\in\mathcal{J}\}}\, d\ln \mathrm{t}^{\varepsilon}_i\big\}_i \qquad\qquad \overline{\mathbf{J}}\odot d\mathbf{t}^b = \big\{\mathbb{1}_{\{i\in\mathcal{J},j\notin\mathcal{J}\}} d\mathrm{t}^b_{ij}\big\}_{ij}$$

with $\mathrm{J} = \mathrm{J}_i = \mathbb{1}_{\{i\in\mathcal{J}\}}$, and $\mathbf{J} \equiv \mathrm{J}_{ki} = \mathbb{1}_{\{i\in\mathcal{J},j\notin\mathcal{J}\}}$.

The welfare decomposition of individual country i, defined as $\mathcal{U}_i = u(\{c_{ij}\}_j)$ the indirect utility is computed as:

$$\frac{d\mathcal{U}_i}{u'(c_i)} = \eta^c_i d\ln \mathrm{p}_i + \Big[-\eta^c_i\bar{\gamma}_i\tfrac{1}{\bar{\nu}} - \eta^c_i s^e_i s^f_i + \eta^\pi_i(1{+}\tfrac{1}{\bar{\nu}})\Big] d\ln q^f - \Big[\eta^c_i s^e_i(s^c_i{+}s^r_i){+}\eta^\pi_i\tfrac{1}{\nu} + 1\Big] d\ln \mathbb{P}_i$$

where $\eta^c_i = \frac{y_i \mathrm{p}_i}{x_i}$, with $x_i = c_i\mathbb{P}_i$ is the ratio of final good output in comparison to consumption – which can also come from energy rent. The counterpart is $\eta^\pi_i = \frac{\pi^f_i}{x_i}$. The energy share $s^e_i = \frac{e_i q^e_i}{y_i \mathrm{p}_i}$ and the share of oil-gas/coal/renewable $s^\ell_i = \frac{e^\ell_i q^\ell_i}{e_i q^e_i}$ governs the impact of energy prices. The aggregate supply elasticity $\bar{\nu} = \big(\sum_i \lambda^x_i \nu_i^{-1}\big)^{-1}$ represents the oil-gas supply curve, and the climate damage $\bar{\gamma}_i = \gamma(T_i{-}T^\star_i)T_i\, s^{E/S}$ is represented in a static fashion – with $\mathcal{E}$ the emission of that period and $s^{E/S} = \mathcal{E}/\mathcal{S}$ with $\mathcal{S}$ the carbon concentration in the atmosphere.

We observe that most of the impacts arise through aggregate quantity of emissions and fossil fuels consumption, which then affect world prices q^f. For conciseness, I express all the General Equilibrium effects on fossil quantities as a function of price q^f: $d\ln q^f \approx \bar{\nu} d\ln E^f + \ldots$

The countries affected the most by a change in equilibrium quantity of fossil fuels consumed E^f, price q^f, and thus by carbon taxation, are the countries with high sensitivity to $d\ln q^f$. A reduction in fossil demand benefits the countries that have large damages from climate changes $\bar{\gamma}_i$, as well as large energy share from fossil s^f_i: this latter effect dampens the cost of taxation: if a larger coalition lower energy demand, it benefits other countries through a reduction in fossil price. This is sometimes called in the literature the "energy price leakage effect". However, this decrease in price hurt the fossil fuel producers as it dries out their energy rent as summarized by $\eta^\pi_i(1{+}\frac{1}{\bar{\nu}})$. Moreover, there are many additional equilibrium effects through trade and good prices p_i and $\mathbb{P}_i$ as we see below.

To see the direct effect of carbon taxation – at the intensive margin – and the extensive margin effect of the size of the club J_i, I simplify the model further to obtain an analytical formula for the fossil price. In the following, I assume that the energy mix is concentrated on oil and gas $s^f_i = 1, s^c_i = s^r_i = 0$. The details of the derivation are provided in Appendix D:

$$d\ln q^f = -\frac{\bar{\nu}}{1{+}\bar{\gamma}{+}\mathbb{C}\mathrm{ov}_i(\widetilde{\lambda}^f_i, \bar{\gamma}_i) + \bar{\nu}\overline{\lambda}^{\sigma,f}} \sum_i \widetilde{\lambda}^f_i \mathrm{J}_i d\mathrm{t}^\varepsilon + \sum_i \beta_i d\ln \mathrm{p}_i$$

with the market share $\lambda^f_i = \frac{\mathcal{P}_i e^f_i}{E^f}$ for fossil, and weighted by elasticity $\widetilde{\lambda}^f_i = \lambda^f_i \frac{\sigma^y}{1-s^e_i}$ and its average $\overline{\lambda}^{\sigma,f} = \sum_i \widetilde{\lambda}^f_i \frac{\sigma^y}{1-s^e_i}$.

The energy curve expressed here q^f is affected by climate change: more emissions imply larger damages $\bar{\gamma} = \sum_i \bar{\gamma}_i$, which in turn reduce energy demand and hence emissions. The price impact of taxation is higher – analogous to the slope of the demand curve – as seen in the denominator of the first term. Moreover, the covariance term indicates that if the large energy producers – with a larger share of the market $\lambda_i^f = \widetilde{\lambda}_i^f = \lambda_i^f \frac{\sigma^y}{1-s_i^e}$ and high elasticity σ – are also the most affected by climate change, this effect is stronger and the demand curve is even steeper and more inelastic.

Moreover, carbon taxation t_j^ε and tariffs t_{ij}^b have large trade and leakage effect, through general equilibrium impact of on y_i and p_i

I can compute the change in price in general equilibrium:

$$d\ln \mathrm{p} = \mathbf{A}^{-1}\big[-(\mathbf{I}-\mathbf{T}\odot v^y)\alpha^{y,qf} + \mathbf{T}(v^{e^x}\odot\tfrac{1}{\nu}+v^{e^f}\tfrac{\sigma^y}{1-s^e}+v^{ne}) - \Big((\mathbf{I}-\mathbf{T}\odot v^y)\alpha^{y,z}-\tfrac{\sigma^y}{1-s^e}\Big)\bar{\gamma}\tfrac{1}{\bar{\nu}}\big]d\ln q^f$$
$$+\big[-(\mathbf{I}-\mathbf{T}\odot v^y)\alpha^{y,qf} + \mathbf{T}(v^{e^f}\odot\tfrac{\sigma^y}{1-s^e})\big]\odot \mathrm{J} d\ln \mathrm{t}^\varepsilon + \theta\big(\mathbf{TS}\odot\mathbf{J}\odot d\ln \mathrm{t}^b - \mathbf{T}(\mathbb{1}+\mathbf{S}')\odot(\mathbf{J}\odot d\ln \mathrm{t}^b)'\big)$$

with parameters: $\mathbf{S}$ for the trade share matrix, $\mathbf{T}$ income flow matrix, θ, Armington CES. Moreover, the general equilibrium (and leakage) effects are summarized in a complicated matrix $\mathbf{A}$ that summarizes the fact that the price p_i also affects energy demand, oil-gas extraction, energy trade balance and output. Further description can be found in the Appendix D.

6.3 Optimal climate agreement

In this section, I describe the design of the optimal climate agreement. The climate club that maximizes the world's welfare is a large coalition with all the country at the exception of Russia. Moreover, the carbon tax for members of the club is lowered below \$100, and tariffs are set at a moderate rate of 50% for non-members. This outcome balances the intensive margin-extensive margin tradeoff of this policy design.

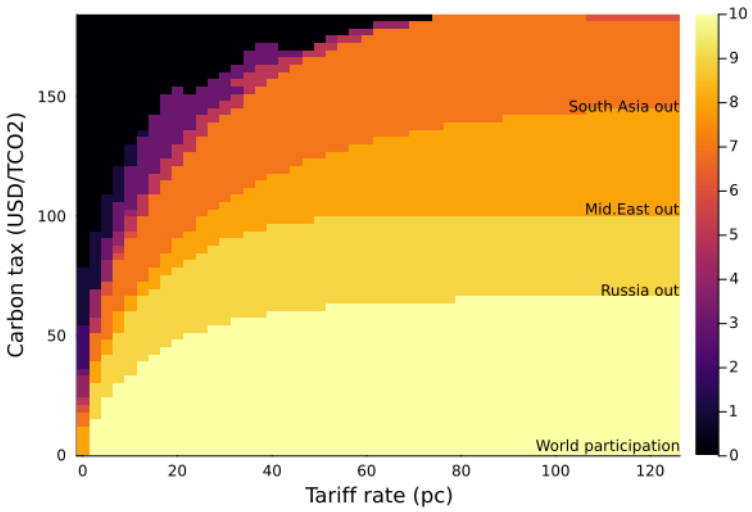

Figure 10: Participation: Intensive and extensive margin trade-off for agreement design: t^ε on y-axis, t^b on x-axis

At the intensive margin, increasing carbon taxation reduces fossil fuel use and emissions for the countries participating in a climate agreement. As a result, aggregate welfare is increased until the optimal carbon tax $t^{\varepsilon} = \$147$ is reached. However, at the extensive margin, a higher tax reduces participation as free-riding incentives increase with the cost of taxation. If the tax becomes too high, individual countries deviate and leave the agreement, which raises world's emissions. In Figure 10, I show this phenomenon, where I plot the maximum participation that can be achieved depending on the choice of the levels of carbon tax for club members on the y-axis and the tariffs that are imposed on non-members on the x-axis.

For tax under \$50, the cost of carbon abatement is low, and it is relatively costless for countries to participate in an agreement. For higher taxes, we observe that the first region to deviate is Russia and former Soviet Republics. Then leaves Middle-Eastern countries and South Asia. For an even larger tax, closer to \$200, Sub-Saharan Africa would also exit the climate agreement. These decisions originate from the tradeoff explained in Section 6.1. Indeed, those countries have a high cost of distortionary carbon taxation, either because they are producers of oil and gas, like Russia and Gulf countries, or because they consume a significant part of their energy mix in coal, like India and Africa. This compares to the cost of tariffs, which are relatively small for these four regions, at least for tariffs below 150% in the case of Africa and South Asia.

Another lesson from this analysis is that trade policy is a key strategic instrument to undermine free-riding. Indeed, absent tariff retaliation, with $t^b < 5\%$, the gains from unilateral deviation prevail over the costs of climate actions, and no carbon tax above $t^{\varepsilon} > \$50$ could be implemented for a large enough set of countries. This result is discussed in Nordhaus (2015) and I recover this effect in this quantitative model. However, if moderate tariffs spur participation for low carbon taxes, this incentive effect vanishes quickly as the carbon tax increases. Since the gains from trade are bounded – and small for some countries like Middle-East and Russia – there is a limit to what carbon policy can achieve.

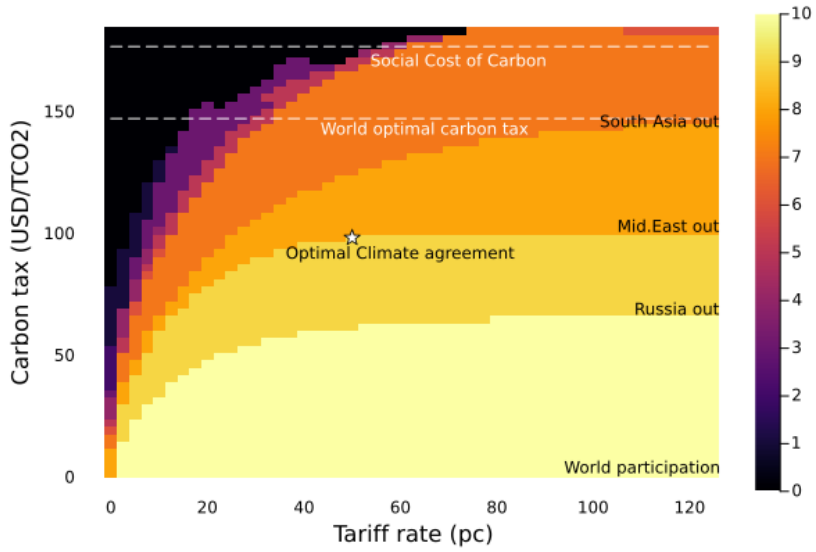

Figure 11: Optimal climate agreement
t^{ε} on y-axis, t^b on x-axis

Now, we turn to the design of optimal climate agreement, and the decision of carbon taxes and tariffs. We note that the optimal Second-Best policy with a large carbon tax $t^{\varepsilon} = \$147$ and complete participation is not achievable in a climate club. As shown in the dotted lines in Figure 11, it corresponds to an area where South-Asia, Middle-East, Russia, and former Soviet countries would all exit the agreement. As a result, the optimal agreement that would maximize welfare is such that the carbon tax is lowered from \$147 to \$98: this incentivizes the participation of South Asia and the Middle-East.

It is optimal to leave Russia outside the agreement. Reducing the carbon tax to accommodate Russia's participation to the agreement necessitates a large fall in climate effort. A decrease of the tax from \$98 to around \$50 increases emissions of the entire world. This compromises the implementation of effective solutions for global warming, and would lower aggregate welfare.

This optimal climate agreement realizes close to 90% of the welfare gains of the optimal-policy without endogenous participation, as seen in Figure 7 of Section 5.1.2. In **??**, I plot global welfare for the different values of the carbon tax and tariffs (t^{ε}, t^{b}), in consumption equivalent, relative to welfare in the competitive equilibrium with $(t^{\varepsilon}, t^{b}) = (0, 0)$. Welfare increases non-monotonically in the carbon tax as carbon emissions and global temperature are reduced. However, when participation declines because countries unilaterally exit the agreement, the deviating countries go back to their status-quo policies, raising their emissions, which decreases discontinuously global welfare. The optimal agreement achieves almost 5% of consumption equivalent welfare gains, close to 90% of the welfare gains attained in the Second-Best where all countries are participating exogenously absent free-riding.

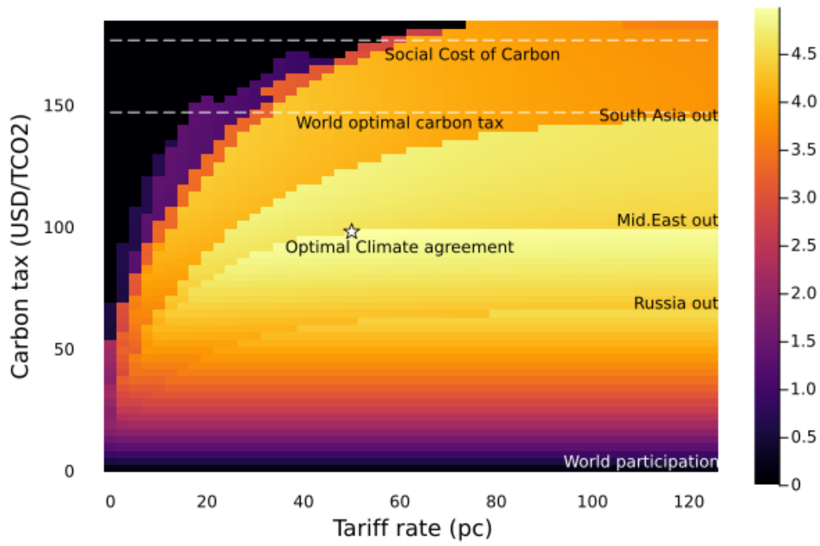

Figure 12: Welfare over different climate agreements
t^{ε} on y-axis, t^{b} on x-axis

Note that tariffs increase have a very moderate impact on welfare. In fact, they have a strong impact on individual countries' utility if they are outside the club. However, since such countries

– e.g. Russia, Middle-East, and South-Asia – do not have much weight in the welfare criterion we used, especially considering the Negishi weights in Figure 2, this has a limited influence on global welfare. It mainly has a strong influence on participation through the impact on the country's outside options and welfare.

Changing the level of the carbon tax is fundamental for participation and the optimal design of the agreement, creating a *Laffer curve* for emissions and welfare. Figure 14 and Figure 13 plot the change in global welfare and carbon emissions varying the carbon tax, and keeping the tariffs fixed at the optimum, $t^b = 50\%$. Raising the carbon tax reduces emissions and improves welfare up to the point where participation declines. It is therefore optimal to "share the burden" of the carbon tax on a larger set of countries. In the optimal agreement, where all the countries in the world except Russia are included, one can reduce emissions from 35 to below 24 $GtCO_2$ per year, a decline of 32% compared to the competitive – Business-as-Usual – equilibrium.

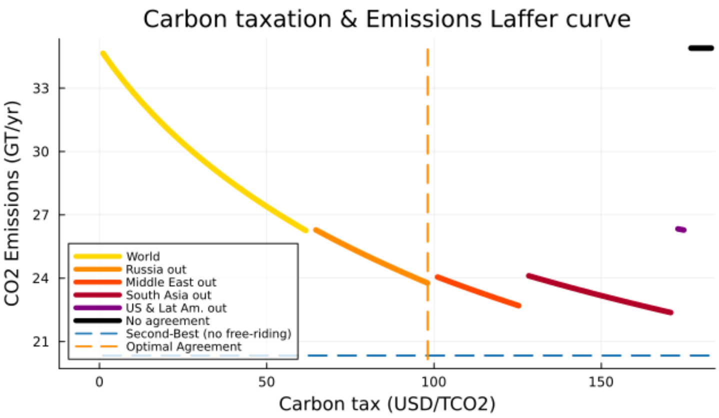

Figure 13: Global Emissions (yearly)
for different carbon taxes a given tariff $t^b = 50\%$

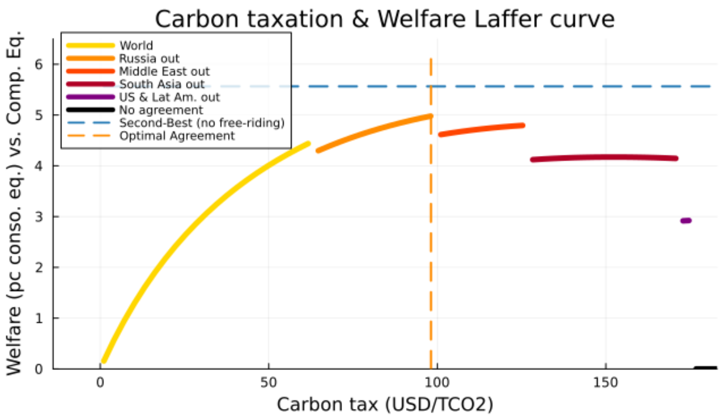

Figure 14: Global welfare (percent consumption eq.),
compared to the Competitive Equilibrium
for a given tariff $t^b = 50\%$

Note that the optimal climate agreement does not result in the maximum emission reduction that can be achieved. If we choose a carbon tax of $\$170/tCO_2$, which corresponds roughly to the Social Cost of Carbon of the Second-Best equilibrium, we can shrink emissions further by more than 35% from 35 to around 22.5 $GtCO_2$. However, this implies that the very affected countries – Russia, Middle-East, South-Asia – all exit the agreements. The remaining countries, which are the developed economies, Europe, the Americas, and East Asia, all have to bear a much higher cost of taxation. This agreement is still stable due to the enforcement power of tariffs, but the negative welfare impact of taxation for those countries is now much larger.

The difference in welfare between those two cases is sizable: the optimal climate agreement achieves a 5% welfare gain while a club with a more restricted set of countries and larger tax reaches only a 4.2% welfare gain. Making a smaller set of countries bear the cost of taxation is detrimental to their welfare: developed countries consume larger quantities of energy, and developing economies have a higher cost of distortion as their production and consumption are scarce – especially if they are affected by climate change. The agreement is stable because the cost of tariffs is larger, enforcing this cooperation. However, it is beneficial to work at the extensive margin to reduce the distortionary carbon tax, foster participation to share the costs of fighting climate change, and lower the tariffs on non-members to promote free trade.

Because of endogenous participation, welfare and emissions are indeed different metrics that provide contrasting insights on what should be the optimal policy. In the next graphs, Figure 15 and Figure 16, I summarize, for the different equilibria we considered above, respectively the global emissions in Gigatons of CO_2 and welfare in consumption equivalent difference compared to the competitive equilibrium. Clearly, the First Best has the lowest emissions – $18.5GtCO_2$, a reduction of 47% relative to the Business-as-Usual scenario – and the maximum welfare – 16% of consumption equivalent change. In this case, the planner has access to unlimited instruments: it uses transfers to redistribute across countries, which offset the negative general equilibrium effects of taxation and allows the increase in carbon taxes and a further reduction in global carbon emissions. In contrast, in the Second-Best, these redistributive instruments are not available, which makes the welfare gains much smaller at 5.6%, although emissions are only slightly higher.

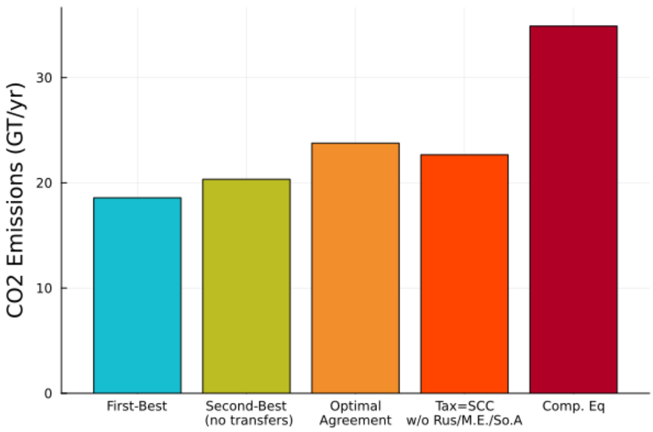

Figure 15: Global Emissions (yearly) comparison across equilibria

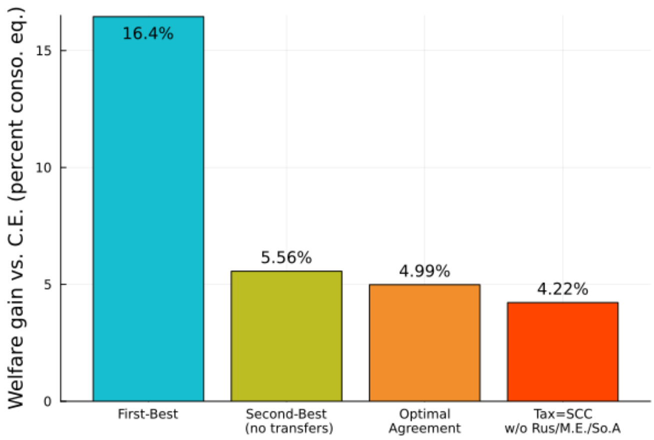

Figure 16: Global welfare (percent consumption eq.), comparison across equilibria

In these two benchmarks, we assume away the free-riding problem, which constrains the achievable policy and carbon reduction. I now compare the two equilibria considered above: the optimal agreement with all the countries except Russia reaches only a reduction of 32% of carbon emissions. This reaches the maximum welfare but not the minimum emissions, while a club with larger taxes loses on welfare by increasing the tax burden because of lower participation in the agreement.

The analysis of the potential welfare gains of the First-Best highlights that transfers can serve as strong instruments to offset the negative effects of the uniform carbon taxation and tariffs, and I investigate if we can provide such welfare improvements with transfers Section 7.1 and with fossil-fuels specific tariffs in Section 7.2.

6.4 Coalition building

The proposed optimal agreement was chosen by the social planner/designer. However, can this agreement be achieved by coalition building? Can a sequence of countries joining the climate agreement in turn reach this agreement? This relates to the question of which country has the most interest in joining such a club.

I investigate if this climate agreement can be constructed, with a sequence of "rounds" of our static equilibrium: At each round (n), each country decide to enter or not depending on the welfare gain:

$$\Delta_i \mathcal{U}_i(\mathbb{J}^{(n)}) = \mathcal{U}_i(\mathbb{J}^{(n)} \cup \{i\}, \mathrm{t}^{\varepsilon}, \mathrm{t}^{b}) - \mathcal{U}_i(\mathbb{J}^{(n)} \setminus \{i\}, \mathrm{t}^{\varepsilon}, \mathrm{t}^{b})$$

For now, the construction is evaluated at the optimal carbon tax $\mathrm{t}^{\varepsilon} = 98\$$, and tariff $\mathrm{t}^{b} = 50\%$ and perform this sequential procedure – which is a direct output in our CDCP algorithm / squeezing procedure in Section 3.2. This experiment is inspired by an analogous exercise in Farrokhi and Lashkaripour (2024).

The result of this exercise is a sequence building up to the optimal climate agreement:

- Round 1: European Union
- Round 2: China, South East Asia (Asean)
- Round 3: North-America, South-Asia, Subsaharian Africa,
  Advanced East Asia, Latin America
- Round 4: Middle-East

$\notin$ Stay out of the agreement: Russia+CIS

The European Union has the best interest in reducing climate change, being positively affected by a decrease in the fossil-fuel price, consuming a small share of coal in their energy mix, and being wealthy enough to suffer less from the energy taxation cost in their production. In the second round, China, an important trading partner of European countries, and Southeast Asia, which has one of the highest gains in fighting global warming, in turn, join the climate agreement. In the third round, most other countries, which have large gains from trade, join the climate club to avoid retaliatory tariffs. Lastly, the Middle East also joins to be able to trade with the rest of the world.

7 Extensions: the impact of additional policy instruments and retaliation

In this section, I propose extensions to our baseline climate agreement by suggesting additional instruments to improve the allocation. By proposing simple policy that could be achievable in practice, I investigate if we can improve on the optimal climate agreement presented above.

7.1 Transfers and loss and damage funds

One of the major policy proposals of the COP28 in Dubai is the idea of *loss and damage funds* to compensate the countries particularly affected by global warming. I propose a simple implementation in this framework. Given that the club is implementing a substantial carbon tax t^{ε}, one practical proposal is to redistribute a share of the revenues of this tax through lump-sum transfer across countries.

In the baseline agreement, tax revenues are redistributed to the household of the country paying the tax: $\mathrm{t}_i^{ls} = \mathrm{t}^{\varepsilon}(\xi^f e_i^f + \xi^c e_i^c)$. Here, the exercise allocates a share α^{ε} to a "loss-and-damage fund" which then redistributes those revenues equally across countries, with a simple rule:

$$\mathrm{t}_i^{ls} = (1-\alpha^{\varepsilon})\mathrm{t}^{\varepsilon}(\xi^f e_i^f + \xi^c e_i^c) + \alpha^{\varepsilon}\frac{1}{\mathcal{P}}\sum_{j\in\mathbb{J}}\mathcal{P}_j \mathrm{t}^{\varepsilon}(\xi^f e_j^f + \xi^c e_j^c)\ , \qquad \forall\, i \in \mathbb{J}$$

In practice, it transfers from large emitters – the developed economies – to low emitters – developing economies that tend to be more vulnerable to climate change.

I then choose the optimal share α^{ε} to maximize global welfare $\mathcal{W}$, which is computed to be $\alpha^{\varepsilon,\star} = 15\%$. The result of that experiment are shown in Figure 17.

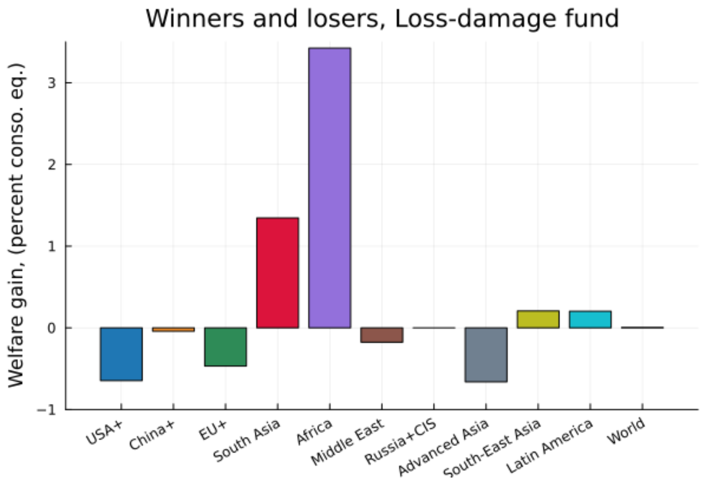

Figure 17: Welfare across countres
Optimal loss and damage fund: $\alpha^{\varepsilon,\star} = 15\%$

The optimal loss-and-damages fund proposed does not provide particularly large welfare gains in aggregate – around 0.006%. This policy only redistributes lump-sum from high to low-energy users. The welfare costs are around 0.6% for Europe, the United States, and advanced countries in Asia and Oceania. Because the welfare function is biased toward these advanced countries, this explains why the optimal effort for transfers is quite low at $\alpha^{\varepsilon,\star} = 15\%$. A larger contribution would be detrimental to their welfare, which is not optimal for the planner.

However, even small loss-and-damage funds redistribution is particularly welfare-improving for South Asia and Africa who gain respectively 1.4 and 3.3% welfare gains – and Southeast Asia and Latin America to a smaller magnitude. Those regions are small contributors to the global climate externality, and such transfers would allow them to lower the cost of climate change through adaptation and dampen the redistributive cost of carbon taxation.

7.2 Fossil-fuels specific tariffs

In this section, I relax the assumption on free-trade on energy. In the current climate club, the members of the club only impose penalty tariffs on the final goods traded by the firm, and not on energy imports. This is empirically relevant, c.f. Shapiro (2021) and Copeland et al. (2021): Inputs are more emission-intensives but trade policy is biased against downstream goods. Moreover, in the context of this model, fossil-fuel energy inputs are not carbon-intensive *per-se*, it is their use – i.e. the burning – of those fossil fuels in production that is carbon-intensive. As a result, Carbon-Border-Adjustment mechanisms would typically only impose a tariff on the "scope-1/scope-2" carbon footprint of fossil fuel extraction – and not the "scope-3" of its use along the downstream supply chain.

In our climate club setting, these tariffs are also strategic to incentivize participation. Therefore, I propose an alternative mechanism where the club members impose a tariff on the fossil fuel exports of the countries outside the club. The tariff is an import tax on the energy import from non-participants of the form:

$$\mathrm{t}_{ij}^{bf} = \beta \xi^f \mathrm{t}^\varepsilon \mathbb{1}\{i \in \mathbb{J}, j \notin \mathbb{J}\}$$

In the following graph Figure 18, I plot over different values of β the welfare impact for members of the club and the non-members – which is Russia here – to see if this strategic tariff can provide enough incentive for Russia to join the club with a carbon-tax as in the second best.

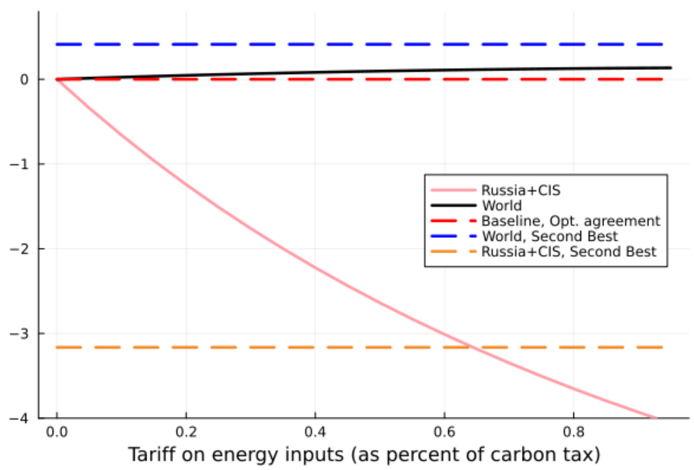

Figure 18: Welfare across countres
Optimal loss and damage fund: $\alpha^{\varepsilon,\star} = 15\%$

We see that with fossil-fuel-specific tariffs of $\beta = 60\%$ of the carbon tax, the second-best allocation can be achieved through incentive effect, and welfare can be maximized.

7.3 Retaliation

In this section, I consider an extension of the game-theoretical setting where countries outside the climate club can also act strategically. Until now, the countries outside the agreement were passive, setting policy $\{t_i^{\varepsilon}, t_{ij}^b\} = \{0, 0\}_{i \notin \mathbb{J}}$.

I now relax this assumption by conducting the following exercise: All the countries outside the club $j \notin \mathbb{J}$ impose a tariffs on club members i as a retaliation from any trade policy they would be targeted with. That is,

$$t_{ji}^b = \kappa t_{ij}^b \ , \qquad\qquad \forall i \in \mathbb{J} \ ,$$

where $\kappa \in (0, 1]$ is an exogenous parameter that represents the extent of the retaliation. In future extensions, I plan to make this decision endogenous, creating a multiplayer game between a climate coalition and a fringe of non-members.

I perform this exercise for different values of κ. The results are displayed in Figure 19, Figure 20 and Figure 21, respectively for $\kappa = 25, 50\%, 100\%$.

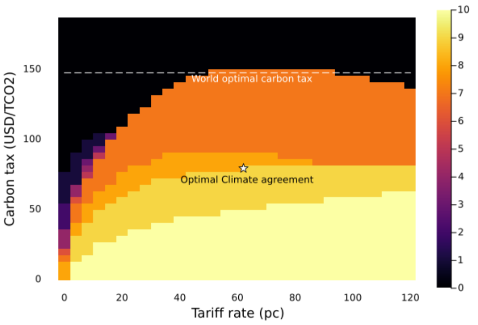

Figure 19: $\kappa = 0.25$

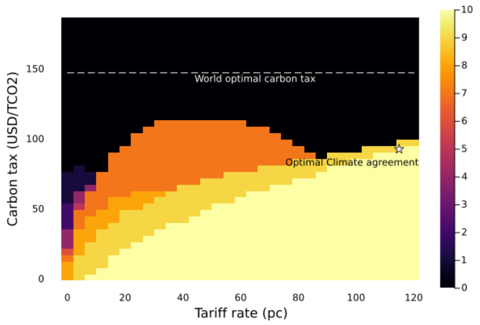

Figure 20: $\kappa = 0.5$

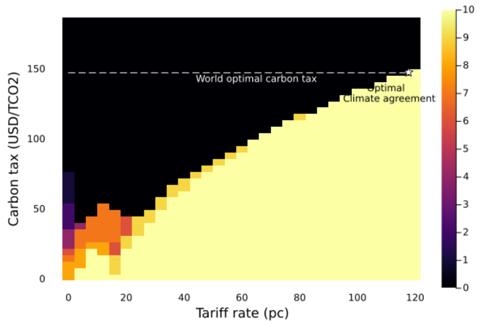

Figure 21: $\kappa = 1.0$

For moderate value of retaliation of non-members, i.e. $\kappa \in (0, 0.4)$, we note that the climate club more constrained than before: the achievable carbon tax is slightly lower, and the tariff needed to enforce it need to be raised. For example for $\kappa = 0.25$, the optimal carbon tax t^{ε} is less than $\$78/tCO_2$, and the tariff is $t^b = 64\%$ – instead of \$98 carbon tax and 50% tariff in $\kappa = 0$ case.

However, the larger the retaliation the larger the cost of trade disruption for both members *and* non members. The countries with the largest gains from trade would still choose optimal to participate in the agreements, which make the cost of being outside larger as κ grows above 0.5.

When κ become very large, it is optimal for the climate club to engage in an aggressive trade war, which push non-member to finally join the agreement. For example, when $\kappa = 1$, we see that the climate club recover its enforcement ability. It can even incentivize complete participation, for a carbon tax up to the optimal level $t^{\varepsilon} = \$147/tCO_2$, for large tariffs of $t^b = 118\%$. Importantly, in equilibrium with complete participation, no country pays these tariffs.

These findings underscore that understanding the underpinnings of trade tariff strategic behavior – beyond the simple terms-of-trade manipulation motives – is key to designing climate agreements.

8 Conclusion

This paper examines the design of an optimal climate agreement in the presence of free-riding incentives and redistributive effects. I develop a multi-country Integrated Assessment Model (IAM) that incorporates international trade in goods and energy markets for fossil fuels. This model accounts for heterogeneity across countries in terms of their vulnerability to climate change, income levels, energy mix, and positions as exporters or importers of goods and energy.

The analysis focuses on a global social planner's problem of maximizing world welfare through a climate agreement comprising three key elements: (1) the set of countries included in the agreement (the "climate club"), (2) an optimal level of carbon tax imposed on club members, and (3) an optimal level of trade tariffs imposed on non-member countries. I consider Nash equilibria where countries make strategic decisions about their participation, either unilaterally or through coalition deviations.

This study reveals a crucial trade-off between intensive and extensive margins in designing the optimal climate agreement. A small coalition of countries can implement high carbon taxes, achieving significant emissions reductions. However, a more extensive club with broader participation may be necessary for effectively combating global climate change, albeit at the cost of lower carbon taxes.

The main findings is first that the optimal climate club includes all countries except Russia, with a moderate carbon tax of \$100 per ton of CO_2 and a 50% tariff on goods from non-participants. To increase participation, it is beneficial to reduce the carbon tax by 35% from the globally optimal level of \$150 per ton of CO_2. This allows for the inclusion of Middle Eastern countries and several developing economies in South Asia and Africa. Excluding fossil fuel producers like Russia from the agreement is optimal, as their welfare costs from carbon taxation are too high to justify inclusion at any reasonable tax rate. Trade policy, particularly the threat of tariffs, is a key strategic instrument for undermining free-riding and incentivizing participation. However, its effectiveness diminishes as carbon taxes increase. Additional policy instruments, such as transfers through a "loss and damage" fund or fossil-fuel-specific tariffs, can improve the climate agreement and push the carbon tax closer to the second-best allocation.

In conclusion, this research underscores the complexity of designing effective climate agreements in a world of heterogeneous countries with divergent interests. It demonstrates that while a universal agreement with globally optimal carbon taxation may be unattainable due to free-riding incentives and redistributive effects, carefully designed climate clubs with strategic use of trade policy can achieve significant progress in global climate action.

References

Adao, Rodrigo, Arnaud Costinot and Dave Donaldson (2023), 'Putting Quantitative Models to the Test: An Application to Trump's Trade War'.

Anderson, James E. (1979), 'A theoretical foundation for the gravity equation', *The American Economic Review* **69**(1), 106–116.

Anthoff, David, Cameron Hepburn and Richard SJ Tol (2009), 'Equity weighting and the marginal damage costs of climate change', *Ecological Economics* **68**(3), 836–849.

Antras, Pol, Teresa C. Fort, Agustin Gutierrez and Felix Tintelnot (2024), 'Trade Policy and Global Sourcing: An Efficiency Rationale for Tariff Escalation', *Journal of Political Economy Macroeconomics* **2**(1), 1–44. Publisher: The University of Chicago Press.

Arkolakis, Costas, Arnaud Costinot and Andres Rodriguez-Clare (2012), 'New Trade Models, Same Old Gains?', *American Economic Review* **102**(1), 94–130.

Arkolakis, Costas and Conor Walsh (2023), 'Clean Growth'.

Arkolakis, Costas, Fabian Eckert and Rowan Shi (2023), 'Combinatorial discrete choice: A quantitative model of multinational location decisions'.

Atkeson, Andrew and Patrick J Kehoe (1999), 'Models of energy use: Putty-putty versus putty-clay', *American Economic Review* **89**(4), 1028–1043.

Barrage, Lint and William Nordhaus (2024), 'Policies, projections, and the social cost of carbon: Results from the DICE-2023 model', *Proceedings of the National Academy of Sciences* **121**(13), e2312030121. Publisher: Proceedings of the National Academy of Sciences.

Barrett, Scott (1994), 'Self-Enforcing International Environmental Agreements', *Oxford Economic Papers* **46**(Supplement 1), 878–894.

Barrett, Scott (2001), Strategic Environmental Policy and International Trade, *in* 'The Economics of International Trade and the Environment', CRC Press. Num Pages: 11.

Barrett, Scott (2003), *Environment and Statecraft : The Strategy of Environmental Treaty-Making: The Strategy of Environmental Treaty-Making*, OUP Oxford. Google-Books-ID: uqrey86neSIC.

Barrett, Scott (2013), 'Climate treaties and approaching catastrophes', *Journal of Environmental Economics and Management* **66**(2), 235–250.

Batabyal, Amitrajeet, ed. (2000), *The Economics of International Environmental Agreements*, Routledge, London.

Belfiori, Elisa, Daniel Carroll and Sewon Hur (2024), 'Unequal Climate Policy in an Unequal World'.

Belfiori, Maria Elisa (2018), 'Climate change and intergenerational equity: Revisiting the uniform taxation principle on carbon energy inputs', *Energy Policy* **121**, 292–299.

Bhandari, Anmol, David Evans, Mikhail Golosov and Thomas J. Sargent (2021*a*), 'Inequality, Business Cycles, and Monetary-Fiscal Policy', *Econometrica* **89**(6), 2559–2599.

Bhandari, Anmol, David Evans, Mikhail Golosov and Tom Sargent (2021*b*), 'Efficiency, Insurance, and Redistribution Effects of Government Policies'.

Bhattarai, Keshab, Sushanta K. Mallick and Bo Yang (2021), 'Are global spillovers complementary or competitive? Need for international policy coordination', *Journal of International Money and Finance* **110**, 102291.

Bilal, Adrien and Diego R Känzig (2024), 'The macroeconomic impact of climate change: Global vs. local temperature'.

Bilal, Adrien and Esteban Rossi-Hansberg (2023), 'Anticipating Climate Change Across the United States'.

Bohringer, Christoph, Jared C. Carbone and Thomas F. Rutherford (2012), 'Unilateral climate policy design: Efficiency and equity implications of alternative instruments to reduce carbon leakage', *Energy Economics* **34**, S208–S217.

Bohringer, Christoph, Jared C. Carbone and Thomas F. Rutherford (2016), 'The Strategic Value of Carbon Tariffs', *American Economic Journal: Economic Policy* **8**(1), 28–51.

Bolt, Jutta and Jan Luiten van Zanden (2023), 'Maddison-style estimates of the evolution of the world economy: A new 2023 update', *Journal of Economic Surveys* **n/a**(n/a).

Bornstein, Gideon, Per Krusell and Sergio Rebelo (2023), 'A World Equilibrium Model of the Oil Market', *The Review of Economic Studies* **90**(1), 132–164.

Bourany, Thomas (2022), 'Energy shocks and aggregate fluctuations'.

Bourany, Thomas (2024), 'The inequality of climate change and optimal energy policy'.

Burke, Marshall, Solomon M Hsiang and Edward Miguel (2015), 'Global non-linear effect of temperature on economic production', *Nature* **527**(7577), 235–239.

Carleton, Tamma A, Amir Jina, Michael T Delgado, Michael Greenstone, Trevor Houser, Solomon M Hsiang, Andrew Hultgren, Robert E Kopp, Kelly E McCusker, Ishan B Nath, James Rising, Ashwin Rode, Hee Kwon Seo, Arvid Viaene, Jiacan Yuan and Alice Tianbo Zhang (2020), Valuing the Global Mortality Consequences of Climate Change Accounting for Adaptation Costs and Benefits, Working Paper 27599, National Bureau of Economic Research. Series: Working Paper Series.

Castro-Vincenzi, Juanma, Gaurav Khanna, Nicolas Morales and Nitya Pandalai-Nayar (2024), 'Weathering the Storm: Supply Chains and Climate Risk'.

Chari, V. V., Juan Pablo Nicolini and Pedro Teles (2023), 'Optimal Cooperative Taxation in the Global Economy', *Journal of Political Economy* **131**(1), 95–130. Publisher: The University of Chicago Press.

Chen, Zhiqi (1997), 'Negotiating an Agreement on Global Warming: A Theoretical Analysis', *Journal of Environmental Economics and Management* **32**(2), 170–188.

Chichilnisky, Graciela and Geoffrey Heal (1994), 'Who should abate carbon emissions?: An international viewpoint', *Economics Letters* **44**(4), 443–449.

Clausing, Kimberly A. and Catherine Wolfram (2023), 'Carbon Border Adjustments, Climate Clubs, and Subsidy Races When Climate Policies Vary', *Journal of Economic Perspectives* **37**(3), 137–162.

Conte, Bruno, Klaus Desmet, Dávid Krisztián Nagy and Esteban Rossi-Hansberg (2021), 'Local sectoral specialization in a warming world', *Journal of Economic Geography* **21**(4), 493–530.

Conte, Maddalena, Pierre Cotterlaz, Thierry Mayer et al. (2022), 'The cepii gravity database'.

Copeland, Brian R., Joseph S. Shapiro and M. Scott Taylor (2021), 'Globalization and the Environment'.

Copeland, Brian R and M Scott Taylor (2004), 'Trade, growth, and the environment', *Journal of Economic literature* **42**(1), 7–71.

Costinot, Arnaud, Dave Donaldson, Jonathan Vogel and Ivan Werning (2015), 'Comparative Advantage and Optimal Trade Policy *', *The Quarterly Journal of Economics* **130**(2), 659–702.

Cruz, Jose-Luis and Esteban Rossi-Hansberg (2022), 'Local Carbon Policy'.

Cruz, Jose-Luis and Esteban Rossi-Hansberg (2024), 'The Economic Geography of Global Warming', *The Review of Economic Studies* **91**(2), 899–939.

Davila, Eduardo and Andreas Schaab (2023), 'Optimal Monetary Policy with Heterogeneous Agents: Discretion, Commitment, and Timeless Policy'.

Davila, Eduardo and Ansgar Walther (2022), 'Corrective Regulation with Imperfect Instruments'.

Dietz, Simon, Frederick van der Ploeg, Armon Rezai and Frank Venmans (2021), 'Are Economists Getting Climate Dynamics Right and Does It Matter?', *Journal of the Association of Environmental and Resource Economists* **8**(5), 895–921. Publisher: The University of Chicago Press.

Dingel, Jonathan I., Kyle C. Meng and Solomon M. Hsiang (2019), 'Spatial Correlation, Trade, and Inequality: Evidence from the Global Climate'.

Douenne, Thomas, Albert Jan Hummel and Marcelo Pedroni (2023), 'Optimal Fiscal Policy in a Climate-Economy Model with Heterogeneous Households'.

Energy Institute (2024), 'Statistical review of world energy'.

Ernst, Anne, Natascha Hinterlang, Alexander Mahle and Nikolai StÃ¤hler (2023), 'Carbon pricing, border adjustment and climate clubs: Options for international cooperation', *Journal of International Economics* **144**, 103772.

Falkner, Robert (2016), 'The Paris Agreement and the new logic of international climate politics', *International' Affairs* **92**(5), 1107–1125.

Falkner, Robert, Naghmeh Nasiritousi and Gunilla Reischl (2022), 'Climate clubs: politically feasible and desirable?', *Climate Policy* **22**(4), 480–487.

Farrokhi, Farid and Ahmad Lashkaripour (2024), 'Can trade policy mitigate climate change', *Econometrica* .

Folini, Doris, Aleksandra Friedl, Felix Kübler and Simon Scheidegger (2024), 'The climate in climate economics', *Review of Economic Studies* p. rdae011.

Fontagne, Lionel and Katheline Schubert (2023), 'The Economics of Border Carbon Adjustment: Rationale and Impacts of Compensating for Carbon at the Border', *Annual Review of Economics* **15**(Volume 15, 2023), 389–424. Publisher: Annual Reviews.

Gentile, Claudia (2024), 'Relying on intermittency: Clean energy, storage, and innovation in a macro climate model'.

Golosov, Mikhail, John Hassler, Per Krusell and Aleh Tsyvinski (2014), 'Optimal Taxes on Fossil Fuel in General Equilibrium', *Econometrica* **82**(1), 41–88.

Hagen, Achim and Jan Schneider (2021), 'Trade sanctions and the stability of climate coalitions', *Journal of Environmental Economics and Management* **109**, 102504.

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

Harstad, Bård (2023), 'Pledge-and-Review Bargaining: from Kyoto to Paris', *The Economic Journal* **133**(651), 1181–1216.

Harstad, Bård, Francesco Lancia and Alessia Russo (2022), 'Prices vs. quantities for self-enforcing agreements', *Journal of Environmental Economics and Management* **111**, 102595.

Hassler, John, Per Krusell and Anthony A Smith (2016), Environmental macroeconomics, *in* 'Handbook of macroeconomics', Vol. 2, Elsevier, pp. 1893–2008.

Hassler, John, Per Krusell and Conny Olovsson (2010), 'Oil monopoly and the climate', *American Economic Review* **100**(2), 460–64.

Hassler, John, Per Krusell and Conny Olovsson (2019), Directed technical change as a response to natural-resource scarcity, Working Paper 375, Sveriges Riksbank Working Paper Series.

Hassler, John, Per Krusell and Conny Olovsson (2021), 'Presidential Address 2020 Suboptimal Climate Policy', *Journal of the European Economic Association* **19**(6), 2895–2928.

Hillebrand, Elmar and Marten Hillebrand (2019), 'Optimal climate policies in a dynamic multi-country equilibrium model', *Journal of Economic Theory* **179**, 200–239.

Hoel, Michael (1991), 'Global environmental problems: The effects of unilateral actions taken by one country', *Journal of Environmental Economics and Management* **20**(1), 55–70.

Hoel, Michael (1994), 'Efficient Climate Policy in the Presence of Free Riders', *Journal of Environmental Economics and Management* **27**(3), 259–274.

Hsiao, Allan (2022), 'Coordination and Commitment in International Climate Action'.

IPCC, Pörtner, H.-O., D.C. Roberts, H. Adams, I. Adelekan, C. Adler, R. Adrian, P. Aldunce, E. Ali, R. Ara Begum, B. Bednar Friedl, R. Bezner Kerr, R. Biesbroek, J. Birkmann, K. Bowen, M.A. Caretta, J. Carnicer, E. Castellanos, T.S. Cheong, W. Chow, G. Cisse G. Cisse and Z. Zaiton Ibrahim (2022), *Climate Change 2022: Impacts, Adaptation and Vulnerability*, Technical Summary, Cambridge University Press, Cambridge, UK and New York, USA.

Iverson, Terrence (2024), 'Tiered Climate Clubs: Global Abatement Without Global Agreement *'.

Jia, Panle (2008), 'What happens when wal-mart comes to town: An empirical analysis of the discount retailing industry', *Econometrica* **76**(6), 1263–1316.

Kleinman, Benny, Ernest Liu and Stephen J Redding (2020), 'International friends and enemies', *Forthcoming, American Economic Journal: Macroeconomics* .

Kleinman, Benny, Ernest Liu, Stephen J Redding and Motohiro Yogo (2023), 'Neoclassical growth in an interdependent world'.

Kortum, Samuel S. and David A. Weisbach (2021), 'Optimal Unilateral Carbon Policy'.

Kotlikoff, Laurence, Felix Kubler, Andrey Polbin, Jeffrey Sachs and Simon Scheidegger (2021), 'Making Carbon Taxation a Generational Win Win', *International Economic Review* **62**(1), 3–46. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/iere.12483.

Kotlikoff, Laurence J., Felix Kubler, Andrey Polbin and Simon Scheidegger (2021), 'Can Today's and Tomorrow's World Uniformly Gain from Carbon Taxation?'.

Krusell, Per and Anthony A. Smith (2022), 'Climate change around the world'.

Le Grand, François, Alaïs Martin-Baillon and Xavier Ragot (2024), 'Should monetary policy care about redistribution? optimal monetary and fiscal policy with heterogeneous agents'.

Maggi, G. (2016), Issue Linkage, *in* 'Handbook of Commercial Policy', Vol. 1, Elsevier, pp. 513–564.

Maggi, Giovanni and Robert W. Staiger (2022), 'International Climate Agreements and the Scream of Greta'.

Mahdiraji, Hannan Amoozad, Elham Razghandi and Adel Hatami-Marbini (2021), 'Overlapping coalition formation in game theory: A state-of-the-art review', *Expert Systems with Applications* **174**, 114752.

Nordhaus, William (2015), 'Climate Clubs: Overcoming Free-riding in International Climate Policy', *American Economic Review* **105**(4), 1339–1370.

Nordhaus, William (2021), 'Climate Club Futures: On the Effectiveness of Future Climate Clubs', *Cowles Foundation Discussion Papers* .

Nordhaus, William D (2011), 'Estimates of the social cost of carbon: background and results from the rice-2011 model'.

Nordhaus, William D. and Zili Yang (1996), 'A Regional Dynamic General-Equilibrium Model of Alternative Climate-Change Strategies', *The American Economic Review* **86**(4), 741–765. Publisher: American Economic Association.

Okada, Akira (2023), 'A dynamic climate negotiation game achieving full cooperation', *Journal of Economic Behavior & Organization* **214**, 657–669.

Ossa, Ralph (2014), 'Trade Wars and Trade Talks with Data', *American Economic Review* **104**(12), 4104–4146.

Papageorgiou, Chris, Marianne Saam and Patrick Schulte (2017), 'Substitution between Clean and Dirty Energy Inputs: A Macroeconomic Perspective', *The Review of Economics and Statistics* **99**(2), 281–290.
URL: *https://doi.org/10.1162/REST_a_00592*

Pickering, Jonathan, Steve Vanderheiden and Seumas Miller (2012), '"If Equity's In, We're Out": Scope for Fairness in the Next Global Climate Agreement', *Ethics & International Affairs* **26**(4), 423–443.

Ray, Debraj and Rajiv Vohra (2015), Chapter 5 - Coalition Formation, *in* H. P.Young and S.Zamir, eds, 'Handbook of Game Theory with Economic Applications', Vol. 4, Elsevier, pp. 239–326.

Rudik, Ivan, Gary Lyn, Weiliang Tan and Ariel Ortiz-Bobea (2021), 'The Economic Effects of Climate Change in Dynamic Spatial Equilibrium'.

Shapiro, Joseph S (2021), 'The Environmental Bias of Trade Policy*', *The Quarterly Journal of Economics* **136**(2), 831–886.

Tarr, David G., Dmitrii E. Kuznetsov, Indra Overland and Roman Vakulchuk (2023), 'Why carbon border adjustment mechanisms will not save the planet but a climate club and subsidies for transformative green technologies may', *Energy Economics* **122**, 106695.

Weisbach, David A., Samuel Kortum, Michael Wang and Yujia Yao (2023), 'Trade, Leakage, and the Design of a Carbon Tax', *Environmental and Energy Policy and the Economy* **4**, 43–90. Publisher: The University of Chicago Press.

Weitzman, Martin L. (2003), Prices vs. Quantities 1 , 2, *in* 'The Theory and Practice of Command and Control in Environmental Policy', Routledge. Num Pages: 15.

Weitzman, Martin L (2012), 'Ghg targets as insurance against catastrophic climate damages', *Journal of Public Economic Theory* **14**(2), 221–244.

Weitzman, Martin L. (2015), 'Internalizing the Climate Externality: Can a Uniform Price Commitment Help?', *Economics of Energy & Environmental Policy* **4**(2), 37–50. Publisher: International Association for Energy Economics.

Appendix

A Calibration

Table 2: Baseline calibration

Technology & Energy markets			
α	0.35	Capital share in $F(\cdot)$	Capital/Output ratio
ϵ	0.12	Energy share in $F(\cdot)$	Energy cost share (8.5%)
σ^y	0.3	Elasticity capital-labor vs. energy	Complementarity in production (c.f. Bourany 2022)
ω^f	0.56	Fossil energy share in $e(\cdot)$	Oil-gas/Energy ratio
ω^c	0.27	Coal energy share in $e(\cdot)$	Coal/Energy ratio
ω^r	0.17	Non-carbon energy share in $e(\cdot)$	Non-carbon/Energy ratio
σ^e	2.0	Elasticity fossil-coal-non-carbon	Slight substitutability & Study by Stern
δ	0.06	Depreciation rate	Investment/Output ratio
$\bar{g}$	$0.01^\star$	Long run TFP growth	Conservative estimate for growth
Preferences & Time horizon			
ρ	0.03	HH Discount factor	Long term interest rate & usual calib. in IAMs
η	1.5	IES / Risk aversion	Standard calibration
n	0.0035	Long run population growth	Conservative estimate for growth
ω_i	1	Pareto weights	Uniforms / Utilitarian Social Planner
ω_i	$1/u'(c_i)$	Pareto weights	Negishi / Status-quo Social Planner
T	400	Time horizon	Time for climate system to stabilize
Climate parameters			
ξ^f	2.761	Emission factor – Oil & natural gas	Conversion 1 $MTOE \Rightarrow 1\ MT\ CO_2$
ξ^c	3.961	Emission factor – Coal	Conversion 1 $MTOE \Rightarrow 1\ MT\ CO_2$
χ	$2.3/1e6$	Climate sensitivity	Pulse experiment: $100\,GtC \equiv 0.23°C$ medium-term warming
δ_s	0.0004	Carbon exit from atmosphere	Pulse experiment: $100\,GtC \equiv 0.15°C$ long-term warming
ζ	0.027	Growth rate, Carbon Capture and Storage	Starting after 2100, Follows Krusell Smith (2022)
$\gamma^{\oplus}$	0.003406	Damage sensitivity	Nordhaus' DICE
$\gamma^{\ominus}$	$0.3\times\gamma^{\oplus}$	Damage sensitivity	Nordhaus' DICE & Rudik et al (2022)
α^T	0.5	Weight historical climate for optimal temp.	Marginal damage correlated with initial temp.
$T^\star$	14.5	Optimal yearly temperature	Average spring temperature / Developed economies

A.1 Additional calibration graphs

A.1.1 Quantification – Trade shares

We displayed the trade share from the data in Figure 22 and how we calibrate the trade model.

Armington Trade model and trade shares:

$$s_{ij} \equiv \frac{c_{ij}p_{ij}}{c_i\mathbb{P}_i} = a_{ij}\frac{((1+\mathrm{t}_{ij})\tau_{ij}\mathrm{p}_j)^{1-\theta}}{\sum_k a_{ik}((1+\mathrm{t}_{ik})\tau_{ik}\mathrm{p}_k)^{1-\theta}}$$

We estimate a gravity regression, and CES $\theta = 5.63$. The Iceberg cost τ_{ij} are projection of geographical distance $\log\tau_{ij} = \beta\log d_{ij}$. The preference parameters a_{ij} identified as remaining variation in the trade share s_{ij}. As a results, both τ_{ij} and a_{ij} are policy invariant in our climate agreement setting. The description of the procedure is detailed in Section 4.3.

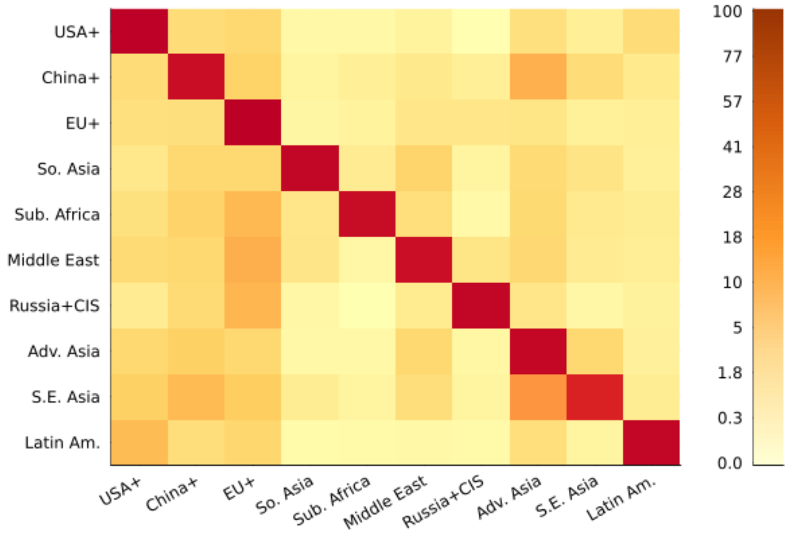

Figure 22: Trade shares as mesured in Conte et al. (2022)

A.1.2 Climate system and pulse experiment

This pulse experiment, from Dietz et al. (2021), summarizes how our climate model should be calibrated to replicate larger scale IAMs like CMIP5.

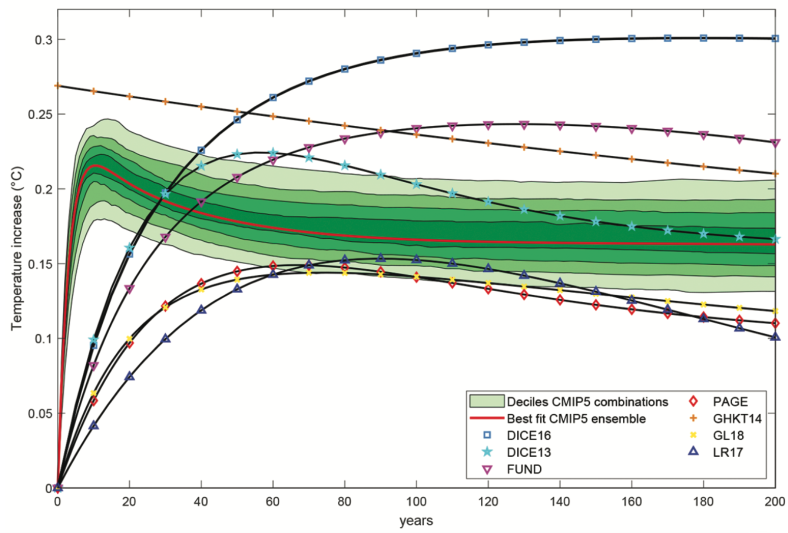

Figure 23: Pulse experiment: $\mathcal{E} = 100GtCO_2$ at $t = 0$
Comparison across different IAMs, Dietz et al. (2021)

B Model

B.1 Model environment

- Gov. policies $t_i = \{\mathrm{t}^{\varepsilon}_{it}, \mathrm{t}^{b}_{ijt}, \mathrm{t}^{ls}_{it}\}$
- State: $s_i = \{k_{it_0}, T_{it}, \mathcal{R}_{it}\}_t$,
- Agents (HH/firms) controls $c_i = \{c_{it}, c_{ijt}, k_{it}, e^f_{it}, e^c_{it}, e^r_{it}, e^x_{it}\}_t$
- Eq prices: $p = \mathrm{p}_{it}, w_{it}, q^f_t, q^c_{it}, q^r_{it}$
- Lagrange multipliers / costates: $\lambda_i = \{\lambda^w_{it}, \lambda^S_{it}, \lambda^T_{it}\}$
- Local welfare vs Global welfare

$$\mathcal{U}_i = \max_{c_i} \int_0^T e^{-\bar{\rho}_i t} \mathcal{P}_i u(c_{it}, T_{it}) dt$$
$$\mathcal{W} = \max_{\{c_i\}_i} \sum_{i \in \mathbb{I}} \omega_i \mathcal{P}_i \int_0^T e^{-\bar{\rho}_i t} u(c_{it}, T_{it}) dt$$

- Assumptions:
  - weak separability of utility

$$u(c_{it}, T_{it}) = \tilde{u}(c_i) \mathcal{D}^u(T_{it}) = \frac{c_{it}^{1-\eta}}{1-\eta} \big(\tilde{\mathcal{D}}^u(T_{it})\big)^{1-\eta}$$

  - DICE damage functions $\mathcal{D}^u_i(T)$ and $\mathcal{D}^y_i(T)$

$$\mathcal{D}^u_i(T) = e^{-\frac{\gamma^c}{2}(T - T^\star_i)^2} \qquad \Rightarrow \qquad \mathcal{D}^{u\prime}_i(T) = -\mathcal{D}^u_i(T) \gamma^c (T - T^\star_i)$$

  - CES demand / Armington structure, price of imports $p_{ij} = \tau_{ij}(1+\mathrm{t}^b_{ij})\mathrm{p}_j$

$$u(\{c_{ij}\}_j, T_i)) = u(c_i, T_i) \qquad c_i = \Big(\sum_j a_{ij}^{\frac{1}{\theta}} c_{ij}^{\frac{\theta-1}{\theta}} \Big)^{\frac{\theta}{\theta-1}}$$

- Heterogenous discount rate: $\bar{\rho}_i = \rho - n_i - (1-\eta)\bar{g}_i$
- Climate system:

$$\dot{\mathcal{S}}_t = \zeta_t \mathcal{E}_t - \delta_s \mathcal{S}_t = e^{-ot}[\xi^f e^f_{it} + \xi^c e^c_{it}] - \delta_s \mathcal{S}_t$$
$$\mathcal{S}_t = \mathcal{S}_0 e^{-\delta_s t} + \int_0^t e^{-\delta_s(t-u)} e^{(n+\bar{g}-o)u} (\xi^f e^f_{iu} + \xi^c e^c_{iu}) du$$
$$T_{it} = T_{it_0} + \Delta_i \chi \mathcal{S}_t$$

with $o = g_\zeta \mathbb{1}\{t > 2100\}$ the rate of growth of additional abatement due to CCS after 2100.

B.2 Summary, model setting

- Expenditure by household:
$$\sum_j c_{ijt}\tau_{ij}(1+\mathrm{t}^b_{ij})\mathrm{p}_{jt} = c_{it}\mathbb{P}_{it}$$

- Final good firm problem: pay for labor and capital and buys three energy inputs:
$$\pi^g_{it} = \mathrm{p}_i\mathcal{D}^y(T_{it})z_{it}F(\ell_i, k_{it}, e^f_{it}, e^c_{it}, e^r_{it}) - w_{it}\ell_i - r_t k_{it} - (q^f_t+\xi^f \mathrm{t}^s_{it})e^f_{it} - (q^c_t+\xi^c\mathrm{t}^s_{it})e^c_{it} - q^r_{it}e^r_{it} = 0$$

- Fossil Energy firm profit:
$$\mathcal{P}_i\pi^f_{it} = q^f_t e^x_{it} - \mathcal{C}^f_i(e^x_{it}, \mathcal{R}_{it})\mathcal{P}_i$$

- Budget constraint for the household:
replace labor income, and divide by price index (analog of "real" quantities). Reminder that capital expenditure are made in the final consumption good bundle.

$$\begin{aligned}
c_{it}\mathbb{P}_{it} + \big(\dot{k}_{it} + (n_i + \bar{g}_i + \delta)\big)\mathbb{P}_{it} &= w_{it}\ell_i + r_t k_{it} + \pi^f_t + \mathrm{t}^{ls}_i \\
0 &= \frac{\mathrm{p}_{it}}{\mathbb{P}_{it}}\mathcal{D}^y(T_{it})z_{it}F(k_{it}, e^f_{it}, e^c_{it}, e^r_{it}) - (n_i + \bar{g}_i + \delta)k_{it} \\
&\quad + \frac{1}{\mathcal{P}_i\mathbb{P}_{it}}\Big[q^f_t e^x_{it} - \nu(e^x_{it}, \mathcal{R}_{it})\Big] - \frac{(q^f_t+\xi^f\mathrm{t}^s_{it})}{\mathbb{P}_{it}}e^f_{it} - \frac{(q^c_t+\xi^c\mathrm{t}^s_{it})}{\mathbb{P}_{it}}e^c_{it} - \frac{q^r_{it}}{\mathbb{P}_{it}}e^r_{it} - c_{it} + \frac{\mathrm{t}^{ls}_{it}}{\mathbb{P}_{it}} - \dot{k}_{it}
\end{aligned}$$

- CES / Armington trade model, with price of imports $p_{ij} = \tau_{ij}(1+\mathrm{t}^b_{ij})\mathrm{p}_j$

$$u(\{c_{ij}\}_j, T_i)) = u(c_i)\,\mathcal{D}^u_i(T_i) \qquad c_i = \Big(\sum_j a_{ij}^{\frac{1}{\theta}} c_{ij}^{\frac{\theta-1}{\theta}}\Big)^{\frac{\theta}{\theta-1}}$$

$$FOC \qquad [c_{ij}] \qquad\qquad u'(c_i)\mathcal{D}^u_i(T_i)\, c_i^{\frac{1}{\theta}} a_{ij}^{\frac{1}{\theta}} c_{ij}^{-\frac{1}{\theta}} = p_{ij}\lambda^w_i$$

Price index:
$$\mathbb{P}_i = \Big(\sum_j a_{ij}p_{ij}^{1-\theta}\Big)^{\frac{1}{1-\theta}} = \Big(\sum_j a_{ij}(\tau_{ij}(1+\mathrm{t}^b_{ij})\mathrm{p}_j)^{1-\theta}\Big)^{\frac{1}{1-\theta}}$$

Demand system:
$$\Rightarrow \qquad \frac{c_{ij}}{c_i} = a_{ij}\big(\frac{\mathbb{P}_i}{p_{ij}}\big)^{\theta} \qquad \Rightarrow \qquad \frac{c_{ij}p_{ij}}{c_i\mathbb{P}_i} = a_{ij}\big(\frac{p_{ij}}{\mathbb{P}_i}\big)^{1-\theta}$$
$$s_{ij} = \frac{c_{ij}p_{ij}}{c_i\mathbb{P}_{it}} = a_{ij}\frac{p_{ij}^{1-\theta}}{\sum_k a_{ik}p_{ik}^{1-\theta}} = a_{ij}\frac{((1+\mathrm{t}_{ij})\tau_{ij}\mathrm{p}_j)^{1-\theta}}{\sum_k a_{ik}((1+\mathrm{t}_{ik})\tau_{ik}\mathrm{p}_k)^{1-\theta}}$$

Aggregating, we obtain the marginal value of "wealth":
$$\lambda^w_i = \frac{u'(c_i)\mathcal{D}^u_i(T_i)}{\mathbb{P}_i}$$

B.2.1 Market clearing

We reexpress the market clearing, for good i in expenditure terms. The time subscripts are removed for conciseness.

$$\mathcal{P}_i y_i = \overline{\mathcal{D}}_i(\{T_{it}\}) z_i F(k_i, e_i) = \sum_{k\in\mathbb{I}} \tau_{ki}\mathcal{P}_k c_{ki} + \sum_{k\in\mathbb{I}} \mathcal{P}_k \tau_{ki}(x^f_{ki} + x^c_{ki} + x^r_{ki} + x^k_{ki})$$

Rewriting in expenditure, multiplying by p_j, and using the fact that the input choice is identical between

$$\begin{aligned}
\mathcal{P}_i y_i \mathrm{p}_i &= \sum_{k\in\mathbb{I}} \mathcal{P}_k \tau_{ki}(x^f_{ki} + x^c_{ki} + x^r_{ki} + x^k_{ki})\mathrm{p}_i \\
&= \sum_{k\in\mathbb{I}} \mathcal{P}_k \frac{1}{1+\mathrm{t}^b_{ki}} \tau_{ki}(1+\mathrm{t}^b_{ki})\mathrm{p}_i(c_{ki} + x^f_{ki} + x^c_{ki} + x^r_{ki} + x^k_{ki}) \\
&= \sum_{k\in\mathbb{I}} \mathcal{P}_k \frac{1}{1+\mathrm{t}^b_{ki}} s_{ki}(c_i + \mathcal{C}^f_i(e^x_i) + z^c_i e^c_i + z^r_i e^r_i + (n_k + \bar{g}_k + \delta)k_k)\mathbb{P}_k
\end{aligned}$$

Using the budget constraint to replace c_i

$$\begin{aligned}
\mathcal{P}_i y_i \mathrm{p}_i = \sum_{k\in\mathbb{I}} \frac{\mathcal{P}_k s_{ki}}{1+\mathrm{t}^b_{ki}} \Big(& y_k \mathrm{p}_k - (q^f + \xi^f \mathrm{t}^\varepsilon_k) e^f_k - (q^c_k + \xi^c \mathrm{t}^\varepsilon) - q^r_k e^r_k - (n_i + \bar{g}_i + \delta) k_{it}\mathbb{P}_k + \Big[q^f e^x_k - \mathcal{C}^f_k(e^x_k)\mathbb{P}_k \Big] + \mathrm{t}^{ls}_k \\
& + \mathcal{C}^f_i(e^x_k)\mathbb{P}_k + z^c_k e^c_k \mathbb{P}_k + z^r_k e^r_k \mathbb{P}_k + (n_k + \bar{g}_k + \delta)k_k\mathbb{P}_k \Big) \\
\mathcal{P}_i y_i \mathrm{p}_i = \sum_{k\in\mathbb{I}} \frac{\mathcal{P}_k s_{ki}}{1+\mathrm{t}^b_{ki}} \Big(& y_k \mathrm{p}_k + q^f (e^x_k - e^f_k) + \widetilde{\mathrm{t}}^{ls}_k \Big) = \sum_{k\in\mathbb{I}} \frac{s_{ki}}{1+\mathrm{t}^b_{ki}} \mathcal{P}_k(\widetilde{v}_k + \widetilde{\mathrm{t}}^{ls}_k)
\end{aligned}$$

where $\widetilde{v}_k = y_k \mathrm{p}_k + q^f (e^x_k - e^f_k)$ represent the revenues of country k in terms of production and energy export and the lump-sum transfers of the tariffs: $\widetilde{\mathrm{t}}^{ls}_k = \sum_j \mathrm{t}^b_{kj}\tau_{kj}(c_{kj} + x^f_{kj} + x^c_{kj} + x^r_{kj} + x^k_{kj})$.

We see that the lump-sum transfer also depends on the quantities. To be able to express the market in expenditure, we solve:

$$\begin{aligned}
\mathcal{P}_i p_i y_i &= \sum_i \frac{s_{ki}}{1+\mathrm{t}^b_{ki}} \mathcal{P}_k [\widetilde{v}_k + \widetilde{\mathrm{t}}^{ls}_k] \\
v_k &:= \widetilde{v}_k + \widetilde{\mathrm{t}}^{ls}_k \\
\widetilde{\mathrm{t}}^{ls}_k &= \sum_j \mathrm{t}^b_{kj}\tau_{kj} \underbrace{(c_{kj} + x^f_{kj} + x^r_{kj} + x^c_{kj} + x^k_{kj})}_{=:x_{kj}} \mathrm{p}_k \qquad\qquad x_{kj} = \frac{s_{kj} v_k}{(1+\mathrm{t}_{kj})\tau_{kj}\mathrm{p}_k}
\end{aligned}$$

As a result, we solve the "fixed point" for v_i as follow:

$$\begin{aligned}
v_i &= \widetilde{v}_i + v_i \sum_j \frac{\mathrm{t}^b_{ij}}{1+\mathrm{t}^b_{ij}} s_{ij} \\
v_i &= \frac{1}{1 - \sum_j \frac{\mathrm{t}^b_{ij}}{1+\mathrm{t}^b_{ij}} s_{ij}} \widetilde{v}_i = m_i \widetilde{v}_i \qquad \text{with} \qquad m_i = \frac{1}{1 - \sum_j \frac{\mathrm{t}^b_{ij}}{1+\mathrm{t}^b_{ij}} s_{ij}}
\end{aligned}$$

To conclude, the market clearing writes:

$$\mathcal{P}_i p_i y_i = \sum_i \frac{s_{ki}}{1+\mathrm{t}^b_{ki}} \mathcal{P}_k m_k \Big[y_k \mathrm{p}_k + q^f \left(e^x_k - e^f_k\right) \Big]$$

B.3 Making the dynamic model stationary

We solve the optimization problem of the household – who own the firms. This is a dynamic problem, since climate changes over time, with emissions $\mathcal{E}_t$. We express the Lagrangian for the problem as a finite-horizon problem, and we take the finite horizon $T \to \infty$.

$$\begin{aligned}
\mathcal{L}(s_i, c_i, \lambda_i) = & \int_0^T e^{-\bar{\rho}_i t} \mathcal{P}_i u(c_{it}, T_{it}) dt + \int_0^T e^{-\bar{\rho}_i t} \mathcal{P}_i \lambda^w_{it} \Big(\mathrm{p}_i \mathcal{D}^y(T_{it}) z_{it} F(k_{it}, e^f_{it}, e^c_{it}, e^r_{it}) - (n_i + \bar{g}_i + \delta) k_{it} \\
& \frac{1}{\mathcal{P}_i} \Big[q^f_t e^x_{it} - \nu(e^x_{it}, \mathcal{R}_{it}) \Big] - (q^f_t + \xi^f \mathrm{t}^s_{it}) e^f_{it} - (q^c_t + \xi^c \mathrm{t}^s_{it}) e^c_{it} - q^r_{it} e^r_{it} - c_{it} \mathbb{P}_{it} + \mathrm{t}^{ls}_i - \dot{k}_{it} \Big) dt \\
& + \int_0^T e^{-\bar{\rho}_i t} \mathcal{P}_i \lambda^S_{it} \Big[\mathcal{S}_0 e^{-\delta_s t} + \int_0^t e^{-\delta_s (t-u)} e^{(n+\bar{g}-o)u} (\xi^f e^f_{iu} + \xi^c e^c_{iu}) du - \mathcal{S}_t \Big] dt \\
& + \int_0^T e^{-\bar{\rho}_i t} \mathcal{P}_i \lambda^T_{it} \Big(\Delta_i \chi \mathcal{S}_t - (T_{it} - T_{it_0}) \Big) dt
\end{aligned}$$

We impose the "constraint" that all the economic controls need to be constant over time, $c_{it} = c_i \ \forall i$. As a result, all the equilibrium prices are also constant over time $p_{it} = p_i \ \forall i$.

$$\begin{aligned}
\mathcal{L}(s_i, c_i, \lambda_i) = & \mathcal{P}_i u(c_i) \underbrace{\int_0^T e^{-\bar{\rho}_i t} \mathcal{D}^u(T_{it}) dt}_{=\overline{\mathcal{D}}^u(\{T_{it}\}_t)} + \mathcal{P}_i \Big(\int_0^T e^{-\bar{\rho}_i t} \lambda^w_{it} \mathcal{D}^y(T_{it}) dt \Big) \mathrm{p}_i z_{it} F(k_i, e^f_i, e^c_i, e^r_i) \ - \int_0^T e^{-\bar{\rho}_i t} \dot{k}_{it} \lambda^w_{it} dt \\
& + \frac{1}{\mathcal{P}_i} [q^f e^x_i - \nu(e^x_i, \mathcal{R}_i)] \underbrace{\int_0^T e^{-\bar{\rho}_i t} \lambda^w_{it} dt}_{=\overline{\lambda^w_i}} - \mathcal{P}_i \overline{\lambda^w_i} \Big((q^f + \xi^f \mathrm{t}^\varepsilon_i) e^f_i + (q^c + \xi^c \mathrm{t}^\varepsilon_i) e^c_i + q^r_i e^r_i + (n_i + \bar{g}_i + \delta) k_i + c_i \mathbb{P}_i + \mathrm{t}^{ls}_i \Big) \\
& + \mathcal{P}_i \int_0^T e^{-\bar{\rho}_i t} \lambda^S_{it} \mathcal{S}_t dt + \mathcal{P}_i \int_0^T e^{-\bar{\rho}_i t} \lambda^T_{it} \Big(\Delta_i \chi \mathcal{S}_t - (T_{it} - T_{it_0}) \Big) dt
\end{aligned}$$

with $\mathcal{S}_t = \mathcal{S}_0 e^{-\delta_s t} + \int_0^t e^{-\delta_s (t-u)} e^{(n+\bar{g}-o)u} (\xi^f e^f_{iu} + \xi^c e^c_{iu}) du$

Optimality conditions:

- Consumption $[c_i]$:

$$u'(c_i) \overline{\mathcal{D}}^u(\{T_{it}\}_t) = \overline{\lambda}^w_{it} \mathbb{P}_i$$

- Energy choices $[e^k_i]$, for $k \in \{f, c, r\}$:

$$\begin{aligned}
\overline{\lambda}^w_i (\mathrm{p}_i MPe^f_i - q^f + \xi^f \mathrm{t}^\varepsilon_i)) &= 0 \\
\overline{\lambda}^w_i (\mathrm{p}_i MPe^c_i - q^c + \xi^c \mathrm{t}^\varepsilon_i) &= 0 \\
\overline{\lambda}^w_i (\mathrm{p}_i MPe^r_i - q^r) &= 0
\end{aligned}$$

- Capital choice $[k_{it}]$

$$\begin{aligned}
\dot{\lambda}^w_{it} &= \lambda^w_{it}(\mathrm{p}_i MPk_i - \delta - \eta\bar{g}_i - \rho) \\
\mathrm{p}_i MPk_i - \delta = \bar{r} &= \eta\bar{g}_i + \rho \qquad \Rightarrow \qquad \dot{\lambda}^w_{it} = 0 \quad \& \quad \lambda^w_{it} = \lambda^w_{it'} = \lambda^w_i \\
\overline{\lambda}^w_{iT} &= \int_0^T e^{-\bar{\rho}_i t}\lambda^w_{it}dt = \frac{1}{\bar{\rho}_i}(1 - e^{-\bar{\rho}_i T})\lambda^w_i
\end{aligned}$$

- Fossil production choice $[e^x_i]$:

$$\overline{\lambda}^w_i (q^f - \mathcal{C}_{e^x}(e^x_i, \mathcal{R}_i)\mathbb{P}_i) = 0$$

- Stock of carbon in the atmosphere $[\mathcal{S}_t]$

$$\lambda^S_{it} = \Delta_i \chi \lambda^T_{it}$$

- Local temperatures $[T_i]$

$$\begin{aligned}
\lambda^T_{it} &= \mathcal{D}^{u'}(T_{it})u(c_i) + \mathcal{D}^{y'}(T_{it})z_i F(k_i, e_i)\lambda^w_i \\
&= -(T_{it} - T^\star_i)(\gamma^c_i c_{it} + \gamma^y_i y_{it})\lambda^w_i \qquad [w/\text{DICE damage fcts + CRRA pref}]
\end{aligned}$$

- Market clearing, energy

$$\sum_{i\in\mathbb{I}} e^x_i = \sum_{i\in\mathbb{I}} \mathcal{P}_i e^f_i \qquad e^c_i = \bar{e}^c_i \qquad e^r_i = \bar{e}^r_i$$

- Market clearing, good i

$$\begin{aligned}
y_i = \overline{\mathcal{D}}_i(\{T_{it}\})z_i F(k_i, e_i) &= \sum_{k\in\mathbb{I}} \tau_{ki} c_{ki} + \sum_{k\in\mathbb{I}} \tau_{ki}(x^f_{ki} + x^c_{ki} + x^r_{ki}) \\
\mathrm{p}_i \underbrace{y_i}_{=\mathcal{D}(T_i)z_i F(\cdot)} &= \sum_{k\in\mathbb{I}} \frac{s_{ki}}{1+\mathrm{t}^b_{ki}} (\mathrm{p}_k y_k + q^f(e^x_k - e^f_k) + \mathrm{t}^{ls}_k)
\end{aligned}$$

B.3.1 Social Cost of Carbon and present discounted value of damages

In Integrated Assessment, we want to measure the present-discounted value of damages – for country i – of one ton of carbon emitted in the atmosphere at time $\mathcal{S}_u$ over all its "lifetime" in the atmosphere $t \in [u, T]$

As a result, with discounting :

$$\begin{aligned}
\lambda^{S,pv}_{iu} &= \int_u^T e^{-\bar{\rho}_i(t-u)}e^{-\delta_s(t-u)}\lambda^S_{it}dt = \int_u^T e^{-(\bar{\rho}_i+\delta_s)(t-u)}\Delta_i\chi\lambda^T_{it}dt \\
&= u(c_i)\Delta_i\chi \int_u^T e^{-(\bar{\rho}_i+\delta_s)(t-u)}\mathcal{D}^{u'}(T_{it})dt + \lambda^w_i z_i F(k_i, e_i)\Delta_i\chi \int_u^T e^{-(\bar{\rho}_i+\delta_s)(t-u)}\mathcal{D}^{y'}(T_{it})dt
\end{aligned}$$

If the cost of carbon is constant (supposing that temperature is stable $T_{it} \to \bar{T}_i$) then the welfare cost of one ton of carbon writes:

$$\lim_{T_{it}\to\bar{T}_i} \lambda_{iu}^{S,pv} = u(c_i)\frac{\chi\Delta_i}{\bar{\rho}_i+\delta_s}(1-e^{-(\bar{\rho}_i+\delta_s)(T-u)})\mathcal{D}^{u'}(\bar{T}_i) + \lambda_i^w z_i F(k_i,e_i)\frac{\chi\Delta_i}{\bar{\rho}_i+\delta_s}\mathcal{D}^{y'}(\bar{T}_i)(1-e^{-(\bar{\rho}_i+\delta_s)(T-u)})$$

$$\lim_{T_{it}\to\bar{T}_i, T\to\infty} \lambda_{iu}^{S,pv} = \bar{\lambda}_i^S = \frac{\chi\Delta_i}{\bar{\rho}_i+\delta_s}\Big(u(c_i)\mathcal{D}^{u'}(\bar{T}_i) + \mathcal{D}^{y'}(\bar{T}_i)\lambda_i^w z_i F(k_i,e_i)\Big)$$

The local cost of carbon LCC_i of emitting one ton $[\varepsilon_{it}]$ summarizes the damages of a ton emitted – per effective capita unit! – at time t by country i. It accounts for the damages occurred between t and T. We measure it in monetary unit by dividing it by marginal value of wealth λ_{it}^w at time t – the time of the emission.

$$\begin{aligned} LCC_{it} &= -\frac{\frac{\partial \mathcal{V}_{it}}{\partial \varepsilon_{it}}}{\frac{\partial \mathcal{V}_{it}}{\partial c_{it}}} = -\frac{1}{\lambda_{it}^w} e^{(n+\bar{g})t}\int_t^T e^{-(\bar{\rho}_i+\delta_s)(s-t)}\mathcal{P}_i\lambda_{is}^S ds \\ &= -e^{(n+\bar{g})t}\frac{\lambda_{it}^{S,pv}}{\lambda_{it}^w} \end{aligned}$$

Now, we were trying to measure the model in stationary form, by taking the present discounted value of the welfare costs and the marginal value of wealth. The stationary local cost of carbon LCC_i writes:

$$LCC_i = -\frac{\overline{\lambda}_i^s}{\overline{\lambda}_i^w} = -\frac{\int_0^T e^{-\bar{\rho}_i t}e^{(n+\bar{g})t}\lambda_{it}^{S,pv}dt}{\int_0^T e^{-\bar{\rho}_i t}\lambda_{it}^w dt}$$

The numerator can be rearranged:

$$\begin{aligned} \int_0^T e^{-\bar{\rho}_i t}\lambda_{it}^{S,pv}dt &= \int_0^T e^{-\bar{\rho}_i t}\int_t^T e^{-\bar{\rho}_i(s-t)}e^{-\delta_s(s-t)}\lambda_{is}^S ds\ dt \\ &= \int_0^T e^{-\bar{\rho}_i t}\int_t^T e^{-(\bar{\rho}_i+\delta_s)(s-t)}\Delta_i\chi\lambda_{is}^T ds\ dt \\ &= \Delta_i\chi\mathcal{P}_i\int_0^T\int_t^T e^{-\bar{\rho}_i s}e^{-\delta_s(s-t)}\lambda_{is}^T ds\ dt \\ &= \Delta_i\chi\mathcal{P}_i\Big[u(c_i)\int_0^T\int_t^T e^{-\bar{\rho}_i s}e^{-\delta_s(s-t)}\mathcal{D}^{u'}(T_{it})ds\ dt + \lambda_i^w z_i F(k_i,e_i)\int_0^T\int_t^T e^{-\bar{\rho}_i s}e^{-\delta_s(s-t)}\mathcal{D}^{y'}(T_{it})ds\ dt\Big] \end{aligned}$$

We see that the "dynamic marginal cost" can be isolated from the other economic variables $y_i, c_i, e_i, \lambda_i^w$. These are the object we will use when considering optimal climate policy.

Example for policy

To give an example for policy, remember that the LCC_i summarize the future cost of climate change:

$$LCC_i = -\frac{\overline{\lambda}_i^s}{\overline{\lambda}_i^w} = -\frac{\int_0^T e^{-\bar{\rho}_i t}e^{(n+\bar{g})t}\lambda_{it}^{S,pv}dt}{\int_0^T e^{-\bar{\rho}_i t}\lambda_{it}^w dt}$$

Suppose one conduct the unilateral climate policy, choosing yearly oil consumption per (effective) capita $[e_i^f]$, internalizing the climate externality ε_u at every period u, and considering that the revenue of the carbon tax is redistributed lump-sum $\mathrm{t}_i^{ls} = \xi^f \mathrm{t}_i^s e_i^f$. The FOC for e_i^f becomes:

$$\begin{aligned} & \mathcal{P}_i \overline{\lambda}_i^w (\xi^f \mathrm{t}_i^s) + \mathcal{P}_i \int_0^T \int_u^T e^{-\bar{\rho}_i t} \lambda_{it}^S e^{-\delta_s(t-u)} e^{(n_i+\bar{g}_i)u} \xi^f dt\ du = 0 \\ \Rightarrow \qquad & \mathrm{t}_i^s = LCC_i \end{aligned}$$

the optimal unilateral carbon tax is the local cost of carbon for country i. This is the standard Pigouvian result and we will see how to conduct the policy at the global level and accounting for redistribution effects and endogenous participation.

B.3.2 Summary: Climate model

Here, we summarize the climate model, and express the present-discounted damages $\overline{\mathcal{D}}(\mathcal{E})$, normalized by discounting $\bar{\rho}_i = \rho - n - (1-\eta)\bar{g}_i$

$$\begin{aligned} \mathcal{S}_t &= \mathcal{S}_0 e^{-\delta_s t} + \int_0^t e^{-\delta_s(t-u)} e^{(n+\bar{g})u} \mathcal{E} du & T_{it} &= T_{it_0} + \Delta_i \chi \mathcal{S}_t \\ \overline{\mathcal{D}}^u(\mathcal{E}) &= \frac{\bar{\rho}_i}{1-e^{-\bar{\rho}_i T}} \int_0^T e^{-\bar{\rho}_i t} \mathcal{D}^u(T_{it}) dt & \overline{\mathcal{D}}^y(\mathcal{E}) &= \frac{\bar{\rho}_i}{1-e^{-\bar{\rho}_i T}} \int_0^T e^{-\bar{\rho}_i t} \mathcal{D}^y(T_{it}) dt \\ \lambda_i^w &= \frac{\bar{\rho}_i}{1-e^{-\bar{\rho}_i T}} \overline{\lambda}_{iT}^w = \frac{\bar{\rho}_i}{1-e^{-\bar{\rho}_i T}} \int_0^T e^{-\bar{\rho}_i t} \lambda_{it}^w dt \end{aligned}$$

$$\overline{\lambda}_{iT}^w LCC_i = \Delta_i \chi \mathcal{P}_i \Big[u(c_i) \int_0^T \int_t^T e^{-\bar{\rho}_i s} e^{-\delta_s(s-t)} \mathcal{D}^{u'}(T_{it}) ds\ dt + \lambda_i^w z_i F(k_i, e_i) \int_0^T \int_t^T e^{-\bar{\rho}_i s} e^{-\delta_s(s-t)} \mathcal{D}^{y'}(T_{it}) ds\ dt \Big]$$

$$\mathcal{E} = \sum_{i \in \mathbb{I}} \mathcal{P}_i (\xi^f e_i^f + \xi^c e_i^c)$$

B.3.3 Summary: Economic model

$$\begin{aligned} u'(c_i) \overline{\mathcal{D}}^u(\mathcal{E}) &= \overline{\lambda}_{it}^w \mathbb{P}_i \\ \mathrm{p}_i MPe_i^f &= q^f + \xi^f \mathrm{t}_i^\varepsilon & \mathrm{p}_i MPe_i^c &= q^c + \xi^c \mathrm{t}_i^\varepsilon \\ \mathrm{p}_i MPe_i^r &= q^r & \mathrm{p}_i MPk_i - \delta &= \bar{r} = \eta \bar{g}_i + \rho \\ q^f &= \mathcal{C}_{e^x}(e_i^x, \mathcal{R}_i) \mathbb{P}_i \\ \sum_{i \in \mathbb{I}} e_i^x &= \sum_{i \in \mathbb{I}} \mathcal{P}_i e_i^f & e_i^c = \bar{e}_i^c &= z_i^c \mathbb{P}_i \qquad e_i^r = \bar{e}_i^r = z_i^r \mathbb{P}_i \\ \overline{\mathcal{D}}_i(\mathcal{E}) z_i F(k_i, e_i) &= \sum_{k \in \mathbb{I}} \tau_{ki} c_{ki} + \sum_{k \in \mathbb{I}} \tau_{ki} (x_{ki}^f + x_{ki}^c + x_{ki}^r + x_{ki}^k) \\ \mathcal{E} &= \sum_i \mathcal{P}_i (\xi^f e_i^f + \xi^c e_i^c) \end{aligned}$$

C Policy

In this section, we provide details on the three policy benchmark, considered in section Section 5.1 and Section 5.2 of the main text. We cover first the optimal allocation when the planner only accounts for resources constraints – the First-Best. Then, we turn to the Ramsey allocation when the planner is constrained and is not allowed cross-countries transfers nor bilateral tariffs, and can only choose carbon taxation. In the last section, we consider unilateral policy, which is a benchmark policy in Nash equilibrium when countries do not cooperate and choose their carbon taxation and trade tariffs to maximize their country's utility.

C.1 First Best

In this allocation, the planner chooses $\boldsymbol{x} = \{c_{ij}, x^{\ell}_{ij}, e^{\ell}_i\}$, i.e. the traded good for consumption c_{ij}, for energy inputs for the production of in fossil x^f_{ij}, coal x^c_{ij}, non-carbon x^r_{ij} or capital x^k_{ij}, and the energy demand, in fossil e^f_i, coal e^c_i and non-carbon e^r_i.

The welfare criterion the planner maximizes is:

$$\mathcal{W} = \sum_i \omega_i \mathcal{P}_i u(\{c_{ij}\}_j)\overline{\mathcal{D}}^u(\mathcal{E})$$

The Planner Lagrangian – in the First-Best allocation – writes:

$$\begin{aligned}\mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda}) = &\sum_i \omega_i \mathcal{P}_i u(\{c_{ij}\}_j)\overline{\mathcal{D}}^u(\mathcal{E}) + \overline{\lambda}\mu^f\big[\sum_{i\in\mathbb{I}} e^x_i - \mathcal{P}_i e^f_i\big] + \sum_{\mathbb{I}} \overline{\lambda}\mu^c_i[\bar{e}^c_i - \mathcal{P}_i e^c_i] + \sum_{\mathbb{I}} \overline{\lambda}\mu^r_i[\bar{e}^r_i - \mathcal{P}_i e^r_i] \\ &+ \sum_{\mathbb{I}} \mathcal{P}_i\omega_i\phi^{\varepsilon}_i\big(\mathcal{E} - \sum_{i\in\mathbb{I}} \mathcal{P}_i(\xi^f e^f_i + \xi^c e^c_i)\big) + \sum_i \omega_i\mu_i\overline{\lambda}\big[\mathcal{P}_i z_i \overline{\mathcal{D}}^y(\mathcal{E})F(\ell_i,k_i,e_i) - \sum_{k\in\mathbb{I}} \mathcal{P}_k\tau_{ki}(c_{ki} + x^f_{ki} + x^c_{ki} + x^r_{ki})\big]\end{aligned}$$

where we rescale the multipliers for the market clearing for good $\overline{\lambda}\mu_i$, for fossil energy $\overline{\lambda}\mu^f_i$, coal energy $\overline{\lambda}\mu^c_i$ and non-carbon energy $\overline{\lambda}\mu^r_i$ by the constant $\overline{\lambda}$ to simplify the comparison with the decentralized equilibrium.

The problem being convex, we write the optimality conditions for each of the controls:

- Consumption

$$[c_{ij}] \qquad \omega_i\mathcal{P}_i u'(c_i)c_i^{1/\theta}a_{ij}^{1/\theta}c_{ij}^{-1/\theta} = \mathcal{P}_i\tau_{ij}\omega_j\mu_j\overline{\lambda}$$

$$c_{ij} = a_{ij}c_i\Big(\tau_{ij}\omega_j\mu_j\frac{\overline{\lambda}}{\omega_i u'(c_i)}\Big)^{-\theta}$$

To get the ideal "price" index, we aggregate:

$$
\begin{aligned}
\Rightarrow \qquad && c_{ij}^{(\theta-1)/\theta} &= [\omega_i u'(c_i)]^{\theta-1} c_i^{(\theta-1)/\theta} a_{ij}^{(\theta-1)/\theta} (\tau_{ij}\omega_j\mu_j\overline{\lambda})^{1-\theta} \\
\Rightarrow \qquad & c_i^{\frac{\theta-1}{\theta}} = \sum_j a_{ij}^{1/\theta} & c_{ij}^{(\theta-1)/\theta} &= [\omega_i u'(c_i)]^{\theta-1} c_i^{(\theta-1)/\theta} \sum_j a_{ij} (\tau_{ij}\omega_j\mu_j\overline{\lambda})^{1-\theta} \\
&& \omega_i u'(c_i) &= \overline{\lambda}\Big[\sum_j a_{ij}(\tau_{ij}\omega_j\mu_j)^{1-\theta}\Big]^{\frac{1}{1-\theta}}
\end{aligned}
$$

- Energy inputs $\bar{e}_i^\ell$ and x_{ij}^ℓ

$$
\begin{aligned}
[x_{ij}] \qquad & \overline{\lambda}\mu_i^\ell g'(x_i^\ell) x_i^{1/\theta} a_{ij}^{1/\theta} x_{ij}^{-1/\theta} &= \tau_{ij}\omega_j\mu_j\overline{\lambda} \\
\Rightarrow \qquad & x_{ij}^{(\theta-1)/\theta} &= [\mu_i^\ell g'(x_i^\ell)]^{\theta-1} (x_i^\ell)^{(\theta-1)/\theta} a_{ij}^{(\theta-1)/\theta} (\tau_{ij}\omega_j\mu_j)^{1-\theta} \\
\Rightarrow \qquad & (x_i^\ell)^{\frac{\theta-1}{\theta}} = \sum_j a_{ij}^{1/\theta} (x_{ij}^\ell)^{(\theta-1)/\theta} &= [\mu_i^\ell]^{\theta-1} (x_i^\ell)^{(\theta-1)/\theta} \sum_j a_{ij}(\tau_{ij}\omega_j\mu_j)^{1-\theta} \\
& \mu_i^\ell g'(x_i^\ell) &= \Big[\sum_j a_{ij}(\tau_{ij}\omega_j\mu_j)^{1-\theta}\Big]^{\frac{1}{1-\theta}}
\end{aligned}
$$

- Energy demand e_i^ℓ

$$
\begin{aligned}
[e_i^f] \qquad & \omega_i \mathcal{P}_i \mu_i \overline{\lambda} MPe_i^f &= \mathcal{P}_i \overline{\lambda}\mu^f + \mathcal{P}_i \xi^f \lambda^{\mathcal{S}} \\
\Rightarrow \qquad & \omega_i \mu_i MPe_i &= \mu^f + \xi^\ell \frac{\phi^{\mathcal{E}}}{\overline{\lambda}} \\
[e_i^c] \qquad & \omega_i \mu_i MPe_i^f &= \mu_i^c + \xi^c \frac{\phi^{\mathcal{E}}}{\overline{\lambda}} \\
[e_i^r] \qquad & \omega_i \mu_i MPe_i^f &= \mu_i^r
\end{aligned}
$$

- Climate damage through carbon emissions $\mathcal{E}$

$$
[\mathcal{E}] \qquad \phi^\varepsilon = \sum_{\mathbb{I}} \mathcal{P}_i \omega_i \phi_i^\varepsilon = -\sum_{\mathbb{I}} \mathcal{P}_i \omega_i \Big[u(c_i)\overline{\mathcal{D}}^{u\prime}(\mathcal{E}) + \overline{\lambda}\mu_i \mathcal{D}_i^{y\prime}(\mathcal{E}) z_i F(e_i, \ell_i)\Big]
$$

Decentralization

We now look at how this planner allocation can be decentralized in the competitive equilibrium.

First, we note the that social cost of carbon is formulated with the multipliers:

$$
SCC = -\frac{\frac{\partial \mathcal{W}}{\partial \mathcal{E}}}{\frac{\partial \mathcal{W}}{c_i}} = \frac{\phi^{\mathcal{E}}}{\overline{\lambda}}
$$

where we recognize that the multiplier $\phi^{\mathcal{E}}$ is the welfare value of one additional ton of carbon (the welfare cost comes from the minus sign), and $\overline{\lambda}$ the average marginal utility of consumption – or marginal value of wealth.

Indeed, the First-Best allocation equalizes marginal utilities through the condition:

$$\overline{\lambda} = \frac{\omega_i u(c_i)\overline{\mathcal{D}}_i^u(\mathcal{E})}{\mathbb{P}_i} = \frac{\omega_j u(c_j)\overline{\mathcal{D}}_j^u(\mathcal{E})}{\mathbb{P}_j} \qquad \forall\ i,j \in \mathbb{I}$$

This implies large redistribution, using lump-sum transfers, such that

$$\begin{aligned} c_i &= u'^{-1}\big(\overline{\lambda}\mathbb{P}_i/\overline{\mathcal{D}}_i^u(\mathcal{E})\big), \forall i \in \mathbb{I} \\ &= w_i\ell_i + \pi_i^f + \mathrm{t}_i^{ls} \end{aligned}$$

In that cases, the transfers t_i^{ls} are designed, such that the consumptions are equalized. This implies redistribution, as $\mathrm{t}_i^{ls} < 0$ for some countries and $\mathrm{t}_i^{ls} > 0$ for some other countries.

The price p_i, output subsidy t_i^y (or inputs subsidy) and tariffs t_{ij}^b in the allocation are determined such that the FOC in the goods demand (for consumption and energy inputs) are satisfied:

$$\begin{aligned} (1+\mathrm{t}_i^y)\mathrm{p}_i &= \omega_i\mu_i \\ (1+\mathrm{t}_{ij}^b)\mathrm{p}_j &= \omega_j\mu_j \\ \mathbb{P}_i &= \big[\sum_j a_{ij}(\tau_{ij}(1+\mathrm{t}_{ij}^b)\mathrm{p}_j)^{1-\theta}\big]^{\frac{1}{1-\theta}} \\ \mathbb{P}_i &= \big[\sum_j a_{ij}(\tau_{ij}\omega_j\mu_j)^{1-\theta}\big]^{\frac{1}{1-\theta}} \end{aligned}$$

A priori, there could be multiple sets of $\{\mathrm{t}_i^y, \mathrm{t}_{ij}^b\}$ such that these conditions are met. It can not be characterized further, because the prices p_i in the Armington model are endogenous objects that depend on the demand and market clearing of each good, and can not be expressed analytically. If the conditions above are satisfied, the energy prices are simply the multipliers :

$$q^f = \mu^f \qquad q_i^c = \mu_i^c \qquad q_i^r = \mu_i^r$$

Finally, the optimal tax is simply the Social Cost of Carbon (SCC)

$$\mathrm{t}^\varepsilon = \frac{\phi^{\mathcal{E}}}{\overline{\lambda}} = -\sum_{\mathbb{I}} \mathcal{P}_i\omega_i\Big[\frac{u(c_i)}{u'(c_i)}\mathbb{P}_i\overline{\mathcal{D}}^{u'}(\mathcal{E}) + \mathcal{D}_i^{y'}(\mathcal{E})z_iF(e_i,\ell_i)\mathrm{p}_i\Big] > 0$$

The optimal carbon tax is the Pigouvian level that summarizes the marginal cost of climate change for all countries i.

We will see now how that results changes when the transfers and other instruments (like tariffs or subsidies) are constrained and prevented to do redistribution.

C.2 Second best: Ramsey policy with constrained instruments

In this allocation, the Ramsey planner again chooses $\boldsymbol{x} = \{c_{ij}, x^\ell_{ij}, e^\ell_i\}$, i.e. the traded good for consumption c_{ij}, for energy inputs for the production of in fossil x^f_{ij}, coal x^c_{ij}, non-carbon x^r_{ij} or capital x^k_{ij}, and the energy demand, in fossil e^f_i, coal e^c_i and non-carbon e^r_i, as well as the carbon tax t^ε and the prices $\boldsymbol{p} = \{\mathrm{p}_i, q^f, q^c_i, q^r_i\}_i$. However, the allocation and prices are constrained to be a competitive equilibrium: in that case, the planner is restricted to choose controls that respect the individual optimality conditions.

We use the same multipliers: $\boldsymbol{\lambda} = \{\mu_i, \mu^c_i, \mu^r_i\}$ and μ^f, ϕ^ε for the market clearing clearing of the final goods, the coal, renewable and fossil energy, and the carbon emissions. We add the constraints that are satisfied in competitive equilibria: λ_i for the budget constraint, ϕ^c for the consumption decision, θ^ℓ_i for the production quantity (supply) choice of energy firms $\ell = f, c, r$ for fossil, coal and renewable of country i, υ^ℓ_i for the quantity (demand) of energy ℓ chosen by the good firm, η_{ij} for the consumption choice for imports j by the household in i, ϑ^ℓ_{ij} for the import choice for inputs from j for the energy firm j. Note that all the multipliers are normalized by ω_i, $\mathcal{P}_i$, and prices or quantity, to simplify optimal policies formulas.

As a result, the controls are $\boldsymbol{x} = \{c_{ij}, x^\ell_{ij}, e^\ell_i, \mathrm{p}_i, q^f, q^c_i, q^r_i, \mathrm{t}^\varepsilon\}_i$ and the multipliers are $\boldsymbol{\lambda} = \{\lambda_i, \mu_i, \mu^c_i, \mu^r_i, \mu^f, \phi^c_i, \theta^\ell_i, \upsilon^\ell_i, \phi^\varepsilon, \eta_{ij}, \vartheta^\ell_{ij}\}_{\ell,i,j}$.

We see that, the Ramsey planner, in choosing t^ε, with other instruments fixed at baseline value t^b_{ij} need to account for many redistributive effects through all the agents decisions.

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = {} & \sum_{\mathbb{I}} \omega_i \mathcal{P}_i u(c_i) \overline{\mathcal{D}}^u_i(\mathcal{E}) + \sum_{\mathbb{I}} \omega_i \mathcal{P}_i \lambda_i \Big(\mathrm{p}_i \overline{\mathcal{D}}^y(\mathcal{E})_i F(\ell_i, k_i, e^f_i, e^c_i, e^r_i) + \frac{1}{\mathcal{P}_i} [q^f g^f(x^\ell_i) - \sum_j x^\ell_{ij} \tau_{ij} \mathrm{p}_j (1+\mathrm{t}^b_{ij})] \\
& + \sum_\ell \{ q^\ell g^\ell(x^\ell_i) - \sum_j x^\ell_{ij} \tau_{ij} \mathrm{p}_j (1+\mathrm{t}^b_{ij}) \} - ((q^f + \xi^f \mathrm{t}^\varepsilon_i) e^f_i + (q^c + \xi^c \mathrm{t}^\varepsilon_i) e^c_i + q^r_i e^r_i + (n_i + \bar{g}_i + \delta) k_i + c_i \mathbb{P}_i + \mathrm{t}^{ls}_i) \Big) \\
& + \sum_{\mathbb{I}} \omega_i \mathrm{p}_i \mu_i \Big(\mathcal{P}_i \overline{\mathcal{D}}^y(\mathcal{E}) z_i F(\ell_i, k_i, e^i) - \sum_{k \in \mathbb{I}} \mathcal{P}_k \tau_{ki} c_{ki} + \sum_{k \in \mathbb{I}} \tau_{ki} (x^f_{ki} + x^c_{ki} + x^r_{ki}) \Big) \\
& + \mu^f q^f [\sum_{i \in \mathbb{I}} e^x_i - \mathcal{P}_i e^f_i] + \sum_{\mathbb{I}} \omega_i \mu^c_i q^c_i (\bar{e}^c_i - \mathcal{P}_i e^c_i) + \sum_{\mathbb{I}} \omega_i \mu^r_i q^r_i (\bar{e}^r_i - \mathcal{P}_i e^r_i) \\
& + \sum_{i \in \mathbb{I}} \omega_i \mathcal{P}_i \phi^\varepsilon_i [\mathcal{E} - \sum_{i \in \mathbb{I}} \mathcal{P}_i (\xi^f e^f_i + \xi^c e^c_i)] + \sum_{\mathbb{I}} \omega_i \mathcal{P}_i \phi^c_i (\mathbb{P}_i \lambda^h_i - u'(c_i) \overline{\mathcal{D}}^u(\mathcal{E})) + \sum_{\ell \in \{f,c,r\}} \sum_{\mathbb{I}} \omega_i \theta^\ell_{it} (\mathbb{P}_{it} - q^\ell_{it} g'(x^\ell_{it})) \\
& + \sum_{i \in \mathbb{I}} \omega_i \mathcal{P}_i \Big(\upsilon^f_i [q^f + \xi^f \mathrm{t}^\varepsilon - \mathrm{p}_i MPe^f_i] + \upsilon^c_i [q^c_i + \xi^c \mathrm{t}^\varepsilon - \mathrm{p}_i MPe^c_i] + \upsilon^r_i [q^r - \mathrm{p}_i MPe^r_i] + \upsilon^k_i [\underbrace{\rho + \eta \bar{g}_i}_{r_i + \bar{g} + n} + \delta - \mathrm{p}_i MPk_i] \Big) \\
& + \sum_{i,j \in \mathbb{I}} \omega_i \mathcal{P}_i \eta_{ij} c_{ij} [(1+\mathrm{t}_{ij}) \tau_{ij} \mathrm{p}_j - \mathbb{P}_i c_i^{\frac{1}{\theta}} a_{ij}^{\frac{1}{\theta}} c_{ij}^{-\frac{1}{\theta}}] \\
& + \sum_{\ell \in \{f,c,r,k\}} \sum_{i,j \in \mathbb{I}} \omega_i \vartheta^\ell_{ij} x^\ell_{ij} [(1+\mathrm{t}_{ij}) \tau_{ij} \mathrm{p}_j - \mathbb{P}_i (x^\ell_i)^{\frac{1}{\theta}} a_{ij}^{\frac{1}{\theta}} (x^\ell_{ij})^{-\frac{1}{\theta}}]
\end{aligned}
$$

Let us go over the optimality conditions of the planner. Note, that – the problem being statics/stationary – the planner does not distort the consumption/saving decision of the household, which implies $\phi_i^c = 0$ – as that can be seen by optimizing over the household marginal value of wealth λ_i^h.

The optimality conditions writes:

- Consumption: c_{ij}

$$\omega_i \mathcal{P}_i u'(c_i) c_i^{\frac{1}{\theta}} a_{ij}^{\frac{1}{\theta}} c_{ij}^{-\frac{1}{\theta}} - \omega_i \mathcal{P}_i \lambda_i \tau_{ij} \mathrm{p}_j - \omega_j \mathcal{P}_i \mu_j \tau_{ij} \mathrm{p}_j + \omega_i \mathcal{P}_i c_{ij} \eta_{ij} \frac{1}{\theta} \frac{\tau_{ij}(1+\mathrm{t}_{ij})\mathrm{p}_j}{c_{ij}} (1 - s_{ij}) = 0$$

$$c_{ij} = a_{ij} c_i \Big((\tau_{ij} \mathrm{p}_j) [1 + \frac{\omega_j}{\omega_i} \frac{\mu_j}{\lambda_i} - \frac{\eta_{ij}}{\lambda_i} \frac{1}{\theta} (1+\mathrm{t}_{ij}^b)(1-s_{ij})] \Big)^{-\theta} \underbrace{\big(\frac{u'(c_i)}{\lambda_i} \big)}_{=\mathbb{P}_i}^{\theta}$$

$$u'(c_i) = \lambda_i \Big(\sum_j a_{ij} (\tau_{ij} \mathrm{p}_j)^{1-\theta} \underbrace{[1 + \frac{\omega_j}{\omega_i} \frac{\mu_j}{\lambda_i} - \frac{\eta_{ij}}{\lambda_i} \frac{1}{\theta} (1+\mathrm{t}_{ij})(1-s_{ij})]^{1-\theta}}_{=1+\mathrm{t}_{ij}} \Big)^{\frac{1}{1-\theta}} = \lambda_i \mathbb{P}_i$$

We see that the the consumption choice c_{ij} is distorted due to (i) the fact that demand for good j change the market clearing of country j, hence with shadow value μ_j, and (ii) the FOC is distorted with value η_{ij}.
If η_{ij} is positive, planner would like to relax the FOC $\mathrm{p}_j(1+\mathrm{t}_{ij}) - u'(c_{ij})$ implying it would like to increase the price.
To give intuition for the good demand distortion, let us give an expression for η_{ij}:

$$\eta_{ij} \frac{1}{\theta} \tau_{ij} (1+\mathrm{t}_{ij}) \mathrm{p}_j (1 - s_{ij}) = u'(c_i) c_i^{\frac{1}{\theta}} a_{ij}^{\frac{1}{\theta}} c_{ij}^{-\frac{1}{\theta}} - \lambda_i \tau_{ij} \mathrm{p}_j - \frac{\omega_j}{\omega_i} \mathcal{P}_i \mu_j \tau_{ij} \mathrm{p}_j$$

$$\Rightarrow \qquad \eta_{ij} = \frac{\theta}{(1-s_{ij})} \Big(\frac{u'(c_i)}{\mathbb{P}_i} \frac{1}{\lambda_i} - \frac{1 + \frac{\omega_j}{\omega_i} \frac{\mu_j}{\lambda_i}}{1+\mathrm{t}_{ij}^b} \Big)$$

The distortion is positive $\eta_{ij} > 0$ for redistributive reasons, related to the budget of i and the market clearing of j. If $u'(c_i)/\mathbb{P}_i > \lambda_i$ and $\mathrm{t}_{ij}^b < \frac{\omega_j}{\omega_i} \frac{\mu_j}{\lambda_i}$, then the planner would like to distort the FOC by increasing the bilateral cost $(1+\mathrm{t}_{ij}^b)\tau_{ij}\mathrm{p}_j$.

If tariffs are set optimally, we have $\eta_{ij} = 0$, – from the above equation – we obtain that $\lambda_i + \frac{\omega_j}{\omega_i} \mu_j = \frac{u'(c_i)}{\mathbb{P}_i}(1+\mathrm{t}_{ij})$ and hence

$$1+\mathrm{t}_{ij}^b = 1 + \frac{\omega_j}{\omega_i} \frac{\mu_j}{\lambda_i}$$

for a hypothetical optimal tariffs on consumption imports. By consequence, we would also obtain naturally that $\frac{u'(c_i)}{\mathbb{P}_i} = \lambda_i$. However, for arbitrary policies t_{ij}^b, the FOC of the household is distorted and $\eta_{ij} \neq 0$.

- Price p_i

$$
\begin{aligned}
&\omega_i \mathcal{P}_i \lambda_i \overline{\mathcal{D}}^y(\mathcal{E}) F(\ell_i, k_i, e_i) - \sum_k \omega_k \mathcal{P}_k \lambda_k \tau_{ki} c_{ki} - \sum_\ell \sum_k \omega_k \mathcal{P}_k \lambda_k \tau_{ki} x_{ki}^\ell \\
&- \sum_{i \in \mathbb{I}} \omega_i \, \mathcal{P}_i \big[\upsilon_i^f MPe_i^f + \upsilon_i^c MPe_i^c + \upsilon_i^r MPe_i^r\big] + \sum_{\ell \in \{f,c,r\}} \sum_k \omega_k \theta_{kt}^\ell \Big(\frac{\tau_{ki}(1+\mathrm{t}_{ki})\mathrm{p}_i}{\mathbb{P}_k} \frac{\partial \mathbb{P}_k}{\partial \mathrm{p}_i}\Big) \\
&+ \sum_k \omega_k \mathcal{P}_k c_{ki} \eta_{ki} \big[\tau_{ki}(1+\mathrm{t}_{ki}) - \frac{\tau_{ki}(1+\mathrm{t}_{ki})\mathrm{p}_i}{\mathbb{P}_k} \frac{\partial \mathbb{P}_k}{\partial \mathrm{p}_i}\big] \\
&+ \sum_{\ell \in \{f,c,r,k\}} \sum_k \omega_k \mathcal{P}_k x_{ki}^\ell \vartheta_{ki}^\ell \tau_{ki}(1+\mathrm{t}_{ki})[1 - s_{ki}] = 0
\end{aligned}
$$

$$
\begin{aligned}
&\omega_i \mathcal{P}_i \lambda_i \overline{\mathcal{D}}^y(\mathcal{E}) F(k_i, e_i) - \sum_k \omega_k \mathcal{P}_k \lambda_k \big[\tau_{ki} c_{ki} + \sum_\ell \tau_{ki} x_{ki}^\ell\big] - \sum_{i \in \mathbb{I}} \omega_i \, \mathcal{P}_i \big[\upsilon_i^f MPe_i^f + \upsilon_i^c MPe_i^c + \upsilon_i^r MPe_i^r + \upsilon_i^k MPe_i^k\big] \\
&+ \sum_{\ell \in \{f,c,r\}} \sum_k \omega_k \theta_{kt}^\ell (\tau_{ki}(1+\mathrm{t}_{ki}) s_{ki}) + \sum_k \omega_k \mathcal{P}_k \tau_{ki}(1+\mathrm{t}_{ki})[1-s_{ki}] \Big(c_{ki}\eta_{ki} + \sum_{\ell \in \{f,c,r,k\}} x_{ki}^\ell \vartheta_{ki}^\ell\Big) = 0
\end{aligned}
$$

This balances out all the redistributive effects: through λ_i on supply from i and on k's demand λ_k, the distortionary effects on energy choice υ^ℓ and energy production θ_i^ℓ, and the distortion on the bilateral import good choice η_{ki} for consumption and ϑ_{ki}^ℓ for inputs in energy inputs.

- Energy inputs: x_{ij}^ℓ, for $\ell = \{f, c, r, k\}$

$$
\begin{aligned}
\omega_i[\lambda_i + \mu_i^\ell] q_i^\ell g'(x_i^\ell)(x_i^\ell)^{\frac{1}{\theta}} a_{ij}^{\frac{1}{\theta}} (x_{ij}^\ell)^{-\frac{1}{\theta}} &- \omega_i \theta_i^\ell q_i^\ell g''(x_i^\ell)(x_i^\ell)^{\frac{1}{\theta}} a_{ij}^{\frac{1}{\theta}} (x_{ij}^\ell)^{-\frac{1}{\theta}} \\
&- \omega_i \lambda_i \tau_{ij} \mathrm{p}_j - \omega_j \mu_j \tau_{ij} \mathrm{p}_j - \omega_i \vartheta_{ij} \frac{1}{\theta} \tau_{ij}(1+\mathrm{t}_{ij})\mathrm{p}_j (s_{ij}-1) = 0 \\
\{\mu_i^\ell - \theta_i^\ell \frac{g''(x_i^\ell)}{g'(x_i^\ell)}\} q_i^\ell g'(x_i^\ell)(x_i^\ell)^{\frac{1}{\theta}} a_{ij}^{\frac{1}{\theta}} (x_{ij}^\ell)^{-\frac{1}{\theta}} &= \tau_{ij} \mathrm{p}_j \big[\mathrm{t}_{ij}^b \lambda_i + \frac{\omega_j}{\omega_i} \mu_j - \vartheta_{ij} \frac{1}{\theta}(1+\mathrm{t}_{ij}^b)(1 - s_{ij})\big] \\
\{\mu_i^\ell - \theta_i^\ell \frac{g''(x_i^\ell)}{g'(x_i^\ell)}\} &= \frac{1}{1+\mathrm{t}_{ij}^b} \big[\mathrm{t}_{ij}^b \lambda_i + \frac{\omega_j}{\omega_i} \mu_j - \vartheta_{ij} \frac{1}{\theta}(1+\mathrm{t}_{ij}^b)(1 - s_{ij})\big]
\end{aligned}
$$

As for the consumption good above, this input choice x_{ij}^ℓ for energy production – which resembles a production networks/supply chain problem – bring additional distortions for each energy price ℓ, i.e. ϑ_{ij}^ℓ, which we can reexpress:

$$
\vartheta_{ij}^\ell = \frac{\theta}{1-s_{ij}} \Big[\{\mu_i^\ell - \theta_i^\ell \frac{g''(x_i^\ell)}{g'(x_i^\ell)}\} - \frac{1 + \frac{\omega_j}{\omega_i} \frac{\mu_j}{\lambda_i}}{1+\mathrm{t}_{ij}^b}\Big]
$$

which resemble the expression for η_{ij} for consumption good. This time the distortion for energy inputs ℓ from j are distorted if the shadow value of the market clearing for that energy sources μ_i^ℓ outweighs the distortion from the supply of that energy θ_i^ℓ – weighted by supply elasticity, related to g''/g', in the case tariffs are such that $\mathrm{t}_{ij}^b < \frac{\omega_j}{\omega_i} \frac{\mu_j}{\lambda_i}$.

Again, in the hypothetical case, where tariffs are set optimally, we obtain $\vartheta_{ij}^{\ell} = 0$ and thus:

$$1+\mathrm{t}_{ij}^{b} = 1 + \frac{\omega_j}{\omega_i}\frac{\mu_j}{\lambda_i}$$

and therefore: $\mu_i^{\ell} = \theta_i^{\ell}\frac{g''(x_i^{\ell})}{g'(x_i^{\ell})}$. However, in the standard case where tariffs are set suboptimally, we have $\vartheta_{ij}^{\ell} \neq 0$.

- Price q_i^{ℓ}, for coal and renewable $\ell = r, c$

$$\omega_i \mathcal{P}_i \lambda_i \Big[\frac{1}{\mathcal{P}_i} g(x_i^{\ell}) - e_i^{\ell}\Big] + \omega_i(\mathcal{P}_i \upsilon_i^{\ell} - g'(x_i^{\ell})\theta_i^{\ell}) = 0 \qquad\qquad \Rightarrow \qquad \mathcal{P}_i \upsilon_i^{\ell} = g'(x_i^{\ell})\theta_i^{\ell}$$

since the market clearing is local at the country level, there are no redistributive effect across countries and the distortion of demand υ_i^{ℓ} equates the distortion of supply θ_i^{ℓ}.

- Price q^f, for oil/gas

$$\sum_{\mathbb{I}} \omega_i \mathcal{P}_i \lambda_i [\frac{1}{\mathcal{P}_i} g(x_i^f) - e_i^f] + \sum_i \omega_i(\mathcal{P}_i \upsilon_i^f - g'(x_i^f)\theta_i^f) = 0$$

At the difference of the FOC for q_i^{ℓ}, for local energy sources, the oil-gas is traded internationally and therefore, changing its price has redistributive effects between countries depending on net-exports $g(x_i^{\ell}) - \mathcal{P}_i e_i^f$, through the covariance between those net-exports and the marginal value of income λ_i:

$$\sum_{\mathbb{I}} \omega_i \mathcal{P}_i \lambda_i [\frac{1}{\mathcal{P}_i} g(x_i^f) - e_i^f] = \mathbb{C}\mathrm{ov}(\omega_i \lambda_i, g(x_i^f) - \mathcal{P}_i e_i^f)$$

- Energy demand e_i^{ℓ}

$$\begin{aligned} \omega_i \mathcal{P}_i \lambda_i (\mathrm{p}_i MPe_i^{\ell} - q^{\ell}) + \omega_i \mathcal{P}_i \mu_i \mathrm{p}_i MPe_i^{\ell} - \omega_i q_i^{\ell} \mu_i^{\ell} \mathcal{P}_i - \phi^{\varepsilon} \mathcal{P}_i \xi^{\ell} \\ - \sum_{\ell'} \omega_i \mathcal{P}_i \mathrm{p}_i \upsilon_i^{\ell'} \partial_{e_i^{\ell}} MPe_i^{\ell'} = 0 \end{aligned}$$

We see that the energy demand choice by the planner internalize multiple effects that will be key in the formulation of the carbon tax: First it internalizes the climate externality, as summarized by the multiplier ϕ^{ε}. Second, it also accounts for the redistributive effect through the change on the energy market clearing μ_i^{ℓ} for that particular energy source. Third, it also distorts the FOC of the firm in all its energy and input sourcing ℓ', as summarized by the multipliers $\upsilon_i^{\ell'}$, and weighted by the terms $\partial_{e_i^{\ell}} MPe_i^{\ell'}$ which relates to the cross elasticity between energies ℓ and ℓ'. Moreover, it internalizes the effects that energy use has on good production, through multiplier μ_i. All these effects are detailed in more details below.
Let us more specific about each energy sources.

Fossil:

$$\omega_i\lambda_i\xi^f \mathrm{t}^\varepsilon + \omega_i\mu_i \mathrm{p}_i MPe_i^f - q^f\mu^f - \phi^\varepsilon\xi^f - \sum_{\ell'}\omega_i \mathrm{p}_i \upsilon_i^{\ell'}\partial_{e_i^f} MPe_i^{\ell'} = 0$$

Coal:

$$\omega_i\lambda_i\xi^c \mathrm{t}^\varepsilon + \omega_i\mu_i \mathrm{p}_i MPe_i^c - \omega_i q_i^c\mu_i^c - \phi^\varepsilon\xi^c - \sum_{\ell'}\omega_i \mathrm{p}_i \upsilon_i^{\ell'}\partial_{e_i^c} MPe_i^{\ell'} = 0$$

Renewable / non-carbon

$$\omega_i\mu_i \mathrm{p}_i MPe_i^r - \omega_i q_i^r\mu^f - \sum_{\ell'}\omega_i \mathrm{p}_i \upsilon_i^{\ell'}\partial_{e_i^r} MPe_i^{\ell'} = 0$$

- Carbon tax t_i^ε:

  $$\sum_{\mathbb{I}}\omega_i \mathcal{P}_i(\upsilon_i^f\xi^f + \upsilon_i^c\xi^c) = 0$$

  The choice of the optimal carbon tax is a uniform tax that does not impose any additional aggregate distortion on the world economy. As a result, the sum of the country-levels distortions sum to zero: a positive distortion – multiplier $\upsilon_i^f > 0$ – need to be compensated by a negative distortion $\upsilon_{i'}^c < 0$ for another country, or across energy sources.

- Climate damage:

  $$\begin{aligned}\phi^\varepsilon &= \sum_{\mathbb{I}}\mathcal{P}_i\omega_i\phi_i^\varepsilon \\ &= -\sum_{\mathbb{I}}\mathcal{P}_i\omega_i\Big[u(c_i)\overline{\mathcal{D}}^{u\prime}(\mathcal{E}) + (\lambda_i+\mu_i)\mathcal{D}_i^{y\prime}(\mathcal{E})\mathrm{p}_i z_i F(e_i,\ell_i,k_i) - \mathcal{D}_i^{y\prime}(\mathcal{E})\mathrm{p}_i\sum_{\ell'}\upsilon_i^{\ell'}MPe_i^{\ell'}\Big]\end{aligned}$$

  The marginal cost of climate change can be summarized by ϕ^ε and it internalizes the direct cost $\mathcal{D}_i^{y\prime}(\mathcal{E})$ and $\mathcal{D}_i^{u\prime}(\mathcal{E})$ of climate change on income – hence the multiplier λ_i – but also the effects of climate on good production μ_i and on the distortion of the energy demand optimality υ_i^ℓ.

Reformulation of the carbon tax

We take the example of the carbon tax on fossil fuels (oil-gas) to provide details the formulation of the tax:

$$\omega_i\lambda_i\xi^f \mathrm{t}^\varepsilon + \omega_i\mu_i \mathrm{p}_i MPe_i^f - q^f\mu^f - \phi^\varepsilon\xi^f - \sum_{\ell'}\omega_i \mathrm{p}_i \upsilon_i^{\ell'}\partial_{e_i^f} MPe_i^{\ell'} = 0$$

$$\underbrace{\sum_i\omega_i\widehat{\mathcal{P}}_i\lambda_i}_{=\overline{\lambda}}\,\xi^f\mathrm{t}^\varepsilon = \xi^f\phi^\varepsilon + q^f\mu^f - (q^f+\xi^f\mathrm{t}^\varepsilon)\sum_i\omega_i\widehat{\mathcal{P}}_i\mu_i + \sum_{\ell'}\sum_i\omega_i\widehat{\mathcal{P}}_i\upsilon_i^{\ell'}\;q_i^{\ell'}\frac{\partial_{e_i^f}MPe_i^{\ell'}}{MPe_i^{\ell'}}$$

Which gives, when aggregating over all countries i and rescaling the multipliers for the good market clearing $\widehat{\mu}_i$, the ones for energy $e_i^{\ell'}$ distortion $\widehat{\upsilon}_i^{\ell'}$, a formula for the carbon tax:

$$\xi^f \mathrm{t}^\varepsilon = \xi^f \underbrace{\frac{\phi^\varepsilon}{\overline{\lambda}}}_{=\text{SCC}} + \underbrace{q^f \frac{\mu^f}{\overline{\lambda}}}_{\substack{E^f \text{ supply} \\ \text{redistribut°}}} -(q^f + \xi^f \mathrm{t}^\varepsilon) \sum_i \underbrace{\widehat{\mu}_i}_{\substack{y_i \text{Trade} \\ \text{redistribut°}}} - \sum_{\ell'} \sum_i \underbrace{\widehat{\upsilon}_i^{\ell'} \; q_i^{\ell'} \frac{\partial_{e_i^f} MPe_i^{\ell'}}{MPe_i^{\ell'}}}_{\substack{e_i^{\ell'} \text{ demand} \\ \text{distort°}}}$$

We now use the functional forms assumptions in our model to simplify this formula further.

Simplifying the formula

Assumptions, for formula in the paper

- Rewriting the social cost of carbon (SCC):

$$SCC = \frac{\phi^\varepsilon}{\overline{\lambda}} = \frac{\mathcal{P} \sum_i \omega_i \widehat{\mathcal{P}}_i \phi_i^\varepsilon}{\sum_i \omega_i \widehat{\mathcal{P}}_i \lambda_i} = \mathcal{P} \sum_i \frac{\omega_i \widehat{\mathcal{P}}_i \lambda_i}{\sum_i \omega_i \widehat{\mathcal{P}}_i \lambda_i} \frac{\phi_i^\varepsilon}{\lambda_i} = \mathcal{P} \sum_i \frac{\omega_i \widehat{\mathcal{P}}_i \lambda_i}{\sum_i \omega_i \widehat{\mathcal{P}}_i \lambda_i} LCC_i$$

$$\phi_i^\varepsilon = -\Big[u(c_i) \overline{\mathcal{D}}^{u\prime}(\mathcal{E}) + (\lambda_i + \mu_i) \mathcal{D}_i^{y\prime}(\mathcal{E}) \mathrm{p}_i z_i F(e_i, \ell_i) \dots \Big]$$

$$\text{with} \qquad \lambda_i = \frac{u'(c_i) \overline{\mathcal{D}}^u(\mathcal{E})}{\mathbb{P}_i} \qquad\qquad \text{and CRRA} \qquad u'(c_i) = \frac{c_i^{1-\eta}}{1-\eta}$$

$$\text{damage} \qquad \overline{\mathcal{D}}^u(\mathcal{E}) = \Big(\widetilde{\overline{\mathcal{D}}^{u}}{}'(\mathcal{E}) \Big)^{1-\eta} \qquad\qquad \overline{\mathcal{D}}^{u\prime}(\mathcal{E}) = (1-\eta) (\widetilde{\overline{\mathcal{D}}^{u}}(\mathcal{E}))^{-\eta} \widetilde{\overline{\mathcal{D}}^{u}}{}'(\mathcal{E})$$

$$\frac{u(c_i) \overline{\mathcal{D}}^{u\prime}(\mathcal{E})}{u'(c_i) \overline{\mathcal{D}}^u(\mathcal{E})} = \frac{c_i^{1-\eta}}{1-\eta} \frac{1}{\widetilde{\overline{\mathcal{D}}^{u}}(\mathcal{E})^{1-\eta}} (1-\eta) (\widetilde{\overline{\mathcal{D}}^{u}}(\mathcal{E}))^{-\eta} \widetilde{\overline{\mathcal{D}}^{u}}{}'(\mathcal{E}) = \frac{c_i \widetilde{\overline{\mathcal{D}}^{u}}{}'(\mathcal{E})}{\widetilde{\overline{\mathcal{D}}^{u}}(\mathcal{E})}$$

$$\frac{\phi_i^\varepsilon}{\overline{\lambda}} = -\Big[\frac{\lambda_i \mathbb{P}_i c_i}{\overline{\lambda}} \frac{\widetilde{\overline{\mathcal{D}}^{u}}{}'(\mathcal{E})}{\widetilde{\overline{\mathcal{D}}^{u}}(\mathcal{E})} + \frac{\lambda_i + \mu_i}{\overline{\lambda}} \frac{\overline{\mathcal{D}}^{y\prime}(\mathcal{E})}{\overline{\mathcal{D}}^y(\mathcal{E})} \mathrm{p}_i \mathcal{D}_i^y(\mathcal{E}) z_i F(e_i, \ell_i) \dots \Big]$$

- Nordhaus DICE quadratic damage function and simple climate system: c.f. above.

$$\mathcal{D}^y(T - T^\star) = e^{-\frac{\gamma^y}{2}(T - T_i^\star)^2} \qquad \Rightarrow \qquad \mathcal{D}_i^{y\prime}(T - T^\star) = -\mathcal{D}_i^y(T - T^\star) \gamma^y (T - T_i^\star)$$

$$\dot{\mathcal{S}}_t = \mathcal{E} - \delta_s \mathcal{S}_t$$

$$T_{it} = T_{it_0} + \Delta\chi \mathcal{S}_t$$

$$\frac{\overline{\mathcal{D}}^{y\prime}(\mathcal{E})}{\overline{\mathcal{D}}^y(\mathcal{E})} \to_{t \to \infty, T_{it} \to T_i} -\frac{\Delta\chi}{\rho - n + (1-\eta)\bar{g} + \delta_s} \gamma^y (T_i - T_i^\star)$$

$$\frac{\phi_i^\varepsilon}{\overline{\lambda}} = \frac{\Delta\chi (T_i - T_i^\star)}{\rho - n + (1-\eta)\bar{g} + \delta_s} \Big[\frac{\lambda_i}{\overline{\lambda}} \mathbb{P}_i c_i \gamma^c + \frac{\lambda_i + \mu_i}{\overline{\lambda}} \mathrm{p}_i y_i \gamma^y - \gamma^y \big(\frac{\upsilon_i^f}{\overline{\lambda}} (q^f + \xi^f \mathrm{t}^\varepsilon) + \frac{\upsilon_i^c}{\overline{\lambda}} (q_i^c + \xi^c \mathrm{t}^\varepsilon) + \frac{\upsilon_i^r}{\overline{\lambda}} q_i^r + \frac{\upsilon_i^k}{\overline{\lambda}} (\rho + \eta \bar{g}) \big) \Big]$$

The Local cost of carbon, for country i if $\omega_i = 1, \omega_j = 0$

$$LCC_i = \frac{\phi_i^\varepsilon}{\lambda_i} = \frac{\Delta\chi (T_i - T_i^\star)}{\rho - n + (1-\eta)\bar{g} + \delta_s} \Big[\mathbb{P}_i c_i \gamma^c + (1 + \frac{\mu_i}{\lambda_i}) \mathrm{p}_i y_i \gamma^y - \gamma^y \big(\frac{\upsilon_i^f}{\lambda_i} (q^f + \xi^f \mathrm{t}^\varepsilon) + \frac{\upsilon_i^c}{\lambda_i} (q_i^c + \xi^c \mathrm{t}^\varepsilon) + \frac{\upsilon_i^r}{\lambda_i} q_i^r + \frac{\upsilon_i^k}{\lambda_i} (\rho + \eta \bar{g}) \big) \Big]$$

Reexpressing the total global social cost of carbon:

$$\begin{aligned}
SCC &= \mathcal{P}\sum_i \omega_i \widehat{\mathcal{P}}_i \frac{\phi_i^\varepsilon}{\bar{\lambda}} = \mathcal{P}\sum_i \widehat{\lambda}_i LCC_i \\
&= \mathcal{P}\sum_i \frac{\Delta\chi(T_i - T_i^\star)}{\rho - n + (1-\eta)\bar{g} + \delta_s}\Big[\widehat{\lambda}_i \mathbb{P}_i c_i \gamma^c + (\widehat{\lambda}_i + \widehat{\mu}_i)\mathrm{p}_i y_i \gamma^y \\
&\qquad - \gamma^y\big(\widehat{v}_i^f(q^f + \xi^f \mathrm{t}^\varepsilon) + \widehat{v}_i^c(q_i^c + \xi^c \mathrm{t}^\varepsilon) + \widehat{v}_i^r q_i^r + \widehat{v}_i^k(\rho + \eta\bar{g})\big)\Big]
\end{aligned}$$

with the rescaled multipliers for the budget constraint: $\widehat{\lambda}_i = \frac{\omega_i \widehat{\mathcal{P}}_i \lambda_i}{\bar{\lambda}} = \frac{\omega_i \widehat{\mathcal{P}}_i \lambda_i}{\sum_i \omega_i \widehat{\mathcal{P}}_i \lambda_i}$, the multiplier the FOC demand $\widehat{v}_i = \frac{\omega_i \widehat{\mathcal{P}}_i v_i}{\bar{\lambda}}$, for the multiplier for market clearing for good: $\widehat{\mu}_i = \frac{\omega_i \widehat{\mathcal{P}}_i \mu_i}{\bar{\lambda}}$ and population share $\widehat{\mathcal{P}}_i = \frac{\mathcal{P}_i}{\mathcal{P}}$

- Isoelastic energy supply curve: $x^f = \frac{\bar{\nu}_i}{1+\nu_i}\Big(\frac{e_i^x}{\mathcal{R}_i}\Big)^{1+\nu_i}\mathcal{R}_i$

$$g_i(x) = \mathcal{R}_i^{\frac{\nu_i}{1+\nu_i}} x^{\frac{1}{1+\nu_i}} \qquad\qquad \frac{g''(x^f)}{g'(x^f)} = -\frac{\nu_i}{1+\nu_i}\frac{1}{x_i}$$

As a result, FOC of energy inputs $[x_i^f]$ becomes:

$$\mu^f = \omega_i \theta_i^f \frac{g''(x^f)}{g'(x^f)} = \omega_i \theta_i^f \frac{-\nu_i}{1+\nu_i}\frac{1}{x_i^f} \qquad\qquad \omega_i \theta_i^f = -\frac{1+\nu_i}{\nu_i} x_i^f \mu^f$$

Moreover, the FOC of energy price $[q^f]$ becomes:

$$\begin{aligned}
\sum_i \omega_i \theta_i^f g'(x^f) &= -\mu^f \sum_i \frac{1+\nu_i}{\nu_i} x_i^f g'(x_i^f) \\
&= -\mu^f \sum_i \frac{1+\nu_i}{\nu_i}\frac{\bar{\nu}_i}{1+\nu_i}\Big(\frac{e_i^x}{\mathcal{R}_i}\Big)^{1+\nu_i}\mathcal{R}_i\Big(\frac{e_i^x}{\mathcal{R}_i}\Big)^{-\nu_i}\bar{\nu}_i^{-1} \\
\sum_i \omega_i \theta_i^f g'(x^f) &= -\mu^f \sum_i \frac{e_i^x}{\nu_i} = -\mu^f / \Big(\frac{1}{E^f}\underbrace{\big(\sum_i \frac{e_i^x}{E^f}\frac{1}{\nu_i}\big)^{-1}}_{=\bar{\nu}}\Big)
\end{aligned}$$

As a result, we have, with the aggregate elasticity $\bar{\nu}$

$$\begin{aligned}
\mu^f &= \frac{\bar{\nu}}{E^f}\sum_i \omega_i \mathcal{P}_i \lambda_i \big(e_i^f - \frac{e_i^x}{\mathcal{P}_i}\big) \qquad\qquad \bar{\nu} = \big(\sum_i \tfrac{e_i^x}{E^f}\nu_i^{-1}\big)^{-1} \\
\frac{\mu^f}{\bar{\lambda}} &= \mathcal{P}\frac{\bar{\nu}}{E^f}\sum_i \frac{\omega_i \widehat{\mathcal{P}}_i \lambda_i}{\bar{\lambda}}\big(e_i^f - \frac{e_i^x}{\mathcal{P}_i}\big)
\end{aligned}$$

- Nested CES framework:

$$\text{Energy} \quad e_i = \Big(\sum_\ell (\omega^\ell)^{\frac{1}{\sigma_e}}(e_i^\ell)^{\frac{\sigma_e - 1}{\sigma_e}}\Big)^{\frac{\sigma_e}{\sigma_e - 1}} \qquad\qquad \text{Output} \quad y_i = \Big((1-\varepsilon)^{\frac{1}{\sigma}}(e_i)^{\frac{\sigma-1}{\sigma}} + \varepsilon^{\frac{1}{\sigma}}(k_i^\alpha \ell_i^{1-\alpha})^{\frac{\sigma-1}{\sigma}}\Big)^{\frac{\sigma}{\sigma-1}}$$

FOC for fossil energy demand:

$$\begin{aligned}
\bar{\upsilon}_i^f &= [\upsilon_i^f \partial_{e^f} MPe_i^f + \upsilon_i^c \partial_{e^f} MPe_i^c + \upsilon_i^r \partial_{e^f} MPe_i^r + \upsilon_i^k \partial_{e^f} MPk_i] \\
&= \frac{1}{e_i^f}\Big[- \upsilon_i^f (q^f + \xi^f \mathrm{t}^\varepsilon)[\frac{1-s^f}{\sigma^e} + s^f \frac{1-s^e}{\sigma^y}] + \upsilon_i^c (q_i^c + \xi^f \mathrm{t}^\varepsilon) s_i^f [\frac{1}{\sigma^e} - \frac{1-s^e}{\sigma^y}] + \upsilon_i^r q_i^r s_i^f [\frac{1}{\sigma^e} - \frac{1-s^e}{\sigma^y}] + \upsilon_i^k (r^\star + \bar{\delta}) \frac{s_i^{e^r/y}}{\sigma^y}\Big]
\end{aligned}$$

and when normalizing by $\overline{\lambda}$

$$\widehat{\upsilon}_i^f = \frac{1}{e_i^f}\Big[- \widehat{\upsilon}_i^f (q^f + \xi^f \mathrm{t}^\varepsilon)[\frac{1-s^f}{\sigma^e} + s^f \frac{1-s^e}{\sigma^y}] + \widehat{\upsilon}_i^c (q_i^c + \xi^f \mathrm{t}^\varepsilon) s_i^f [\frac{1}{\sigma^e} - \frac{1-s^e}{\sigma^y}] + \widehat{\upsilon}_i^r q_i^r s_i^f [\frac{1}{\sigma^e} - \frac{1-s^e}{\sigma^y}] + \widehat{\upsilon}_i^k (r^\star + \bar{\delta}) \frac{s_i^{e^r/y}}{\sigma^y}\Big]$$

If the production function only contains fossil (oil-gas) in energy $\omega^f = 1$, then we obtain:

$$\bar{\upsilon}_i^f = -\frac{q^f + \xi^f \mathrm{t}^\varepsilon}{e_i^f} \upsilon_i^f [\frac{1-s^e}{\sigma^y}] \qquad\qquad s_i^e = \frac{q^e e_i}{\mathrm{p}_i y_i}$$

Proposition: Using these assumptions, we can reexpress the carbon tax:

$$\begin{aligned}
\xi^f \mathrm{t}^\varepsilon &= \xi^f \mathcal{P} \sum_i \widehat{\lambda}_i LCC_i + q^f \mathcal{P} \frac{\bar{\nu}}{E^f} \sum_i \widehat{\lambda}_i (e_i^f - \frac{e_i^x}{\mathcal{P}_i}) - (q^f + \xi^f \mathrm{t}^\varepsilon) \sum_i \widehat{\mu}_i - \sum_i \widehat{\bar{\upsilon}}_i^f \\
\xi^f \mathrm{t}^\varepsilon &= \xi^f \underbrace{\mathcal{P}\mathbb{E}_i[LCC_i] + \mathcal{P}\mathbb{C}\mathrm{ov}_i(\widehat{\lambda}_i, LCC_i)}_{=\text{Social Cost of Carbon}} + \underbrace{q^f \mathcal{P} \frac{\bar{\nu}}{E^f} \mathbb{C}\mathrm{ov}_i(\widehat{\lambda}_i, e_i^f - \frac{e_i^x}{\mathcal{P}_i})}_{\substack{E^f \text{ supply} \\ \text{redistribut}^\circ}} - (q^f + \xi^f \mathrm{t}^\varepsilon) \underbrace{\mathbb{E}_i[\widehat{\mu}_i]}_{\substack{y_i \text{Trade} \\ \text{redistribut}^\circ}} - \underbrace{\mathbb{E}_i[\widehat{\bar{\upsilon}}_i^f]}_{\substack{e_i^{\ell'} \text{ demand} \\ \text{distort}^\circ}}
\end{aligned}$$

With the demand distortion of fossil fuels $\widehat{\bar{\upsilon}}_i^f$

$$\begin{aligned}
\widehat{\bar{\upsilon}}_i^f &= \frac{1}{e_i^f}\Big[- \widehat{\upsilon}_i^f (q^f + \xi^f \mathrm{t}^\varepsilon)[\frac{1-s^f}{\sigma^e} + s^f \frac{1-s^e}{\sigma^y}] + \widehat{\upsilon}_i^c (q_i^c + \xi^f \mathrm{t}^\varepsilon) s_i^f [\frac{1}{\sigma^e} - \frac{1-s^e}{\sigma^y}] \\
&\qquad + \widehat{\upsilon}_i^r q_i^r s_i^f [\frac{1}{\sigma^e} - \frac{1-s^e}{\sigma^y}] + \widehat{\upsilon}_i^k (r^\star + \bar{\delta}) \frac{s_i^{e^r/y}}{\sigma^y}\Big] \\
\widehat{\bar{\upsilon}}_i^f &= -\frac{q^f + \xi^f \mathrm{t}^\varepsilon}{e_i^f} \widehat{\upsilon}_i^f [\frac{1-s_i^e}{\sigma^y}] \qquad\qquad \text{if } \ s^f = 1, s^r = s^c = 0
\end{aligned}$$

and the social cost of carbon SCC as:

$$\begin{aligned}
SCC &= \mathcal{P} \sum_i \widehat{\lambda}_i LCC_i = \mathcal{P}\mathbb{E}_i[LCC_i] + \mathcal{P}\mathbb{C}\mathrm{ov}_i(\widehat{\lambda}_i, LCC_i) \\
&= \mathcal{P} \sum_i \frac{\Delta\chi(T_i - T_i^\star)}{\rho - n + (1-\eta)\bar{g} + \delta_s} \Big[\widehat{\lambda}_i \mathbb{P}_i c_i \gamma^c + (\widehat{\lambda}_i + \widehat{\mu}_i) \mathrm{p}_i y_i \gamma^y \\
&\qquad - \gamma^y \big(\widehat{\upsilon}_i^f (q^f + \xi^f \mathrm{t}^\varepsilon) + \widehat{\upsilon}_i^c (q_i^c + \xi^c \mathrm{t}^\varepsilon) + \widehat{\upsilon}_i^r q_i^r + \widehat{\upsilon}_i^k (\rho + \eta \bar{g})\big)\Big]
\end{aligned}$$

with the rescaled multipliers for the budget constraint: $\widehat{\lambda}_i = \frac{\omega_i \widehat{\mathcal{P}}_i \lambda_i}{\overline{\lambda}} = \frac{\omega_i \widehat{\mathcal{P}}_i \lambda_i}{\sum_i \omega_i \widehat{\mathcal{P}}_i \lambda_i}$, the multiplier the FOC demand $\widehat{\upsilon}_i = \frac{\omega_i \widehat{\mathcal{P}}_i \upsilon_i}{\overline{\lambda}}$, for the multiplier for market clearing for good: $\widehat{\mu}_i = \frac{\omega_i \widehat{\mathcal{P}}_i \mu_i}{\overline{\lambda}}$

Simplifying the multiplier for the FOC for energy demand

$$\xi^f \mathrm{t}^\varepsilon = \xi^f \mathcal{P}\Big(\mathbb{E}_j[LCC_j] + \mathbb{C}\mathrm{ov}_j(\widehat{\lambda}_j, LCC_j)\Big) + \mathcal{P}\frac{q^f \bar{\nu}}{E}\mathbb{C}\mathrm{ov}_j(\widehat{\lambda}_j, e^f_j - \frac{e^x_j}{\mathcal{P}_j})$$
$$- (q^f + \xi^f \mathrm{t}^\varepsilon)\mathbb{E}_j[\widehat{\mu}_j] - (q^f + \xi^f \mathrm{t}^\varepsilon)\mathbb{C}\mathrm{ov}(\widehat{\upsilon}^f_i, \frac{1-s^e}{\sigma e^f_i})$$

where the last equality comes from the fact that $\mathbb{E}_i[\widehat{\upsilon}^f_i] = \sum_i \widehat{\upsilon}^f_i = 0$ by the assumption that there is no aggregate distortion – only individual distortion – when the uniform carbon tax is set at the world level and $\mathbb{E}_i[e^f_i - \frac{e^x_i}{\mathcal{P}_i}] = \sum_i e^f_i - \frac{e^x_i}{\mathcal{P}_i} = 0$ by market clearing on the fossil energy market.

To investigate the demand distortion $\widehat{\upsilon}^f_i$ further, we can see that, with the planner's FOC for energy demand, the individual distortion becomes:

$$\omega_i \mathcal{P}_i \upsilon^f_i = \frac{\mathcal{P}_i}{q^f + \xi^f \mathrm{t}^\varepsilon} \frac{\sigma^y e^f_i}{1-s^e_i}\Big[\xi^f \phi^S + q^f \mu^f - \omega_i \mu_i (q^f + \xi^f \mathrm{t}^\varepsilon) - \omega_i \lambda_i \xi^f \mathrm{t}^\varepsilon\Big]$$

The distortion is higher if the welfare motives in terms of cost of climate change ϕ^S, supply redistribution μ^f and trade effect $-\mu_i$ etc. outweighs the welfare cost of the carbon tax $\lambda_i \mathrm{t}^\varepsilon$. As mentioned earlier, the averate/aggregate distortion is null, so – in the case where fossil (oil-gas) is the only energy – we obtain:

$$\sum_i \omega_i \mathcal{P}_i \upsilon^f_i = 0$$

$$\frac{1}{q^f + \xi^f \mathrm{t}^\varepsilon}(\xi^f \phi^S + q^f \mu^f)\sum_i \mathcal{P}_i \frac{\sigma^y e^f_i}{1-s^e_i} - \frac{1}{q^f + \xi^f \mathrm{t}^\varepsilon}\xi^f \mathrm{t}^\varepsilon \sum_i \omega_i \mathcal{P}_i \lambda_i \frac{\sigma^y e^f_i}{1-s^e_i} - \frac{q^f + \xi^f \mathrm{t}^\varepsilon}{q^f + \xi^f \mathrm{t}^\varepsilon}\sum_i \omega_i \mathcal{P}_i \mu_i \frac{\sigma^y e^f_i}{1-s^e_i} = 0$$

$$\xi^f \mathrm{t}^\varepsilon \sum_i \omega_i \mathcal{P}_i \lambda_i \frac{\sigma^y e^f_i}{1-s^e_i} = (\xi^f \phi^S + q^f \mu^f)\sum_i \mathcal{P}_i \frac{\sigma^y e^f_i}{1-s^e_i} - (q^f + \xi^f \mathrm{t}^\varepsilon)\sum_i \omega_i \mathcal{P}_i \mu_i \frac{\sigma^y e^f_i}{1-s^e_i}$$

Divide both side by $\overline{\lambda} = \sum_i \omega_i \widehat{\mathcal{P}}_i \lambda_i$ and by $E^{s,\sigma} = \sum_i \widehat{\mathcal{P}}_i \frac{\sigma^y e^f_i}{1-s^e_i}$ and $\widehat{e}^{s,\sigma}_i = \frac{\frac{\sigma^y e^f_i}{1-s^e_i}}{\sum_i \widehat{\mathcal{P}}_i \frac{\sigma^y e^f_i}{1-s^e_i}}$, it becomes;

$$\xi^f \mathrm{t}^\varepsilon \sum_i \frac{\omega_i \widehat{\mathcal{P}}_i \lambda_i}{\overline{\lambda}} \frac{1}{E^{s,\sigma}} \frac{\sigma^y e^f_i}{1-s^e_i} = (\xi^f \frac{\phi^S}{\overline{\lambda}} + q^f \frac{\mu^f}{\overline{\lambda}})\frac{1}{E^{s,\sigma}}\sum_i \widehat{\mathcal{P}}_i \frac{\sigma^y e^f_i}{1-s^e_i} - (q^f + \xi^f \mathrm{t}^\varepsilon)\sum_i \frac{\omega_i \widehat{\mathcal{P}}_i \mu_i}{\overline{\lambda}} \frac{1}{E^{s,\sigma}} \frac{\sigma^y e^f_i}{1-s^e_i}$$

$$\xi^f \mathrm{t}^\varepsilon = \frac{1}{1 + \mathbb{C}\mathrm{ov}_j(\widehat{\lambda}_j, \widehat{e}^{s,\sigma}_i)}\Big[\xi^f \mathcal{P} SCC + \mathcal{P}\frac{q^f \bar{\nu}}{E}\mathbb{C}\mathrm{ov}_j(\widehat{\lambda}_j, e^f_j - \frac{e^x_j}{\mathcal{P}_j}) - (q^f + \xi^f \mathrm{t}^\varepsilon)\mathbb{C}\mathrm{ov}_j(\widehat{\mu}_j, \widehat{e}^{s,\sigma}_i)\Big]$$

It implies that the carbon tax is dampened if the Planner put larger social weights on countries that use a lot of energy e^f_i, with a higher elasticity σ^y, and as a larger share of production s^e_i, resulting in $\mathbb{C}\mathrm{ov}_j(\widehat{\lambda}_j, \widehat{e}^{s,\sigma}_i) > 0$. If this covariance is negative, then the carbon tax is amplified and the planner would optimally choose a higher carbon tax.

C.3 Unilaleral policy

In this allocation, the Ramsey planner now maximizes country i's welfare $\mathcal{P}_i\mathcal{U}_i$, choosing the allocation in country i. It chooses $\boldsymbol{x}_i = \{c_{ij}, x^\ell_{ij}, e^\ell_i\}$, i.e. the traded good for consumption c_{ij}, for energy inputs for the production of in fossil x^f_{ij}, coal x^c_{ij}, non-carbon x^r_{ij} or capital x^k_{ij}, and the energy demand, in fossil e^f_i, coal e^c_i and non-carbon e^r_i, and the prices $\boldsymbol{p}_i = \{\mathrm{p}_i, q^f, q^c_i, q^r_i\}$. For the policy instruments, they choose the country i carbon tax t^ε_i as well as the set of trade tariffs $\{\mathrm{t}^b_{ij}\}_j$ against country j. Moreover, we consider the Nash equilibrium, and the planner i take as given the policies of the other countries j, i.e. $\boldsymbol{x}_j = \{c_{jk}, x^\ell_{jk}, e^\ell_j, \mathrm{t}_j\}_j, \forall j \neq i$.

Again, the allocation and prices are constrained to be a competitive equilibrium and the planner chooses controls that respect the individual optimality conditions. However, due to all the general equilibrium and redistributive effects, we need to take a stance of what the planner internalizes. I make the assumption that the planner in country i only internalizes the optimality conditions (FOC) of its own country, but still internalize the market clearing of other countries when it involves the good traded by country i.

As a result, it uses the multiplier for the market clearing for the good traded from j, i.e. $\mu^{(i)}_j$: so $\mu^{(i)}_j$ represents the shadow value of relaxing country j market clearing (i.e. producing one extra unit of variety j), as internalized by planner i. Similarly, oil-gas is traded internationally so $\mu^{f,(i)}$ represents the shadow value of producing one extra unit of oil-gas (in the market, so produced by any country), as internalized by planner i. However, the market clearing of j for coal and renewable are not affected directly, only the market in i, which we keep denoting μ^c_i and μ^r_i.

Moreover, for climate damage ϕ^ε_i is the shadow cost of one additional ton of carbon emissions in the atmosphere as seen by country i. For the other multipliers, we use the same as before: λ_i for the budget constraint, ϕ^c for the consumption decision, θ^ℓ_i for the production quantity (supply) choice of energy firms $\ell = f, c, r$ for fossil, coal and renewable of country i, υ^ℓ_i for the quantity (demand) of energy ℓ chosen by the good firm, η_{ij} for the consumption choice for imports j by the household in i, ϑ^ℓ_{ij} for the import choice for inputs from j for the energy firm j. We are only considering those values for country i as the planner does not affect directly the distortion in country j.

As a result, the controls are $\boldsymbol{x}_i = \{c_{ij}, x^\ell_{ij}, e^\ell_i, \mathrm{p}_i, q^f, q^c_i, q^r_i, \mathrm{t}^\varepsilon_i, \mathrm{t}^b_{ij}\}_j$ and the multipliers are $\boldsymbol{\lambda}_i = \{\lambda_i, \mu^{(i)}_j, \mu^c_i, \mu^r_i, \mu^{f,(i)}, \phi^c_i, \theta^\ell_i, \upsilon^\ell_i, \phi^\varepsilon, \eta_{ij}, \vartheta^\ell_{ij}\}_{\ell,j}$.

The planner i's Lagrangian writes:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{x}_i, \boldsymbol{\lambda}_i) &= \mathcal{P}_i u(c_i)\overline{\mathcal{D}}_i^u(\mathcal{E}) + \mathcal{P}_i\lambda_i\Big(\mathrm{p}_i\overline{\mathcal{D}}^y(\mathcal{E})f(\ell_i, k_i, e_i^f, e_i^c, e_i^r) + \frac{1}{\mathcal{P}_i}[q^f g^f(x_i^\ell) - \sum_j x_{ij}^\ell \tau_{ij}\mathrm{p}_j(1+\mathrm{t}_{ij}^b)] \\
&\quad + \sum_\ell \{q^\ell g^\ell(x_i^\ell) - \sum_j x_{ij}^\ell\tau_{ij}\mathrm{p}_j(1+\mathrm{t}_{ij}^b)\} - ((q^f+\xi^f\mathrm{t}_i^\varepsilon)e_i^f + (q^c+\xi^c\mathrm{t}_i^\varepsilon)e_i^c + q_i^r e_i^r + (n_i+\bar{g}_i+\delta)k_i + c_i\mathbb{P}_i + \mathrm{t}_i^{ls})\Big) \\
&\quad + \sum_{\mathbb{I}} \omega_i^{(i)}\mathrm{p}_i\mu_i^{(i)}\Big(\mathcal{P}_i\mathcal{D}_i(T_i)z_i f(e_i,\ell_i) - \sum_{k\in\mathbb{I}}\mathcal{P}_k\tau_{ki}c_{ki} + \sum_{k\in\mathbb{I}}\tau_{ki}(x_{ki}^f + x_{ki}^c + x_{ki}^r)\Big) \\
&\quad + \mu^{f(i)}q^f[\sum_{i\in\mathbb{I}} e_i^x - \mathcal{P}_i e_i^f] + \mu_i^c q_i^c(\bar{e}_i^c - \mathcal{P}_i e_i^c) + \mu_i^r q_i^r(\bar{e}_i^r - \mathcal{P}_i e_i^r) \\
&\quad + \mathcal{P}_i\phi_i^\varepsilon[\mathcal{S} - \sum_{i\in\mathbb{I}}\mathcal{P}_i(\xi^f e_i^f + \xi^c e_i^c)] + \mathcal{P}_i\phi_i^c(\mathbb{P}_i\lambda_i^h - u'(c_i)\overline{\mathcal{D}}^u(\mathcal{E})) + \sum_{\ell\in\{f,c,r\}}\theta_i^\ell(\mathbb{P}_i - q_i^\ell g'(x_i^\ell)) \\
&\quad + \mathcal{P}_i\Big(\upsilon_i^f[(q^f+\xi^f\mathrm{t}^\varepsilon) - \mathrm{p}_i MPe_i^f] + \upsilon_i^c[(q_i^c+\xi^c\mathrm{t}^\varepsilon) - \mathrm{p}_i MPe_i^c] + \upsilon_i^r[q_i^r - \mathrm{p}_i MPe_i^r] + \upsilon_i^k[\underbrace{\rho+\eta\bar{g}_i}_{r_i+\bar{g}+n} - \mathrm{p}_i MPk_i - \delta]\Big) \\
&\quad + \sum_{j\in\mathbb{I}}\mathcal{P}_i\eta_{ij}c_{ij}[(1+\mathrm{t}_{ij})\tau_{ij}\mathrm{p}_j - \mathbb{P}_i c_i^{\frac{1}{\theta}} a_{ij}^{\frac{1}{\theta}} c_{ij}^{-\frac{1}{\theta}}] + \sum_{\ell\in\{f,c,r\}}\sum_{j\in\mathbb{I}}\vartheta_{ij}^\ell x_{ij}^\ell[(1+\mathrm{t}_{ij})\tau_{ij}\mathrm{p}_j - \mathbb{P}_i(x_i^\ell)^{\frac{1}{\theta}} a_{ij}^{\frac{1}{\theta}}(x_{ij}^\ell)^{-\frac{1}{\theta}}]
\end{aligned}$$

where $\omega_j^{(i)}$ are the weights planner i puts on market clearing j, and $\omega_i = 1$.

The First Order Conditions of the control over home variables $c_{ij}, x_{ij}^\ell, e_i^\ell, \mathrm{p}_i$ write:

- Consumption c_{ij}

$$\mathcal{P}_i u'(c_i)c_i^{\frac{1}{\theta}} a_{ij}^{\frac{1}{\theta}} c_{ij}^{-\frac{1}{\theta}} - \mathcal{P}_i\lambda_i\tau_{ij}\mathrm{p}_j - \omega_j^{(i)}\mathcal{P}_i\mu_j^{(i)}\tau_{ij}\mathrm{p}_j + \omega_i\mathcal{P}_i c_{ij}\eta_{ij}\frac{1}{\theta}\frac{\tau_{ij}(1+\mathrm{t}_{ij})\mathrm{p}_j}{c_{ij}}(s_{ij}-1)$$

$$u'(c_i) = \lambda_i\Big(\sum_j a_{ij}(\tau_{ij}\mathrm{p}_j)^{1-\theta}\underbrace{[1 + \omega_j^{(i)}\frac{\mu_j^{(i)}}{\lambda_i} - \frac{\eta_{ij}}{\lambda_i}\frac{1}{\theta}(1+\mathrm{t}_{ij}^b)(1-s_{ij})]}_{=1+\mathrm{t}_{ij}^b}{}^{1-\theta}\Big)^{\frac{1}{1-\theta}}$$

As before, if t_{ij}^b are set optimally, we obtain:

$$1 + \mathrm{t}_{ij}^b = 1 + \omega_j^{(i)}\frac{\mu_j^{(i)}}{\lambda_i}$$

The optimal tariff on j internalizes the change in demand through the market clearing of good j and uses this shadow value to set the tariffs: if $\mu_j^{(i)} > 0$, planner i would like to relax the constraint for j and thus reduce demand, hence setting a positive tariff t_{ij}^b.

Moreover, in this context, the distortionary effect on the ij trade, denoted by η_{ij}, is zero. Indeed, now, since the planner can manipulate that decisions freely, they choose to avoid causing a distortion $\eta_{ij} = 0$.

- Energy inputs: x_{ij}^ℓ

$$q_i^\ell g'(x_i^\ell)(x_i^\ell)^{\frac{1}{\theta}} a_{ij}^{\frac{1}{\theta}}(x_{ij}^\ell)^{-\frac{1}{\theta}} - \theta_i^\ell q_i^\ell g''(x_i^\ell)(x_i^\ell)^{\frac{1}{\theta}} a_{ij}^{\frac{1}{\theta}}(x_{ij}^\ell)^{-\frac{1}{\theta}} - \lambda_i \tau_{ij}\mathrm{p}_j - \omega_j^{(i)}\mu_j\tau_{ij}\mathrm{p}_j - \vartheta_{ij}\frac{1}{\theta}\tau_{ij}(1+\mathrm{t}_{ij})\mathrm{p}_j(s_{ij}-1) =$$

$$\{\mu_i^\ell - \theta_i^\ell \frac{g''(x_i^\ell)}{g'(x_i^\ell)}\} q_i^\ell g'(x_i^\ell)(x_i^\ell)^{\frac{1}{\theta}} a_{ij}^{\frac{1}{\theta}}(x_{ij}^\ell)^{-\frac{1}{\theta}} = \tau_{ij}\mathrm{p}_j[\mathrm{t}_{ij}^b\lambda_i + \omega_j^{(i)}\mu_j^{(i)} - \vartheta_{ij}\frac{1}{\theta}(1+\mathrm{t}_{ij}^b)(1-s_{ij})]$$

$$\{\mu_i^\ell - \theta_i^\ell \frac{g''(x_i^\ell)}{g'(x_i^\ell)}\} = \frac{1}{1+\mathrm{t}_{ij}^b}\big[-\mathrm{t}_{ij}^b\lambda_i + \omega_j^{(i)}\mu_j^{(i)} - \vartheta_{ij}\frac{1}{\theta}(1+\mathrm{t}_{ij}^b)(1-s_{ij})\big]$$

Choosing tariffs optimally yield again:

$$\mathrm{t}_{ij}^b = \omega_j^{(i)}\frac{\mu_j^{(i)}}{\lambda_i}$$

$$\vartheta_{ij} = 0 \qquad\qquad \mu_i^\ell = \theta_i^\ell\frac{g''(x_i^\ell)}{g'(x_i^\ell)}$$

The optimal tariff is the same as above, and there is no distortion on the FOC for good inputs for energy productions, i.e. $\vartheta_{ij} = 0$.
Moreover, the shadow value of the market clearing for energy i equals the distortion from the supply of that energy θ_i^ℓ – weighted by the supply elasticity, related to g''/g'.
For oil-gas, since the market is global and the supply curve is strictly convex, we get:

$$\mu^{f(i)} = \theta_i^f\frac{g''(x_i^f)}{g'(x_i^f)}$$

For coal and renewable energy this implies that $\mu_i^c = \mu_i^r = 0$ and there is no distortion of the market clearing for these local, again because the planner can completely control the demand and supply for those energy sources.

- Price q_i^ℓ, first with non-oil/gas, $\ell = r$ or $\ell = c$, the local market implies:

$$\mathcal{P}_i\lambda_i[\frac{1}{\mathcal{P}_i}g(x_i^\ell) - e_i^\ell] + (\mathcal{P}_i\upsilon_i^\ell - g'(x_i^\ell)\theta_{it}^\ell) = 0 \qquad\Rightarrow\qquad \mathcal{P}_i\upsilon_i^\ell = g'(x_i^\ell)\theta_{it}^\ell$$

- Price q^f, for oil/gas

$$\lambda_i[g(x_i^f) - \mathcal{P}_i e_i^f] + (\mathcal{P}_i\upsilon_i^f - g'(x_i^f)\theta_i^f) = 0$$

This implies that the total net distortion, between demand $\mathcal{P}_i\upsilon_i^f$ and supply θ_i^f equals the net-import $\mathcal{P}_i e_i^f - g(x_i^f)$.

- Tariffs t_{ij}^b

$$\sum_{\ell\in\{f,c,r\}}\theta_i^\ell(\frac{\tau_{ij}(1+\mathrm{t}_{ij})\mathrm{p}_j}{\mathbb{P}_i}\frac{\partial\mathbb{P}_i}{\partial(1+\mathrm{t}_{ij}^b)}) + \mathcal{P}_i\eta_{ij}(\tau_{ij}\mathrm{p}_j - \frac{\tau_{ij}(1+\mathrm{t}_{ij})\mathrm{p}_j}{\mathbb{P}_i}\frac{\partial\mathbb{P}_i}{\partial(1+\mathrm{t}_{ij}^b)}) + \sum_{\ell\in\{f,c,r,k\}}\vartheta_{ij}^\ell x_{ij}^\ell(\tau_{ij}\mathrm{p}_j - \frac{\tau_{ij}(1+\mathrm{t}_{ij})\mathrm{p}_j}{\mathbb{P}_i}\frac{\partial\mathbb{P}_i}{\partial(1+\mathrm{t}_{ij}^b)}) = 0$$

$$\sum_{\ell\in\{f,c,r\}}\theta_i^\ell s_{ij} + \eta_{ij}c_{ij}\mathcal{P}_i(1-s_{ij}) + \sum_{\ell\in\{f,c,r,k\}}\vartheta_{ij}^\ell x_{ij}^\ell(1-s_{ij}) = 0$$

As we saw that the tariff is set optimally, we have no distortion for each good sourcing: consumption $\eta_{ij}=0$ and energy inputs $\vartheta^{\ell}_{ij}=0, \forall \ell$.
This implies that for all the goods sourced, we obtain:

$$\sum_{\ell\in\{f,c,r\}} \theta^{\ell}_i = 0$$

the shadow values of the supply distortions are summing to zero at the country level: since all energy source from the same input bundle at price $\mathbb{P}_i$, some supply are distorted positively while some are distorted negatively.

- Price p_i – the planner can only control the price p_i from local good i

$$\begin{aligned}
&\mathcal{P}_i\lambda_i\overline{\mathcal{D}}^y(\mathcal{E})f(k_i,e_i) - \mathcal{P}_i\lambda_i\big[\tau_{ii}c_{ii} + \sum_{\ell}\tau_{ii}x^{\ell}_{ii}\big] - \mathcal{P}_i\big[\upsilon^f_i MPe^f_i + \upsilon^c_i MPe^c_i + \upsilon^r_i MPe^r_i + \upsilon^k_i MPe^k_i\big] \\
&\quad + \sum_{\ell\in\{f,c,r,k\}} \theta^{\ell}_i\big(\tau_{ii}(1+\mathrm{t}_{ii})s_{ii}\big) + \mathcal{P}_i\tau_{ii}(1+\mathrm{t}_{ii})[1-s_{ii}]\Big(c_{ii}\eta_{ii} + \sum_{\ell\in\{f,c,r,k\}} x^{\ell}_{ii}\vartheta^{\ell}_{ii}\Big) = 0 \\
\Rightarrow \qquad & \lambda_i\big[\overline{\mathcal{D}}^y(\mathcal{E})f(\ell_i,k_i,e_i) - \tau_{ii}(c_{ii} + \sum_{\ell} x^{\ell}_{ii})\big] - \big[\upsilon^f_i MPe^f_i + \upsilon^c_i MPe^c_i + \upsilon^r_i MPe^r_i + \upsilon^k_i MPe^k_i\big] = 0
\end{aligned}$$

since we have no distortion good sourcings: $\eta_{ij}=\vartheta^{\ell}_{ij}=0, \forall\ell$, and use $\sum_{\ell\in\{f,c,r\}}\theta^{\ell}_i=0$, we have that the planner would like to balance out different effects.
This condition is akin to a terms-of-trade manipulation: the planner would like to increase the price p_i, as it increases its purchasing power $\lambda_i \mathrm{p}_i y_i$ via income, and allow to demand more energy υ^{ℓ}_i, but at the same time balance out the cost for its own household and energy inputs $\lambda_i(c_{ii}+\sum_{\ell} x^{\ell}_{ii})$.

- Climate damage per capita ϕ^{ε}_i from changing emissions $\mathcal{E}$:

$$\begin{aligned}
\phi^{\varepsilon}_i = \Big[&u(c_i)\overline{\mathcal{D}}^{u\prime}_i(\mathcal{E}) + (\lambda_i+\mu^{(i)}_i)\mathcal{D}^{y\prime}_i(\mathcal{E})\mathrm{p}_i z_i F(\ell_i,k_i,e_i) - \mathcal{D}^{y\prime}_i(\mathcal{E})\mathrm{p}_i\{\upsilon^f_i MPe^f_i + \upsilon^c_i MPe^c_i + \upsilon^r_i MPe^r_i + \upsilon^k_i MPe^k_i\} \\
&+ \sum_j \omega^{(i)}_j \mu^{(i)}_j \mathcal{D}^{y\prime}_j(\mathcal{E})\mathrm{p}_j z_j F(\ell_j,k_j,e_j)\Big]
\end{aligned}$$

This represents the local social cost of carbon in welfare units. Climate change has different impact on country i: it affects utility $u(\cdot)$, affect production and thus budget $\lambda_i y_i$. It also distorts the input demand υ^{ℓ}_i through its impact on firm productivity.

Moreover, it also impacts production of good i through market clearing $\mu^{(i)}_i$. Interestingly, climate also impacts the goods production from countries j, which affects indirectly the country i through imports i: as a result, the planner i does indirectly account for the impact of climate on other countries j through international trade – because it cares of its own imported consumption c_{ij} through value $\mu^{(i)}_j$

This channel is novel when computing the social cost of carbon and I plan to investigate further how local social of carbon can be correlated across countries through international trade, an idea also discussed in Dingel et al. (2019).

- Carbon tax $\mathrm{t}_i^{\varepsilon}$:

$$\mathcal{P}_i(\upsilon_i^f \xi^f + \upsilon_i^c \xi^c) = 0$$

given that the planner has a single tax instruments for both energy inputs – fossil (oil/gas) and coal – they would like to avoid create distortion at the country level, and hence the value of these two distortions should offset each others. In practice, the distortion for oil is positive while the one for coal is negative: if the planner had two instruments it would set a higher tax on oil and a lower one of coal, to attenuate the distortionary effects.

- Energy demand e_i^ℓ

$$\mathcal{P}_i\lambda_i(\mathrm{p}_i MPe_i^\ell - q^f) + \mathcal{P}_i\mu_i^{(i)}\mathrm{p}_i MPe_i^\ell - q_i^\ell\mu_i^\ell\mathcal{P}_i - (\mathcal{P}_i\phi_i^\varepsilon)\mathcal{P}_i\xi^\ell - \sum_{\ell'}\mathcal{P}_i\mathrm{p}_i\upsilon_i^{\ell'}\partial_{e_i^\ell}MPe_i^{\ell'} = 0$$

Oil-gas:

$$\lambda_i\xi^f\mathrm{t}_i^\varepsilon + \mu_i^{(i)}\mathrm{p}_i MPe_i^f - q^f\mu_i^{f(i)} - (\mathcal{P}_i\phi_i^\varepsilon)\xi^f - \sum_{\ell'}\mathrm{p}_i\upsilon_i^{\ell'}\partial_{e_i^f}MPe_i^{\ell'} = 0$$

Coal

$$\lambda_i\xi^c\mathrm{t}_i^\varepsilon + \mu_i^{(i)}\mathrm{p}_i MPe_i^c - q_i^c\mu_i^c - (\mathcal{P}_i\phi_i^\varepsilon)\xi^c - \sum_{\ell'}\mathcal{P}_i\mathrm{p}_i\upsilon_i^{\ell'}\partial_{e_i^c}MPe_i^{\ell'} = 0$$

Renewable:

$$\mu_i^{(i)}\mathrm{p}_i MPe_i^r - q_i^r\mu_i^r - \sum_{\ell'}\mathcal{P}_i\mathrm{p}_i\upsilon_i^{\ell'}\partial_{e_i^r}MPe_i^{\ell'} = 0$$

Using these last conditions, we can express the optimal carbon tax.

Unilateral carbon tax

From the optimality condition for energy we can express the carbon tax in function of different motives. Let us consider the example of oil and gas:

$$\lambda_i\xi^f\mathrm{t}_i^\varepsilon = -\mu_i^{(i)}\mathrm{p}_i MPe_i^f + q^f\mu_i^{f(i)} + (\mathcal{P}_i\phi_i^\varepsilon)\xi^f + \sum_{\ell'}\mathrm{p}_i\upsilon_i^{\ell'}\partial_{e_i^f}MPe_i^{\ell'}$$

$$\begin{aligned}\xi^f\mathrm{t}_i^\varepsilon &= -\frac{\mu_i^{(i)}}{\lambda_i}q^f + \frac{\mu_i^{f(i)}}{\lambda_i}q^f + (\mathcal{P}_i\phi_i^\varepsilon)\xi^f \\ &\quad + [\mathrm{p}_i\frac{\upsilon_i^f}{\lambda_i}\partial_{e_i^f}MPe_i^f + \frac{\upsilon_i^c}{\lambda_i}\partial_{e_i^f}MPe_i^c + \frac{\upsilon_i^r}{\lambda_i}\partial_{e_i^f}MPe_i^r + \frac{\upsilon_i^k}{\lambda_i}\partial_{e_i^f}MPe_i^k]\end{aligned}$$

We see that the carbon balance four different effects – as in the global Second-Best but this time at the country level:

First, a climate externality Pigouvian motive $\mathcal{P}_i\phi_i^\varepsilon/\lambda_i$: the tax needs to be larger to account for the local cost of carbon defined as :

$$LCC_i = \mathcal{P}_i\frac{\phi^\varepsilon}{\lambda_i}$$

Moreover, this scale with population: if the country is larger $\mathcal{P}_i$, the tax internalize more the damage effect that energy consumption has on its' population.

Second, a distortionary effect: it would like to offset the distortion effect it has across the different energy inputs, and this weights by cross-elasticities, which relates to the terms $\partial_{e_i^f} MPe_i^\ell$.

Third, a redistributive term linked to the energy market clearing $\mu^{f(i)}$, if the planner i would like to relax the market for oil, or reduce oil demand, i.e. a positive $\mu^{f(i)}$, it would set a higher tax to lower its own demand

Fourth, a terms-of-trade manipulation effect, linked to the market clearing of it's own good $\mu_i^{(i)}$. This term is usually positive: the planner would like to increase its own production and to that purpose it could lower the carbon tax, or even subsidize carbon (!) to manipulate terms-of-trade.

Finally, all these terms are weighted by the country own marginal utility of consumption $\lambda_i = u'(c_i)/\mathbb{P}_i$. They are amplified if the country is higher income and dampened for lower income/consumption countries.

To provide even more intuitions, let us consider that the production only use labor and fossil (oil/gas) energy and use our isoelastic supply function for oil and gas.

In that case, the carbon tax is solely a tax on oil: there is no demand distortion effect and $v_i^f = 0$ since the tax can completely offset distortion. The second distortionary effect drops out.

Unfortunately, terms-of-trade effect $\mu_j^{(i)}$ and $\mu_i^{(i)}$ can not be expressed in closed-form easily as they depend on the general equilibrium effects and international trade in the Armington model, usually not tractable

Moreover, we have the isoelastic supply curve $e_i^x = g(x_i^f)$ that implies the term:

$$\frac{g''(x_i^f)}{g'(x_i^f)} = -\frac{\nu_i}{1+\nu_i}\frac{1}{x_i}$$

From the optimality condition for energy inputs goods x_i^f we had:

$$\mu^{f(i)} = \theta_i^f \frac{g''(x_i^f)}{g'(x_i^f)} = -\frac{\nu_i}{1+\nu_i}\frac{1}{x_i^f}\theta_i^f$$

where ν_i is the oil-gas inverse elasticity for country i, and θ_i^f is the supply distortion.
From the optimality condition for q^f we have, with simplification thanks to isoelastic supply:

$$\begin{aligned}
\theta_i^f &= \frac{1}{g'(x_i^f)}\lambda_i(e_i^x - \mathcal{P}_i e_i^f -) \\
\mu^{f(i)} &= -\frac{\nu_i}{1+\nu_i}\frac{1}{x_i^f}\theta_i^f = -\frac{\nu_i}{1+\nu_i}\frac{1}{x_i^f}\frac{1}{g'(x_i^f)}\lambda_i(e_i^x - \mathcal{P}_i e_i^f) = -\frac{\nu_i}{1+\nu_i}\frac{\bar{\nu}_i}{e_i^x}\frac{1+\nu_i}{\bar{\nu}_i}\lambda_i(e_i^x - \mathcal{P}_i e_i^f) \\
\mu^{f(i)} &= \frac{\nu_i}{e_i^x}\lambda_i(\mathcal{P}_i e_i^f - e_i^x)
\end{aligned}$$

As a result, this redistribution terms is positive for net-importers and negative for exporters. This is intuitive: energy dependent countries would like to reduce their dependence to imports, by taxing the energy (and redistributing lump-sum those revenues) they hope to reduce the equilibrium price q^f to benefit for better terms-of-trade. This is the same logic as the global social planner of Appendix C.2, as a smaller scale, and weighted by the country i production e_i^x and inverse elasticity ν_i: a more inelastic supply, with large ν_i would amplify this effect: as usual in Ramsey taxation, it is more relevant to tax inelastic supply goods.

As a result the unilateral carbon tax can write:

$$\xi^f \mathrm{t}_i^\varepsilon = -q^f \frac{\mu_i^{(i)}}{\lambda_i} + q^f \nu_i \frac{\mathcal{P}_i e_i^f - e_i^x}{e_i^x} + \xi^f LCC_i$$

As a result, we see that the optimal carbon tax can become a *subsidy* if the terms-of-trade manipulation motive $\mu_i^{(i)}$ is large enough, if the energy-supply redistribution term is negative, for example for oil-gas exporters, and if the local cost of carbon LCC_i is small.

D Welfare decomposition

D.1 Change in welfare – experiment

We compute the change in welfare, linearizing the model around an equilibrium where $\mathrm{t}^{\varepsilon} = \bar{\mathrm{t}}^{\varepsilon} = 0$ and $\mathrm{t}^b_{ij} = \bar{\mathrm{t}}^b_{ij} = 0$. This corresponds to the competitive equilibrium, since policies are identical to the "status-quo". We consider a climate agreement $\mathcal{J}$ of J countries, which are indifferent of being in the club or not, since the policy $(\mathrm{t}^{\varepsilon}_i, \mathrm{t}^b_{ij}) = (0,0)$ does not change the equilibrium. We consider a deviation where we increase those policy instruments by a small amount. To save on notation, we denote $d\ln z_i = \frac{dz_i}{z_i}$ – with a slight abuse of notation.

As a result, the policy change we consider is $d\ln \mathrm{t}^{\varepsilon}_i = \frac{d\mathrm{t}^{\varepsilon}_i}{1+\mathrm{t}^{\varepsilon}_i} = d\mathrm{t}^f_i$ where we consider a multiplicative carbon tax on fossil fuel $q^f_i(1+\mathrm{t}^{\varepsilon}_i)$. Similarly we consider a tariff change : $d\ln \mathrm{t}^b_{ij} = \frac{d\mathrm{t}^b_{ij}}{1+\mathrm{t}^b_{ij}} = d\mathrm{t}^b_{ij}$. In matrix notation, these changes in carbon tax are noted

$$\mathbf{J} d\mathbf{t}^{\varepsilon} = \left\{\mathbb{1}_{\{i\in\mathcal{J}\}}\, d\ln \mathrm{t}^{\varepsilon}_i\right\}_i \qquad \overline{\mathbf{J}} \odot d\mathbf{t}^b = \left\{\mathbb{1}_{\{i\in\mathcal{J}, j\notin\mathcal{J}\}} d\mathrm{t}^b_{ij}\right\}_{ij}$$

with $\mathbf{J} = \mathrm{J}_i = \mathbb{1}_{\{i\in\mathcal{J}\}}$, and $\overline{\mathbf{J}} \equiv \mathrm{J}_{ki} = \mathbb{1}_{\{i\in\mathcal{J}, j\notin\mathcal{J}\}}$.

This section is inspired Kleinman et al. (2020), where we follow the same steps, using a richer model, with trade a la Armington, energy in production and carbon and trade policy instruments.

We compute the welfare of individual country i, defined as the indirect utility $\mathcal{U}_i = u(\{c_{ij}\}_j)$, changes as:

$$d\mathcal{U}_i = du\Big(\frac{x_i}{\mathbb{P}_i}\Big) = u'(c_i)\Big(\frac{dx_i}{\mathbb{P}_i} - \frac{x_i}{\mathbb{P}_i}\frac{d\mathbb{P}_i}{\mathbb{P}_i}\Big) = u'(c_i)c_i\Big(\frac{dx_i}{x_i} - \frac{d\mathbb{P}_i}{\mathbb{P}_i}\Big)$$

with $x_i = c_i\mathbb{P}_i$ the consumption expenditure.

D.2 Model summary

First let us summarize the model, as presented above

$$\begin{aligned}
c_i\mathbb{P}_i = x_i &= w_i\ell_i + \pi^x_i + \mathrm{t}^{ls}_i = \mathrm{p}_i z_i \mathcal{D}_i(T_i)F(e_i,\ell_i) - q^e_t e_i + \frac{1}{\mathcal{P}_i}\Big(q^e e^x_i - \mathbb{P}_i\mathcal{C}^f(e^x_i,\mathcal{R}_i)\Big) + \mathrm{t}^{ls}_i \\
\mathcal{P}_i \mathrm{p}_i y_i &= \sum_{k\in\mathbb{I}} \mathcal{P}_k s_{ki}\frac{v_k}{1+\mathrm{t}_{ki}} \\
\widetilde{v}_i &= \mathrm{p}_i y_i + q^f(e^x_i/\mathcal{P}_i - e^f_i) \\
v_i &= m_i\widetilde{v}_i = \mathrm{p}_i y_i + q^f(e^x_i/\mathcal{P}_i - e^f_i) + \mathrm{t}^{ls}_i \\
\pi^f_i &= \frac{1}{\mathcal{P}_i}\frac{\nu\bar{\nu}^{-1/\nu}}{1+\nu}\mathcal{R}_i(q^f)^{1+\frac{1}{\nu}}\mathbb{P}_i^{-1/\nu} \\
\sum_k \mathcal{P}_i e^f_i &= \sum_k e^x_i = (q^f)^{1/\nu}\sum_k \mathcal{R}_i\bar{\nu}_i^{-1/\nu}\mathbb{P}_i^{-1/\nu} \\
F_i(\varepsilon(e^f,e^c,e^r),\ell) &= \Big[(1-\epsilon)^{\frac{1}{\sigma_y}}\big(\bar{k}^{\alpha}\ell^{1-\alpha}\big)^{\frac{\sigma_y-1}{\sigma_y}} + \epsilon^{\frac{1}{\sigma_y}}\big(z^e_i\ \varepsilon_i(e^f,e^c,e^r)\big)^{\frac{\sigma_y-1}{\sigma_y}}\Big]^{\frac{\sigma_y}{\sigma_y-1}} \\
\varepsilon(e^f,e^c,e^r) &= \Big[(\omega^f_i)^{\frac{1}{\sigma_e}}(e^f)^{\frac{\sigma_e-1}{\sigma_e}} + (\omega^c_i)^{\frac{1}{\sigma_e}}(e^c)^{\frac{\sigma_e-1}{\sigma_e}} + (\omega^r_i)^{\frac{1}{\sigma_e}}(e^r)^{\frac{\sigma_e-1}{\sigma_e}}\Big]^{\frac{\sigma_e}{\sigma_e-1}}
\end{aligned}$$

D.3 Production

Starting from the budget constraint:

$$c_i\mathbb{P}_i = x_i = \mathrm{p}_i z_i \mathcal{D}_i(T_i)F(e_i,\ell_i) - q^e e_i + \frac{1}{\mathcal{P}_i}\Big(q^f e_i^x - \mathbb{P}_i\mathcal{C}^f(e_i^x,\mathcal{R}_i)\Big) + \mathrm{t}_i^{ls}$$
$$= \mathrm{p}_i z_i \mathcal{D}_i(T_i)F(e_i,\ell_i) - \big(q^f(1+\mathrm{t}_i^\varepsilon)e_i^f + q_i^c(1+\mathrm{t}_i^\varepsilon)e_i^c + q_i^r e_i^r\big) + \frac{1}{\mathcal{P}_i}\Big(q^e e_i^x - \mathbb{P}_i\mathcal{C}^f(e_i^x,\mathcal{R}_i)\Big) + \widetilde{\mathrm{t}}_i^{ls} + q^f \mathrm{t}_i^\varepsilon e_i^f + q_i^c \mathrm{t}_i^\varepsilon e_i^c$$

Since, the revenues of the carbon-tax are redistributed lump-sum to the Household, we do not see any direct redistributive effect of carbon taxation, as the terms $q^f \mathrm{t}_i^\varepsilon e_i^f$ cancel out. Moreover,

We define the shares what are relevant for the decomposition:

- Energy share in production: $s_i^e = \frac{e_i q_i^e}{y_i \mathrm{p}_i}$
- Fossil share in energy mix $s_i^f = \frac{e_i^f q^f}{e_i q_i^e}$ and similarly $s_i^c = \frac{e_i^c q_i^c}{e_i q_i^e}$ and $s_i^r = \frac{e_i^r q_i^r}{e_i q_i^e}$
- Production share/rent share in GDP: $\eta_i^y = \frac{y_i \mathrm{p}_i}{y_i \mathrm{p}_i + \pi_i^f} = 1 - \eta_i^\pi$
- Consumption share in GDP: $\eta_i^c = \frac{x_i}{y_i \mathrm{p}_i + \pi_i^f}$
- Consumption as a ratio of output: $s_i^c = \frac{c_i \mathbb{P}_i}{y_i \mathrm{p}_i} = \frac{x_i}{y_i\mathrm{p}_i+\pi_i^f}\frac{y_i\mathrm{p}_i+\pi_i^f}{y_i\mathrm{p}_i} = \frac{\eta_i^c}{1-\eta_i^\pi} = \frac{\eta_i^c}{\eta_i^y}$,
- Energy share as a ratio of consumption: $\frac{e_i q_i^e}{x_i} = \frac{e_i q_i^e}{y_i \mathrm{p}_i}\frac{y_i\mathrm{p}_i}{y_i\mathrm{p}_i+\pi_i^f}\frac{y_i\mathrm{p}_i+\pi_i^f}{x_i} = s_i^e\frac{\eta_i^y}{\eta_i^c}$
- Profit share as a ratio of consumption: $\frac{\pi_i^f}{x_i} = \frac{\pi_i^f}{y_i\mathrm{p}_i+\pi_i^f}\frac{y_i\mathrm{p}_i+\pi_i^f}{x_i} = \frac{\eta_i^\pi}{\eta_i^c}$

Taking the first-order expansion of the budget constraint, we obtain:

$$\frac{dc_i}{c_i} = \frac{dx_i}{x_i} - \frac{d\mathbb{P}_i}{\mathbb{P}_i} = \frac{\mathrm{p}_i y_i}{x_i}\Big(\frac{d\mathrm{p}_i}{\mathrm{p}_i} + \frac{dy_i}{y_i}\Big) - \frac{e_i q_i^e}{x_i}\Big(\frac{e_i^f q^f}{e_i q_i^e}\Big(\frac{de^f}{e^f}+\frac{dq^f}{q^f}\Big) + \frac{e_i^c q^c}{e_i q_i^e}\Big(\frac{de^c}{e^c}+\frac{dq^c}{q^c}\Big) + \frac{e_i^r q^r}{e_i q_i^e}\Big(\frac{de^r}{e^r}+\frac{dq^r}{q^r}\Big)\Big)$$
$$+ \frac{\pi^f}{x_i}\frac{d\pi_i^f}{\pi^f} + \frac{\widetilde{\mathrm{t}}_i^{ls}}{x_i}\Big(\frac{d\widetilde{\mathrm{t}}_i^{ls}}{\widetilde{\mathrm{t}}_i^{ls}}\Big) \quad - \frac{d\mathbb{P}_i}{\mathbb{P}_i}$$

$$\frac{dc_i}{c_i} = \frac{\eta_i^y}{\eta_i^c}\Big(\frac{d\mathrm{p}_i}{\mathrm{p}_i} + \frac{dy_i}{y_i}\Big) - s_i^e\frac{\eta_i^y}{\eta_i^c}\Big(s_i^f\Big(\frac{de^f}{e^f}+\frac{dq^f}{q^f}\Big) + s_i^c\Big(\frac{de^c}{e^c}+\frac{dq^c}{q^c}\Big) + s_i^r\Big(\frac{de^r}{e^r}+\frac{dq^r}{q^r}\Big)\Big)$$
$$+ \frac{\eta_i^\pi}{\eta_i^c}\frac{d\pi_i^f}{\pi^f} + \frac{\widetilde{\mathrm{t}}_i^{ls}}{x_i}\Big(\frac{d\widetilde{\mathrm{t}}_i^{ls}}{\widetilde{\mathrm{t}}_i^{ls}}\Big) \quad - \frac{d\mathbb{P}_i}{\mathbb{P}_i}\Big)$$

First, using output changes, approximating the production function – c.f. the logic in Farhi-Baqaee (2021):

$$\frac{dy_i}{y_i} = \frac{d\mathcal{D}_i}{\mathcal{D}_i} + \frac{MPe_i e_i}{y_i}\frac{de_i}{e_i} = \frac{d\mathcal{D}_i}{\mathcal{D}_i} + s_i^e\Big[s_i^f\frac{de_i^f}{e_i^f} + s_i^r\frac{de_i^r}{e_i^r}\Big]$$

we see that the Hulten's theorem imply an first-order impact of a change in energy price that scale with the share of energy in production s_i^e, which is typically around $5-10\%$ and the share of fossils in the energy mix s_i^f, s_i^c, which sum to above 85%.

Second, using fossil energy firm problem and the profit change,

$$\frac{d\pi_i^f}{\pi^f} = \Big((1+\tfrac{1}{\nu_i})\frac{dq^f}{q^f} - \frac{1}{\nu_i}\frac{d\mathbb{P}_i}{\mathbb{P}_i}\Big)$$

The energy rent is affected by changes in the aggregate fossil energy price dq^f. Since the cost also depends on imported inputs, the prices of goods $\mathbb{P}_i$ also matter for profit and welfare.

Third, using the production of coal and renewable, which are simply $q_i^c = \upsilon_i^c \mathbb{P}_i$ and $q_i^r = \upsilon_i^r \mathbb{P}_i$, we get

$$\frac{dq^r}{q^r} = \frac{d\mathbb{P}_i}{\mathbb{P}_i} \qquad \text{and} \qquad \frac{dq^c}{q^c} = \frac{d\mathbb{P}_i}{\mathbb{P}_i}$$

the price of both coal and renewable energy are directly exposed to changes in the price of imports.

Accounting for these different effects simplify dramatically the change in consumption:

$$\begin{aligned}
\frac{dc_i}{c_i} &= \frac{\eta_i^y}{\eta_i^c}\Big(\frac{d\mathrm{p}_i}{\mathrm{p}_i} + \frac{dy_i}{y_i}\Big) - s_i^e \frac{\eta_i^y}{\eta_i^c}\Big(s_i^f\big(\frac{de^f}{e^f} + \frac{dq^f}{q^f}\big) + s_i^c\big(\frac{de^c}{e^c} + \frac{dq^c}{q^c}\big) + s_i^r\big(\frac{de^r}{e^r} + \frac{dq^r}{q^r}\big)\Big) \\
&\qquad + \frac{\eta_i^\pi}{\eta_i^c}\Big(1+\frac{1}{\nu}\Big)\Big(\frac{dq^f}{q^f} - \frac{1}{1+\nu}\frac{d\mathbb{P}_i}{\mathbb{P}_i}\Big) + \frac{d\widetilde{\mathrm{t}}_i^{ls}}{x_i}\big) \quad - \frac{d\mathbb{P}_i}{\mathbb{P}_i} \\
&= \frac{\eta_i^y}{\eta_i^c}\Big(\frac{d\mathrm{p}_i}{\mathrm{p}_i} + \frac{d\mathcal{D}_i}{\mathcal{D}_i}\Big) - \frac{\eta_i^y}{\eta_i^c} s_i^e\Big[s_i^f \frac{dq^f}{q^f} + s_i^c \frac{dq^c}{q^c} + s_i^r \frac{dq^r}{q^r}\Big] + \frac{\eta_i^\pi}{\eta_i^c}\Big(1+\frac{1}{\nu}\Big)\Big(\frac{dq^f}{q^f} - \frac{1}{1+\nu}\frac{d\mathbb{P}_i}{\mathbb{P}_i}\Big) + \frac{d\widetilde{\mathrm{t}}_i^{ls}}{x_i} \quad - \frac{d\mathbb{P}_i}{\mathbb{P}_i} \\
\frac{dc_i}{c_i} &= \frac{\eta_i^y}{\eta_i^c}\Big(\frac{d\mathrm{p}_i}{\mathrm{p}_i} + \frac{d\mathcal{D}_i}{\mathcal{D}_i}\Big) - \frac{\eta_i^y}{\eta_i^c} s_i^e s_i^f \frac{dq^f}{q^f} + \frac{\eta_i^\pi}{\eta_i^c}\Big(1+\frac{1}{\nu}\Big)\frac{dq^f}{q^f} - \Big[\frac{\eta_i^y}{\eta_i^c} s_i^e (s_i^c + s_i^r) + \frac{\eta_i^\pi}{\eta_i^c}\frac{1}{\nu} + 1\Big]\frac{d\mathbb{P}_i}{\mathbb{P}_i} + \frac{d\widetilde{\mathrm{t}}_i^{ls}}{x_i}
\end{aligned}$$

D.4 Climate externality

To see the positive influence of carbon taxation on climate, we unpack the damage $d\mathcal{D}_i$. In this section, we simplify the climate system by considering the following static model:

$$\begin{aligned}
T_i &= \Delta_i \mathcal{T} = \Delta_i \chi \mathcal{S} \\
\mathcal{S} &= \mathcal{S}_0 + \mathcal{E} = \mathcal{S}_0 + E^f + \hat{\xi}^c E^c
\end{aligned}$$

where $\mathcal{E} = E^f + \xi^c E^c$ is a representation of long-term emission. If the static model represent a long-period of time – say 50 or 100 years – we consider $\mathcal{E}$ scale up the annual emissions $\sum_i e_i^f + e_i^c$ due to oil-gas and coal. Moreover, the premium ξ^c represents the premium in carbon emission due to coal $\hat{\xi}^c = \xi^c - \xi^f$, and we therefore normalize the unit of energy to the carbon content of oil and gas.

Using Nordhaus' damage function,

$$\mathcal{D}_i = e^{-\frac{\gamma}{2}(T_i - T_i^\star)^2} \bar{z}$$

A linear approximation implies:

$$\frac{d\mathcal{D}_i}{\mathcal{D}_i} = -\gamma(T_i - T_i^\star) dT_i = -\gamma(T_i - T_i^\star) T_i \frac{dT_i}{T_i}$$

Regarding the change in temperature caused by emissions, we get:

$$
\begin{aligned}
& dT_i = \Delta_i \chi d\mathcal{S} = \Delta_i \chi (dE^f + \hat{\xi}^c dE^c) \\
\Rightarrow \qquad & \frac{dT_i}{T_i} = \frac{\Delta_i \chi d\mathcal{E}}{\Delta_i \chi (\mathcal{S}_0 + \mathcal{E})} = s^{E/S}\Big(s^{f/E}\frac{dE^f}{E^f} + s^{c/E}\frac{dE^c}{E^c}\Big) \\
& \qquad \text{with } s^{E/S} = \frac{\mathcal{E}}{\mathcal{S}_0 + \mathcal{E}} \qquad s^{f/E} = \frac{E^f}{\mathcal{E}} \qquad s^{c/E} = \hat{\xi}^c \frac{E^c}{\mathcal{E}}
\end{aligned}
$$

As a result, to summarize, the change in damage depends on the total energy used in fossil (oil-gas) and coal.

$$
d\ln \mathcal{D}_i = -\bar{\gamma}_i \big(s^{f/E} d\ln E^f + s^{c/E} d\ln E^c\big) \qquad \bar{\gamma}_i = \gamma(T_i - T_i^\star) T_i \, s^{E/S}
$$

where $\bar{\gamma}_i$ summarize in one parameter the heterogeneous impact of climate change.

D.5 Energy markets

We now turn to energy where the demand and equilibrium effect on prices will be of first-oder importance for our welfare decomposition.

Energy demand

To examine the demand side of the market, we compute the elasticities of demand for each energy source, which are determined jointly by the firm First-Order Conditions. Thanks to our nested CES formulation, we can compute the elasticity $\varepsilon^\ell_{q^k} = \frac{\partial e_i^\ell}{\partial q^k}\frac{q^k}{e_i^\ell}$ as :

$$
\begin{bmatrix} \varepsilon^f_{q^f} & \varepsilon^f_{q^c} & \varepsilon^f_{q^r} \\ \varepsilon^c_{q^f} & \varepsilon^c_{q^c} & \varepsilon^c_{q^r} \\ \varepsilon^r_{q^f} & \varepsilon^r_{q^c} & \varepsilon^r_{q^r} \end{bmatrix} = (\widetilde{H}^e)^{-1} = -\frac{\sigma^y}{1-s^e} \begin{bmatrix} s^f & s^c & s^r \\ s^f & s^c & s^r \\ s^f & s^c & s^r \end{bmatrix} + \sigma^e \begin{bmatrix} -(1-s^f) & s^c & s^r \\ s^f & -(1-s^c) & s^r \\ s^f & s^c & -(1-s^r) \end{bmatrix}
$$

where the first part correspond to the change in aggregate price of energy q^e, since $\frac{\partial q_i^e}{\partial q^k}\frac{q^k}{q_i^e} = s_i^k$, which reduces demands for overall energy, according to elasticity $\frac{\sigma^y}{1-s_i^e}$ where s_i^e is the cost share of energy and σ^y the elasticity between energy and other inputs. Second, the later part summarizes the substitution effect across energy sources, negative along the diagonal and positive out of diagonal, due to positive cross-elasticity in the CES framework.

Moreover, the energy demand also depends on aggregate TFP (and hence climate damage), and the price level at which the final good is sold. As a result, the productivity elasticities and the final good price elasticity write:

$$
\begin{bmatrix} \varepsilon^f_z \\ \varepsilon^c_z \\ \varepsilon^r_z \end{bmatrix} = \frac{\sigma^y}{1-s^e} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad \begin{bmatrix} \varepsilon^f_p \\ \varepsilon^c_p \\ \varepsilon^r_p \end{bmatrix} = \frac{\sigma^y}{1-s^e} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}
$$

which again is standard in the Nested CES framework.

As a result, we can express the energy demand as a function of the other endogenous variables:

$$
\begin{aligned}
d\ln e_i^f &= -(\frac{\sigma^y}{1-s_i^e}s_i^f + (1-s_i^f)\sigma^e)[d\ln q^f + \mathrm{J}_i d\ln \mathrm{t}^\varepsilon] + (\sigma^e - \frac{\sigma^y}{1-s_i^e})s_i^c[d\ln q_i^c + \hat{\xi}^c \mathrm{J}_i d\ln \mathrm{t}^\varepsilon] + (\sigma^e - \frac{\sigma^y}{1-s_i^e})s_i^r d\ln q_i^r \\
&\quad + \frac{\sigma^y}{1-s^e}d\ln \mathcal{D}_i + \frac{\sigma^y}{1-s^e}d\ln \mathrm{p}_i \\
d\ln e_i^c &= (\sigma^e - \frac{\sigma^y}{1-s_i^e})s_i^f[d\ln q^f + \mathrm{J}_i d\ln \mathrm{t}^\varepsilon] - (\frac{\sigma^y}{1-s_i^e}s_i^c + (1-s_i^c)\sigma^e)[d\ln q_i^c + \hat{\xi}^c \mathrm{J}_i d\ln \mathrm{t}^\varepsilon] + (\sigma^e - \frac{\sigma^y}{1-s_i^e})s_i^r d\ln q_i^r \\
&\quad + \frac{\sigma^y}{1-s^e}d\ln \mathcal{D}_i + \frac{\sigma^y}{1-s^e}d\ln \mathrm{p}_i \\
d\ln e_i^r &= (\sigma^e - \frac{\sigma^y}{1-s_i^e})s_i^f[d\ln q^f + \mathrm{J}_i d\ln \mathrm{t}^\varepsilon] + (\sigma^e - \frac{\sigma^y}{1-s_i^e})s_i^c[d\ln q_i^c + \hat{\xi}^c \mathrm{J}_i d\ln \mathrm{t}^\varepsilon] - (\frac{\sigma^y}{1-s_i^e}s_i^r + (1-s_i^r)\sigma^e)d\ln q_i^r \\
&\quad + \frac{\sigma^y}{1-s^e}d\ln \mathcal{D}_i + \frac{\sigma^y}{1-s^e}d\ln \mathrm{p}_i
\end{aligned}
$$

Those endogenous energy demand can be reintegrated in the production function to obtain, the change in output, as function of prices of good, energies and productivity:

$$
\begin{aligned}
d\ln y_i &= d\ln \mathcal{D}_i + s_i^e[s_i^f d\ln e_i^f + s_i^r d\ln e_i^c + s_i^r d\ln e_i^r] \\
&= (1+\frac{s_i^e\sigma^y}{1-s_i^e})d\ln \mathcal{D}_i + \frac{s_i^e\sigma^y}{1-s_i^e}d\ln \mathrm{p}_i - s_i^e\frac{\sigma^y}{1-s_i^e}s_i^f[d\ln q^f + d\ln \mathrm{t}_i^\varepsilon] \\
&\qquad - s_i^e\frac{\sigma^y}{1-s_i^e}s_i^r[d\ln q_i^c + \hat{\xi}^c d\ln \mathrm{t}_i^\varepsilon] - s_i^e\frac{\sigma^y}{1-s_i^e}s_i^r d\ln q_i^r \\
d\ln y_i &= \alpha^{y,z}d\ln z_i + \alpha^{y,p}d\ln \mathrm{p}_i - \alpha^{y,qf}[d\ln q^f + d\ln \mathrm{t}_i^\varepsilon] - \alpha^{y,qc}[d\ln q_i^c + d\ln \mathrm{t}_i^\varepsilon] - \alpha^{y,qr}d\ln q_i^r \\
\alpha_i^{y,z} &= 1+\frac{s_i^e\sigma^y}{1-s_i^e} \qquad \alpha_i^{y,p} = \frac{s_i^e\sigma^y}{1-s_i^e} \\
\alpha_i^{y,qf} &= s_i^e\frac{\sigma^y}{1-s^e}s_i^f \qquad \alpha_i^{y,qc} = s_i^e\frac{\sigma^y}{1-s^e}s_i^c \qquad \alpha_i^{y,qr} = s_i^e\frac{\sigma^y}{1-s_i^e}s_i^r
\end{aligned}
$$

$$
d\ln y_i = \alpha^{y,z}d\ln z_i + \alpha^{y,p}d\ln \mathrm{p}_i - \alpha^{y,qf}d\ln q^f - (\alpha^{y,qf}+\hat{\xi}^c\alpha^{y,qc})d\ln \mathrm{t}_i^\varepsilon - (\alpha^{y,qc}+\alpha^{y,qr})d\ln \mathbb{P}_i
$$

where this last equation uses the supply curve of coal and renewable. We can see the exposure of country i's output of carbon tax: $\alpha^{y,qf}+\hat{\xi}^c\alpha^{y,qc} = s_i^e\frac{\sigma^y}{1-s^e}(s_i^f + \hat{\xi}^c s_i^c)$, through the price and substitution effect of oil, gas and coal.

Fossil energy market

The energy demand in fossil is the sum of individual countries demand, where we denote the share of country i in global production $\lambda_i^f = \frac{\mathcal{P}_i e_i^f}{E^f}$

$$
\begin{aligned}
dE^f &= \sum_i \mathcal{P}_i de_i^f \\
d\ln E^f &= \sum_i \lambda_i^f d\ln e_i^f \\
&= -\sum_i \lambda_i^f \big(\frac{\sigma^y}{1-s_i^e} s_i^f + (1-s_i^f)\sigma^e\big)[d\ln q^f + \mathrm{J}_i d\ln \mathrm{t}^\varepsilon] + \sum_i \lambda_i^f \big(\sigma^e - \frac{\sigma^y}{1-s_i^e}\big) s_i^c [d\ln q_i^c + \hat{\xi}^c \mathrm{J}_i d\ln \mathrm{t}^\varepsilon] \\
&\qquad + \sum_i \lambda_i^f \big(\sigma^e - \frac{\sigma^y}{1-s_i^e}\big) s_i^r d\ln q_i^r + \sum_i \lambda_i^f \frac{\sigma^y}{1-s_i^e} d\ln \mathcal{D}_i + \sum_i \lambda_i^f \frac{\sigma^y}{1-s_i^e} d\ln \mathrm{p}_i
\end{aligned}
$$

We see that carbon taxation decreases demand for oil and gas by substitution, but can also increases it if the substitution away for coal is strong enough. The first effect dominate the second – up to the first order – if:

$$
\overline{\lambda}_{\mathcal{J}}^{\sigma,f} := \sum_{i\in\mathcal{J}} \lambda_i^f \big(\frac{\sigma^y}{1-s_i^e} s_i^f + (1-s_i^f)\sigma^e\big) > \sum_{i\in\mathcal{J}} \lambda_i^f \big(\sigma^e - \frac{\sigma^y}{1-s_i^e}\big) s_i^c \hat{\xi}^c =: \overline{\lambda}_{\mathcal{J}}^{\sigma,c}
$$

which depend, among others, on the covariance $\mathbb{C}\mathrm{ov}_i(\lambda_i^f, 1-s_i^f)$ and $\mathbb{C}\mathrm{ov}_i(\lambda_i^f, s_i^c)$, since the substitution effect is stronger than the income effect $\sigma^e > \sigma^y/(1-s_i^e)$, in most empirically-relevant cases.

Now, the energy supply curve can also be recasted as the sum of individual extraction $E^f = \sum_i \mathcal{P}_i e_i^f = \sum_i e_i^x$, and, with the share of fossil production $\lambda_i^x = e_i^x/E^f$, it hence derives as follow:

$$
\begin{aligned}
e_i^x &= (q^f)^{1/\nu_i} \mathcal{R}_i \bar{\nu}_i^{-1/\nu_i} \mathbb{P}_i^{-1/\nu_i} \\
d\ln E^f &= \sum_i \lambda_i^x d\ln e_i^x = \sum_i \lambda_i^x \frac{1}{\nu_i}[d\ln q^f - d\ln \mathbb{P}_i] \\
\Rightarrow \qquad d\ln q^f &= \bar{\nu} d\ln E^f + \sum_i \lambda_i^x \frac{\bar{\nu}}{\nu_i} d\ln \mathbb{P}_i
\end{aligned}
$$

with the aggregate supply elasticity $\bar{\nu} = \big(\sum_i \lambda_i^x \nu_i^{-1}\big)^{-1}$, that we already encountered in the second best optimal Ramsey policy.

Now, replacing the energy demand quantity $d\ln E^f$ into the energy supply/price curve, we obtain:

$$
\begin{aligned}
d\ln q^f &= \bar{\nu} d\ln E^f + \sum_i \lambda_i^x \frac{\bar{\nu}}{\nu_i} d\ln \mathbb{P}_i \\
&\qquad + \bar{\nu} \sum_i \lambda_i^f \big(\sigma^e - \frac{\sigma^y}{1-s_i^e}\big) s_i^r d\ln q_i^r + \bar{\nu} \sum_i \lambda_i^f \frac{\sigma^y}{1-s_i^e}[d\ln \mathcal{D}_i + d\ln \mathrm{p}_i] + \sum_i \lambda_i^x \frac{\bar{\nu}}{\nu_i} d\ln \mathbb{P}_i \\
d\ln q^f &= -\frac{\bar{\nu}}{1+\bar{\nu}\overline{\lambda}^{\sigma,f}} \sum_i \lambda_i^{\sigma,f} \mathrm{J}_i d\ln \mathrm{t}^\varepsilon + \frac{\bar{\nu}}{1+\bar{\nu}\overline{\lambda}^{\sigma,f}} \sum_i \lambda_i^f \big(\sigma^e - \frac{\sigma^y}{1-s_i^e}\big) s_i^c [d\ln q_i^c + \hat{\xi}^c \mathrm{J}_i d\ln \mathrm{t}^\varepsilon] \\
&\qquad + \frac{\bar{\nu}}{1+\bar{\nu}\overline{\lambda}^{\sigma,f}} \sum_i \lambda_i^f \big(\sigma^e - \frac{\sigma^y}{1-s_i^e}\big) s_i^r d\ln q_i^r + \frac{\bar{\nu}}{1+\bar{\nu}\overline{\lambda}^{\sigma,f}} \sum_i \lambda_i^f \frac{\sigma^y}{1-s_i^e}[d\ln \mathcal{D}_i + d\ln \mathrm{p}_i] + \frac{1}{1+\bar{\nu}\overline{\lambda}^{\sigma,f}} \sum_i \lambda_i^x \frac{\bar{\nu}}{\nu_i} d\ln \mathbb{P}_i
\end{aligned}
$$

where $\overline{\lambda}^{\sigma,f} = \overline{\lambda}^{\sigma,f}_{\mathbb{I}}$, for $\mathbb{I}$ the whole world, and $\overline{\lambda}^{\sigma,f}_{\mathcal{J}} := \sum_{i\in\mathcal{J}} \lambda^f_i \big(\frac{\sigma^y}{1-s^e_i} s^f_i + (1-s^f_i)\sigma^e\big)$. Replacing the known terms, prices of renewable, coal, this factor, we obtain:

$$d\ln q^f = \frac{\bar{\nu}}{1+\bar{\nu}\overline{\lambda}^{\sigma,f}}\big[-\overline{\lambda}^{\sigma,f}_{\mathcal{J}} + \overline{\lambda}^{\sigma,c}_{\mathcal{J}}\big] d\ln \mathrm{t}^{\varepsilon} + \frac{1}{1+\bar{\nu}\overline{\lambda}^{\sigma,f}} \sum_i \big(\bar{\nu}\lambda^{\sigma,c}_i + \bar{\nu}\lambda^{\sigma,r}_i + \lambda^x_i \frac{\bar{\nu}}{\nu_i}\big) d\ln \mathbb{P}_i$$
$$+ \frac{\bar{\nu}}{1+\bar{\nu}\overline{\lambda}^{\sigma,f}} \sum_i \lambda^f_i \frac{\sigma^y}{1-s^e_i}\big[d\ln \mathcal{D}_i + d\ln \mathrm{p}_i\big]$$

As before we see that carbon taxation decrease the oil-gas energy price if $\overline{\lambda}^{\sigma,f}_{\mathcal{J}} > \overline{\lambda}^{\sigma,c}_{\mathcal{J}}$. Moreover, we see that a change in the price index $d\ln \mathcal{P}_i$ of all the countries change the aggregate price of oil and gas because it both increases the price of renewable and coal, increases demand for oil-gas by substitutions – the terms $\bar{\nu}\lambda^{\sigma,c}_i$ and $\bar{\nu}\lambda^{\sigma,r}_i$ – and it also increases the price of the input – through the term $\lambda^x_i \frac{\bar{\nu}}{\nu_i}$.

D.6 Trade *à la* Armington

To investigate how the price indices $\mathcal{P}_i$ and the good price p_i are determined, we should how consider the market of goods.

$$\mathcal{P}_i \mathrm{p}_i y_i = \sum_{k\in\mathbb{I}} \mathcal{P}_k s_{ki} \frac{v_k}{1+\mathrm{t}_{ki}}$$

Using the CES framework, we obtain that:

$$\mathbb{P}_i = \Big(\sum_j a_{ij}(\tau_{ij}(1+\mathrm{t}^b_{ij})\mathrm{p}_j)^{1-\theta}\Big)^{\frac{1}{1-\theta}}$$
$$\frac{d\mathbb{P}_i}{\mathbb{P}_i} = \sum_j s_{ij}\Big(\frac{d\mathrm{p}_j}{\mathrm{p}_j} + \frac{d\mathrm{t}^b_{ij}}{1+\mathrm{t}^b_{ij}}\Big)$$
$$s_{ij} = \frac{c_{ij}(1+\mathrm{t}_{ij})\tau_{ij}\mathrm{p}_j}{\sum_k c_{ik}(1+\mathrm{t}_{ik})\tau_{ik}\mathrm{p}_k} = a_{ij}\frac{\big((1+\mathrm{t}_{ij})\tau_{ij}\mathrm{p}_j\big)^{1-\theta}}{\sum_k \big((1+\mathrm{t}_{ik})\tau_{ik}\mathrm{p}_k\big)^{1-\theta}} = \Big(\frac{(1+\mathrm{t}_{ij})\tau_{ij}\mathrm{p}_j}{\mathbb{P}_i}\Big)^{1-\theta}$$
$$\frac{ds_{ij}}{s_{ij}} = (\theta-1)\Big(\frac{d\mathbb{P}_i}{\mathbb{P}_i} - \big(\frac{d\mathrm{p}_j}{\mathrm{p}_j} + \frac{d\mathrm{t}^b_{ij}}{1+\mathrm{t}^b_{ij}}\big)\Big)$$
$$\frac{ds_{ij}}{s_{ij}} = (\theta-1)\Big(\sum_k s_{ik}\big(\frac{d\mathrm{p}_k}{\mathrm{p}_k} + \frac{d\mathrm{t}^b_{ik}}{1+\mathrm{t}^b_{ik}}\big) - \big(\frac{d\mathrm{p}_j}{\mathrm{p}_j} + \frac{d\mathrm{t}^b_{ij}}{1+\mathrm{t}^b_{ij}}\big)\Big)$$

Using those formulas, the market clearing linearizes as follow:

$$\frac{d\big[\frac{s_{ij}m_i\widetilde{v}_i}{1+t_{ij}}\big]}{\frac{s_{ij}m_i\widetilde{v}_i}{1+\mathrm{t}_{ij}}} = \Big[d\ln\widetilde{v}_i + \theta\sum_k \big(s_{ik}d\ln \mathrm{t}_{ik} - (1+s_{ij})d\ln \mathrm{t}_{ij}\big) + (\theta-1)\sum_{k\neq j}\big(s_{ik}d\ln \mathrm{p}_k - d\ln \mathrm{p}_j\big)\Big]$$

with $\quad \widetilde{v}_i = \mathrm{p}_i y_i + q^f(e^x_i/\mathcal{P}_i - e^f_i)$

This implies:

$$
\begin{aligned}
\mathcal{P}_i\widetilde{v}_i(\frac{d\mathrm{p}_i}{d\mathrm{p}_i}+\frac{dy_i}{y_i}) &= \sum_k \mathcal{P}_k\frac{s_{ki}m_k\widetilde{v}_k}{1+\mathrm{t}_{ki}}d\ln[\frac{s_{ki}m_k\widetilde{v}_k}{1+t_{ki}}] \\
(\frac{d\mathrm{p}_i}{d\mathrm{p}_i}+\frac{dy_i}{y_i}) &= \sum_k \frac{\mathcal{P}_k\widetilde{v}_k}{\mathcal{P}_i\widetilde{v}_i}s_{ik}\Big[d\ln\widetilde{v}_k+\theta\sum_h\big(s_{kh}d\ln\mathrm{t}_{kh}-(1+s_{ki})d\ln\mathrm{t}_{ki}\big)+(\theta-1)\sum_h\big(s_{kh}d\ln\mathrm{p}_h-d\ln\mathrm{p}_i\big)\Big] \\
(\frac{d\mathrm{p}_i}{d\mathrm{p}_i}+\frac{dy_i}{y_i}) &= \sum_k \boldsymbol{t}_{ik}\Big[(\frac{\mathrm{p}_k y_k}{v_k})(d\ln\mathrm{p}_k+d\ln y_k)+\frac{q^f e^x_k}{v_k}d\ln e^x_k-\frac{q^f e^f_k}{v_k}d\ln e^f_k+\frac{q^f(e^x_k-e^f_k)}{v_k}d\ln q^f \\
&\quad +\theta\sum_h\big(s_{kh}d\ln\mathrm{t}_{kh}-(1+s_{ki})d\ln\mathrm{t}_{ki}\big)+(\theta-1)\sum_h\big(s_{kh}d\ln\mathrm{p}_h-d\ln\mathrm{p}_i\big)\Big]
\end{aligned}
$$

with $\boldsymbol{t}_{ik}=\frac{\mathcal{P}_k\widetilde{v}_k}{\mathcal{P}_i\widetilde{v}_i}s_{ki}$, which is analogous to the same matrix in Kleinman et al. (2020). This implies, that rewritten in matrix notation, we get:

$$
\begin{aligned}
d\ln(\mathrm{p}y) &= \mathbf{T}v^y d\ln(\mathrm{p}y)+\mathbf{T}v^{e^x}d\ln e^x_k-\mathbf{T}v^{e^f}d\ln e^f_k+\mathbf{T}v^{ne}d\ln q^f+(\theta-1)(\mathbf{TS}-\mathbf{I})d\ln\mathrm{p} \\
&\quad +\theta\big(\mathbf{TS}\odot\mathbf{J}\odot d\ln\mathrm{t}^b-\mathbf{T}(\mathbb{1}+\mathbf{S}')\odot(\mathbf{J}\odot d\ln\mathrm{t}^b)'\big) \\
\Big[(\mathbf{I}-\mathbf{T}\odot v^y)-(\theta-1)(\mathbf{TS}-\mathbf{T}')\Big]d\ln\mathrm{p} &= (\mathbf{I}-\mathbf{T}\odot v^y)d\ln y+\mathbf{T}v^{e^x}d\ln e^x_k-\mathbf{T}v^{e^f}d\ln e^f_k+\mathbf{T}v^{ne}d\ln q^f \\
&\quad +\theta\big(\mathbf{TS}\odot\mathbf{J}\odot d\ln\mathrm{t}^b-\mathbf{T}(\mathbb{1}+\mathbf{S}')\odot(\mathbf{J}\odot d\ln\mathrm{t}^b)'\big)
\end{aligned}
$$

with $v^{e^x}=\frac{q^f e^x_i}{v_i}$, $v^{e^f}=\frac{q^f e^f_i}{v_i}$ and $v^{ne}=\frac{q^f(e^x_i-e^f_i)}{v_k}$.

D.7 Back to welfare

From above, we saw that welfare writes as:

$$
\frac{d\mathcal{U}_i}{u'(c_i)c_i}=\frac{dc_i}{c_i}=\frac{\eta^y_i}{\eta^c_i}\Big(\frac{d\mathrm{p}_i}{\mathrm{p}_i}+\frac{d\mathcal{D}_i}{\mathcal{D}_i}\Big)-\frac{\eta^y_i}{\eta^c_i}s^e_i s^f_i\frac{dq^f}{q^f}+\frac{\eta^\pi_i}{\eta^c_i}\Big(1+\frac{1}{\nu}\Big)\frac{dq^f}{q^f}-\Big[\frac{\eta^y_i}{\eta^c_i}s^e_i(s^c_i+s^r_i)+\frac{\eta^\pi_i}{\eta^c_i}\frac{1}{\nu}+1\Big]\frac{d\mathbb{P}_i}{\mathbb{P}_i}+\frac{d\widetilde{\mathrm{t}}^{ls}_i}{x_i}
$$

with the damage

$$
d\ln\mathcal{D}_i=-\bar{\gamma}_i\big(s^{f/E}d\ln E^f+s^{c/E}d\ln E^c\big)
$$

With the oil-gas energy price:

$$
\begin{aligned}
d\ln q^f &= \bar{\nu}d\ln E^f+\sum_i\lambda^x_i\frac{\bar{\nu}}{\nu_i}d\ln\mathbb{P}_i \\
d\ln q^f &= \frac{\bar{\nu}}{1+\bar{\nu}\overline{\lambda}^{\sigma,f}}\big[-\overline{\lambda}^{\sigma,f}_{\mathcal{J}}+\overline{\lambda}^{\sigma,c}_{\mathcal{J}}\big]d\ln\mathrm{t}^\varepsilon+\frac{\bar{\nu}}{1+\bar{\nu}\overline{\lambda}^{\sigma,f}}\sum_i\big(\lambda^{\sigma,c}_i+\lambda^{\sigma,r}_i+\lambda^x_i\frac{1}{\nu_i}\big)d\ln\mathbb{P}_i \\
&\quad +\frac{\bar{\nu}}{1+\bar{\nu}\overline{\lambda}^{\sigma,f}}\sum_i\lambda^f_i\frac{\sigma^y}{1-s^e_i}[d\ln\mathcal{D}_i+d\ln\mathrm{p}_i] \\
d\ln E^f &= \frac{1}{1+\bar{\nu}\overline{\lambda}^{\sigma,f}}\big[-\overline{\lambda}^{\sigma,f}_{\mathcal{J}}+\overline{\lambda}^{\sigma,c}_{\mathcal{J}}\big]d\ln\mathrm{t}^\varepsilon+\frac{1}{1+\bar{\nu}\overline{\lambda}^{\sigma,f}}\sum_i\big(\lambda^{\sigma,c}_i+\lambda^{\sigma,r}_i\big)d\ln\mathbb{P}_i+\frac{1}{1+\bar{\nu}\overline{\lambda}^{\sigma,f}}\sum_i\lambda^f_i\frac{\sigma^y}{1-s^e_i}[d\ln\mathcal{D}_i+d\ln\mathrm{p}_i]
\end{aligned}
$$

D.8 Further simplification

To simplify the welfare formula even further, in the following we consider that energy is only composed of oil-gas. In practice, oil and gas compose the largest share of energy, with oil representing close to 35% of energy use and natural gas close to 20% at the world level.

We consider that $s_i^f = 1$ and $s_i^r = s_i^c = 0$ in all the formulas above. We also consider that oil-gas supply only uses the local good, which makes isolated from international trade – making the extraction $e_i^x = (q^f)^{1/\nu_i}\mathcal{R}_i\bar{\nu}_i^{-1/\nu_i}\mathrm{p}_i^{-1/\nu_i}$ instead of $\mathbb{P}_i^{-1/\nu_i}$.

Those two assumptions simplify our setting dramatically. The previous welfare decomposition reduces to :

$$\frac{d\mathcal{U}_i}{u'(c_i)c_i} = \frac{dc_i}{c_i} = [\frac{\eta_i^y}{\eta_i^c} - \frac{\eta_i^\pi}{\eta_i^c}\frac{1}{\nu_i}]\frac{d\mathrm{p}_i}{\mathrm{p}_i} + \frac{\eta_i^y}{\eta_i^c}\frac{d\mathcal{D}_i}{\mathcal{D}_i} - \frac{\eta_i^y}{\eta_i^c}s_i^e\frac{dq^f}{q^f} + \frac{\eta_i^\pi}{\eta_i^c}\Big(1+\frac{1}{\nu}\Big)\frac{dq^f}{q^f} - \frac{d\mathbb{P}_i}{\mathbb{P}_i} + \frac{d\widetilde{\mathrm{t}}_i^{ls}}{x_i}$$

where the damage rewrite:

$$d\ln\mathcal{D}_i = -\bar{\gamma}_i d\ln E^f$$

with the average damage is defined as $\bar{\gamma} = \sum_i \bar{\gamma}_i$. And the oil-gas demand curve write:

$$\begin{aligned}
d\ln E^f &= \sum_i \lambda_i^f d\ln e_i^f \\
&= -\sum_i \lambda_i^f \frac{\sigma^y}{1-s_i^e}[d\ln q^f + \mathrm{J}_i d\ln \mathrm{t}^\varepsilon] + \sum_i \lambda_i^f \frac{\sigma^y}{1-s_i^e} d\ln\mathcal{D}_i + \sum_i \lambda_i^f \frac{\sigma^y}{1-s_i^e} d\ln \mathrm{p}_i \\
&= -\sum_i \widetilde{\lambda}_i^f [d\ln q^f + \mathrm{J}_i d\ln \mathrm{t}^\varepsilon] + \sum_i \widetilde{\lambda}_i^f d\ln\mathcal{D}_i + \sum_i \widetilde{\lambda}_i^f d\ln \mathrm{p}_i
\end{aligned}$$

where, to simplify notations, we denote $\widetilde{\lambda}_i^f = \lambda_i^f \frac{\sigma^y}{1-s_i^e}$, and it's average $\overline{\lambda}^{\sigma,f} = \sum_i \widetilde{\lambda}_i^f \frac{\sigma^y}{1-s_i^e}$. As a result, the demand now rewrites:

$$d\ln E^f = \frac{1}{1+\bar{\gamma}+\mathbb{C}\mathrm{ov}_i(\widetilde{\lambda}_i^f, \bar{\gamma}_i)}\Big[-\sum_i \widetilde{\lambda}_i^f [d\ln q^f + \mathrm{J}_i d\ln \mathrm{t}^\varepsilon] + \sum_i \widetilde{\lambda}_i^f d\ln \mathrm{p}_i\Big]$$

We can see that the energy demand curve is affected by climate change: more emission imply larger damage, which in turn reduce energy demand and hence emissions. Moreover, the covariance term indicates that if the large energy producers (with a larger share of the market, and high elasticity σ) are also the most affected by climate change, this effect is stronger and the demand curve is even steeper / more inelastic.

Recasting this logic in the supply curve, we obtain:

$$
\begin{aligned}
d\ln q^f &= \bar{\nu} d\ln E^f + \sum_i \lambda_i^x \frac{\bar{\nu}}{\nu_i} d\ln \mathrm{p}_i \\
d\ln E^f &= -\sum_i \widetilde{\lambda}_i^f [d\ln q^f + \mathrm{J}_i d\ln \mathrm{t}^\varepsilon] - \sum_i \widetilde{\lambda}_i^f \bar{\gamma}_i d\ln E^f + \sum_i \widetilde{\lambda}_i^f d\ln \mathrm{p}_i \\
[1 + \sum_i \widetilde{\lambda}_i^f \bar{\gamma}_i + \bar{\nu}\sum_i \widetilde{\lambda}_i^f] d\ln E^f &= -\sum_i \widetilde{\lambda}_i^f \mathrm{J}_i d\ln \mathrm{t}^\varepsilon + \sum_i \widetilde{\lambda}_i^f d\ln \mathrm{p}_i - (\sum_i \widetilde{\lambda}_i^f) \sum_i \lambda_i^x \frac{\bar{\nu}}{\nu_i} d\ln \mathrm{p}_i \\
d\ln E^f &= \frac{1}{1+\bar{\gamma}+\mathbb{C}\mathrm{ov}_i(\widetilde{\lambda}_i^f, \bar{\gamma}_i) + \bar{\nu}\overline{\lambda}^{\sigma,f}} \Big[-\sum_i \widetilde{\lambda}_i^f \mathrm{J}_i d\ln \mathrm{t}^\varepsilon + \sum_i (\widetilde{\lambda}_i^f - \overline{\lambda}^{\sigma,f} \lambda_i^x \tfrac{\bar{\nu}}{\nu_i}) d\ln \mathrm{p}_i \Big] \\
d\ln q^f &= -\frac{\bar{\nu}\sum_i \widetilde{\lambda}_i^f \mathrm{J}_i d\ln \mathrm{t}^\varepsilon}{1+\bar{\gamma}+\mathbb{C}\mathrm{ov}_i(\widetilde{\lambda}_i^f, \bar{\gamma}_i) + \bar{\nu}\overline{\lambda}^{\sigma,f}} + \sum_i \beta_i d\ln \mathrm{p}_i \\
\text{with} \qquad \beta_i &= \lambda_i^x \frac{\bar{\nu}}{\nu_i} + \frac{\bar{\nu}}{1+\bar{\gamma}+\mathbb{C}\mathrm{ov}_i(\widetilde{\lambda}_i^f, \bar{\gamma}_i) + \bar{\nu}\overline{\lambda}^{\sigma,f}} (\widetilde{\lambda}_i^f - \overline{\lambda}^{\sigma,f} \lambda_i^x \tfrac{\bar{\nu}}{\nu_i})
\end{aligned}
$$

Welfare rewrites:

$$
\begin{aligned}
\frac{d\mathcal{U}_i}{u'(c_i)c_i} &= [\frac{\eta_i^y}{\eta_i^c} - \frac{\eta_i^\pi}{\eta_i^c}\frac{1}{\nu_i}] \frac{d\mathrm{p}_i}{\mathrm{p}_i} - \frac{\eta_i^y}{\eta_i^c} \bar{\gamma}_i d\ln E_i^f - \frac{\eta_i^y}{\eta_i^c} s_i^e \frac{dq^f}{q^f} + \frac{\eta_i^\pi}{\eta_i^c}\Big(1+\frac{1}{\nu}\Big)\frac{dq^f}{q^f} - \frac{d\mathbb{P}_i}{\mathbb{P}_i} + \frac{d\widetilde{\mathrm{t}}_i^{ls}}{x_i} \\
&= [\frac{\eta_i^y}{\eta_i^c} - \frac{\eta_i^\pi}{\eta_i^c}\frac{1}{\nu_i}] d\ln \mathrm{p}_i + \Big[-\frac{\eta_i^y}{\eta_i^c}\bar{\gamma}_i \frac{1}{\bar{\nu}} - \frac{\eta_i^y}{\eta_i^c} s_i^e + \frac{\eta_i^\pi}{\eta_i^c}\Big(1+\frac{1}{\nu}\Big)\Big] d\ln q^f + \frac{\eta_i^y}{\eta_i^c}\bar{\gamma}_i \sum_i \lambda_i^x \frac{1}{\nu_i} d\ln \mathrm{p}_i - d\ln \mathbb{P}_i + d\widetilde{\mathrm{t}}_i^{ls}/x_i
\end{aligned}
$$

Trade :

$$
\begin{aligned}
(\frac{d\mathrm{p}_i}{d\mathrm{p}_i} + \frac{dy_i}{y_i}) &= \sum_k \boldsymbol{t}_{ik} \Big[(\frac{\mathrm{p}_k y_k}{v_k})(d\ln \mathrm{p}_k + d\ln y_k) + \frac{q^f e_k^x}{v_k} d\ln e_k^x - \frac{q^f e_k^f}{v_k} d\ln e_k^f + \frac{q^f(e_k^x - e_k^f)}{v_k} d\ln q^f \\
&\quad + \theta \sum_h \big(s_{kh} d\ln \mathrm{t}_{kh} - (1+s_{ki}) d\ln \mathrm{t}_{ki}\big) + (\theta-1)\sum_h \big(s_{kh} d\ln \mathrm{p}_h - d\ln \mathrm{p}_i\big)\Big] \\
\Big[(\mathbf{I} - \mathbf{T}\odot v^y) - (\theta-1)(\mathbf{TS} - \mathbf{T}')\Big] d\ln \mathrm{p} &= (\mathbf{I} - \mathbf{T}\odot v^y) d\ln y + \mathbf{T} v^{e^x} d\ln e_k^x - \mathbf{T} v^{e^f} d\ln e_k^f + \mathbf{T} v^{ne} d\ln q^f \\
&\quad + \theta\big(\mathbf{TS}\odot\mathbf{J}\odot d\ln \mathrm{t}^b - \mathbf{T}(\mathbb{1}+\mathbf{S}')\odot(\mathbf{J}\odot d\ln \mathrm{t}^b)'\big)
\end{aligned}
$$

Output:

$$
\begin{aligned}
d\ln y_i &= \alpha^{y,z} d\ln \mathcal{D}_i + \alpha^{y,p} d\ln \mathrm{p}_i - \alpha^{y,qf}[d\ln q^f + d\ln \mathrm{t}_i^\varepsilon] \\
\alpha_i^{y,z} &= 1+\frac{s_i^e \sigma^y}{1-s_i^e} \qquad \alpha_i^{y,p} = \frac{s_i^e \sigma^y}{1-s_i^e} \qquad \alpha_i^{y,qf} = s_i^e \frac{\sigma^y}{1-s^e} s_i^f
\end{aligned}
$$

Energy demand:

$$
d\ln e_i^f = -\frac{\sigma^y}{1-s_i^e}[d\ln q^f + \mathrm{J}_i d\ln \mathrm{t}^\varepsilon] + \frac{\sigma^y}{1-s^e} d\ln \mathcal{D}_i + \frac{\sigma^y}{1-s^e} d\ln \mathrm{p}_i
$$

Energy supply:

$$
d\ln e_i^x = \frac{1}{\nu_i}[d\ln q^f - d\ln \mathrm{p}_i]
$$

$$\Big[(\mathbf{I}-\mathbf{T}\odot v^{y})[\mathbf{I}-\alpha^{y,p}\odot\mathbf{I}]+\mathbf{T}(v^{e^{x}}\odot\tfrac{1}{\nu})+\mathbf{T}v^{e^{f}}\tfrac{\sigma^{y}}{1-s^{e}}-(\theta-1)(\mathbf{TS}-\mathbf{T}')\Big]d\ln \mathrm{p}=$$

$$(\mathbf{I}-\mathbf{T}\odot v^{y})[\alpha^{y,z}d\ln\mathcal{D}-\alpha^{y,qf}\big[d\ln q^{f}+\mathbf{J}d\ln \mathrm{t}^{\varepsilon}\big]]+\mathbf{T}(v^{e^{x}}\odot\tfrac{1}{\nu})d\ln q^{f}+\mathbf{T}v^{e^{f}}(\tfrac{\sigma^{y}}{1-s_{i}^{e}}[d\ln q^{f}+\mathrm{J}_{i}d\ln \mathrm{t}^{\varepsilon}]-\tfrac{\sigma^{y}}{1-s^{e}})d\ln\mathcal{D}_{i}+\mathbf{T}v^{n}$$

$$+\,\theta\big(\mathbf{TS}\odot\mathbf{J}\odot d\ln \mathrm{t}^{b}-\mathbf{T}(\mathbb{1}+\mathbf{S}')\odot(\mathbf{J}\odot d\ln \mathrm{t}^{b})'\big)$$

$$=\Big((\mathbf{I}-\mathbf{T}\odot v^{y})\alpha^{y,z}-\tfrac{\sigma^{y}}{1-s^{e}}\Big)d\ln\mathcal{D}_{i}+\big[-(\mathbf{I}-\mathbf{T}\odot v^{y})\alpha^{y,qf}+\mathbf{T}(v^{e^{x}}\odot\tfrac{1}{\nu}+v^{e^{f}}\tfrac{\sigma^{y}}{1-s^{e}}+v^{ne})\big]d\ln q^{f}$$

$$+\big[-(\mathbf{I}-\mathbf{T}\odot v^{y})\alpha^{y,qf}+\mathbf{T}(v^{e^{f}}\odot\tfrac{\sigma^{y}}{1-s^{e}}\odot\mathrm{J})\big]d\ln \mathrm{t}^{\varepsilon}+\theta\big(\mathbf{TS}\odot\mathbf{J}\odot d\ln \mathrm{t}^{b}-\mathbf{T}(\mathbb{1}+\mathbf{S}')\odot(\mathbf{J}\odot d\ln \mathrm{t}^{b})'\big)$$

$$\Big[(\mathbf{I}-\mathbf{T}\odot v^{y})[\mathbf{I}-\alpha^{y,p}\odot\mathbf{I}]+\mathbf{T}(v^{e^{x}}\odot\tfrac{1}{\nu})+\mathbf{T}v^{e^{f}}\tfrac{\sigma^{y}}{1-s^{e}}-(\theta-1)(\mathbf{TS}-\mathbf{T}')-\Big((\mathbf{I}-\mathbf{T}\odot v^{y})\alpha^{y,z}-\tfrac{\sigma^{y}}{1-s^{e}}\Big)\odot\bar{\gamma}\mathbf{I}\odot(\tfrac{\lambda^{x}}{\nu})'\Big]d\ln \mathrm{p}=$$

$$\big[-(\mathbf{I}-\mathbf{T}\odot v^{y})\alpha^{y,qf}+\mathbf{T}(v^{e^{x}}\odot\tfrac{1}{\nu}+v^{e^{f}}\tfrac{\sigma^{y}}{1-s^{e}}+v^{ne})-\Big((\mathbf{I}-\mathbf{T}\odot v^{y})\alpha^{y,z}-\tfrac{\sigma^{y}}{1-s^{e}}\Big)\bar{\gamma}\tfrac{1}{\bar{\nu}}\big]d\ln q^{f}$$

$$+\big[-(\mathbf{I}-\mathbf{T}\odot v^{y})\alpha^{y,qf}+\mathbf{T}(v^{e^{f}}\odot\tfrac{\sigma^{y}}{1-s^{e}})\big]\odot\mathrm{J}d\ln \mathrm{t}^{\varepsilon}+\theta\big(\mathbf{TS}\odot\mathbf{J}\odot d\ln \mathrm{t}^{b}-\mathbf{T}(\mathbb{1}+\mathbf{S}')\odot(\mathbf{J}\odot d\ln \mathrm{t}^{b})'\big)$$

which gives:

$$d\ln \mathrm{p}=\mathbf{A}\big[-(\mathbf{I}-\mathbf{T}\odot v^{y})\alpha^{y,qf}+\mathbf{T}(v^{e^{x}}\odot\tfrac{1}{\nu}+v^{e^{f}}\tfrac{\sigma^{y}}{1-s^{e}}+v^{ne})-\Big((\mathbf{I}-\mathbf{T}\odot v^{y})\alpha^{y,z}-\tfrac{\sigma^{y}}{1-s^{e}}\Big)\bar{\gamma}\tfrac{1}{\bar{\nu}}\big]d\ln q^{f}$$

$$+\,\mathbf{A}\big[-(\mathbf{I}-\mathbf{T}\odot v^{y})\alpha^{y,qf}+\mathbf{T}(v^{e^{f}}\odot\tfrac{\sigma^{y}}{1-s^{e}})\big]\odot\mathrm{J}d\ln \mathrm{t}^{\varepsilon}+\theta\big(\mathbf{TS}\odot\mathbf{J}\odot d\ln \mathrm{t}^{b}-\mathbf{T}(\mathbb{1}+\mathbf{S}')\odot(\mathbf{J}\odot d\ln \mathrm{t}^{b})'\big)$$

with

$$\mathbf{A}=\Big[(\mathbf{I}-\mathbf{T}\odot v^{y})[\mathbf{I}-\alpha^{y,p}\odot\mathbf{I}]+\mathbf{T}(v^{e^{x}}\odot\tfrac{1}{\nu})+\mathbf{T}v^{e^{f}}\tfrac{\sigma^{y}}{1-s^{e}}-(\theta-1)(\mathbf{TS}-\mathbf{T}')-\Big((\mathbf{I}-\mathbf{T}\odot v^{y})\alpha^{y,z}-\tfrac{\sigma^{y}}{1-s^{e}}\Big)\odot\bar{\gamma}\mathbf{I}\odot(\tfrac{\lambda^{x}}{\nu})'\Big]^{-1}$$

represents all the feedback general equilibrium effects because changes in prices affect input choices of energy, exports of fossil fuels, energy production, and vulnerability to the rest of the world.

As a result, the price of energy q^{f} affects the price of country i because (i) it increases the income and hence demand if i is an exporter of fossil increases, (ii) it decreases the energy inputs e_{i}^{f} of all countries due to increased prices, (iii) it reflects an increase of energy demand and global emissions, which affects world damage and lowers the price of the most vulnerable countries. The second effect (ii) is also a direct impact of carbon taxation $d\mathrm{t}^{\varepsilon}$ on the energy demand of all countries, directly affecting output and prices. Finally, tariffs $d\mathrm{t}_{ij}^{b}$ affect country i's demand changing terms of trade and prices in general equilibrium.

E Climate agreement – Solution method

In this section, we detail the general formulation for the climate agreement design, which join together a policy choice and a combinatorial discrete choice problem.

E.1 Inner problem and Outer problem

The world welfare maximization requires to separate the maxima between the choice of countries and the choice of policy instruments $(\mathbb{J}, \mathbf{t})$. We can split the choice in two ways:

$$\begin{aligned}\max_{\mathbb{J},\mathbf{t}} \sum_i \omega_i \mathcal{U}_i(\mathbb{J},\mathbf{t}) &= \max_{\mathbb{J}} \max_{\mathbf{t}(\mathbb{J})} \sum_i \omega_i \mathcal{U}_i(\mathbb{J},\mathbf{t}) \\ &= \max_{\mathbf{t}} \max_{\mathbb{J}(\mathbf{t})} \sum_i \omega_i \mathcal{U}_i(\mathbb{J},\mathbf{t})\end{aligned}$$

where $\mathcal{U}_i(\mathbb{J},\mathbf{t})$ the indirect utility can also be written with

$$\mathcal{U}_i(\mathbb{J},\mathbf{t}) = \mathcal{U}_i(\mathbb{J},\mathrm{t}^f,\mathrm{t}^b) = \mathcal{U}_i(I_i,\{I_j\}_{j\neq i},\mathrm{t}^f,\mathrm{t}^b)$$

with indicators $I_i = \mathbb{1}\{i \in \mathbb{J}\}$.

Depending on how we split the joint problem, it leads to different treatments of the problem.

Combinatorial problem, first representation

$$\begin{aligned}&\max_{\mathcal{I}\in\mathcal{P}(\mathbb{I}),\,\mathbf{t}\in\mathbb{T}} \mathcal{W}(\mathcal{I},\mathbf{t}) = \max_{\mathcal{I}} \max_{\mathbf{t}\in\mathbb{T}} \sum_i \omega_i u_i(\mathcal{I},\mathbf{t}) && (P_1)\\ &s.t. \qquad u_i(\mathcal{I},\mathbf{t}) \geq u_i(\mathcal{I}\backslash\{i\},\mathbf{t}) \qquad \forall\, i\in\mathcal{I} \qquad [\lambda_i] \\ &\qquad\quad g_k(\mathcal{I},\mathbf{t}) \geq 0 \qquad [\mu_k]\end{aligned}$$

Define the marginal gain of $j \in \mathcal{I}$ for utility of agent i

$$\Delta_j u_i(\mathcal{I},\mathbf{t}) = u_i(\mathcal{I}\cup\{j\},\mathbf{t}) - u_i(\mathcal{I}\backslash\{j\},\mathbf{t})$$

Lagrangian function

$$\mathcal{L}(\mathcal{I},\mathbf{t},\boldsymbol{\lambda},\boldsymbol{\mu}) = \mathcal{W}(\mathcal{I},\mathbf{t}) + \sum_{i\in\mathcal{I}} \lambda_i \Delta_i u_i(\mathcal{I},\mathbf{t}) + \sum_k \mu_k g_k(\mathcal{I},\mathbf{t})$$

Rewriting

$$\max_{\mathcal{I}} \max_{\mathbf{t}} \min_{\boldsymbol{\lambda},\boldsymbol{\mu}} \mathcal{L}(\mathcal{I},\mathbf{t},\boldsymbol{\lambda},\boldsymbol{\mu}) \leq \max_{\mathcal{I}} \min_{\boldsymbol{\lambda},\boldsymbol{\mu}} \max_{\mathbf{t}} \mathcal{L}(\mathcal{I},\mathbf{t},\lambda)$$

If $\mathcal{W}$ is concave in $\mathbf{t}$, for every $\mathcal{I}$, where the optimal $\mathbf{t}$ and λ are implicit functions of $\mathcal{I}$

$$\begin{aligned}(P_1(\mathcal{I})) = \max_{\mathbf{t}} \min_{\boldsymbol{\lambda},\boldsymbol{\mu}} \mathcal{L}(\mathcal{I},\mathbf{t},\boldsymbol{\lambda}) &= \min_{\boldsymbol{\lambda}} \max_{\mathbf{t}} \mathcal{L}(\mathcal{I},\mathbf{t},\boldsymbol{\lambda},\boldsymbol{\mu}) \\ &= \min_{\boldsymbol{\lambda},\boldsymbol{\mu}} \max_{\mathbf{t}} \sum_{i\in\mathbb{I}} \omega_i u_i(\mathcal{I},\mathbf{t}) + \sum_{i\in\mathcal{I}} \lambda_i \Delta_i u_i(\mathcal{I},\mathbf{t}) + \sum_k \mu_k g_k(\mathcal{I},\mathbf{t})\end{aligned}$$

Step 1: First, if the constraint functions $\mathbf{t} \to \Delta_i u_i(\mathcal{I}, \mathbf{t})$ and $\mathbf{t} \to g_k(\mathcal{I}, \mathbf{t})$ maps into $\mathbb{R}_+$, then we can solve:

$$\begin{aligned} \mathbf{t}^\star \qquad\qquad s.t. \qquad & \sum_{i\in\mathbb{I}} \omega_i D_{\mathbf{t}} u_i(\mathcal{I}, \mathbf{t}^\star) + \sum_{i\in\mathcal{I}} \lambda_i D_{\mathbf{t}} \Delta_i u_i(\mathcal{I}, \mathbf{t}^\star) + \sum_k \mu_k g_k(\mathcal{I}, \mathbf{t}) = 0 \\ & \sum_{i\in\mathcal{I}} \lambda_i \Delta_i u_i(\mathcal{I}, \mathbf{t}^\star) = 0 \\ & \sum_k \mu_k g_k(\mathcal{I}, \mathbf{t}^\star) = 0 \end{aligned}$$

where the two last equations are the complementary slackness

Step 2: Given $\mathbf{t}^\star$, dual problem:

$$\min_{\boldsymbol{\lambda},\boldsymbol{\mu}} \sum_{i\in\mathbb{I}} \omega_i u_i(\mathcal{I}, \mathbf{t}) + \sum_{i\in\mathcal{I}} \lambda_i \Delta_i u_i(\mathcal{I}, \mathbf{t}) + \sum_k \mu_k g_k(\mathcal{I}, \mathbf{t})$$

Step 3: Given $\mathbf{t}^\star(\mathcal{I}), \boldsymbol{\lambda}, \boldsymbol{\mu}$, we have to solve the combinatorial discrete choice problem:

$$\max_{\mathcal{I}} \mathcal{L}(\mathcal{I}, \mathbf{t}^\star(\mathcal{I}), \boldsymbol{\lambda}(\mathcal{I}), \boldsymbol{\mu}(\mathcal{I}))$$

Problem: costly to search for $\boldsymbol{\lambda}(\mathcal{I}), \boldsymbol{\mu}(\mathcal{I})$ for every combination of countries $\mathcal{I}$.

Second representation:

$$\begin{aligned} \max_{\mathcal{I}\in\mathcal{P}(\mathbb{I}),\, \mathbf{t}\in\mathbb{T}} \mathcal{W}(\mathcal{I}, \mathbf{t}) &= \max_{\mathbf{t}\in\mathbb{T}} \max_{\mathcal{I}} \sum_i \omega_i u_i(\mathcal{I}, \mathbf{t}) && (P_2) \\ s.t. \qquad u_i(\mathcal{I}, \mathbf{t}) &\geq u_i(\mathcal{I}\backslash\{i\}, \mathbf{t}) \qquad \forall\, i \in \mathcal{I} && [\lambda_i] \\ g_k(\mathcal{I}, \mathbf{t}) &\geq 0 && [\mu_k] \end{aligned}$$

Rewriting, defining

$$\mathcal{I}(\mathbf{t}) = \{\mathcal{I} \subset \mathbb{I} \mid \Delta_i u_i(\mathcal{I}, \mathbf{t}) \geq 0 \qquad \forall i \in \mathcal{I}\}$$

Lagrangian:

$$\mathcal{L}(\mathbf{t}, \boldsymbol{\mu}) = \sum_i \omega_i u_i(\mathcal{I}(\mathbf{t}), \mathbf{t}) + \sum_k \mu_k g_k(\mathcal{I}(\mathbf{t}), \mathbf{t})$$

Issue: the problem becomes noncontinuous. So can not use calculus of variation/optimization

E.2 Combinatorial discrete choice problem

Second method: Squeezing procedure for CDCP with Participation Constraints Second, since full enumeration is costly, I provide an alternative algorithm inspired by methods used in the international trade literature to solve combinatorial discrete choice problems. The additional difficulty that needs to be considered is the presence of participation constraints. In this section, we only consider unilateral deviations. The idea behind this method is greatly inspired by Arkolakis, Eckert and Shi (2023) and Farrokhi and Lashkaripour (2024).

The idea is to build iteratively sets that are lower bound $\underline{\mathcal{J}}$ and upper bound $\overline{\mathcal{J}}$ sets for the optimal coalition $\mathbb{J}$: $\underline{\mathcal{J}}$ is a subset which *includes* all the countries that we know to be part of the optimal set $\mathbb{J}$ and $\overline{\mathcal{J}}$ is a superset, such that it excludes the countries that we know are not part of the optimal set. The set $\overline{\mathcal{J}}\backslash\underline{\mathcal{J}}$ is the set of potential countries. The natural starting point is $\underline{\mathcal{J}} = \varnothing$, $\overline{\mathcal{J}} = \mathbb{I}$.

The squeezing step in standard CDCP is a mapping from $\mathcal{J}$ to members that bring a positive marginal value to the objective $\mathcal{W}(\mathbb{J}) := \mathcal{W}(\mathbb{J}, \mathbf{t})$. The modification needed in our setting with participation constraints is that the country also needs have marginal *individual* value $\mathcal{U}_i(\mathcal{J}) = \mathcal{U}_i(\mathcal{J}, \mathbf{t})$ to be part of the coalition:

$$\Phi(\mathcal{J}, \mathbf{t}) \equiv \{ j \in \mathbb{I} \,|\, \Delta_j \mathcal{W}(\mathcal{J}, \mathbf{t}) > 0 \;\&\; \Delta_j \mathcal{U}_j(\mathcal{J}, \mathbf{t}) > 0 \} \tag{21}$$

where the marginal values for global welfare and individual welfare are

$$\Delta_j \mathcal{W}(\mathcal{J}, \mathbf{t}) \equiv \mathcal{W}(\mathcal{J} \cup \{j\}, \mathbf{t}) - \mathcal{W}(\mathcal{J}\backslash\{j\}, \mathbf{t}) = \sum_{i \in \mathbb{I}} \mathcal{P}_i \omega_i \big(\mathcal{U}_i(\mathcal{J} \cup \{j\}, \mathbf{t}) - \mathcal{U}_i(\mathcal{J}\backslash\{j\}, \mathbf{t}) \big)$$
$$\Delta_j \mathcal{U}_i(\mathcal{J}, \mathbf{t}) \equiv \mathcal{U}_i(\mathcal{J} \cup \{j\}, \mathbf{t}) - \mathcal{U}_i(\mathcal{J}\backslash\{j\}, \mathbf{t})$$

The iterative procedure builds the lower bound $\underline{\mathcal{J}}$ and upper bound $\overline{\mathcal{J}}$ by successive application of the squeezing step.

$$\underline{\mathcal{J}}^{(k+1)} = \Phi(\underline{\mathcal{J}}^{(k)}, \mathbf{t}) \qquad\qquad \overline{\mathcal{J}}^{(k+1)} = \Phi(\overline{\mathcal{J}}^{(k)}, \mathbf{t}) \tag{22}$$

Under some conditions – complementarity, as defined next section – this sequential procedure yields two sets $\underline{\mathcal{J}}$ and $\overline{\mathcal{J}}$ such that $\underline{\mathcal{J}} \subseteq \mathbb{J} \subseteq \overline{\mathcal{J}}$. In some cases $\underline{\mathcal{J}} = \overline{\mathcal{J}} = \mathbb{J}$ which gives the optimal coalition. If not, with $\overline{\mathcal{J}}\backslash\underline{\mathcal{J}} = \mathcal{J}^{pot}$, we find the optimal coalition by searching exhaustively over all coalitions in:

$$\mathcal{J}^{rem} = \{ \mathcal{J} \,|\, \mathcal{J} = \underline{\mathcal{J}} \cup \hat{\mathcal{J}}, \text{ with } \hat{\mathcal{J}} \in \mathcal{P}(\mathcal{J}^{pot}) \}$$

Applicability of the squeezing procedure

From the combinatorial discrete choice literature, Arkolakis, Eckert and Shi (2023), we know that the squeezing procedure applies in cases where the model exhibit "complementarity" or single-crossing differences in choices.

Indeed, we say that the objective $\mathcal{W}(\mathcal{J})$ obeys the property of *single crossing differences in choice* (SCD-C) from below if:

$$\Delta_j \mathcal{W}(\mathcal{J}) \geq 0 \quad \Rightarrow \quad \Delta_j \mathcal{W}(\mathcal{J}') \geq 0 \qquad \text{for } \mathcal{J} \subset \mathcal{J}' \quad \& \quad j \in \mathbb{I}$$

A simple sufficient condition for SCD-C, from below to be respected is that the marginal value of the objective is monotone in the set $\mathcal{J}$, also called "complementarity":

$$\Delta_j \mathcal{W}(\mathcal{J}) \leq \Delta_j \mathcal{W}(\mathcal{J}') \qquad \text{for } \mathcal{J} \subseteq \mathcal{J}' \qquad \& \quad j \in \mathbb{I}$$

Theorem (Arkolakis, Eckert and Shi (2023)) The SCD-C from below is *sufficient* for the application of squeezing algorithm to yield $\underline{\mathcal{J}} \subseteq \mathbb{J} \subseteq \overline{\mathcal{J}}$ in standard CDCPs.

In this setting, considering participation constraints requires to adjust the welfare objective, from $\mathcal{W}(\mathbb{J})$ to $\widetilde{\mathcal{W}}(\mathbb{J})$ as in eq. (17). In this context, the single crossing differences in choice with participation constraints (SCD-C, PC) take an intricate form, which we detail below:

The ...

The following condition is sufficient for (SCD-C, PC) and provides intuitions on the trade-offs at play in the construction of the optimal coalition:

$$\begin{cases} \Delta_j \mathcal{W}(\mathcal{J}, \mathbf{t}) \leq \Delta_j \mathcal{W}(\mathcal{J}', \mathbf{t}) \\ 0 \leq \Delta_i \mathcal{U}_i(\mathcal{J} \cup \{j\}, \mathbf{t}) \leq \Delta_i \mathcal{U}_i(\mathcal{J}' \cup \{j\}, \mathbf{t}) \qquad \forall\, i \in \mathcal{J} \cup \{j\} \;\&\; i \in \mathcal{J}' \cup \{j\} \\ 0 \leq \Delta_i \mathcal{U}_i(\mathcal{J}, \mathbf{t}) \leq \Delta_i \mathcal{U}_i(\mathcal{J}', \mathbf{t}) \qquad \forall\, i \in \mathcal{J} \;\&\; i \in \mathcal{J}' \end{cases} \tag{23}$$

or

$$\begin{cases} 0 \leq \Delta_i \mathcal{U}_i(\mathcal{J} \cup \{j\}, \mathbf{t}) \leq \Delta_i \mathcal{U}_i(\mathcal{J}' \cup \{j\}, \mathbf{t}) \qquad \forall\, i \in \mathcal{J} \cup \{j\} \;\&\; i \in \mathcal{J}' \cup \{j\} \\ \exists i \in \mathcal{J} \Delta_i \mathcal{U}_i(\mathcal{J}, \mathbf{t}) < 0 \quad \& \quad \exists i \in \mathcal{J}' \quad \Delta_i \mathcal{U}_i(\mathcal{J}', \mathbf{t}) < 0 \end{cases} \tag{24}$$

$$\text{for } \forall\, \mathcal{J} \subseteq \mathcal{J}' \subseteq \mathbb{I} \qquad \& \quad j \in \mathbb{I}$$

This sufficient condition states either one of these cases is verified: (1) in the first case of eq. (23), the marginal welfare $\Delta_j \mathcal{W}$ is monotone in $\mathcal{J}$: the welfare gain of adding country j grows with the size of the coalition $\mathcal{J}$. Moreover, the participation constraint of each member i is still respected when we include country j, and this monotonically in the coalition, from $\mathcal{J}$ to $\mathcal{J}'$, and the coalition is also stable without j. (2) In the second case of eq. (24), we do not require any condition of global welfare, but the participation constraint of each member i is respected when including country j, while it is violated when j is not present in $\mathcal{J}$ and $\mathcal{J}'$. Either one of these two conditions needs to be respected for every pairs of sets $\mathcal{J} \subset \mathcal{J}'$ and every country j.

This condition, as well as its weaker counterpart in appendix (SCD-C-PC), are sufficient conditions for SCD-C from below for $\widetilde{\mathcal{W}}$. It shows that the requirements for coalition building are much stronger as they need to verify if adding marginal members still satisfies the participation constraints of all the incumbent members. In this context, the modified squeezing steps account for such constraint and thus:

Theorem The SCD-C-PC from below is *sufficient* for the application of modified squeezing algorithm, i.e. successive application of eq. (21), starting from $\{\varnothing, \mathbb{I}\}$ and eq. (22), to yield bounding sets $\underline{\mathcal{J}} \subseteq \mathbb{J} \subseteq \overline{\mathcal{J}}$ in CDCPs with participation constraints.

One of the advantages of this setting is that, for a small number of countries $\#\mathbb{I} \approx 10$, we can evaluate numerically if the sufficient conditions mentioned above are satisfied. The disadvantage of the model displayed above is that the large amount of heterogeneity, general equilibrium effects through energy markets, and international trade, prevent the simple evaluation of those sufficient conditions analytically.