YALE UNIVERSITY

SENIOR ESSAY IN ECONOMICS

April 6, 2022

Unilateral Carbon Policies and Multilateral Coalitions

An Analysis of Coalition Stability under the Optimal Unilateral Policy

Author:

MICHAEL BARRESI[1]

Advisor:

PROF. SAMUEL KORTUM

Abstract

Kortum and Weisbach (2021) propose a general equilibrium model for carbon taxation and derive the optimal policy as a function of marginal damages from emissions. This paper serves as an extension of Kortum and Weisbach by combining the general structure of Kortum and Weisbach with the Eaton and Kortum (2002) model of trade. Our results generalize Kortum and Weisbach to a multi-country world where the coalition consists of many heterogeneous countries. Like Kortum and Weisbach, our optimal policy consists of i) an extraction tax, ii) a partial border adjustment that partially shifts the tax imposition downstream and reduces leakage, and iii) an export policy that simultaneously exploits inelastic demand for goods amongst noncoalition countries and expands the export margin. Because the export policy functions as an external penalty incentivizing participation, we interpret coalitions implementing the optimal policy as climate clubs akin to those orignally proposed by Nordhaus (2015). We find that coalitions that implement the optimal unilateral policy can be unstable and that coalition instability has a considerable impact on the effectiveness of the optimal unilateral policy in reducing global emissions.

[1]I would like to express my sincerest gratitude to Professor Kortum, whose invaluable assistance, support, and mentorship over the past two years as my professor, research advisor, and thesis advisor made this paper possible. I would also like to thank my parents, sister, uncle, grandparents, and friends, including Chris, Robert, Mark, Daniel, and Michelle, who supported and encouraged me throughout this process. Finally, I'd like to thank Professor Patrick Devlin, who taught me to find curiosity at the intersection of rigor and intuition, and the late Bob Turansky, who taught me that knowledge without understanding is hollow. All errors are my own.

1 Introduction

The atmosphere knows no borders: when individual countries burn fossil fuels and release carbon dioxide, the entire world experiences the damages, representing a signficant negative externality on the global economy. However, in a world with 195 countries that range from $\sim 0.00005\%$ to $\sim 25\%$ of global GDP, there is a large wedge between locally optimal emissions and globally optimal emissions. This causes a *free-rider problem*: countries are able to enjoy the benefits of the abatement efforts of other countries without implementing any policy of their own, incentivizing countries to emit higher quantities of energy than globally optimal.

Kortum and Weisbach (2021) provides a response to these difficulties. It derives the optimal carbon policy for a taxing region, called Home, in a two-country model in which the non-taxing region, called Foreign, behaves passively, as if in competitive equilibrium. Their solution optimizes the welfare of Home by spreading out a carbon tax over fossil fuel extraction and final goods production in order to minimize overall leakage, an effect where energy supply and demand in Foreign increases due to Home's tax policy. Kortum and Weisbach's solution also includes a tax on final goods exported from Home to Foreign that arises as a dominant pricing strategy in a static, one-period model of trade. Although not their original purpose, the export policy of Home serves a second function as an incentive for Foreign to join Home and implement a joint policy. The idea is that by joining the coalition and implementing the globally optimal policy alongside Home, Foreign can stop Home from enforcing the export penalty, which improves their terms of trade.

In the two-country model, it is difficult to disentangle the incidental terms-of-trade benefits of participation from the taxation costs of participation. Our paper resolves this obstacle by integrating the Kortum and Weisbach model of carbon taxation with the Eaton and Kortum (2002) model of trade. Our model functions as an explicit extension of Kortum and Weisbach to a multi-country world, where Home and Foreign are coalitions comprised of heterogeneous countries each facing their own unique participation decision.[2] We derive the analytic general equilibrium solution for the optimal carbon policy in the multi-country world, generalizing the results of Kortum and Weisbach. Like Kortum and Weisbach, we find that the optimal policy consists of i) a Pigouvian tax on fossil fuel extraction equal to the collective marginal damages of emissions for all countries within the coalition, ii) a partial border adjustment on energy imported by coalition countries from noncoalition countries, iii) a border adjustment, applied at the same rate as the border adjustment on energy, on the energy embodied in the production of final goods imported by coalition countries from noncoalition countries, and iv) an export tax and subsidy policy designed to simultaneously exploit inelastic demand for goods in noncoalition countries and expand low-carbon exports from coalition countries.

Because our optimal tax policy is a generalization of Kortum and Weisbach, we do not repeat their in-depth analysis of the policy. Instead, we leverage the multi-country model to interpret the coalition's export policy (iv) as an external penalty that orients a coalition implementing the optimal tax policy as a climate club, similar to the ones outlined in Nordhaus (2015).[3] The export policy provides an incentive for every country in the world to participate in the coalition, convincing countries currently participating to stay and convincing countries not currently participating to join. Throughout the paper, we analyze the extent to which this trade incentive combats the free-rider problem. Calibrating our results to OECD and IEA data, we quantify the welfare of each country under the optimal policy and identify three primary determinants of a country's willingness to participate in a coalition:[4]

1. Countries that are highly autarkic are less incentivized to participate in a coalition because the coalition's export policy (iv) plays a less impactful role for these countries.

[2] Although we formally cite Kortum and Weisbach (2021), our paper is a more direct extension of an earlier draft of that paper, Weisbach, Kortum, and Wang (2020).

[3] Unless otherwise specified, further references to Nordhaus refer to his 2015 paper.

[4] All code and raw data is available at https://drive.google.com/drive/folders/1l-mnj6bqEXLSGyTOu1M0KpG5s0UkgkpS?usp=sharing.

2. Countries that are highly intensive in energy extraction are less incentivized to participate in a coalition since the extraction tax (i) is more costly for these countries.

3. Countries that are very small are less incentivized to participate in a coalition because there is a larger wedge between the locally-optimal and coalition-optimal tax policy for these countries, amplifying the temptation to free-ride.

Our results are extremely sensitive to the marginal damages from emissions for countries within the coalition, since the marginal damages from emissions directly control the extraction tax (i) and heavily influence the rest of the optimal policy (ii), (iii), (iv). We find that the relative importance of determinants 1, 2, and 3 varies depending on the marginal damages from emissions. At small marginal damages from emissions, the extent to which a country is autarkic is the dominant determinant; at moderate to high marginal damages, a country's intensity in energy extraction is the dominant determinant; at very high to extreme marginal damages, the size of a country's economy is the dominant determinant.

Because marginal damages from emissions play such a critical role in determining which countries are incentivized to participate in a climate coalition, we continue our analysis by determining the largest stable coalition as a function of emissions aversion, a parameter we use to normalize marginal damages of emissions and serves as a rough proxy for the social cost of carbon.[5] [6] As emissions aversion increases, countries within the coalition become more concerned about carbon emissions, so they choose to tax carbon more heavily. The aggressiveness of the coalition's policy at high levels of emissions aversion exacerbates the free-rider problem, because it amplifies the wedge between the locally-optimal and the coalition-optimal policy. As a result, we find that countries become less incentivized to join the coalition and coalitions become less stable as emissions aversion increases. Eventually, all but the world's largest countries refuse to participate in the coalition, because countries

[5]We define the largest stable coalition as the largest coalition for which every country within the coalition wishes to stay in the coalition, and every country not in the coalition does not wish to join.

[6]See section 4.1.2 for more details on how marginal damages from emissions and emissions aversion relate to the social cost of carbon.

whose own economies make up an insufficiently large portion of the coalition would rather suffer the trade penalties than internalize the full externality of the entire coalition. Further, because the incidence of the coalition's policy is highly unequal, the critical emissions aversion threshold at which each country leaves the coalition varies dramatically according to determinants 1, 2, and 3. Our simulation suggests that coalition instability makes it very difficult to generate global emissions reductions beyond about 40%. Moreover, the vast majority of emissions reductions are achieved at low to moderate levels of emissions aversion when global participation is sustainable. At high levels of emissions aversion, virtually all of the gains achieved by a marginal increase in emissions aversion are offset as countries leave the coalition.

The paper proceeds as follows. In the remainder of Section 1, we discuss the key ideas and prior literature surrounding two primary schools of thought about how to overcome the free-rider problem— optimal unilateral policies and climate clubs– and explain how our model sits at the intersection of these two schools. Section 2 sets up our multi-country model, applying the Eaton and Kortum model of trade to Kortum and Weisbach. In Section 3, we solve the Coalition Planner's problem for our setup and discuss the set of taxes that yield the optimal outcome in equilibrium. In Section 4, we calibrate the theoretical solution to the data and analyze the implications of the optimal policy on the welfare of individual countries, coalition stability, and the potential for a coalition implementing the optimal policy to reduce global emissions. Section 5 briefly summarizes our results and discusses their implications. Section 6 concludes.

1.1 Background and Prior Literature: Two Schools of Thought

There are two primary schools of thought about how to overcome the free-riding problem and reduce global carbon emissions. The first school, reflecting the struggles of global negotiations, gives up on achieving global cooperation. Instead, it focuses on the second-best option: it assumes that some, but not all of the world, implements a unilaterally-optimal policy and seeks to optimize that policy by minimizing leakage. The second school maintains hope for global cooperation above and beyond the loose agreements of the Kyoto Protocol and Paris Accords, suggesting that widespread and substantial cooperation is possible with a more targeted and aggressive tax policy. This school promotes the establishment of climate clubs that enforce external penalties on non-members as a means of incentivizing participation. Our paper begins to build a bridge between these two schools. We analyze how effective a climate club consisting of some, but not all of the world, is at incentivizing participation and curbing global emissions when it implements the optimal unilateral policy of that coalition. As such, our paper adds to the literature of both schools.

1.1.1 School 1: Curbing Global Emissions with the Optimal Unilateral Policy

The first school makes no explicit attempt to incentivize countries to participate in the coalition, instead treating the participation status of each country or region as exogenous. Given a starting coalition, this school debates the optimal unilateral tax policy under the assumption that nonparticipating countries will act passively and implement no carbon policy. Because there is a nearly perfect 1-1 relationship between fossil fuel extraction and carbon emissions, in a closed economy, all carbon taxes of equal quantities yield identical economic impacts regardless of their imposition. In an open economy, however, the incidence of the tax may vary depending on its imposition. This is the problem that the optimal unilateral policy school faces: in an open world featuring a taxing region and a non-taxing region, what is the best tax implementation?

This problem is particularly difficult because of the sheer number of possible taxes that the coalition can choose from. Carbon taxes can take the form of extraction taxes, where extractors pay a tax on the fossil fuels they dig out of the ground; production taxes, where producers of final goods pay a tax on the energy they use as an intermediate input; consumption taxes, where consumers of a final good pay a tax on the energy that went into the production of that good; or any combination of the three.[7] Each of these taxes has its own unique implications. Consider, for example, an extraction tax t^E that reduces the price that extractors within the taxing coalition receive to $p^E - t^E$, where p^E is the post-tax price of energy. Assuming the incidence of this extraction tax does not fall entirely on the extractor, the tax will raise p^E above competitive equilibrium levels. If energy can be costlessly traded between the taxing region and the non-taxing region, this must raise the price of energy in the non-taxing region to the new, higher level of p^E. This causes extractors in the non-taxing region to pull more energy out of the ground, diminishing the effects of the extraction tax. A similar effect occurs if a tax is imposed on energy when it is used as an intermediate good in production. A production tax will increase the cost of goods produced within coalition countries relative to the cost of goods produced within noncoalition countries. In an open economy, this shifts the location of production away from the taxing region to the non-taxing region, once again diminishing the effectiveness of the tax. We call the shift in extraction and/or production that occurs due to a carbon tax in an open economy *leakage*. An extraction tax causes extraction leakage by raising the global price of energy, increasing energy extraction in the non-taxing region. A production tax causes production leakage by raising the cost of final goods produced in the taxing region, increasing the imports of goods by the taxing region from the non-taxing region.[8] Leakage is often the primary measure of the efficiency of a tax policy and a central question to tax design (Fowlie 2009).

Countries can also enhance their tax policy via border adjustments. Border adjustments, which are taxes on imports or rebates of taxes paid on exports, can be added to the main carbon tax to minimize leakage. When border adjustments are applied directly to fossil

[7]So far, most carbon policies, such as the European Union Emissions Trading System, have relied on production taxes. Consumption taxes are the least common because they tend to be difficult to implement due to the sheer number of final goods and the complexity of determining the exact quantity of energy that went into their production. The exception to this is "at the pump" gasoline taxes, which represent a form of consumption taxes.

[8]In the carbon tax literature, the term leakage often refers specifically to what we call production leakage. We use the term production leakage to disambiguate between production leakage and extraction leakage, both of which play critical roles in our optimal policy.

fuels or final goods (or both), they shift a carbon policy downstream. For example, consider a border adjustment t^B on energy that is applied whenever energy is imported from the non-taxing region and rebated whenever energy is exported to the non-taxing region. This border adjustment guarantees that producers within the taxing-region always face energy price $p^E + t^B$ and producers outside the taxing-region always face energy price p^E regardless of the location of extraction. If the border adjustment is full— that is, if it is identical to the extraction tax that coalition extractors face— then the extraction tax alongside the border adjustment is identical to a pure production tax. If the border adjustment is partial— that is, less than the extraction tax— then the extraction tax alongside the border adjustment is a hybrid tax on extraction and production.

The optimal unilateral policy school seeks to derive the optimal set of taxes and border adjustments from the myriad of possiblities described above. Markusen (1975) pioneered this school by deriving the optimal Pigouvian tax and border adjustment to curb the pollution of Lake Erie in a two-country model featuring a dirty good and a passive polluting country. Hoel (1994) derives a similar result, finding that the optimal carbon tax in the presence of free riders consists of a mixture of an extraction tax, moderating coalition fossil fuel supply, and a production tax, moderating coalition fossil fuel demand.

Kortum and Weisbach combines Markusen's model of fossil fuel extraction and trade with the Dornbusch, Fischer, and Samuelson (1977) model of production and consumption, letting energy serve as an intermediate good used in the production of final goods. As previously discussed, Kortum and Weisbach find that the optimal policy consists of a mixture of a Pigouvian extraction tax and a border adjustment that partially shifts the tax to production and curbs emissions in the non-taxing region. This result is in line with previous literature, such as Branger and Quiron (2014) and Bohringer et. al (2012), which find that border adjustments diminish overall leakage. The border adjustment reduces production leakage because it causes noncoalition producers to internalize the same cost of energy as coalition producers when exporting goods to the coalition, reducing the coalition's imports of dirty goods. The most innovative aspect of Kortum and Weisbach's solution is that it also seeks to reduce production leakage by increasing coalition exports of clean goods. Kortum and Weisbach's optimal policy features a unique export policy that subsidizes coalition producers, counterbalancing the comparative disadvantage of coalition producers.

1.1.2 School 2: Curbing Global Emissions with Climate Clubs

A second school, most famously associated with Nordhaus (2015), seeks to overcome the free-rider problem by establishing climate clubs. As described by Nordhaus, a climate club consists of an agreement amongst club members to 1) undertake harmonized emissions reductions and 2) penalize non-members. In the Nordhaus model, emissions reductions are achieved by demanding that club members implement policies that produce a target domestic energy price, though it is up to individual countries to choose a specific tax or cap-and-trade policy that achieves this domestic energy price. The climate clubs also impose uniform import tariffs that serve as the external penalties. This trade policy is fundamentally different from the border adjustments featured in the optimal unilateral policy literature both in function and in purpose. Unlike border adjustments, the tariffs imposed by climate clubs are *external penalties* because they exist outside of the public-goods nature of carbon emissions. While only internal penalties help minimize leakage, the climate club relies on external penalties as a stronger incentive for non-members to join the club. Nordhaus finds that external penalties are a necessary and sufficient condition to generate a stable climate club, and that climate clubs can be effective in raising the global energy price to target levels of up to $50 per ton of CO_2. Farrid and Lashkaripour (2021) also finds that climate clubs are highly effective: their model suggests that climate clubs are capable of reducing global emissions by up to 61%, dramatically more than emissions reductions under optimal border adjustments.

1.1.3 Our Model: A Hybrid Approach

Our model bridges the optimal unilateral policy school and the climate club school. As an extension of Kortum and Weisbach, our paper directly adds to the optimal unilateral policy literature. Our results generalize Kortum and Weisbach to a multicountry world, demonstrating that previous findings about the optimal relationship between extraction taxes and border adjustments hold true in the broader case of a unilateral taxing region composed of several heterogeneous countries.

Kortum and Weisbach's solution, generalized here, features an innovative export policy that subsidizes or taxes coalition producers so that they export goods at the cost of production of the cheapest noncoalition producer. When coalition producers are at a comparative disadvantage, the coalition offers a subsidy to expand the coalition's export margin, reducing leakage; when coalition producers are at a comparative advantage, the coalition taxes coalition producers to exploit inelastic demand for goods as a dominant pricing strategy in a one-period game. However, these taxes incidentally harm noncoalition countries, causing the export policy implemented by the coalition to function as an external penalty imposed on noncoalition countries. This penalty is different in implementation than the import tariffs proposed by Nordhaus but has a similar impact: it incentivizes noncoalition countries to join the coalition in order to improve their terms of trade. Consequently, our coalitions function as climate clubs.

Intuitively, our hybrid approach takes a middle ground between the competing perspectives of the optimal unilateral policy school and climate club school. It acknowledges that because uniform global participation is unlikely, a coalition can enhance its effectiveness by implementing a policy that minimizes leakage while also recognizing the fact that coalition participation can be enhanced using external penalties. However, our model is limited to the extent that it only analyzes the effectiveness of a coalition implementing the optimal unilateral policy as a climate club. We optimize for leakage concerns, include external penalties, and observe the impact on coalition stability, but we do not seek to derive the policy that optimally balances leakage concerns and participation concerns.

2 Basic Model

Our model consists of a world W with an arbitrary number of countries and three sectors—services S, energy E, and goods G, that can be consumed, produced, and exported. We use several variables to describe various aggregate measures and superscripts to specify the sector being described. For example $\left(C^S, C^G, C^E\right)$ refer to the consumption of services, goods, and energy, respectively. Production is denoted $\left(Q^S, Q^G, Q^E\right)$, labor used in production $\left(L^S, L^G, L^E\right)$, and net exports $\left(X^S,\ X^G, X^E\right)$. We use lowercase subscripts to denote a country acting as a consumer or a producer. Countries acting as consumers are denoted using the subscript n, while countries acting as producers are denoted using the subscript i. For example, C_n^G refers to country n's consumption of goods, and Q_i^E refers to country i's production, or extraction, of energy. In some circumstances, such as when we evaluate the overall utility of a country, we want to express a country that acts as both a consumer and a producer in the same expression. In these cases, we use a lowercase subscript w. We also use w to denote a secondary consumer or producer when we need to describe multiple countries of the same type in the same expression.

The utility of each country $w \in W$ is a function of C_w^S, C_w^G, and Q_W^E:

$$U_w = C_w^S + \eta_w^{1/\sigma_w} \frac{(C_w^G)^{1-1/\sigma_w} - 1}{1 - 1/\sigma_w} - \varphi_w \sum_{i \in W} Q_i^E$$

In the utility function, the parameter σ_w represents the elasticity of substitution across different types of goods, which we elaborate on in Section 2.2. The parameter η_w represents the relative weight country w places on consumption in its utility function, and φ_w reflects country w's marginal damages from global carbon emissions. Note that U_w depends on $\sum_{i \in W} Q_i^E$, as opposed to just Q_w^E, reflecting fact that country w is harmed by the emissions of every country in the world, not just their own emissions.

2.1 Services Sector

Services production is equal to the labor used in services production:

$$Q_i^S = L_i^S \tag{1}$$

We assume that laborers can move freely between sectors, so that the laborers working in the services, goods, and energy sectors receive the same wage.

Services are traded freely across each country in W. This pins down a common cost of services across W, and by extension, a common cost of labor. We set this cost to be the numeraire, so all laborers earn wage equal to 1 and all services trade at a price of 1. Services consumption is linked to services trade and services production via:

$$\sum_{n \in W} X_{nw}^S + C_w^S = Q_w^S \tag{2}$$

where X_{nw}^S represents the net exports of services from country w to country n.

2.2 Goods Sector

There are a continuum of goods $j \in [0, 1]$, and aggregate consumption over these goods is given by a CES function:

$$C_n^G = \left(\int_0^1 c_n(j)^{(\sigma_n - 1)/\sigma_n} dj \right)^{\sigma_n/(\sigma_n - 1)} \tag{3}$$

Countries produce goods using labor and energy inputs, which they customize for each consumer and each good. We denote the quantity of good j that country i produces for country n as $q_{ni}(j)$, using the first subscript to specify the consumer and the second subscript to specify the producer. We organize multiple subscripts in this manner throughout the paper. The production function $q_{ni}(j)$ is Cobb-Douglas, dependent on the efficiency of country i in producing good j, denoted $A_i(j)$; the labor country i uses to produce good j for country n, denoted $L_{ni}(j)$; and the energy country i uses to produce good j for country n, denoted $C^E_{ni}(j)$:

$$q_{ni}(j) = \frac{A_i(j)}{\nu}\left[L_{ni}(j)\right]^{\alpha}\left[C^E_{ni}(j)\right]^{1-\alpha} \tag{4}$$

Here, $0 < \alpha < 1$ is a parameter representing the unit output elasticity of labor and $\nu = \alpha^{\alpha}(1-\alpha)^{1-\alpha}$. It is convenient to think about production in terms of energy intensity, which we will rely on throughout the paper. Energy intensity is defined as the ratio of energy to labor used in production and can vary for each producer, consumer, and good combination:

$$z_{ni}(j) = \frac{C^E_{ni}(j)}{L_{ni}(j)}$$

We can invert the production function (4) to derive unit energy and labor requirements as a function of energy intensity:

$$e_i(j, z_{ni}(j)) = \frac{\nu}{A_i(j)} z_{ni}(j)^{\alpha}$$

$$l_i(j, z_{ni}(j)) = \frac{e_i(j, z_{ni}(j))}{z_{ni}(j)} = \frac{\nu}{A_i(j)} z_{ni}(j)^{\alpha-1}$$

Following the Eaton and Kortum model of trade, $A_i(j)$ is probabilistic, depending on T_i, which represents the relative efficiency of country i in goods production, and θ, which is univeral to all countries and governs the variance in production efficiency across goods. The CDF of $A_i(j)$ is given by:

$$F^A_i(a) = Pr\left[A_i(j) \leq a\right] = e^{-T_i a^{-\theta}} \tag{5}$$

When a producer i sends goods to a consumer n, it faces an iceberg cost $d_{ni} \geq 1$ that is independent of the specific good j.[9] The amount of good j originating from a producer i that a country n actually consumes, denoted $c_{ni}(j)$, depends on this iceberg cost and the quantity produced:

$$c_{ni}(j) = \frac{q_{ni}(j)}{d_{ni}} \tag{6}$$

Trade in goods can be two-directional for two reasons. Because goods are heterogeneous, country i may have a comparative advantage to country n in the production of some goods and a comparative disadvantage in others. Consequently, country i may export some goods to country n and import others. Furthermore, depending on the energy intensities used in production, country i may be incentivized to crosshaul— both import and export the same good to a country n— if doing so gives country i greater control of the energy intensity used to produce the good in both countries. Because of this, we split up net exports in goods from country i to country n into an exports term and an imports term:

$$X_{ni}^G = V_{ni}^G - V_{in}^G$$

V_{ni}^G and V_{in}^G refer to the total value of goods country i exports to country n and imports from country n, respectively.

2.3 Energy Sector

We assume that for an extractor i, energy deposits are distributed in a continuum D_i, where the amount of energy that country i can extract at a unit labor requirement below a is given by $D_i(a)$. Assuming efficient extraction, the labor country i invests in energy extraction is given by:

$$L_i^E = \int_0^{\bar{a}} a \; dD_i(a)$$

[9] The variable d is used to represent distance, but d_{ni} is best interpreted as a transportation cost of $\frac{(d_{ni}-1)}{100}\%$ on goods produced by country i for country n. While related to physical distances, this variable can encompass other factors as well.

and the amount of energy extracted by country i is:

$$Q_i^E = D_i(\bar{a})$$

Energy can be traded between any two countries n, i at a global energy price p^E with no iceberg cost.

2.4 The Labor Market

Each country w is endowed with a fixed labor supply L_w, and can choose how to allocate that labor supply to the services, goods, and energy sectors subject to the constraint:

$$L_w = L_w^S + L_w^G + L_w^E \tag{7}$$

2.5 The Energy Market

In Section 2.3, we outlined Q_i^E, the energy supply schedule for a producer i. Countries demand energy in order to use it as an intermediate input in the production of final goods. We derive the total amount of energy country i uses in goods production by calculating the amount of energy country i uses to produce each good for each consumer, integrating over all goods $j \in [0, 1]$, and summing over all consuming countries $n \in W$:[10]

$$C_i^E = \sum_{n \in W} \int_0^1 e\big(j, z_{ni}(j)\big) q_{ni}(j) \; dj \tag{8}$$

Because energy can be traded between countries, the energy market does not necessarily need to clear on the country-level. Instead, we have:

$$\sum_{n \in W} X_{nw}^E + C_w^E \leq Q_w^E \tag{9}$$

[10]Here, we use the subscript i because country i is acting as a producer of final goods, even though we are interested in how much energy country i consumes as an intermediate good.

where X_{nw}^E is the net exports of energy from country w to country n. However, the global energy market must clear, imposing a global energy constraint:

$$\sum_{i \in W} Q_i^E = \sum_{n \in W} C_n^E$$

2.6 Trade Balance

Each producer i must ultimately have net 0 trade balance with the rest of the world:[11]

$$\sum_{n \in W} \left(X_{ni}^S + X_{ni}^G + p^E X_{ni}^E \right) = 0 \tag{10}$$

3 Coalition Planner's Problem

3.1 Description and Preliminaries

We now describe a coalition planner's problem, where a social planner seeks to optimize collective welfare for a set of countries $W' \subset W$, while a second set of countries, $W^* = W \backslash W'$ behaves passively as if in competitive equilibrium. Countries within the coalition are denoted (w', n', i'), while countries outside the coalition are denoted (w^*, n^*, i^*).

On the energy supply side, the coalition planner controls how much carbon each coalition country extracts. On the energy demand side, the coalition planner controls consumption quantities and energy intensities for three of the four possible directions of good flow: from coalition producers to coalition consumers, from coalition producers to noncoalition consumers, and from noncoalition producers to coalition consumers. The planner does not directly control good flow from noncoalition producers to noncoalition consumers, but he can indirectly influence this value by setting good flow for the other three directions.

Noncoalition producers sell goods at their cost of production regardless of the consumer's

[11] Because we previously defined $X_{ni}^G = V_{ni}^G - V_{in}^G$, X_{ni}^G represents the net value of goods traded from country i to country n rather than the net quantity of goods. Because of this, we do not need an additional price term scaling X_{ni}^G in the trade balance equation.

coalition status. This price is a function of the cost of energy inputs, the cost of labor inputs, and the iceberg cost:

$$p_{ni^*}(j) = d_{ni^*}\Big[l_{i^*}\big(j, z_{ni^*}(j)\big) + p^E \; e_{i^*}\big(j, z_{ni^*}(j)\big)\Big]$$

The planner dictates the price at which coalition producers sell their goods to noncoalition consumers, though it is bounded by the price the consumer can get from the lowest-cost noncoalition producer. We discuss the coalition planner's choice for this price in Section 3.4.2.

3.2 Formulation

The planner's utility function is a simple sum of the utilities of each country $w' \in W'$:

$$U_{W'} = \sum_{w' \in W'} (C^S_{w'} + \eta_{w'}^{1/\sigma_{w'}} \frac{(C^G_{w'})^{1-1/\sigma_{w'}}}{1 - 1/\sigma_{w'}} - \varphi_{w'} \sum_{i \in W} Q^E_i)$$

This utility function is subject to a series of constraints: each country must satisfy the labor constraint (7), the energy constraint (9), and trade balance constraint (10). We substitute in the labor and trade balance constraints directly by rewriting services consumption using (1), (2) and these constraints. Summing (2) across coalition countries gives:[12]

$$\sum_{w' \in W'} C^S_{w'} = \sum_{w' \in W'} \Big(Q^S_{w'} - \sum_{n^* \in W} X^S_{n^* w'}\Big)$$

[12]Note that trade in services between two coalition countries (n', i') cancels, since $X^S_{n'i'} = -X^S_{n'i'}$, and thus these terms do not appear in the equation

We account for the labor constraints by substituting (1) and (7) in for $Q^S_{w'}$ and account for the trade balance constraints by substituting (10) in for $\sum_{n\in W} X^S_{nw'}$: [13]

$$\sum_{w'\in W'} C^S_{w'} = \sum_{w'\in W'} \left[L_{w'} - L^E_{w'} - L^G_{w'} + \sum_{n^*\in W^*} \left(X^G_{n^*w'} + p^E X^E_{n^*w'} \right) \right] \tag{11}$$

To account for the energy constraints, we add Lagrange Multipliers ($\lambda_{w'}$ or λ_{w^*}), which represent the shadow cost of energy for each country w:

$$\begin{aligned} \mathcal{L} = \sum_{w'\in W'} \Bigg[\eta_{w'}^{1/\sigma_{w'}} \frac{(C^G_{w'})^{1-1/\sigma_{w'}}}{1-1/\sigma_{w'}} - \varphi_{w'} \sum_{i\in W} Q^E_i + L_{w'} - L^G_{w'} - L^E_{w'} + \sum_{n^*\in W^*} (X^G_{n^*w'} + p^E X^E_{n^*w'}) \Bigg] \\ - \sum_{w'\in W'} \left[\lambda_{w'} \left(C^E_{w'} + \sum_{n\in W} X^E_{nw'} - Q^E_{w'} \right) \right] - \sum_{w^*\in W^*} \left[\lambda_{w^*} \left(C^E_{w^*} + \sum_{n\in W} X^E_{nw^*} - Q^E_{w^*} \right) \right] \end{aligned}$$

Solving this Lagrangian will give us the optimal carbon policy of the coalition planner. We do this via four main steps. In Section 3.3, we take the FOCs with respect to energy exports to relate the multitude of energy shadow costs to each other and to the energy price, simplifying our problem to two degrees of freedom– a global energy price and shadow cost of energy. In Section 3.4, we solve what Kortum and Weisbach, call "The Inner Problem."[14] We rewrite the aggregates $C^G_{w'}$, $C^E_{w'}$, $L^G_{w'}$, and $X^G_{n^*w'}$ in terms of the myriad of consumption quantities $c_{ni}(j)$ and energy intensities $z_{ni}(j)$. Temporarily fixing the global energy price and shadow cost of energy, we take the FOC with respect to each consumption quantity and energy intensity to derive their optimal values. Re-aggregating these values yields global energy demand as a function of the global energy price and shadow cost of energy. In Section 3.5, we solve the Outer Problem, which consists of taking the FOC with respect to Q^E_w to derive global energy supply as a function of energy price and shadow cost of ene4rgy. Equating global energy supply and demand gives us one equation linking global energy price to the

[13]Equation (11) is identical if we relax the labor and trade constraints to only hold at the coalition level rather than the country level. This implies that the optimal solution holds for the more general case where trade and labor are unconstrained on an intracoalition basis. For the purposes of this paper, we focus on the more specific case, as it enables an intuitive and practical determination of utility on the country level.

[14]Kortum and Weisbach's solution of the Inner Problem is based off of the original solution strategy of Costinot, Donaldson, Vogel, and Werner (2015).

shadow cost of energy. Taking the FOC with respect to p^E gives us a second relation, solving the coalition planner's problem.

3.3 Energy Trade No-Arbitrage Conditions

To link p^E with the various Lagrange multipliers we establish several energy trade no-arbitrage conditions by taking the FOCs with respect to energy exports. Between two coalition countries n', i', the FOC is:

$$\frac{\partial \mathcal{L}}{\partial X^E_{n'i'}} = \lambda_{n'} - \lambda_{i'}$$

this implies that the shadow cost of energy is universal throughout the coalition:

$$\lambda_{n'} = \lambda_{i'} = \lambda' \tag{12}$$

Similarly, the FOC between two noncoalition countries n^*, i^* is:

$$\frac{\partial \mathcal{L}}{\partial X^E_{n^*i^*}} = \lambda_{n^*} - \lambda_{i^*}$$

so the shadow cost of energy is the same within noncoalition countries as well:

$$\lambda_{n^*} = \lambda_{i^*} = \lambda^* \tag{13}$$

Finally, we consider energy exports from a coalition country i' to a noncoalition country n^*, relating λ' to λ^*:

$$\frac{\partial \mathcal{L}}{\partial X^E_{n^*i'}} = p^E - \lambda_{i'} + \lambda_{n^*}$$

$$\lambda^* = \lambda' - p^E \tag{14}$$

Equations (12)-(14) collapse the multitude of shadow prices into two degrees of freedom—a global energy price p^E and a shadow cost of energy λ'. This provides the necessary relationship to solve the planner's problem by clearing the energy market and taking the first order condition of the Largrangian with respect to energy price.

3.4 The Inner Problem and Energy Demand

We now calculate the demand for energy as an intermediate good used in production in terms of p^E and λ'. To do this, we rewrite the terms C_i^E and L_i^G as aggregate terms dependent on each $c_{ni}(j)$ and $z_{ni}(j)$:

$$C_i^E = \sum_{n \in W} \int_0^1 e_i\big(j, z_{ni}(j)\big) d_{ni} c_{ni}(j)\ dj \tag{15}$$

$$L_i^G = \sum_{n \in W} \int_0^1 l_i\big(j, z_{ni}(j)\big) d_{ni} c_{ni}(j)\ dj \tag{16}$$

We can also expand $X_{n^*i'}^G$:

$$\begin{aligned} X_{n^*i'}^G &= V_{n^*i'}^G - V_{n'i^*}^G \\ &= \int_0^1 p_{n^*i'}(j) c_{n^*i'}(j) dj - \int_0^1 p_{n'i^*}(j) c_{n'i^*}(j) dj \\ &= \int_0^1 p_{n^*i'}(j) c_{n^*i'}(j) dj - \int_0^1 \Big[l_{i^*}\big(j, z_{n'i^*}(j)\big) + p^E e_{i^*}\big(j, z_{n'i^*}(j)\big) \Big] d_{n'i^*} c_{n'i^*}(j)\ dj \end{aligned} \tag{17}$$

Dropping constants and applying equations (3), (14)-(17), we can rewrite the Lagrangian as:

$$
\begin{aligned}
\mathcal{L} = \sum_{w' \in W'} \Bigg\{ & \frac{\eta_{w'}^{1/\sigma_{w'}}}{1 - 1/\sigma_{w'}} \int_0^1 \big(\sum_{i \in W} c_{w'i}(j)^{1-1/\sigma_{w'}} \big) dj - \varphi_{w'} \sum_{i \in W} Q_i^E \\
& - L_{w'}^E - \sum_{n \in W} \Big(\int_0^1 l_{w'}\big(j, z_{nw'}(j)\big) d_{nw'} c_{nw'}(j) dj \Big) \\
& + \sum_{w^* \in W^*} \Bigg[p^E X_{w^* w'}^E + \int_0^1 p_{w^* w'}(j) c_{w^* w'}(j) dj \\
& - \int_0^1 \Big(l_{w^*}\big(j, z_{w'w^*}(j)\big) + p^E e_{w^*}\big(j, z_{w'w^*}(j)\big) \Big) d_{w'w^*} c_{w'w^*}(j) dj \Bigg] \Bigg\} \\
& - \lambda' \sum_{w \in W} \Bigg[\sum_{n \in W} \Big(\int_0^1 e_w\big(j, z_{nw}(j)\big) d_{nw} c_{nw}(j) dj + X_{nw}^E \Big) - Q_w^E \Bigg] \\
& + p^E \sum_{w^* \in W^*} \Bigg[\sum_{n \in W} \Big(\int_0^1 e_{w^*}\big(j, z_{nw^*}(j)\big) d_{nw^*} c_{nw^*}(j) dj + X_{nw^*}^E \Big) - Q_{w^*}^E \Bigg]
\end{aligned}
$$

The first order conditions with respect to energy intensities and consumption will give us global energy demand as a function of p^E and λ'.

3.4.1 Energy Intensities

For a coalition producer, the FOC for energy intensity $z_{ni'}(j)$ is the same regardless of whether the destination country is in the coalition or not.

$$
\begin{aligned}
\frac{\partial \mathcal{L}(j)}{\partial z_{ni'}(j)} &= d_{ni'} c_{ni'}(j) \Big(- \frac{\partial l_{i'}\big(j, z_{ni'}(j)\big)}{\partial z_{ni'}(j)} - \lambda' \frac{\partial e_{i'}\big(j, z_{ni'}(j)\big)}{\partial z_{ni'}(j)} \Big) \\
&= d_{ni'} c_{ni'}(j) \Big((1-\alpha) \frac{\nu}{A_{i'}(j)} z_{ni'}(j)^{\alpha-2} - \lambda' \alpha \frac{\nu}{A_{i'}(j)} z_{ni'}(j)^{\alpha-1} \Big)
\end{aligned}
$$

The FOC for goods produced by a noncoalition country for a coalition consumer is:

$$
\begin{aligned}
\frac{\partial \mathcal{L}(j)}{\partial z_{n'i^*}(j)} &= d_{n'i^*} c_{n'i^*}(j)\Big(- \frac{\partial l_{i^*}\big(j, z_{n'i^*}(j)\big)}{\partial z_{n'i^*}(j)} - (p^E + \lambda' - p^E)\frac{\partial e_{i^*}\big(j, z_{n'i^*}(j)\big)}{\partial z_{n'i^*}(j)}) \\
&= d_{n'i^*} c_{n'i^*}(j)\Big((1-\alpha)\frac{\nu}{A_{i^*}(j)} z_{n'i^*}(j)^{\alpha-2} - \lambda'\alpha\frac{\nu}{A_{i^*}(j)} z_{n'i^*}(j)^{\alpha-1}\Big)
\end{aligned}
$$

Setting these two FOCs equal to zero gives:

$$
z_{ni'}(j) = z_{n'i^*}(j) = \frac{1-\alpha}{\alpha\lambda'} \tag{18}
$$

This implies that, for all planner-controlled trade, the optimal energy intensity is a fixed $z = \frac{1-\alpha}{\alpha\lambda'}$ regardless of consumer, producer, or good.

3.4.2 Consumption

The FOC for consumption from a coalition producer to a coalition consumer is given by:

$$
\begin{aligned}
\frac{\partial \mathcal{L}(j)}{c_{n'i'}(j)} &= \eta_{n'}^{1/\sigma_{n'}}\big(\sum_{w\in W} c_{n'w}(j)\big)^{-1/\sigma_{n'}} - d_{n'i'} l_{i'}\big(j, z_{n'i'}(j)\big) - \lambda' d_{n'i'} e_{i'}\big(j, z_{n'i'}(j)\big) \\
&= \eta_{n'}^{1/\sigma_{n'}}\big(\sum_{w\in W} c_{n'w}(j)\big)^{-1/\sigma_{n'}} - \frac{d_{n'i'}}{A_{i'}(j)}(\lambda')^{1-\alpha}
\end{aligned}
$$

while the FOC for a noncoalition producer i^* and a coalition consumer n', is:

$$
\begin{aligned}
\frac{\partial \mathcal{L}(j)}{c_{n'i^*}(j)} &= \eta_{n'}^{1/\sigma_{n'}}\big(\sum_{w\in W} c_{n'w}(j)\big)^{-1/\sigma_{n'}} - d_{n'i^*} l_{i^*}\big(j, z_{n'i^*}(j)\big) - (p^E + \lambda' - p^E) d_{n'i^*} e_{i^*}\big(j, z_{n'i^*}(j)\big) \\
&= \eta_{n'}^{1/\sigma_{n'}}\big(\sum_{w\in W} c_{n'w}(j)\big)^{-1/\sigma_{n'}} - \frac{d_{n'i^*}}{A_{i^*}(j)}(\lambda')^{1-\alpha}
\end{aligned}
$$

It is useful to interpret the FOCs with respect to $c_{n'i'}(j)$ and $c_{n'i^*}(j)$ as a marginal benefit less a marginal shadow cost. The marginal benefit, denoted $b_{n'i}(j)$, represents the added utility from a marginal increase in $C^G_{n'}$.

$$b_{n'i}(j) = \eta_{n'}^{1/\sigma_{n'}}(\sum_{w\in W} c_{n'w}(j))^{-1/\sigma_{n'}} \tag{19}$$

The marginal shadow cost, denoted $\tilde{p}_{n'i}(j)$, represents the shadow cost of energy and the opportunity cost of using labor (that would otherwise produce services) to produce good j for country n':[15]

$$\tilde{p}_{n'i}(j) = \frac{d_{n'i'}}{A_{i'}(j)}(\lambda')^{1-\alpha} \tag{20}$$

Notice that the marginal benefit is declining in $c_{n'i}(j)$ but does not depend on the country of origin. The marginal shadow cost, on the other hand, is fixed with respect to consumption but does depend on the country of origin through $\frac{d_{n'i}}{A_i(j)}$. These dependencies imply that country n' should consume good j from the producer i that minimizes marginal shadow cost until $\frac{\partial \mathcal{L}(j)}{\partial c_{n'i}(j)} = 0$ and $\frac{\partial \mathcal{L}(j)}{\partial c_{n'w}(j)} \leq 0$ for all $w \in W$. These results are expressed by the following equations:

$$c_{n'i}(j) = I_{n'i}(j)\eta_{n'}\Big(\frac{A_i(j)}{(\lambda')^{1-\alpha}d_{n'i}}\Big)^{\sigma_{n'}} \tag{21}$$

where

$$I_{n'i}(j) = \begin{cases} 1 & \tilde{p}_{n'i}(j) \leq \tilde{p}_{n'w}(j) \text{ for all } w \in W \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

The FOC for goods that coalition countries export to noncoalition countries is more intricate because the planner can set the price at which coalition producers sell goods to noncoalition consumers. To simplify our analysis, we assume inelastic demand for goods amongst noncoalition consumers. Consequently, the coalition planner always wants to charge

[15] We use a tilde to clarify that $\tilde{p}_{ni}(j)$ is a shadow cost, not a literal price.

noncoalition consumers the maximum possible price, which is the cost of production for the cheapest noncoalition producer.[16] This implies that $\min_{i\in W} p_{n^*i}(j) = \min_{i^*\in W^*} p_{n^*i^*}(j)$. Because the planner has no control over noncoalition production, country n^* will consume until its marginal utility with respect to good j consumption equals the cheapest price at which it can purchase good j. This fixes noncoalition consumption:

$$\sum_{i\in W} c_{n^*i}(j) = \eta_{n^*}\big(\min_{i^*\in W^*} p_{n^*i^*}(j)\big)^{-\sigma_{n^*}} \tag{23}$$

The social planner cannot change $\sum_{i\in W} c_{n^*i}(j)$, but he can choose how much of this consumption is supplied by coalition producers as opposed to the cheapest noncoalition producer. Because of this, a change in $c_{n^*i'}(j)$ is not associated with a change in the overall consumption of good j by country n^*, but rather a change in the sourcing of that consumption from i' to the cheapest noncoalition producer. Consequently, a marginal increase in $c_{n^*i'}(j)$ corresponds to a marginal decrease in some $c_{n^*i^*}(j)$. This causes the FOC with respect to $c_{n^*i^*}(j)$ to pick up an additional term reflecting the shadow benefit of this decrease in noncoalition energy usage:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial c_{n^*i'}(j)} = \min_{i^*\in W^*}&\left(p_{n^*i^*}(j) + (\lambda' - p^E)\frac{\partial e_{i^*}\big(j, z_{n^*i^*}(j)\big)}{\partial c_{n^*i^*}(j)}\right)\\ &- d_{n^*i'}\frac{\partial l_{i'}\big(j, z_{ni'}(j)\big)}{\partial c_{n^*i'}(j)} - \lambda' d_{n^*i'}\frac{\partial e_{i'}\big(j, z_{ni'}(j)\big)}{\partial c_{n^*i'}(j)}\end{aligned}$$

The marginal benefit of coalition exports is simply the price they receive for those exports. This is independent of both the producing country i' and of $c_{n^*i'}(j)$:

$$b_{n^*i'}(j) = \min_{i^*\in W^*} p_{n^*i^*}(j) = \min_{i^*\in W^*}\frac{d_{n^*i^*}}{A_{i^*}(j)}(p^E)^{1-\alpha} \tag{24}$$

[16] This also assumes that the coalition planner does not receive a marginal benefit from carbon emissions, or $\sum_{w'\in W'}\varphi_{w'} \geq 0$.

The marginal shadow cost is the opportunity cost of labor and the shadow cost of energy used in production less the shadow benefit of the foreign energy saved. This is independent of consumption $c_{n^*i'}(j)$ but is dependent on the producing country:

$$
\begin{aligned}
\tilde{p}_{n^*i'}(j) &= d_{n^*i'} l_{i'}\big(j, z_{n^*i'}(j)\big) + \lambda' d_{n^*i'} e_{i'}\big(j, z_{n^*i'}(j)\big) - (\lambda' - p^E) \min_{i^* \in W^*} e_{i^*}\big(j, z_{n^*i'}(j)\big) \\
&= \frac{d_{n^*i'}}{A_{i'}(j)} (\lambda')^{1-\alpha} + (1-\alpha) \min_{i^* \in W^*} \frac{d_{n^*i^*}}{A_{i^*}(j)} \frac{p^E - \lambda'}{(p^E)^\alpha}
\end{aligned}
\tag{25}
$$

The planner chooses a coalition country i' to export good j to country n^* if 1) its marginal benefit of production exceeds its marginal shadow cost of production 2) it is the coalition country with the cheapest marginal shadow cost of production. Because neither the marginal shadow benefit nor the marginal shadow cost is dependent on consumption levels, if i' satisfies 1 and 2, i' will export until the demand of country n^* is completely exhausted, leaving no noncoalition production:

$$
c_{n^*i'}(j) = I_{n^*i'}(j)\ \eta_{n^*} \min_{i^* \in W^*} \left(p_{n^*i^*}(j)^{-\sigma_{n^*}}\right) \tag{26}
$$

where

$$
I_{n*i'}(j) = \begin{cases} 1 & \tilde{p}_{n^*i'}(j) \leq b_{n^*i'}(j) \text{ and } \tilde{p}_{n^*i'}(j) \leq \tilde{p}_{n^*w'}(j) \text{ for all } w' \in W' \\ 0 & \text{otherwise} \end{cases} \tag{27}
$$

For each good j and each pair of countries $(n'i')$, $(n'i^*)$, and (n^*i'), equation (18) provides the optimal energy intensity, while equations (19)-(27) provide optimal consumption levels. Table 1 on the next page summarizes these results. By integrating over goods j and summing over every possible producer-consumer pairing, this fully specifies global energy demand as a function of p^E and λ'. In Appendix C we derive closed form expressions for (most) of these values but move onto deriving energy supply here.

Table 1: Results from the Inner Problem

Direction of Good Flow	Marginal Benefit	Marginal Shadow Cost
Coalition to Coalition	$\eta_{n'}^{1/\sigma_{n'}}(\sum_{w\in W} c_{n'w}(j))^{-1/\sigma_{n'}}$	$\frac{d_{n'i'}}{A_{i'}(j)}(\lambda')^{1-\alpha}$
Noncoalition to Coalition	$\eta_{n'}^{1/\sigma_{n'}}(\sum_{w\in W} c_{n'w}(j))^{-1/\sigma_{n'}}$	$\frac{d_{n'i^*}}{A_{i^*}(j)}(\lambda')^{1-\alpha}$
Coalition to Noncoaliton	$\min_{i^*\in W^*} \frac{d_{n^*i^*}}{A_{i^*}(j)}(p^E)^{1-\alpha}$	$\frac{d_{n^*i'}}{A_{i'}(j)}(p^E+t^B)^{1-\alpha} + (1-\alpha)\min_{i^*\in W^*} \frac{d_{n^*i^*}}{A_{i^*}(j)} \frac{\lambda'-p^E}{(p^E)^{\alpha}}$
Noncoaliton to Noncoalition	Not Applicable	Not Applicable

Direction of Good Flow	Quantity	Price
Coalition to Coalition	$I_{n'i'}(j)\eta_{n'}\left(\frac{A_{i'}(j)}{(\lambda')^{1-\alpha}d_{n'i'}}\right)^{\sigma_{n'}}$	Not Applicable
Noncoalition to Coalition	$I_{n'i^*}(j)\eta_{n'}\left(\frac{A_{i^*}(j)}{(\lambda')^{1-\alpha}d_{n'i^*}}\right)^{\sigma_{n'}}$	Cost of Production
Coalition to Noncoalition	$I_{n^*i'}(j)\eta_{n^*}\left(\max_{i^*\in W^*} \frac{A_{i^*}(j)}{(p^E)^{1-\alpha}d_{n^*i^*}}\right)^{\sigma_{n^*}}$	Cheapest Noncoalition Producer's Cost of Production
Noncoalition to Noncoalition	$I_{n^*i^*}(j)\eta_{n^*}\left(\frac{A_{i^*}(j)}{(p^E)^{1-\alpha}d_{n^*i^*}}\right)^{\sigma_{n^*}}$	Cost of Production

Direction of Good Flow	$I_{ni}(j) = 1$ if and only if...
Coalition to Coalition	$\tilde{p}_{n'i'}(j) \le \tilde{p}_{n'w}(j)$ for all $w \in W$
Noncoalition to Coalition	$\tilde{p}_{n'i^*}(j) \le \tilde{p}_{n'w}(j)$ for all $w \in W$
Coalition to Noncoalition	$\tilde{p}_{n^*i'}(j) \le b_{n^*i'}(j)$ and $\tilde{p}_{n^*i'}(j) \le \tilde{p}_{n^*w'}(j)$ for all $w' \in W'$
Noncoalition to Noncoalition	$\tilde{p}_{n^*w'} \ge b_{n^*w'}(j)$ for all $w' \in W'$ and $p_{n^*i^*}(j) \le p_{n^*w^*}(j)$ for all $w^* \in W^*$

3.5 The Outer Problem and Energy Supply

We now turn to calculating the optimal energy supply in terms of p^E and λ', which we achieve by taking the first order condition with respect to $Q^E_{i'}$:

$$\frac{\partial \mathcal{L}}{\partial Q^E_{i'}} = -\sum_{w' \in W'} \varphi_{w'} - \frac{\partial L^E_{i'}}{\partial Q^E_{i'}} + \lambda'$$

which gives energy supplied by coalition countries:

$$Q^E_{i'} = D^{-1}_{i'}(\lambda' - \sum_{w' \in W'} \varphi_{w'}) \tag{28}$$

The planner cannot control noncoalition extraction, so noncoalition energy supply is given by:

$$Q^E_{i^*} = D^{-1}_{i^*}(p^E) \tag{29}$$

Combining (28) and (29) gives global energy supply as a function of p^E and λ': Equating global energy supply and global energy demand gives us one equation linking p^E and λ'. The FOC with respect to p^E gives another.

3.6 Optimizing with Respect to Energy Price

We now optimize our objective function by taking the FOC of the Lagrangian with respect to p^E. Turning back to our Lagrangian:

$$\begin{aligned}\mathcal{L} = &\sum_{w' \in W'} \Big[\eta_{w'}^{1/\sigma_{w'}} \frac{(C^G_{w'})^{1-1/\sigma_{w'}}}{1 - 1/\sigma_{w'}} - \varphi_{w'} \sum_{i \in W} Q^E_i + L_{w'} - L^G_{w'} - L^E_{w'} + \sum_{n^* \in W^*} (X^G_{n^*w'} + p^E X^E_{n^*w'})\Big] \\ &- \lambda' \sum_{w \in W} \big(C^E_w + \sum_{n \in W} X^E_{nw} - Q^E_w\big) + p^E \sum_{w^* \in W^*} \big(C^E_{w^*} + \sum_{n \in W} X^E_{nw^*} - Q^E_{w^*}\big)\end{aligned}$$

Several terms— $C^G_{w'}$, $L^E_{w'}$, X^E_{ni}, and $Q^E_{w'}$— depend on λ' rather than p^E, and thus their partial derivatives with respect to p^E are 0. Ignoring these terms and setting the partial derivative equal to zero gives the following relationship:

$$(\lambda' - p^E) \sum_{w^* \in W^*} \left(\frac{\partial Q^E_{w^*}}{\partial p^E} - \frac{\partial C^E_{w^*}}{\partial p^E} \right) = \sum_{w' \in W'} \left[\frac{\partial L^G_{w'}}{\partial p^E} + \lambda' \frac{\partial C^E_{w'}}{\partial p^E} - \sum_{w^* \in W^*} \left(\frac{\partial X^G_{w^* w'}}{\partial p^E} + X^E_{w^* w'} - \varphi_{w'} \frac{\partial Q^E_{w^*}}{\partial p^E} \right) \right]$$

$$\lambda' - p^E = \frac{\sum_{w' \in W'} \left[\frac{\partial L^G_{w'}}{\partial p^E} + \lambda' \frac{\partial C^E_{w'}}{\partial p^E} - \sum_{w^* \in W^*} \left(\frac{\partial X^G_{w^* w'}}{\partial p^E} + X^E_{w^* w'} - \varphi_{w'} \frac{\partial Q^E_{w^*}}{\partial p^E} \right) \right]}{\sum_{w^* \in W^*} \left(\frac{\partial Q^E_{w^*}}{\partial p^E} - \frac{\partial C^E_{w^*}}{\partial p^E} \right)}$$

Since coalition countries produce with energy intensity $z_{ni'}(j) = \frac{1-\alpha}{\alpha\lambda'}$,

$$L^G_{i'} + \lambda' C^E_{i'} = \frac{\alpha\lambda'}{1-\alpha} C^E_{i'} + \lambda' C^E_{i'} = \frac{\lambda'}{1-\alpha} C^E_{i'}$$

$C^E_{n'i'}$ is also dependent on λ' rather than p^E, so we can further reduce this term:

$$\sum_{i' \in W'} C^E_{i'} = \sum_{i' \in W'} \sum_{n^* \in W^*} C^E_{n^* i'}$$

Given this, we can rewrite our FOC as:

$$\lambda' - p^E = \sum_{w^* \in W^*} \left[\frac{\sum_{w' \in W'} \left(\frac{\lambda'}{1-\alpha} \frac{\partial C^E_{w^* w'}}{\partial p^E} + \varphi_{w'} \frac{\partial Q^E_{w^*}}{\partial p^E} - \frac{\partial X^G_{w^* w'}}{\partial p^E} - X^E_{w^* w'} \right)}{\frac{\partial Q^E_{w^*}}{\partial p^E} - \frac{\partial C^E_{w^*}}{\partial p^E}} \right]$$

To distill this condition, we need to take intermediate derivatives of $C^E_{n^* i'}$, $Q^E_{i^*}$, $X^G_{n^* i'}$, and $C^E_{ni^*}$ with respect to p^E.

3.6.1 Intermediate Derivatives

Differentiating our consumption formulae from Table 1 with respect to p^E gives the partial derivatives for energy demand:

$$\frac{\partial C^E_{n' i^*}}{\partial p^E} = 0$$

$$\frac{\partial C^E_{n^* i'}}{\partial p^E} = -\sigma_{n^*}(1-\alpha) \frac{C^E_{n^* i'}}{p^E}$$

$$\frac{\partial C^E_{n^*i^*}}{\partial p^E} = -\epsilon^D_{n^*} \frac{C^E_{n^*i^*}}{p^E}$$

where $\epsilon^D_{n^*} = \alpha + \sigma_{n^*} - \alpha\ \sigma_{n^*}$ represents the elasticity of energy demand for country n^*. To calculate the partial derivative with respect to noncoalition energy supply, we introduce an additional parameter ϵ^S_{i*}, which represent's country i^*'s elasticity of energy supply:

$$\epsilon^S_{i^*} = \frac{\partial D_{i^*}(p^E)}{\partial p^E} \frac{p^E}{Q^E_{i^*}}$$

Rearranging gives:

$$\frac{\partial Q_{i^*}}{\partial p^E} = \frac{\partial D_{i^*}(p^E)}{\partial p^E} = \epsilon^S_{w^*} \frac{Q^E_{i^*}}{p^E}$$

Finally, we have $X^G_{w^*w'} = V^G_{w^*w'} - V^G_{w'w^*}$. Because noncoalition countries sell goods to coalition countries at their cost of production:

$$V^G_{n'i^*} = L^G_{n'i^*} + p^E C^E_{n'i^*} = \Big(\frac{\alpha\lambda'}{1-\alpha} + p^E\Big) C^E_{n'i^*}$$

$$\frac{\partial V^G_{n'i^*}}{\partial p^E} = C^E_{n'i^*}$$

On the other hand, coalition countries implement the limit pricing strategy when selling goods to noncoalition countries. Because of this, applying (26):

$$\begin{aligned} V^G_{n^*i'} &= \int_0^1 p_{n^*i'}(j) c_{n^*i'}(j) dj \\ &= \int_0^1 I_{n^*i'}(j) \Big(\min_{i^* \in W^*} p_{n^*i^*}(j)\Big)^{1-\sigma_{n^*}} dj \\ &= \int_0^1 I_{n^*i'}(j) (p^E)^{1-\epsilon^D_{n^*}} \Big(\min_{i^* \in W^*} \frac{d_{n^*i^*}}{A_{i^*}(j)}\Big)^{1-\sigma_{n^*}} dj \end{aligned}$$

Differentiating this formula gives:

$$\frac{\partial V^G_{n^*i'}}{\partial p^E} = (1 - \epsilon^D_{n^*}) \frac{V^G_{n^*i'}}{p^E}$$

Combining these two expressions, we have:

$$
\begin{aligned}
\frac{\partial X^G_{w^*w'}}{\partial p^E} =& \frac{\partial V^G_{w^*w'}}{\partial p^E} - \frac{\partial V^G_{w'w^*}}{\partial p^E} \\
=& (1-\epsilon^D_{w^*})\frac{V^G_{w^*w'}}{p^E} - C^E_{w'w^*}
\end{aligned}
\tag{30}
$$

Substituting these partial derivatives into our FOC with respect to p^E and multiplying through by p^E gives:

$$
\begin{aligned}
\lambda' - p^E =& \sum_{w^* \in W^*} \left[\frac{\sum_{w' \in W'} \left(\varphi_{w'} \epsilon^S_{w^*} Q^E_{w^*} + p^E (C^E_{w'w^*} - X^E_{w^*w'}) - \lambda' \sigma_{w^*} C^E_{w^*w'} - (1-\epsilon^D_{w^*}) V^G_{w^*w'} \right)}{\epsilon^S_{w^*} Q^E_{w^*} + \sum_{n^* \in W^*} \epsilon^D_{n^*} C^E_{n^*w^*}} \right] \\
=& \sum_{w^* \in W^*} \left[\frac{p^E \left(Q^E_{w^*} - \sum\limits_{n^* \in W^*} C^E_{n^*w^*} \right) + \sum\limits_{w' \in W'} \left(\varphi_{w'} \epsilon^S_{w^*} Q^E_{w^*} - \lambda' \sigma_{w^*} C^E_{w^*w'} - (1-\epsilon^D_{w^*}) V^G_{w^*w'} \right)}{\epsilon^S_{w^*} Q^E_{w^*} + \sum_{n^* \in W^*} \epsilon^D_{n^*} C^E_{n^*w^*}} \right]
\end{aligned}
$$

This gives us our last equation linking p^E and λ'. Combined with the energy market clearing condition, this completely specifies the optimal policy of the coalition planner.

3.7 Optimal Taxes and Subsidies

We now look at the set of taxes and subsidies that generate the unilaterally optimal solution in equilibrium. We need our optimal set of taxes to accomplish the following:

1. Coalition countries extract $Q^E_{i'} = D^{-1}_{i'}(\lambda' - \sum_{w' \in W'} \varphi_{w'})$.

2. For all countries and all goods, coalition to coalition, coalition to noncoalition, and noncoalition to coalition energy intensity satisfies (18):

$$
z_{n'i'}(j) = z_{n'i^*}(j) = z_{i^*n'}(j) = \frac{1-\alpha}{\alpha\lambda'}
$$

3. Consumption quantities satisfy the optimal values summarized in Table 1.

4. Coalition producers sell goods to noncoalition consumers at the cheapest coalition producer's cost of production.

There are a number of tax mechanisms that can achieve these outcomes. In the paper, we consider one such set of taxes. First, we impose an extraction tax $t^E = \sum_{w' \in W'} \varphi_{w'}$ and a border adjustment $t^B = \lambda' - p^E$ on both noncoalition to coalition energy trade and on the energy embodied in noncoalition to coalition goods trade. The border adjustment raises the cost of energy in the coalition to λ' and implicitly raises the energy cost faced by noncoalition producers to λ' when they are producing goods for coalition consumers. Coalition extractors therefore receive energy price $p^E + t^B - t^E = \lambda' - \sum_{w' \in W'} \varphi_{w'}$, satisfying 1. The border adjustment ensures that the cost-minimizing energy intensity for coalition to coalition, coalition to noncoalition, and noncoalition to coalition goods production is $\frac{1-\alpha}{\alpha \lambda'}$, satisfying 2. Further, it sets the cost of production for all countries i producing for a coalition consumer n' equal to $\frac{d_{n'i}}{A_i(j)}(\lambda')^{1-\alpha}$. Since producers within the producing country compete and sell goods at their cost of production until coalition consumers are no longer willing to buy the good, this satisfies the consumption conditions for coalition consumers.

Lastly, the planner subsidizes/taxes coalition goods exports to noncoalition consumers at a rate $t_{n^*i'}(j)$. If $t_{n^*i'}(j)$ is positive, it is a tax; if it is negative, it is a subsidy.

$$t_{n^*i'}(j) = \max \left\{ \min_{i^* \in W^*} \frac{d_{n^*i^*}}{A_{i^*}(j)}(p^E)^{1-\alpha} - \min_{i' \in W'} \frac{d_{n^*i'}}{A_{i'}(j)}(\lambda')^{1-\alpha}, -(1-\alpha) \min_{i^* \in W^*} \frac{d_{n^*i^*}}{A_{i^*}(j)} \frac{\lambda' - p^E}{(p^E)^\alpha} \right\}$$

The first term of the *max* function taxes/subsidizes coalition producers so that the post-tax cost of production for the cheapest coalition producer equals the cost of production for the cheapest noncoalition producer. This implements the limit pricing strategy discussed in the Inner Problem, and ensures that noncoalition consumers always consume the optimal quantity of goods. The *max* function caps the largest subsidy that can be offered to coalition producers. This ensures that coalition producers only sell goods to noncoalition consumers when the marginal benefit to the coalition outweighs the marginal shadow cost to the coalition. Consequently, this policy causes coalition to noncoalition trade to satisfy 3 and 4, completing our tax policy.

4 Quantitative Illustration

4.1 Procedural Specifications

4.1.1 Calibrating to Data

In this section, we quantify the results outlined in Section 3 and discuss their implications. We follow Kortum and Weisbach by calibrating our competitive equilibrum solution (derived in Appendix A) to real world data. Like Kortum and Weisbach, we calibrate energy demand to carbon embodied in bilateral trade flows data made available by the OECD Trade Embodied in CO_2 (TECO$_2$) database.[17] We calibrate energy extraction to International Energy Agency data on global energy extraction.[18]

We expand the calibrated energy data to quantify the theoretical solution outlined in Section 3. Using this approach, we quantify the optimal extraction tax, border adjustment, and utility changes in terms of a chosen coalition, two universal parameters (α and θ) and three country-specific parameters (σ_w, ϵ_w^S, and φ_w). We use the same values for these

[17] The TECO$_2$ database contains data for much of, but not all of, the world. We calibrate data for each country that data is available for, and aggregate the amount of global energy demand and supply that is unaccounted for into a region called "Rest of World." This us a world consisting of the following countries/regions: Australia, Austria, Belgium, Canada, Chile, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, South Korea, Latvia, Lithuania, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom, United States of America, Argentina, Brazil, Brunei Darussalam, Bulgaria, Cambodia, China, Colombia, Costa Rica, Croatia, Cyprus, Hong Kong, India, Indonesia, Kazakhstan, Malaysia, Malta, Morocco, Peru, Philippines, Romania, Russian Federation, Saudi Arabia, Singapore, South Africa, Taiwan, Thailand, Tunisia, Vietnam, Rest of World

[18] Like Kortum and Weisbach, we first convert energy supply for several nonrenewable energy sources to CO_2 emissions according to default emission factors for stationary combustion from the 2006 IPCC Guidelines for National Greenhouse Gas Inventories. We convert 1 TJ of coal and oil to $94,600$ kg of CO_2, 1TJ of Crude Oil and NGL to $73,300$ kg of CO_2, and 1 TJ of Natural Gas to $56,100$ kg of $C0_2$. This gives us $36,354$ kg of $C0_2$ emissions supplied, roughly 12% more than the $32,361$ kg of $C0_2$ emissions embodied in final goods consumption. As Kortum and Weisbach identify, this is likely due to fossil fuels that are never combusted. In our simulations, we follow Weisbach and Kortum and re-scale global energy supply so that the global energy market clears.

parameters as Kortum and Weisbach, and for simplicity's sake, keep σ and ϵ^S constant across countries:

Table 2: Parameter Values

α	θ	σ_w	ϵ_w^S
0.85	4	1	0.5

4.1.2 Regularizing Marginal Damages

Marginal damages from carbon emissions, denoted φ in Sections 2 and 3, is a critical parameter of the model. Marginal damages directly control the optimal extraction tax, which is a simple sum of the marginal damages of each country in the coalition, and play a large role in determining the optimal energy price and border adjustment as well. Intuitively, it is helpful can think of marginal damages through the following relationship:[19]

$$\varphi_w + p^E = \text{National Social Cost of Carbon for Country } w$$

φ_w represents the damages that carbon emissions inflict on a country that are not internalized into the energy price in competitive equilibrium. The idea is that country w is indifferent between reducing emissions by φ_w or increasing services by φ_w; since services cost a unit price of 1, the value of a unit reduction in emissions is φ_w. The national social cost of carbon is the value of the emissions reduction plus the energy price.

National SCCs are underresearched and poorly understood. However, Nordhaus (2017) notes that they are highly correlated with the size of a country's economy, finding a strong

[19]Critically, we think of φ_w vís a ví the *national* social cost of carbon for country w, not the *global* social cost of carbon. National social costs of carbon are much lower than global social costs of carbon, because they only consider the climate-induced damages to a specific region or country rather than the entire world.

correlation between output and estimates of natioanl SCCs (R = 0.71). Following this logic, we regularize marginal damages of emissions across countries by making it proportional to spending on goods in the competitive equilibrium scenario. We define one parameter, which we call emissions aversion, that holds for the entire world. For each country w, we set:

$$\varphi_w = \text{spending of country } w \times \text{ global carbon emissions aversion parameter}$$

This regularization is an imperfect proxy for a more thorough national SCC estimate, but it greatly simplifies our analysis by reducing emissions aversion down to a single global parameter. In the following sections, we analyze how the optimal policy and utilities of participating and abstaining countries vary alongside changes in the emissions aversion parameter. As a rough reference point for the sections that follow, an emissions aversion equal to $(1, 2, 3, 4, 5)$ corresponds to a social cost of carbon $(316\%, 531\%, 747\%, 963\%, 1179\%)$ of the competitive equilibrium price. However, to avoid overstating this relationship in our highly-stylized model, we speak in terms of emissions aversion rather than social cost of carbon for the following sections.

4.2 Results and Analysis

Our solution for the optimal unilateral tax consists of four elements: i) an extraction tax equal to the collective marginal damages of emissions of coalition countries, ii) a border adjustment on energy imported from noncoalition countries, iii) a border adjustment on the energy embodied in the production of imported goods, and iv) an export policy that exploits inelastic demand for goods in noncoalition countries but expands the export margin of the coalition. This solution is identical to the optimal unilateral tax policy proposed by Kortum and Weisbach, only generalized to a multi-country world. Because of this, we refer readers interested in the features of the optimal policy to Kortum and Weisbach's detailed discussion and do not repeat their analysis here. Instead, we focus on interpreting the optimal policy as the carbon policy of a climate club. Conducting this analysis is impractical in the original two-country framework of Kortum and Weisbach but feasible using the multi-country solution derived in Section 3, since we are able to analyze changes in welfare for individual countries facing a temptation to free-ride.

In our model, all goods are traded at their cost of production in competitive equilibrium. However, the optimal policy specifies an export tax that exploits inelastic demand for goods in noncoalition countries by taxing coalition producers the maximum possible amount when they export goods. More specifically, our solution suggests that the coalition should tax its producers until the final price paid by noncoalition consumers is equal to the price those consumers would face if they had bought the good from the cheapest noncoalition producer. This pricing strategy arises naturally as the dominant strategy in a one-period model of trade. However, while it is not set explicitly for this purpose, the tax functions as the largest possible external penalty that could be imposed on noncoalition members, and consequently incentivizes coalition participation. In some ways, the export tax proposed by our solution is a more realistic external penalty than the uniform import tariffs common in other models of climate clubs, because the intensity of our tax varies dynamically across goods, countries, and clubs in a way that encapsulates relative efficiencies in production and trade patterns. On the other hand, it is more stylized and difficult to implement, since it demands a unique tax for every possible combination of producers, consumers, and goods. Moreover, it is only one external penalty of the many that could be enforced, and our analysis does not compare the relative effectiveness of this penalty as opposed to others.

In the following subsections, we analyze the effectiveness of the coalition-optimal policy in light of the fact that participating countries are free to leave and noncoalition countries are free to join. We begin by defining the participation decision facing each country. Next, we decompose the participation decision into several underlying and competing motivations. We then analyze the characteristics that make a country likely to join a broad coalition. Finally, we look at coalition stability and quantify the extent to which the world's largest stable coalition can reduce global emissions. At each step, we consider how the various effects change as a function of emissions aversion.

4.2.1 Defining Participation Incentives

Given a baseline coalition, each participating country must decide whether or not to leave the coalition, and each abstaining country must decide whether or not to join. We formally define this participation decision via a variable called INOUT Utility Change, which we define as the utility of a country when it participates in a coalition less its utility when it does not participate, fixing the participation statuses of all other countries. INOUT Utility Change for a country w is calculated using the following method:

1. Fix a value for the emissions aversion parameter and keep it constant throughout steps 2-7.

2. Choose a baseline coalition (that includes the country w) and calculate the optimal tax policy for the coalition.

3. Calculate the utility of country w as a coalition country under this tax policy.

4. Assume country w unilaterally leaves the coalition and re-optimize the tax policy for the new coalition.

5. Calculate the utility of country w as a noncoalition country under the new tax policy.

6. Subtract the utility of country w when it is not in the coalition (step 5) from when it is in the coalition (step 3).

7. Scale INOUT Utility Change by the competitive equilibrium spending on goods of country w.

If INOUT Utility Change is positive, then country w wants to participate in the coalition. Otherwise, it wants to abstain.

4.2.2 Interpreting INOUT Utility Change

To help interpret INOUT Utility Change, we decompose the utility of country w into a sum of four distinct surpluses:[20]

$$
\begin{aligned}
U_w =& \Big[\eta_w^{1/\sigma_w} \frac{(C_w^G)^{1-1/\sigma_w}}{1-1/\sigma_w} - \varphi_w \sum_{i \in W} Q_i^E + L_w - L_w^G - L_w^E + \sum_{n \in W} (X_{nw}^G + p^E X_{nw}^E)\Big] \\
=& \Big[\eta_w^{1/\sigma_w} \frac{(C_w^G)^{1-1/\sigma_w}}{1-1/\sigma_w} - \sum_{i \in W} V_{wi}^G\Big] + \\
& \Big[\sum_{n \in W} \big(V_{nw}^G - L_w^G - p^E C_{nw}^E\Big]\big) + \Big[p^E Q_w^E - L_w^E\Big] - \Big[\varphi_w \sum_{i \in W} Q_i^E\Big] \\
=& \Pi_w^C + \Pi_w^Q + \Pi_w^E + \Pi_w^\varphi
\end{aligned}
$$

These surpluses are, in order:

- Consumption Surplus (Π_w^C): The benefit country w derives from goods consumption less the cost to produce and/or purchase those goods.
- Production Surplus (Π_w^Q): The revenue country w receives for goods it produces less the cost of production.
- Extraction Surplus (Π_w^E): The value of energy country w extracts less the cost of labor used in energy extraction.
- Emissions Damages (Π_w^φ): The utility that global carbon emissions produce for country w. Assuming that emissions are harmful rather than beneficial, Π_w^φ is negative.

We can analyze INOUT Utility Change by breaking it up into changes of these surpluses:

$$
\text{INOUT Utility Change} = \frac{\Delta\Pi^C + \Delta\Pi^P + \Delta\Pi^Q + \Delta\Pi^E + \Delta\Pi^\varphi}{\text{Competitive Equilibrium Spending on Goods}}
$$

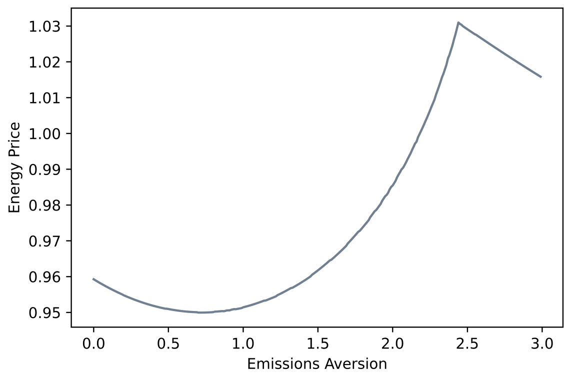
[20]Techinically speaking, this formula for U_w ignores the constant L_w, the labor supply of country w. However, because our analysis focuses on utility changes across scenarios rather than utility values, we can safely ignore L_w in our analysis.

Analyzing $\Delta\Pi^C, \Delta\Pi^Q, \Delta\Pi^E$, and $\Delta\Pi^\varphi$ alongside INOUT Utility Change helps us pinpoint the primary forces influencing INOUT Utility Change.

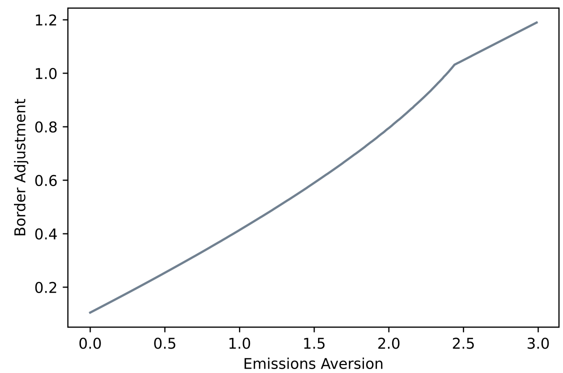
In the remainder of this subsection, we analyze INOUT Utility Change and its components in an example scenario. Setting the OECD as the baseline coalition, in Figures 1a, 1b, and 1c, we plot the optimal global energy price, the optimal border adjustment, and the optimal post-tax energy price for coalition extractors as a function of emissions aversion. In Figure 2, we analyze the United States' decision of whether or not to participate in the coalition by plotting INOUT utility change, consumption, production, and extraction surplus changes, and gains from global emissions reductions as functions of emissions aversion. Note that while we use the United States as the example country for the remainder of this subsection, the trends apply to varying degrees to all countries. The goal is to better understand the underlying motivations behind a country's participation decision, building our intuition for how and why a country's participation decision varies alongside emissions aversion.

Figure 1: Energy Price and Taxes as a Function of Emissions Aversion (Coalition: OECD)

(a) Energy Price



(b) Border Adjustment



(c) Post-Tax Energy Price for Cltn. Extractors

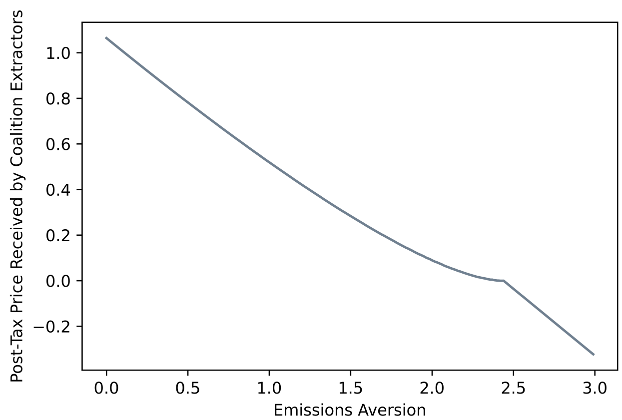
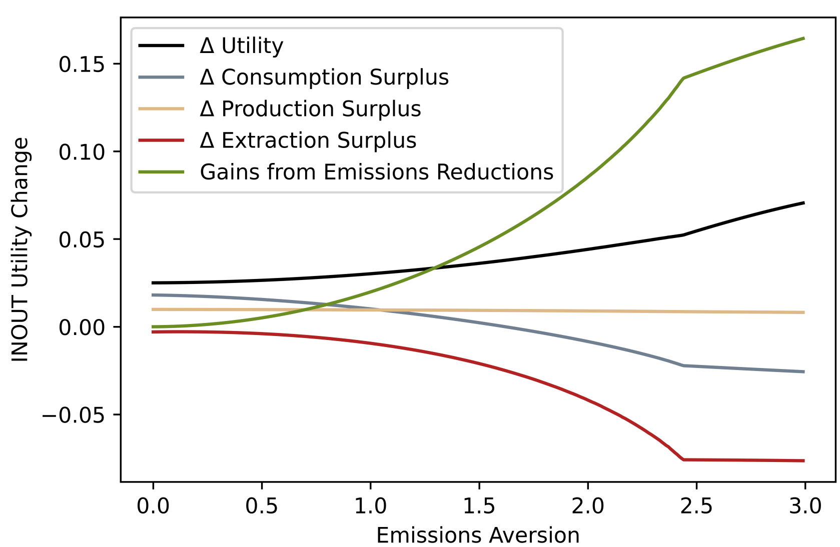


Figure 2: INOUT Utility Change for the USA (Coalition: OECD)



In Figure 2, change in consumption surplus is given by the gray line. Consumption surplus is primarly driven by the prices that the United States must pay for the goods it consumes and declines as emissions aversion increases. The price that coalition consumers must pay for a good is dependent on two factors: the cost of production and the cost of any external taxes that are applied. When the United States participates in the coalition, it must pay the border adjustment on the energy it uses in goods production; since most goods are self-produced, this raises the cost of production for the majority of goods the United States ends up consuming. On the other hand, when the United States abstains from the coalition, it avoids the border adjustment but faces the external penalties applied by coalition countries exporting goods to the United States. At low levels of emissions aversion, the border adjustment is small (Figure 1b), so the external penalties on imports are far more impactful than the only slightly higher cost of production. As a result, the United States faces higher prices when it abstains from the coalition than when it participates, so change in consumption surplus is positive. As emissions aversion increases, however, the cost of the external penalties on imports stays roughly constant, but the border adjustment becomes larger, so the increased cost of production when participating grows in importance. The negative trend between change in consumption surplus and emissions aversion is a consequence of the gradual shift in prominence from the external penalties to the border adjustment as the primary driver of the cost of goods. Eventually, the border adjustment becomes so high that the cost of production effect dominates the terms of trade effect, and change in consumption surplus becomes negative.

Change in production surplus is displayed by the yellow line in Figure 2. It is positive and roughly constant at all levels of emissions aversion. Since in our model, noncoalition countries behave passively, as if in competitive equilibrium, and sell goods at their cost of production, they generate no production surplus. The export policy of the coalition, however, dictates that coalition producers sell goods to noncoalition consumers at the cost of production of the cheapest noncoalition producer. For goods where the coalition has a comparative advantage, the coalition imposes a tax on coalition producers and generates a positive production surplus. For goods where the coalition has a slight comparative disadvantage, the coalition offers a subsidy, losing production surplus, so that a coalition country will produce the good cleanly rather than having a noncoalition country produce the good dirtily. In practice, we find that there are more goods sold at a markup than at a loss, so change in production surplus is positive overall.

Extraction surplus is shown by the red line in Figure 2 and decreases superlinearly with respect to emissions aversion. The severity of this trend is a consequence of the fact that almost all extraction reductions take place amongst coalition extractors. Since wages are set as our numeraire, all countries are incentivized to extract energy until the marginal quantity of labor required to extract any more energy equals the price they receive for that energy. For noncoalition countries, this price is simply the global energy price displayed in Figure 1a, which stays rougly constant throughout the simulation. Because the global energy price is relatively stable, noncoalition extraction surplus is relatively stable. For coalition countries, the tax policy reduces the price extractors receive for energy to the post-tax energy price displayed in Figure 1c, which decreases rapidly. The rapid price decrease has a dual effect on coalition extractors: it simultaneously decreases the quantity of energy that coalition extractors can profitably extract and decreases the price that the extractors receive for the energy that they do extract. The dual effect causes coalition extraction surplus, and by extension change in extraction surplus, to decrease superlinearly with respect to emissions aversion. As emissions aversion climbs to extreme levels, change in extraction surplus suddenly levels off. This occurs becuase the post-tax price for coalition extractors becomes negative, so coalition countries extract no energy and generate no extraction surplus. Coalition extraction surplus cannot decrease any further, so change in extraction surplus past this point varies solely due to the very small changes in noncoalition extraction surplus resulting from the very small changes in the global energy price.

Gains from global emissions reductions are displayed by the green line in Figure 2 and increase superlinearly, balancing the negative effect of change in extraction surplus. When the United States participates in the coalition, global emissions are reduced for two reasons. First, the coalition includes the United States' marginal damages from emissions when determining its optimal tax policy, making its policy more aggressive. Second, the United States extracts the coalition-optimal quantity of energy and produces with the coalition-optimal energy intensity, decreasing its own energy supply and demand, and by extension, global energy supply and demand. Both of these effects are increasing in emissions aversion. Moreover, the greater emissions aversion, the happier the United States is about the reduction in global emissions that its own participation generates. Consequently, an increase in emissions aversion both makes the United States happier about a reduction in global emissions and amplifies the size of global emissions reductions caused by its own participation— the interaction between these effects drives the superlinearity. At extreme levels of emissions aversion, gains from emissions reductions kinks and becomes linear because one of the two factors driving the superlinear relationship abruptly stabilizes. Because the post-tax energy price received by coalition extractors dips below zero (Figure 1c), coalition extraction goes to zero and is unable to be decreased any further. Because of this, the coalition resorts to trying to curb noncoalition extraction by decreasing the energy price. As demonstrated by the relative stability of the global energy price (Figure 1a), this policy is extremely ineffective. Consequently, the United States' participation in the coalition has limited impact on the coalition's effectiveness, so the United States stops being able to further reduce global emissions through its own participation. On the other hand, the United States continues to get happier about the constant level of emissions reductions its participation causes, so the relationship is still positive but linear.

Taking a step back, we can see how these surpluses interact to dictate INOUT Utility Change for the United States as a function of emissions aversion. At low levels of emissions aversion, the extraction tax and border adjustment are small, leading to only minor changes in extraction surplus and gains from emissions reductions. Consequently, the trade benefits from joining the coalition, reflected by the positive changes in consumption and production surplus, are large enough that INOUT Utility Change is positive. As emissions aversion increases, change in extraction surplus decreases superlinearly and gains from emissions reductions increases superlinearly. These two terms eventually dominate the trade benefits from participating in the coalition and are the primary factors influencing INOUT Utility Change at moderate to high levels of emissions aversion. In this scenario, gains from emissions reductions slightly outweigh the loss of extraction surplus, so the United States is incentivized to participate. At extreme levels of emissions aversion, coalition countries extract no energy, causing change in consumption surplus to decrease linearly while gains from emissions damages increases linearly; changes in production and extraction surplus are negligible. For the United States, the gain from emissions reduction effect is greater than the change in consumption surplus effect, so INOUT Utility Change continues to grow even at extreme levels of emissions aversion, and the United States continues to participate in the coalition. While the intuition behind these trends, as well as their general shape, applies to all countries, the relative influence of the trends can vary. In the next subsection, we investigate how these factors depend on the characteristics of a given country, and how that dependence impacts INOUT utility change.

4.2.3 Characteristics Affecting INOUT Utility Change

We now aim to predict which countries are more or less likely to participate in a carbon coalition. Using four competitive equilibrium statistics– spending on goods, import share, export share, and energy extraction– as predictors, we run a multivariate regression to predict INOUT Utility Change.[21] For each predictor, we include an interaction term with emissions aversion. The interaction terms are quadratic with respect to emissions aversion, which

[21] We define import share as the fraction of goods consumed that are imported from other countries and export share as the fraction of goods produced that are exported to other countries. From a theoretical standpoint, only imports from coalition countries and exports to noncoalition countries are relevant. However, ignoring this detail simplifies our model and makes import share and export share independent of the coalition.

we find perform better than linear interaction terms and intuitively capture the superlinear trends in change in consumption surplus, change in extraction surplus, and change in gains from emissions damages shown in Figure 2. To generate data for our model, we randomly select 500 values for emissions aversion from a uniform distribution [0, 2.44] with the OECD as the baseline coalition. The upper bound of this range is intentionally chosen to be the point where the post-tax energy price received by coalition extractors dips below zero, and change in consumption surplus, change in extraction surplus, and change in gains from emissions reductions all kink (see Figure 2). We then randomly select a country for which to calculate INOUT Utility Change. Each predictor is scaled so that the coefficients represent the effect of a one standard-deviation increase in the predictor. In the remainder of this subsection, we break down the intuition behind each of the significant coefficients in the regression.

Table 3: Effect of Competitive Equilibrium Characteristics on INOUT Utility Change

	Dependent variable:
	INOUT_Utility_Change
Emissions Aversion	-0.058
	(0.043)
Competitive Equilibrium Spending	-0.004
	(0.028)
Spending × Emissions Aversion^2	0.135^{***}
	(0.029)
Import Share	0.222^{***}
	(0.037)
Import Share × Emissions Aversion^2	-0.163^{***}
	(0.060)
Export Share	0.309^{***}
	(0.036)
Export Share × Emissions Aversion^2	0.090
	(0.057)
Energy Extraction	0.358^{***}
	(0.028)
Energy Extraction × Emissions Aversion^2	-1.062^{***}
	(0.030)
Constant	0.000
	(0.013)
Observations	500
R^2	0.915
Adjusted R^2	0.914
Residual Std. Error	0.294 (df = 490)
F Statistic	588.406^{***} (df = 9; 490)
Note:	*p<0.1; **p<0.05; ***p<0.01

Larger countries, as measured by goods expenditure (spending) are more incentivized to join a coalition because they would play a larger role in determining the coalition's optimal policy. For each country, the locally optimal carbon policy factors in their own marginal damages from emissions, but not the marginal damages of other countries. In our model, however, no country implements the locally optimal solution: countries either choose to participate in the coalition, in which case they factor in the marginal damages of every country within the coalition, or abstain, in which case they behave passively and do not factor in any marginal damages from emissions. The coalition-optimal policy is closer to locally-optimal for larger countries, because larger countries make up a larger portion of the coalition. For example, the marginal damages for the United States are roughly 45% the collective marginal damages for the OECD. The collective damages for Iceland are roughly 0.02% the collective marginal damages. Because of this, taxing emissions at the coalition-optimal level is much more painful for small countries, like Iceland, than it is for large countries, like the United States. This difference is reflected by the positive coefficient on the spending and aversion interaction term in Table 3 and plays an important role in the simulation in Section 4.2.4.

INOUT Utility Change is positively dependent on import share through change in consumption surplus. Noncoalition countries suffer under the optimal tax policy because coalition countries penalize them via the export taxes. Participating in the coalition removes these penalties, resulting in a positive change in consumption surplus. Countries that import a large portion of goods suffer greater damages from these penalties when abstaining from the coalition, making them more incentivized to join the coalition. This explains the positive coefficient on the non-interaction import share term in Table 3. However, the interaction term between import share and emissions aversion in Table 3 is negative. As emissions aversion increases, the cost of production of coalition countries increases, limiting the ability of the coalition to tax coalition exports and penalize noncoalition countries. As a result, the damages inflicted on noncoalition countries by the external penalties is decreasing with respect to emissions aversion, diminishing the extent to which the external penalties serve as an incentive for heavy importers. Additionally, an increase in export share increases INOUT Utility Change through change in producer surplus. While noncoalition countries always sell goods at their cost of production, coalition countries sell goods to noncoalition consumers at a markup, generating producer surplus. Coalition countries with greater export share generate greater producer surplus, and are therefore more incentivized to be a part of the coalition.

Finally, energy extraction, which we have scaled by competitive equilibrium spending, negatively impacts INOUT Utility Change through extraction surplus. Because the OECD is a net energy importer, at low levels of energy intensity, coalition taxes on noncoalition energy can actually increase coalition extraction in order to improve terms of trade, leading to a positive change in extraction surplus. However, as emissions aversion increases, the coalition's tax policy causes coalition extraction to decrease and eventually go to zero, leading to a negative change in extraction surplus. Both of these effects are more extreme for countries that extract large quantities of energy, explaining the coefficients on the energy extraction terms.

Collectively, these coefficients imply that at low levels of emissions aversion, all countries are incentivized to participate in the coalition. When emissions aversion is zero, for example, countries want to join the coalition because of the trade benefits participation status offers. Joining the coalition increases consumption and production surpluses without changing extraction surplus or significantly increasing emissions, so each country is better off participating than not participating. However, as emissions aversion grow, the coalition's tax policy becomes more aggressive, decreasing the change in consumption and extraction surpluses but increasing the gains from emissions reduction. The effect on extraction surplus is the largest, as indicated by the coefficient on the energy extraction and emissions aversion interaction term. Consequently, countries that extract large amounts of energy relative to their size are the least likely to participate in a climate coalition. Amongst relatively small energy extractors, the other two factors come into play: large countries are more likely to participate due to their impact on global emissions, and countries whose economies are highly dependent on trade are more likely to participate due to the consumption and production surpluses they receive from being in a coalition. We can summarize these findings in the following hypothesis: as emissions aversion increases, the first countries to leave a coalition will be countries who are intensive in energy production. Eventually, small-autarkic countries will follow, while large countries that are highly-dependent on trade will remain in the coalition the longest. In the following subsection, we test this hypothesis on the real world data.

4.2.4 Building Top Down Coalitions

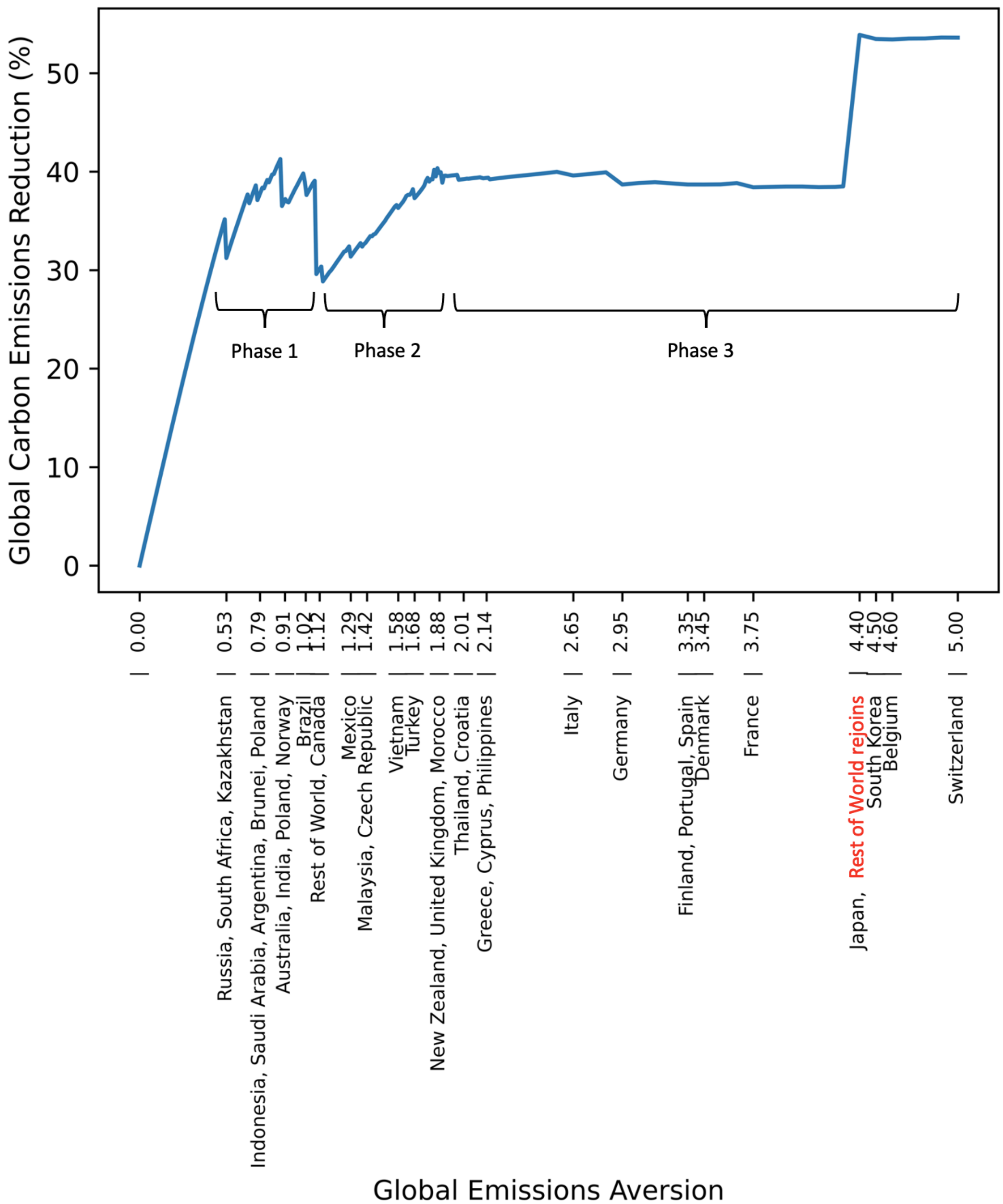
The findings in Section 4.2.3 suggest that as emissions aversion increases, countries become less incentivized to participate in a global coalition (very large countries being the exception). A consequence of this finding is that coalitions implementing the optimal policy under high emissions aversion are inherently less stable because its constituents are more likely to defect and free ride; by extension, stable coalitions implementing the optimal policy with a large emissions aversion tend to be smaller. This diminishes the effectiveness of the coalition's policy, since smaller coalitions are more prone to leakage. In this section, we quantify this intuition by observing global emissions reduction under the largest stable coalition's optimal policy as a function of emissions aversion. In our simulation, we begin with zero emissions aversion, so every country wants to participate in the coalition for the trade benefits. We then slowly inch up emissions aversion. At each step, we check INOUT Utility Change for each country. If INOUT Utility Change is negative for a participating country, we remove it from the coaltiion and re-calculate the optimal tax. If INOUT Utility Change is positive for an abstaining country, we add it to the coalition and re-calculate the optimal tax. We conclude that a coalition is stable at a given emissions aversion when every participating country does not wish to leave and every abstaining country does not wish to join. Because INOUT Utility Change tends to decrease with respect to emissions aversion, we find that most countries "drop out" of the coalition at some threshold emissions aversion that is unique to that country and never come back. We feel highly confident that this algorithm produces the largest stable coalition at each level of emsisions aversion, though we do not formally prove this fact.[22] [23] Figure 3 displays the results of this simulation. The horizontal axis displays emissions aversion, while the vertical axis shows global emissions reductions under the largest stable coalition. The horizontal axis also specifies countries as they "drop out" of the coalition.[24]

[22]We are confident of this fact because INOUT Utility Change tends to decrease with the extraction tax. Larger coalitions have larger collective marginal damages from emissions, and therefore feature larger extraction taxes. This implies that countries are less likely to participate in larger coalitions, and smaller coalitions tend to be more stable. For this reason, it is highly unlikely that adding several countries to the coalition would cause each country's INOUT Utility Change to flip from negative to positive, assuring us that in the vast majority of cases, our algorithm successfully finds the largest stable coalition. Moreover, if this ever is not the case, all countries in question would defect shortly afterwards anyways, so there would be minimal effect on our broader conclusions.

[23]There are a few problems with this algorithm as it is written, so our implementation is slightly different. First, perhaps counterintuitively, the global coalition is stable up to extreme emissions aversions. In the global coalition, there is no border adjustment. When one country defects, the coalition institutes a border adjustment that largely manipulates terms of trade with the defecting country (if the defector is an energy exporter, the border adjustment is set extremely high and the energy price extremely low; if the defector is an energy importer, the border adjustment is set extremely negative and the energy price skyrockets). Within the coalition, this manipulation causes a dramatic one-time utility transfer from energy importers to energy exporters or vice versa, but the defecting country is almost always worse off by defecting. This artificially keeps the global coalition stable. However, if the coalition lacks even one country, the highly manipulative border adjustment is already in place, so the coalition cannot threaten the institution of such a highly manipulative border adjustment to incentivize cooperation amongst all other countries. For the purposes of our simulation, we ignore this phenomenon and allow the first two countries to defect as soon as INOUT Utility Change is negative for both countries when both countries abstain from the coalition. The other technical problem that arises when implementing this algorithm occurs when two countries decide to leave the coalition almost simultaneously, but their decisions are highly intertwined. For example, suppose two countries reach roughly zero INOUT Utility Change at the same time and one imports a large amount of goods from the other (but not vice versa). The importing country may want to stay in the coalition to avoid being exploited by the exporting country's limit pricing strategy. But once the exporting country defects, this is no longer a threat, so the importing country defects alongside it. This presents an opportunity for the exporting country to rejoin the coalition and exploit the importing country. This dynamic is an endless loop, and no stable coalition exists. However, these cases are rare and because INOUT Utility Change is decreasing in emissions aversion, they are short-lived– inching up emissions aversion slightly further always causes both countries to defect. In our results, we assume both countries stay in the coalition until both want to defect in unison. This assumption has only a marginal effect on our results and does not impact on any of the broader conclusions.

[24]Countries that leave the coalition at similar levels of emissions aversion are grouped together on the horizontal axis to make the graph more readable.

Figure 3: Global Emissions Reduction Under Largest Stable Coalition



At low levels of emissions aversion, the carbon policy is not severe enough that the trade benefits from participating outweigh the loss of extraction surplus for every single country in the world. As a result, the global coalition is stable and accomplishes emissions reductions rapidly. At emissions aversion $(\sim 0.15, \sim 0.30, \sim 0.45)$, the global coalition implements a policy that raises the global energy price by $(\sim 12\%, \sim 27\%, \sim 45\%)$ with an extraction tax of about $(\sim 29\%, \sim 51\%, \sim 67\%)$ and global emissions are reduced by $(\sim 10\%, \sim 20\%, \sim 30\%)$. However, beyond this point, the optimal tax policy becomes too extreme and countries begin to leave the coalition. The departure from the coalition occurs in roughly three phases that are outlined in Figure 3. We describe the economic intuition behind each of these phases in the remainder of this subsection.

In the second phase, ranging from emissions aversion $\sim$ 1.1 to $\sim$ 1.9, highly-autarkic countries such as Romania, Tunisia, and Morocco, as well as moderate energy extractors such as the Czech Republic, Vietnam, and New Zealand, leave the coalition. Because the countries leaving the coalition during this phase tend to be smaller and less intensive in energy extraction, they have a smaller impact on emissions reduction when they leave the coalition: we do not see the same dramatic spikes in global emissions that take place when Russia, Saudi Arabia, and Rest of World leave the coalition in Phase 1, so global emissions reduction steadily trend upwards. However, because the largest energy extractors no longer participate in the coalition, extraction leakage, and to a lesser extent production leakage, become much more significant issues. As the coalition attempts to curtail emissions, it increases the global energy price, which increases noncoalition extraction and curtails the gains made by the more aggressive policy. The flatter slope in Phase 2 than in Phase 1 reflects the negative impact of leakage on the effectiveness on the optimal policy.

Phase 3 (emissions aversion greater than $\sim$ 1.9) begins when the post-tax price that coalition extractors receive dips below zero. At this point, coalition countries stop extracting any energy, so the coalition can only influence global emissions by decreasing the global energy price. This is largely ineffective, because it decreases noncoalition energy supply but inadvertently increases noncoalition energy demand, mitigating any progress. Because of this, the coalition decreases the energy price very slowly, and global emissions decrease almost imperceptibly. Meanwhile, the countries left in the coalition, most of whom extract very little energy and depend heavily on trade, are asked to pay a higher and higher border adjustment when they produce goods. This decreases their consumption surplus, and eventually, they leave the coalition as well. During this phase, EU countries such as France, Spain, and Germany leave the coalition, as do Japan and South Korea. These countries leaving the coalition wipe out the imperceptible gains made from decreasing the global energy price so that global emissions increase during most of Phase 3 as they did in Phase 1.

During Phase 3, the energy policy becomes so extreme that the trade benefits of participation become negligible. Moreover, the energy price decreases very slowly, so change in extraction surplus moves only marginally. At this point, two main forces drive further changes in INOUT Utility Change: 1) the border adjustment increases rapidly, increasing the cost of production for coalition producers, which decreases consumption surplus for coalition consumers and 2) emissions aversion continues to grow, so each country is more motivated by their own emissions reductions when they participate in the coalition. Critically, the first driver depends on the border adjustment, and by extension, the marginal damages of the whole coalition. The second driver, while potentially more impactful, only depends on a country's individual marginal damages. Because of this discrepancy, INOUT Utility Change actually increases during Phase 3 for very large countries whose own marginal damages make up a significant fraction of the coalition's collective marginal damages. This effect causes China and the United States to stay in the coalition throughout the simulation and Rest of World to rejoin the coalition during Phase 3, while every other country eventually drops out.[25] Essentially, as emissions aversion reaches extreme levels, size becomes the only determinant of participation. Critically, this effect only occurs because noncoalition

[25]As a reminder, Rest of World functions as a single country in our simulation with energy demand and supply aggregated over all countries that we do not have data for. Because of this, it acts unilaterally, and rejoins the coalition because it is very large.

countries extract and produce at competitive equilibrium levels, not unilaterally optimal levels. If countries factored in their own marginal damages when making extraction and production decisions, they would balance the two drivers optimally themselves, providing no incentive to join the coalition beyond the negligible trade benefits. If this were the case, both the United States and China would eventually leave the coalition, and Rest of World would never rejoin, causing Figure 3 to resemble a Laffer curve similar to the one found in Nordhaus. Because it seems likely that very large countries would implement some sort of unilateral carbon policy if there were truly extreme marginal damages from emissions and therefore would not participate in the coalition, we believe that the emissions reductions achieved during Phase 3 are likely an overestimate.

5 Summary and Implications

5.1 Summary of Key Findings

The optimal tax policy derived in Section 3, a generalization of the solution presented in Kortum and Weisbach, consists of an extraction tax and a partial border adjustment that partially shifts the incidence of the tax downstream to minimize overall leakage. The border adjustment increases the price of energy faced by coalition producers and by noncoalition producers when selling goods to coalition consumers. Since goods consumed by coalition countries trade at their cost of production, consumers internalize this tax, so the incidence falls roughly equally on all countries.[26] The extraction tax, on the other hand, decreases the price that coalition countries receive for energy they extract, causing the incidence of the extraction tax to fall disproportionately on countries whose economies are highly-dependent on energy extraction. This makes energy extractors less amenable to coalitions implementing the optimal policy, a result we confirm in Section 4.2.3.

[26] More specifically, it falls on countries proportionally to their consumption levels and marginal damages from emissions. But because we defined marginal damages from emissions to be proportional to consumption, these effects cancel, and the incidence is roughly constant.

The optimal policy also includes an export tax on goods that arises naturally as the optimal-pricing strategy in a one-period game, but incidentally serves as an external penalty imposed on noncoalition countries. Coalition producers generate a production surplus by selling goods to noncoalition countries at a markup, while noncoalition consumers suffer because they are forced to pay a premium for imported goods. These taxes are applied on a per-good basis, so their impact on a country is directly proportional to how much that country trades. It follows that countries whose economies are heavily dependent on trade are more incentivized to participate in the coalition due to the heightened importance of the external penalties. Section 4.2.3 confirms this intuition, finding that highly-autarkic countries are less incentivized to participate in a global climate coalition.

Throughout Section 4, we unpack how countries' incentives to participate in a climate coalition change as the coalition's carbon policy becomes more aggressive, which we simulate by varying a parameter we call emissions aversion that serves as a rough proxy for the global social cost of carbon. We find that the trade benefits of participating in a coalition implementing external penalties are roughly invariant of emissions aversion, but the negative impact that the border adjustment and extraction tax have on production and extraction efficiency becomes more exaggerated as emissions aversion increases. This result is unsurprising and intuitive— as emissions aversion increases, the wedge between the locally optimal and the coalition optimal quantities of energy extracted and energy used in production grows, making it more painful to adopt the coalition optimal levels. In Section 4.2.4, we find that, as a consequence of this inverse relationship, at a sufficiently large emissions aversion, virtually every single country wants to free ride off of coalition emissions reduction without participating in the coalition themselves, accepting the trade penalties.

The question then becomes when, and not if, a carbon tax policy eventually becomes too extreme for a country to willingly participate. In Section 4.2.4, our simulation suggests that this answer varies significantly depending on the underlying characteristics of the country. Because highly intensive energy extractors shoulder a disproportionate amount of the incidence of the extraction tax, we find that high-intensity energy extractors tend to leave the coalition at lower levels of emissions aversion than other countries. We find a similar result for relatively autarkic countries, since the trade benefits of participating in the coalition are lower for these countries. Of these two underlying determinants, intensity in energy extraction tends to play a more prominent role, especially as emissions aversion increase to moderate to high levels, due to the fact that the loss in extraction surplus from participating grows superlinearly with respect to emissions aversion, a result we discussed in detail in Section 4.2.2.

In Section 4.2.4, we analyzed the stability of a coalition implementing the optimal policy in light of these concerns. We find that, at low levels of emissions aversion, trade penalties are sufficient to motivate every single country to participate in the carbon coalition. Under these low levels of emissions aversion, the tax policy is extremely effective because there is no leakage. Our simulation finds that, at the largest emissions aversion for which the global coalition is stable, global emissions are reduced by more than 35%. At higher levels of emissions aversion, the global coalition becomes unstable, because large energy extractors prefer to abstain from the coalition, suffering the trade penalties but extracting energy at competitive equilibrium levels. This has a dramatic impact on global emissions reductions, especially because the largest stable coalition is subject to a high degree of adverse selection—it is precisely the countries that are most responsible for carbon emissions that refuse to participate in the coalition. In fact, there is a period where making the carbon policy more aggressive by increasing emissions aversion actually increases global emissions because the impact of large energy extractors leaving the coalition outweighs any further reductions made by the countries that remain in the coalition. At very high levels of emissions aversion, all large energy extractors abstain from the coalition, so further increases in emissions aversion become effective at reducing emissions again, albeit at a diminished level because the large bloc of noncoalition energy extractors leads to significant leakage. Our simulation finds that by internalizing very high levels of emissions aversion, global emissions can be reduced by up to 40%, though we have reason to suspect this number is an overestimate (see the discussion in Section 4.2.4).

5.2 Implications: Just how Optimal is the Optimal Policy?

Previous analyses of the optimal unilateral carbon policy, including Markusen, Hoel, and Kortum and Weisbach, have treated participation in the coalition as exogenous. Our paper endogenizes participation, and through this endogenization, paints a far less optimistic picture of the effectiveness of the optimal carbon policy, especially at moderate to high levels of emissions aversion. We find that, beyond a certain critical threshold of emissions aversion, marginal increases in emissions aversion become highly ineffective and sometimes counterproductive at reducing emissions when the coalition implements the optimal policy. This ineffectiveness is a consequence of the fact that more aggressive policies induce countries to leave the coalition, counteracting the gains made by those policies. This result is both surprising and concerning: the fact that coalitions becomes less stable, rather than more stable, at higher levels of emissions aversion undermines the ability of a coalition to dramatically reduce emissions and casts doubt on the hopes for an 11th hour cooperative solution.

It seems likely that the optimal policy could be modified to improve its performance when countries' participation decisions are endogenized and the coalition operates as a climate club. There are several ways a coalition might go about modifying the optimal policy. It could, for example, place greater weight on the production tax as opposed to the extraction tax than has been previously suggested. This would increase overall leakage, but spread out the incidence of the tax more equally in hopes of increasing coalition participation. Alternatively, a coalition could consider utility transfers to extractors if they participate in the coalition, such as paying large energy extractors for energy that they don't pull out of the ground.[27] A third possibility is to simply reflect on the fact that there may not be a "true" value

[27] Note that by introducing surplus transfers into our model, we open ourselves to what Nordhaus refers to as "stab in the back" coalition instability. This type of instability goes beyond the scope of this paper but is an important for consideration for certain modifications of the optimal policy.

for emissions aversion. Any estimate of the social cost of carbon must depend heavily on a discount rate, which is socially-estimated and selected. In this sense, society selects the emissions aversion by selecting the discount rate. Our results indicate that, for climate clubs implementing the optimal unilateral policy, selecting a low discount rate, and by extension a low emissions aversion, might be just as effective at reducing global emissions but spread the tax incidence out more fairly and evenly across a broader, more stable coalition.

6 Conclusion

By generalizing the results of Kortum and Weisbach to a multi-country model, our model begins to bridge the optimal unilateral policy school and the carbon club school. Our solution combines an extraction tax with a border adjustment, supporting previous literature surrounding optimal unilateral tax policy, but features an additional export policy that serves as an external penalty non-members, incentivizing participation. The export policy provides a mechanism by which a coalition implementing the optimal policy functions as a climate club. However, our results suggest that while the optimal policy optimizes the collective welfare of the coalition, it affects countries highly unequally. At moderate to high marginal damages from emissions, the unequal impact of the optimal policy can be so dramatic that certain countries– mostly autarkic countries that are highly intensive in energy extraction– prefer to leave the coalition and suffer the external penalties, dramatically reducing the effectiveness of the coaliton-optimal policy. While our model is highly stylized, its unique approach provides key insights into the effectiveness of a coalition implementing the optimal unilateral policy and operating as a climate club.

However, while our model hints at the weak points of the optimal unilateral policy, it does not directly answer 1) how a coalition implementing an optimal policy can modify its approach to enhance coalition stability and improve its performance as a climate club and 2) how a climate club can standardize its tax imposition in light of the fact that true global cooperation seems unlikely and leakage remains a critical concern. Analyzing these two questions poses an important area of future research.

References

Branger, Frédéric, and Philippe Quiron.(2014), "Climate policy and the 'carbon haven' effect," *Wiley Interdisciplinary Reviews: Climate Change*, 5, 53–71.

Costinot, Arnaud, Dave Donaldson, Jonathan Vogel, and Ivan Werning (2015), "Comparative Advantage and Optimal Trade Policy," *Quarterly Journal of Economics*, 659-702.

Dornbusch, Rudiger, Stanley Fischer, and Paul A. Samuelson (1977)l "Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods," *American Economic Review*, 67, 823-839.

Eaton, Jonathan and Samuel Kortum (2002), "Technology, Geography, and Trade," *Econometrica*, 70, 1741-1779.

Farrokhi, Farid and Ahmad Lashkaripour (2021) "Trade, Firm-Delocation, and Optimal Environmental Policy," unpublished, Purdue University.

Fowlie, Merideth (2009), "Incomplete Environmental Regulation, Imperfect Competition, and Emissions Leakage," *American Economic Journal*, 1, 72-112.

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

IEA (2022), "World energy balances," *IEA World Energy Statistics and Balances (database)*, https://www.iea.org/data-and-statistics

Markusen, James R. (1975), "International Externalities and Optimal Tax Structures," *Journal of International Economics*, 5, 15-29.

Kortum, Samuel and David Weisbach (2015), "Optimal Unilateral Carbon Policy," *Working Paper*, Yale University.

Nordhaus, William (2015), "Climate Clubs: Overcoming Free-Riding in International Climate Policy," *American Economic Review*, 105, 1139-70.

Nordhaus, William (2017), "Revisiting the Social Cost of Carbon," *PNAS*, 114, 1518-1523.

OECD, Trade in Embodied CO2 Database (TECO2), April 2022.

A Competitive Equilibrium

In the competitive equilibrium scenario, producers sell goods at their cost of production:

$$p_{ni}(j) = d_{ni}\Big[l_i\big(j, z_{ni}(j)\big) + p^E e_i\big(j, z_{ni}(j)\big)\Big]$$

Each producer i uses the cost-minimizing energy intensity when producing a good j for a consumer n, which we derive by taking the FOC of $p_{ni}(j)$ with respect to $z_{ni}(j)$

$$\begin{aligned}\frac{\partial p_{ni}(j)}{\partial z_{ni}(j)} &= d_{ni}\Big[\frac{\partial l_i\big(j, z_{ni}(j)\big)}{\partial z_{ni}(j)} + p^E \frac{\partial e_i\big(j, z_{ni}(j)\big)}{\partial z_{ni}(j)}\Big] \\ &= d_{ni}\Big[(1-\alpha)\frac{\nu}{A_i(j)} z_{ni}(j)^{\alpha-2} - p^E \alpha \frac{\nu}{A_i(j)} z_{ni}(j)^{\alpha-1}\Big]\end{aligned}$$

which implies a constant energy intensity

$$z_{ni}(j) = \frac{1-\alpha}{\alpha p^E}$$

that each producer uses regardless of the consumer and specific good j. Given this energy intensity, the price $p_{ni}(j)$ is given by:

$$p_{ni}(j) = \frac{d_{ni}}{A_i(j)}(p^E)^{1-\alpha}$$

Consumers buy goods from the cheapest producer until the price they face equals their marginal production, which implies:

$$c_{ni}(j) = I_{ni}(j)\eta_i\Big(p_{ni}(j)\Big)^{-\sigma_i}$$

$$I_{ni}(j) = \begin{cases} 1 & p_{ni}(j) \leq p_{nw}(j) \text{ for all } w \in W \\ 0 & \text{otherwise} \end{cases}$$

Aggregating these functions via

$$\sum_{i \in W} C_i^E = \sum_{i \in W} \sum_{n \in W} \int_0^\infty c_{ni}(j) dj$$

gives global energy demand as a function of p^E. Each country i extracts energy so that the marginal cost of extraction equals the energy price:

$$Q_i^E = D_i(p^E)$$

Global energy supply is then given by:

$$\sum_{i \in W} Q_i^E = \sum_{i \in W} D_i(p^E)$$

Equating $\sum_{n \in W} C_n^E = \sum_{i \in W} Q_i^E$ pins down a common energy price that solves the competitive equilibrium.

B Global Planner's Problem

We now derive the Global Planner's Problem as a special case of the Coalition Planner's Problem where $W' = W$. Throughout the derivation, we leave out the $'$ denoting a coalition country since all countries are coalition countries. The Planner's objective function is given by:

$$U_W = \sum_{w \in W} (C_w^S + \eta_w^{1/\sigma_w} \frac{(C_w^G)^{1-1/\sigma_w}}{1 - 1/\sigma_w} - \varphi_w \sum_{i \in W} Q_i^E)$$

Global services consumption is linked to global labor usage via:

$$\sum_{w\in W} C_w^S = \sum_{w\in W} \Big(L_w - L_w^G - L_w^E \Big)$$

which we can substitute directly into the objective:

$$U_W = \sum_{w\in W} (\eta_w^{1/\sigma_w\prime} \frac{(C_w^G)^{1-1/\sigma_w}}{1-1/\sigma_w} - \varphi_w Q_W^E + L_w - L_w^G - L_w^E)$$

The planner faces a global energy constraint

$$\sum_{n\in W} C_n^E \leq \sum_{i\in W} Q_i^E$$

which we include using a Lagrange Multiplier:

$$\mathcal{L} = \sum_{w\in W} \big(\eta_w^{1/\sigma_w\prime} \frac{(C_w^G)^{1-1/\sigma_w}}{1-1/\sigma_w} - \varphi_w Q_W^E + L_w - L_w^G - L_w^E - \lambda(C_w^E - Q_w^E)\big)$$

B.1 The Inner Problem and Energy Demand

The Inner Problem is given by:

$$\begin{aligned}\mathcal{L}(j) = \sum_{w\in W} \Bigg\{ &\frac{\eta_w^{1/\sigma_w}}{1-1/\sigma_w} \big(\sum_{i\in W} c_{wi}(j)\big)^{1-1/\sigma_w} \\ &- \sum_{n\in W} \Big[l_w\big(j, z_{nw}(j)\big) d_{nw} c_{nw}(j) - \lambda \sum_{n\in W} e_w\big(j, z_{nw}(j)\big) d_{nw} c_{nw}(j) \Big] \Bigg\}\end{aligned}$$

B.1.1 Energy Intensity

The FOC with respect to $z_{ni}(j)$ is:

$$\frac{\partial \mathcal{L}}{\partial z_{ni}(j)} = -\Big[\frac{\partial l_i\big(j, z_{ni}(j)\big)}{\partial z_{ni}(j)} + \lambda \frac{\partial e_i\big(j, z_{ni}(j)\big)}{\partial z_{ni}(j)}\Big] d_{ni} c_{ni}(j)$$
$$= \Big[(1-\alpha)\frac{\nu}{A_i(j)} z_{ni}(j)^{\alpha-2} - \lambda\alpha \frac{\nu}{A_i(j)} z_{ni}(j)^{\alpha-1}\Big] d_{ni} c_{ni}(j)$$

So the planner optimizes global energy intensity by setting $z_{ni}(j) = \frac{1-\alpha}{\alpha\lambda}$ for all production.

B.1.2 Consumption

The FOC with respect to $c_{ni}(j)$ is:

$$\frac{\partial \mathcal{L}(j)}{c_{ni}(j)} = \eta_n^{1/\sigma_n}\big(\sum_{w\in W} c_{nw}(j)\big)^{-1/\sigma_i} - d_{ni} l_i\big(j, z_{ni}(j)\big) - \lambda d_{ni} e_i\big(j, z_{ni}(j)\big)$$
$$= \eta_n^{1/\sigma_n}\big(\sum_{w\in W} c_{nw}(j)\big)^{-1/\sigma_n} - \frac{d_{ni}}{A_i(j)}(\lambda)^{1-\alpha}$$

The marginal benefit of consumption is given by $b_{ni}(j) = \eta_n^{1/\sigma_n}\big(\sum_{w\in W} c_{nw}(j)\big)^{-1/\sigma_n}$ and the marginal shadow cost is given by $\tilde{p}_{ni}(j) = \frac{d_{ni}}{A_i(j)}(\lambda)^{1-\alpha}$. As with the Coalition Planner's Problem, the marginal benefit is dependent on previous consumption by the consumer n but not on the producer i, and marginal shadow cost is independent of previous consumption but dependent on source country. So, the consumer n should consume good j from the same country i until $\frac{\partial \mathcal{L}}{\partial c_{ni}(j)} = 0$ and $\frac{\partial \mathcal{L}}{\partial c_{nw}(j)} \leq 0$ for all other w. So,

$$c_{ni}(j) = I_{ni}(j)\eta_n\Big(\tilde{p}_{ni}(j)\Big)^{-\sigma_n}$$

$$I_{ni}(j) = \begin{cases} 1 & \tilde{p}_{ni}(j) \leq \tilde{p}_{nw}(j) \text{ for all } w \in W \\ 0 & \text{otherwise} \end{cases}$$

Global energy demand is given by:

$$\sum_{i \in W} C_i^E = \sum_{i \in W} \sum_{n \in W} \int_0^1 c_{ni}(j) dj$$

B.2 The Outer Problem and Energy Supply

Turning back to the outer problem, the FOC with respect to Q_i^E is:

$$\frac{\partial \mathcal{L}}{\partial Q_i^E} = -\sum_{w \in W} \varphi_w - \frac{\partial L_i^E}{\partial Q_i^E} + \lambda$$

So the planner sets energy supply by country i as:

$$Q_i^E = D_i^{-1}(\lambda - \varphi_W)$$

Equating $\sum_{n \in W} C_n^E = \sum_{i \in W} Q_i^E$ pins down a value for λ, solving the global planner's problem.

C Calibration Specifications

We calibrate the theoretical values derived in Section 3 to real world data using the following process. First, we equate our expressions for energy demand and supply in the competitive equilibrium scenario to real world data from the TECO_2 and IEA databases. For convenience, we set energy price equal to 1 for this calibration, implicitly defining one unit of energy as the amount of energy that costs the same as one unit of services in competitive equilibrium. Next, we write optimal energy metrics in terms of competitive equilibrium energy metrics and the optimal tax policy (p^E, t^B, and t^E). We use these values to derive optimal labor and trade aggregates, still in terms of p^E, t^B, and t^E. In the sections that follow, an overbar is always used to indicate a competitive equilibrium aggregate; no overbar indicates an optimal aggregate.

t^E is easily calculated via $t^E = \sum_{w' \in W'} \varphi_{w'}$, where each $\varphi_{w'}$ is a parameter set ahead of time. To calculate p^E and t^B, we begin by assuming $p^E = 1$ and $t^B = 0$. Assuming $t^B = 0$, we update our energy price to clear the energy market. Assuming this energy price, we update the optimal t^B. We repeat this process iteratively until we find a fixed point, which represents the optimal energy price and border adjustment.[28] Given these values (p^E, t^B, t^E), we can quantify the optimal aggregates.

C.1 Optimal Energy Metrics in Terms of Competitive Equilibrium

We calculate closed forms for energy demand using an intermediate aggregate Y_n^G, which represents the post-tax expenditure of consumers within country n on goods. In Appendix D, we derive the closed for for Y_n^G in detail. The results are restated here:

$$\bar{Y_n^G} = \eta_n (\Phi_{nW})^{\frac{\sigma_n - 1}{\theta}} \ \Gamma(\frac{1 - \sigma_n + \theta}{\theta})$$

$$Y_{n'}^G = \eta_{n'} (\Phi_{n'W})^{\frac{\sigma_{n'} - 1}{\theta}} (p^E + t^B)^{1 - \epsilon_{n'}^D} \ \Gamma(\frac{1 - \sigma_{n'} + \theta}{\theta})$$

$$Y_{n^*}^G = \eta_{n^*} (\Phi_{n^*W^*})^{\frac{\sigma_{n^*} - 1}{\theta}} (p^E)^{1 - \epsilon_{n^*}^D} \ \Gamma(\frac{1 - \sigma_{n^*} + \theta}{\theta})$$

In these formulae, $\Phi_{nW} = \sum_{i \in W} (d_{ni})^{-\theta} T_i$, $\Phi_{nW^*} = \sum_{i^* \in W^*} (d_{ni^*})^{-\theta} T_{i^*}$, And Γ is the complete gamma function. Another useful aggregate is π_{ni}, the proportion of goods that country i supplies for country n (derivation found in Appendix A):

$$\bar{\pi_{ni}} = \frac{(d_{ni})^{-\theta} T_i}{\Phi_{nW}}$$

[28] In order to speed up the runtime of our algorithm, we slightly modified the implementation from what is described here, though conceptually the steps are the same. Examples of modifications include starting with an initial $t^B = \frac{t^E}{2}$ rather than $t^B = 0$, which empirically is a better first estimate, and jumping to the middle of two points when the algorithm is slowly nudging them closer and closer together.

$$\pi_{n'i} = \frac{(d_{n'i'})^{-\theta} T_{i'}}{\Phi_{n'W}}$$

$$\begin{aligned}
\pi_{n^*i'} &= \frac{S(p^E, t^B)^{-\theta}\ (d_{n^*i'})^{-\theta} T_{i'}}{S(p^E, t^B)^{-\theta}\ \Phi_{n^*W'} + \Phi_{n^*W^*}} \\
&= \frac{S(p^E, t^B)^{-\theta}\ \bar{\pi}_{n^*i'}}{S(p^E, t^B)^{-\theta}\ \bar{\pi}_{n^*W'} + \bar{\pi}_{n^*W^*}}
\end{aligned}$$

$$\begin{aligned}
\pi_{n^*i^*} &= \frac{(d_{n^*i^*})^{-\theta} T_{i^*}}{S(p^E, t^B)^{-\theta}\ \Phi_{n^*W'} + \Phi_{n^*W^*}} \\
&= \frac{\bar{\pi}_{n^*i^*}}{S(p^E, t^B)^{-\theta}\ \bar{\pi}_{n^*W'} + \bar{\pi}_{n^*W^*}}
\end{aligned}$$

Here, $S(p^E, t^B) = \left(\frac{p^E + t^B}{p^E}\right)^{1-\alpha}\left(1 + (1-\alpha)\frac{t^B}{p^E}\right)^{-1}$ is a scaling factor that increases coalition exports to noncoalition consumers, reflecting the coalition's desire to expand its export margin as a means of gaining more control over the energy used in goods production. We also use the shorthand $\pi_{nW'} = \sum_{i' \in W'} \pi_{ni'}$ and $\pi_{nW^*} = \sum_{i^* \in W^*} \pi_{n^*W^*}$ for convenience. In Appendix D, we derive Y^G_{ni}, the post-tax expenditure of consumers in n on goods produced in country i.

$$\bar{Y}^G_{ni} = \bar{\pi}_{ni} \bar{Y}^G_n$$

$$Y^G_{ni} = \pi_{ni} Y^G_n$$

We use Y^G_{ni} to derive closed form expressions for C^E_{ni}:

$$\bar{Y}^G_{ni} = \bar{L}^G_{ni} + \bar{C}^E_{ni} = \frac{\alpha}{1-\alpha}\bar{C}^E_{ni} + \bar{C}^E_{ni} = \frac{1}{1-\alpha}\bar{C}^E_{ni}$$

$$\iff$$

$$\bar{C}^E_{ni} = (1-\alpha)\bar{\pi}_{ni}\bar{Y}^G_n$$

The derivation is also straightforward for coalition consumers:

$$Y^G_{n'i} = L^G_{n'i} + (p^E + t^B)C^E_{n'i} = \frac{\alpha(p^E + t^B)}{1-\alpha}C^E_{n'i} + (p^E + t^B)C^E_{n'i} = \frac{p^E + t^B}{1-\alpha}C^E_{n'i}$$

$$\iff$$

$$C^E_{n'i} = \frac{1-\alpha}{p^E + t^B}\pi_{n'i}Y^G_{n'}$$

and for noncoalition to noncoalition trade:

$$Y^G_{n^*i^*} = L^G_{n^*i^*} + p^E C^E_{n^*i^*} = \frac{\alpha p^E}{1-\alpha}C^E_{n^*i^*} + p^E C^E_{n^*i^*} = \frac{p^E}{1-\alpha}C^E_{n^*i^*}$$

$$\iff$$

$$C^E_{n^*i^*} = \frac{1-\alpha}{p^E}\pi_{n^*i^*}Y^G_{n^*}$$

Unfortunately, we cannot employ the same trick for coalition to noncoalition trade, because the export taxes enforced by the coalition disconnect coalition energy usage from noncoalition expenditure. Instead, we must derive $C^E_{n^*i'}$ directly (derivation in Appendix D):[29]

$$C^E_{n^*i'}(p^E, t^B) = \Bigg\{ \eta_{n^*}\frac{1-\alpha}{(p^E + t^B)^\alpha}(p^E)^{-\sigma_{n^*}(1-\alpha)}\Big[\frac{\Phi_{n^*W'}}{(\Phi_{n^*W^*})^{-\sigma_{n^*}}}\Big]^{1/\theta} \quad \times$$
$$\int_0^\infty \gamma\Big(\frac{1+\theta}{\theta}, S(p^E, t^B)^{-\theta}\big(\frac{\Phi_{n^*W'}}{\Phi_{n^*W^*}}\big)v\Big)e^{-v}dv\Bigg\}$$

[29]In this expression, γ is the lower incomplete gamma function.

We use these relationships to express the optimal values of energy demand in terms of the competitive equilibrium values:

$$
\begin{aligned}
C^E_{n'i} &= \frac{1-\alpha}{p^E + t^B}\ \pi^G_{ni}\ Y^G_n \\
&= \frac{(\lambda')^{1-\epsilon^D_{n'}}}{p^E + t^B} \bar{C}^E_{n'i}
\end{aligned}
$$

$$
\begin{aligned}
C^E_{n^*i^*} &= \frac{1-\alpha}{p^E + t^B}\ \pi^G_{n^*i^*}\ Y^G_{n^*} \\
&= \frac{(p^E)^{1-\epsilon^D_{n^*}}}{p^E + t^B} \frac{\pi_{n^*i^*}}{\bar{\pi}_{n^*i^*}} \frac{\Phi_{n^*W^*}}{\Phi_{n^*W}} \bar{C}^E_{n^*i^*} \\
&= \frac{(p^E)^{1-\epsilon^D_{n^*}}}{p^E + t^B} \frac{\bar{\pi}_{n^*W^*}}{S(p^E, t^B)\bar{\pi}_{n^*W'} + \bar{\pi}_{n^*W^*}} \bar{C}^E_{n^*i^*}
\end{aligned}
$$

$$
\begin{aligned}
C^E_{n^*i'}(p^E, t^B) = &\Bigg\{ \left[\frac{(p^E)^{-\sigma_{n^*}(1-\alpha)}}{(p^E + t^B)^{\alpha}}\right] \left[\frac{\left(\bar{\pi}_{n^*W^*}\right)^{\sigma_{n^*}}}{\left(\bar{\pi}_{n^*W'}\right)^{1+\theta}}\right]^{\frac{1}{\theta}} \times \\
&\frac{\int_0^{\infty} \gamma\big(\frac{1+\theta}{\theta}, S(p^E, t^B)^{-\theta}(\frac{\pi_{n^*W'}}{\pi_{n^*W^*}})v\big)e^{-v}dv}{\Gamma(\frac{1-\sigma_{n^*}+\theta}{\theta})} \Bigg\} \times \bar{C}^E_{n^*i'}
\end{aligned}
$$

To quantify energy supply, we assume the functional form:

$$D_i(a) = D_i a^{\epsilon^S_i}$$

where D_i is a parameter governing the relative efficiency of country i in extracting energy. Using this functional form:

$$\bar{Q}^E_i = D_i$$

$$
\begin{aligned}
Q^E_{i'} &= D_{i'}(p^E + t^B - t^E)^{\epsilon^S_{i'}} \\
&= (p^E + t^B - t^E)^{\epsilon^S_{i'}} \bar{Q}^E_{i'}
\end{aligned}
$$

$$
\begin{aligned}
Q^E_{i^*}(p^E) &= D_{i^*}(p^E)^{\epsilon^S_{i^*}} \\
&= (p^E)^{\epsilon^S_{i^*}} \bar{Q}^E_{i^*}
\end{aligned}
$$

These equations completely specify optimal energy demand and supply in terms of competitive equilibrium values and parameters, quantifying the optimal energy aggregates.

C.2 Labor Metrics in Terms of Energy Metrics

We can write labor usage in terms of energy usage by applying our formulae for energy intensity:

$$
\bar{L}^G_{ni} = \frac{\alpha}{1-\alpha} \bar{C}^E_{ni}
$$

$$
L^G_{n'i} = \frac{\alpha(p^E + t^B)}{1-\alpha} C^E_{n'i}
$$

$$
L^G_{n^*i'} = \frac{\alpha(p^E + t^B)}{1-\alpha} C^E_{n^*i'}
$$

$$
L^G_{n^*i^*} = \frac{\alpha p^E}{1-\alpha} C^E_{n^*i^*}
$$

The general form of labor used in energy extraction is given by:

$$
L^E_i = \int_0^{\bar{a}} a \; dD_i(a)
$$

Applying our functional form for $D_i(a)$:

$$\begin{aligned} L_i^E(\bar{a}) &= \int_0^{\bar{a}} D_i \epsilon_i^S a^{\epsilon_i^S} \, da \\ &= D_i \frac{\epsilon_i^S}{\epsilon_i^S + 1} (\bar{a})^{\epsilon_i^S + 1} \\ &= \bar{a} \frac{\epsilon_i^S}{\epsilon_i^S + 1} Q_i^E \end{aligned}$$

This implies:

$$\bar{L_i^E} = \frac{\epsilon_i^S}{\epsilon_i^S + 1} \bar{Q_i^E}$$

$$L_{i'}^E = (p^E + t^B - t^E) \frac{\epsilon_{i'}^S}{\epsilon_{i'}^S + 1} Q_{i'}^E$$

$$L_{i^*}^E = p^E \frac{\epsilon_{i^*}^S}{\epsilon_{i^*}^S + 1} Q_{i^*}^E$$

C.3 Trade Value Metrics in Terms of Energy Metrics

In competitive equilibrium, goods trade at their cost of production:

$$\begin{aligned} \bar{V_{ni}^G} &= \bar{C_{ni}^E} + \bar{L_{ni}^E} \\ &= \bar{C_{ni}^E} + \frac{\alpha}{1 - \alpha} \bar{C_{ni}^E} \\ &= \frac{1}{1 - \alpha} C_{ni}^E \end{aligned}$$

Goods also trade at their cost of production for coalition to coalition, noncoalition to coalition, and noncoalition to noncoalition good flow:

$$\begin{aligned} V^G_{n'i'} &= (p^E + t^B)C^E_{n'i'} + L^E_{n'i'} \\ &= (p^E + t^B)C^E_{n'i'} + \frac{\alpha(p^E + t^B)}{1-\alpha}C^E_{n'i'} \\ &= \frac{p^E + t^B}{1-\alpha}C^E_{n'i'} \end{aligned}$$

$$\begin{aligned} V^G_{n'i^*} &= p^E C^E_{n'i^*} + L^E_{n'i^*} \\ &= p^E C^E_{n'i^*} + \frac{\alpha(p^E + t^B)}{1-\alpha}C^E_{n'i^*} \\ &= \frac{p^E + \alpha t^B}{1-\alpha}C^E_{n'i^*} \end{aligned}$$

$$\begin{aligned} V^G_{n^*i^*} &= p^E C^E_{n^*i^*} + L^E_{n^*i^*} \\ &= p^E C^E_{n^*i^*} + \frac{\alpha p^E}{1-\alpha}C_{n^*i^E} \\ &= \frac{p^E}{1-\alpha}C^E_{n^*i^*} \end{aligned}$$

Because the coalition planner applies the export tax to coalition producers exporting goods to noncoalition consumers, there is no underlying connection between the cost of production and the trade value for coalition to noncoalition good flow. However, we can derive their trade value indirectly through noncoalition to noncoalition good flow:

$$
\begin{aligned}
V^G_{n^*i'} &= \pi_{n^*i'} Y^G_{n^*} \\
&= \pi_{n^*i'} \Big(\frac{p^E}{1-\alpha} \frac{1}{\pi_{n^*W^*}} \sum_{i^* \in W^*} C^E_{n^*i^*} \Big) \\
&= \frac{p^E}{1-\alpha} \frac{\pi_{n^*i'}}{\pi_{n^*W^*}} \sum_{i^* \in W^*} C^E_{n^*i^*}
\end{aligned}
$$

C.4 Consumption Metrics in Terms of Energy Metrics

In both the competitive equilibrium and optimal scenarios, each country n consumes good j until their marginal benefit equals $\min_{i \in W} p_{ni}(j)$. This relates $\sum_{i \in W} c_{ni}(j)$ to $\min_{i \in W} p_{ni}(j)$, so we can express C^G_n in terms of expenditure:

$$
\begin{aligned}
C^G_n &= \Big(\int_0^1 \sum_{i \in W} c_{ni}(j)^{(\sigma_n - 1)/\sigma_n} dj \Big)^{\sigma_n/(\sigma_n-1)} \\
&= \Big(\int_0^1 \big(\eta_n \min_{i \in W} p_{ni}(j)^{-\sigma_n} \big)^{(\sigma_n-1)/\sigma_n} dj \Big)^{\sigma_n/(\sigma_n-1)} \\
&= \Big(\eta_n^{-1/\sigma_n} \int_0^1 \eta_n \big(\min_{i \in W} p_{ni}(j) \big)^{1-\sigma_n} dj \Big)^{\sigma_n/(\sigma_n-1)} \\
&= \eta^{1/(1-\sigma_n)} (Y^G_n)^{\sigma_n/(\sigma_n-1)}
\end{aligned}
$$

Raising consumption to the $(\sigma_n - 1)/\sigma_n$ power and multiplying by $\frac{\eta_n^{1/\sigma_n}}{1-1/\sigma_n}$, as we do in the utility function, gives us an expression independent of η_n:

$$
\frac{\eta_n^{1/\sigma_n}}{1 - 1/\sigma_n} \big(C^G_n \big)^{(\sigma_n-1)/\sigma_n} = \frac{\sigma_n}{\sigma_n - 1} Y^G_n
$$

C.5 Calculating Change in Utility

Because our utility function for a country w is dependent on services consumption, which is dependent on labor supply L_w, we cannot directly calculate the utility of each country without calibrating to an additional data source. However, we can calculate the change in utility from the competitive equilibrium scenario to the optimal scenario with only the energy data calibration and our predetermined parameters. Change in utility is given by:

$$
\begin{aligned}
U_w - \bar{U_w} = \Bigg\{ & C_w^S + \eta_w^{1/\sigma_w} \frac{(C_w^G)^{1-1/\sigma_w} - 1}{1 - 1/\sigma_w} - \varphi_w \sum_{i \in W} Q_i^E - \\
& \left[\bar{C_w^S} + \eta_w^{1/\sigma_w} \frac{(\bar{C_w^G})^{1-1/\sigma_w} - 1}{1 - 1/\sigma_w} - \varphi_w \sum_{i \in W} \bar{Q_i^E} \right] \Bigg\} \\
= & \frac{\sigma_w}{\sigma_w - 1} \left(Y_w^G - \bar{Y_w^G} \right) - \varphi_w \sum_{i \in W} \left(Q_i^E - \bar{Q_i^E} \right) \\
& - \left(L_w^G - \bar{L_w^G} + L_w^E - \bar{L_w^E} \right) + \sum_{n \in W} \left(X_{nw}^E - \bar{X_{nw}^E} + X_{nw}^G - \bar{X_{nw}^G} \right)
\end{aligned}
$$

Each value presented in the above expression can be quantified using the relationships derived in sections 9.1-9.4, giving us an explicit formula for the utility change of each country under the optimal policy.

C.5.1 Limit Case

We now consider utility change– specifically the consumption term– for a country n under the limit case $\sigma_n \to 1$. As shown in Appendix C.4, change in utility due to consumption is given by $\frac{\sigma_n}{\sigma_n - 1}\left(Y_n^G - \bar{Y_n^G}\right)$. Applying our expressions for Y_n^G and $Y_{n'}^G$ from Appendix A.1, this simplifies to the following expressions for coalition and noncoalition consumers, respectively:

$$
\frac{\eta_{n'}^{1/\sigma_{n'}}}{1 - 1/\sigma_{n'}} \left[\left(C_{n'}^G \right)^{(\sigma_{n'} - 1)/\sigma_{n'}} - \left(\bar{C_{n'}^G} \right)^{(\sigma_{n'} - 1)/\sigma_{n'}} \right] = \frac{\sigma_{n'}}{\sigma_{n'} - 1} \left((p^E + t^B)^{(1-\alpha)(1-\sigma_{n'})} - 1 \right) \bar{Y_{n'}^G}
$$

$$\frac{\eta_{n^*}^{1/\sigma_{n^*}}}{1-1/\sigma_{n^*}}\Big[\big(C_{n^*}^G\big)^{(\sigma_{n^*}-1)/\sigma_{n^*}} - \big(\bar{C}_{n^*}^G\big)^{(\sigma_{n^*}-1)/\sigma_{n^*}}\Big] = \frac{\sigma_{n^*}}{\sigma_{n^*}-1}\Big(\big(\pi_{n^*\bar{W}^*}\big)^{\frac{\sigma_{n^*}-1}{\theta}}\big(p^E\big)^{(1-\alpha)(1-\sigma_{n^*})} - 1\Big)\bar{Y}_{n^*}^G$$

Taking the limit $\sigma \to 1$, we apply L'Hopital's rule to get:

$$\begin{aligned}\lim_{\sigma_{n'}\to 1}\frac{\eta_{n'}^{1/\sigma_{n'}}}{1-1/\sigma_{n'}}\Big[\big(C_{n'}^G\big)^{(\sigma_{n'}-1)/\sigma_{n'}} - \big(\bar{C}_{n'}^G\big)^{(\sigma_{n'}-1)/\sigma_{n'}}\Big] &= -(1-\alpha)\ln(p^E + t^B)\ \bar{Y}_{n'}^G \\ &= -\ln(p^E + t^B)\sum_{i\in W}\bar{C}_{n'i}^E\end{aligned}$$

$$\begin{aligned}\lim_{\sigma_{n^*}\to 1}\frac{\eta_{n^*}^{1/\sigma_{n^*}}}{1-1/\sigma_{n^*}}\Big[\big(C_{n^*}^G\big)^{(\sigma_{n^*}-1)/\sigma_{n^*}} - \big(\bar{C}_{n^*}^G\big)^{(\sigma_{n^*}-1)/\sigma_{n^*}}\Big] &= -\Big[(1-\alpha)\ln\big(p^E\big) - \frac{1}{\theta}\ln\big(\pi_{n^*\bar{W}^*}\big)\Big]\bar{Y}_{n^*}^G \\ &= -\Big[\ln(p^E) - \frac{\ln(\pi_{n^*\bar{W}^*})}{\theta(1-\alpha)}\Big]\sum_{i\in W}\bar{C}_{n^*i}^E\end{aligned}$$

These formulae are the ones used to calculate utility change for our simulations in Section 4, where we assume $\sigma_w = 1$ for each country w.

D Further Derivations

D.1 Goods Expenditure

Before calculating expenditure, we first calculate a CDF for consumption. In this example, we do this for the competitive equilibrium scenario, and explain the minor variations in the optimal scenario and their implications. In competitive equilibrium, consumption is given by:

$$c_n(j) = \eta_n\Big(\min_{i\in W}\frac{A_i(j)}{(p^E)^{1-\alpha}d_{ni}}\Big)^{\sigma_n}$$

This is probabilistic because of its dependence on $A_i(j)$. For an arbitrary producer i, the CDF of $A_i(j)$ is given by:

$$F_i^A(a) = Pr\big[A_i(j) \leq a\big] = e^{-T_i a^{-\theta}}$$

We want to derive a CDF for $c_n(j)$, which we denote F_n^C:

$$F_n^C(c) = Pr[c_n(j) \leq c]$$

Substituting in our formula for consumption:

$$\begin{aligned} F_n^C &= Pr\Big[\eta_n\big(\min_{i \in W} \frac{A_i(j)}{(p^E)^{1-\alpha} d_{ni}}\big)^{\sigma_n} \leq c\Big] \\ &= Pr\Big[\min_{i \in W} \frac{A_i(j)}{d_{ni}} \leq \big(\frac{c}{\eta_n}\big)^{1/\sigma_n} (p^E)^{1-\alpha}\Big] \\ &= 1 - Pr\Big[\min_{i \in W} \frac{A_i(j)}{d_{ni}} \geq \big(\frac{c}{\eta_n}\big)^{1/\sigma_n} (p^E)^{1-\alpha}\Big] \end{aligned}$$

$\min_{i \in W} \frac{A_i(j)}{d_{ni}} \geq \big(\frac{c}{\eta_n}\big)^{1/\sigma_n} (p^E)^{1-\alpha}$ if and only if for all $i \in W$, $\frac{A_i(j)}{d_{ni}} \geq \big(\frac{c}{\eta_n}\big)^{1/\sigma_n} (p^E)^{1-\alpha}$. Because of this, we can rewrite our CDF as a product:

$$\begin{aligned} F_n^c &= 1 - \Pi_{i \in W} Pr\Big[\frac{A_i(j)}{d_{ni}} \geq \big(\frac{c}{\eta_n}\big)^{1/\sigma_n} (p^E)^{1-\alpha}\Big] \\ &= 1 - \Pi_{i \in W}\Bigg(1 - Pr\Big[A_i(j) \leq \big(\frac{c}{\eta_n}\big)^{1/\sigma_n} (p^E)^{1-\alpha} d_{ni}\Big]\Bigg) \\ &= 1 - \Pi_{i \in W} e^{-T_i[(c/\eta_n)^{1/\sigma_n}(p^E)^{1-\alpha} d_{ni}]^{-\theta}} \\ &= 1 - e^{-[(c/\eta_n)^{1/\sigma_n}(p^E)^{1-\alpha}]^{-\theta}\Phi_{nW}} \end{aligned}$$

In competitive equilibrium, consumers consume until their marginal utility equals the price they face. That is, $\min_{i\in W} p_{ni}(j) = \big(c_n(j)/\eta_n\big)^{1-1/\sigma_n}$. Total expenditure is then given by:

$$Y_n^G = \int_0^1 \min_{i\in W} p_{ni}(j)c_{ni}(j)dj = \int_0^1 \eta_n\big(c_n(j)/\eta_n\big)^{1-1/\sigma_n}dj$$

Since there are infinitely many goods, we can apply the law of large numbers:

$$Y_n^G = \mathbb{E}\Big[Y_n^G\Big] = \mathbb{E}\Big[\eta_n\big(c_{ni}(j)/\eta_n\big)^{1-1/\sigma_n}\Big] = \eta_n\int_0^\infty \big(c/\eta_n\big)^{1-1/\sigma_n}dF_n^C$$

We can apply our CDF F_n^C, letting $u = [(c/\eta_n)^{1/\sigma_n}(p^E)^{1-\alpha}]^{-\theta}\Phi_{nW}$:

$$Y_n^G = \eta_n\int_0^\infty (c/\eta_n)^{1-1/\sigma_n}e^{-u}du$$

Inverting our u-substitution, we have:

$$(c/\eta_n) = u^{-\sigma_n/\theta}(p^E)^{-\sigma_n(1-\alpha)}(\Phi_{nW})^{\sigma_n/\theta}$$

$$(c/\eta_n)^{1-1/\sigma_n} = u^{(1-\sigma_n)/\theta}(p^E)^{1-\epsilon_n^D}(\Phi_{nW})^{(\sigma_n-1)/\theta}$$

Substituting this into our expression for Y_n^G:

$$\begin{aligned}
Y_n^G &= \eta_n\int_0^\infty u^{(1-\sigma_n)/\theta}(p^E)^{1-\epsilon_n^D}(\Phi_{nW})^{(\sigma_n-1)/\theta}e^{-u}du \\
&= \eta_n(p^E)^{1-\epsilon_n^D}(\Phi_{nW})^{(\sigma_n-1)/\theta}\int_0^\infty u^{(1-\sigma_n)/\theta}e^{-u}du \\
&= \eta_n(p^E)^{1-\epsilon_n^D}\ (\Phi_{nW})^{(\sigma_n-1)/\theta}\ \Gamma\Big[\frac{1-\sigma_n+\theta}{\theta}\Big]
\end{aligned}$$

In the competitive equilibrium, we normalize $p^E = 1$, so:

$$\bar{Y_n^G} - \eta_n\ (\Phi_{nW})^{(\sigma_n-1)/\theta}\ \Gamma\Big[\frac{1-\sigma_n+\theta}{\theta}\Big]$$

The derivations for expenditure in the coalition-optimal equilibrium are almost identical. For coalition countries, the only difference is that $(p^E + t^B)$ takes the places of p^E:

$$Y^G_{n'} = \eta_{n'}(p^E + t^B)^{1-\epsilon^D_{n'}} \ (\Phi_{n'W})^{(\sigma_{n'}-1)/\theta} \ \Gamma\Big[\frac{1-\sigma_{n'}+\theta}{\theta}\Big]$$

For noncoalition countries, the only difference is that they always pay the cost of production of the cheapest noncoalition producer for a good, even if a coalition producer could manufacture that good more efficiently. This replaces Φ_{n^*W} with $\Phi_{n^*W^*}$:

$$Y^G_{n^*} = \eta_{n^*}(p^E)^{1-\epsilon^D_{n^*}} \ (\Phi_{n^*W^*})^{(\sigma_{n^*}-1)/\theta} \ \Gamma\Big[\frac{1-\sigma_{n^*}+\theta}{\theta}\Big]$$

D.2 Trade Shares

The trade share of a producer i for a consumer n, denoted π_{ni} represents the portion of goods $j \in [0,1]$ that country i produces for country n. We begin by deriving the competitive equilibrium value and then discuss the changes to the derivation in the optimal equilibrium. Applying the law of large numbers, we can write this as:

$$\pi_{ni} = \mathbb{E}\big[I_{ni}(j)\big]$$

In competitive equilibrium, country i produces good j for country n if and only if it is the cheapest producer.

$$\begin{aligned}
\pi_{ni} &= \mathbb{E}\big[I_{ni}(j)\big] \\
&= Pr\big[p_{ni}(j) \leq p_{nw}(j) \text{ for all } w \in W, w \neq i\big] \\
&= Pr\Big[\frac{d_{ni}}{A_{ni}(j)} \leq \frac{d_{nw}}{A_w(j)} \text{ for all } w \in W, w \neq i\Big] \\
&= Pr\Big[\frac{d_{nw}}{d_{ni}} A_{ni}(j) \geq A_w(j) \text{ for all } w \in W, w \neq i\Big]
\end{aligned}$$

We can rewrite this probability as a product:

$$\pi_{ni} = \Pi_{w \in W \setminus i} Pr\big[\frac{d_{nw}}{d_{ni}} A_{ni}(j) \geq A_w(j)\big]$$

And then as an integral, using conditional probabilities:

$$\begin{aligned}
\pi_{ni} &= \int_0^\infty \Pi_{w \in W \setminus i} Pr\big[A_w(j) \leq \frac{d_{nw}}{d_{ni}} A_i(j) \mid A_i(j) = a\big] dF_i^A \\
&= \int_0^\infty \Pi_{w \in W \setminus i} e^{-T_w(\frac{d_{nw}}{d_{ni}} a)^{-\theta}} dF_i^A \\
&= \int_0^\infty \theta a^{-\theta-1} T_i e^{-T_i a^{-\theta}} \; \Pi_{w \in W \setminus i} e^{-T_w(\frac{d_{nw}}{d_{ni}} a)^{-\theta}} da \\
&= \int_0^\infty \theta a^{-\theta-1} T_i e^{-(a/d_{ni})^{-\theta} \sum_{w \in W} T_w (d_{nw})^{-\theta}} da \\
&= \int_0^\infty \theta a^{-\theta-1} T_i e^{-\Phi_{nW}(a/d_{ni})^{-\theta}} da
\end{aligned}$$

We now apply a u-substitution, where $u = \Phi_{nW}(a/d_{ni})^{-\theta}$. Then $du = \Phi_{nW} \; (d_{ni})^{\theta} \theta a^{-\theta-1} da$ and $da = \frac{1}{\Phi_{nW}} \; \frac{1}{\theta} (d_{ni})^{-\theta} a^{\theta+1} du$

$$\begin{aligned}
\pi_{ni} &= \frac{T_i (d_{ni})^{-\theta}}{\Phi_{nW}} \int_0^\infty e^{-u} du \\
&= \frac{T_i (d_{ni})^{-\theta}}{\Phi_{nW}}
\end{aligned}$$

In the coalition-optimal equilibrium, countries producing goods for coalition countries face energy price $p^E + t^B$, but otherwise the scenario is identical. Because $\bar{\pi_{ni}}$ is independent of the energy price, $\pi_{n'i} = \bar{\pi_{ni}} = \frac{T_i (d_{ni})^{-\theta}}{\Phi_{nW}}$. The proportion of goods supplied to noncoalition consumers is more complicated due to the additional term in the FOC. The coalition's subsidy policy sometimes encourages coalition producers to sell goods to noncoalition consumers even though they have a higher cost of production than a noncoalition producer. Coalition consumers only sell goods to noncoalition producers if the maximum subsidy offered by the planner lowers the post-subsidy cost of production below the cost of production of the cheapest noncoalition producer. This constraint is represented by:

$$I_{n^*i'}(j) = \begin{cases} 1 & t_{n^*i'}(j) \geq \min_{i' \in W'} \min_{i^* \in W^*} p_{n^*i^*}(j) - \frac{d_{n^*i'}}{A_{i'}(j)}(p^E + t^B)^{1-\alpha} \\ & \text{and } \frac{d_{n^*i'}}{A_{i'}(j)} \leq \frac{d_{n^*w'}}{A_{w'}(j)} \text{ for all } w' \in W' \\ \\ 0 & \text{otherwise} \end{cases}$$

Substituting our formulae for $p_{n^*i^*}(j)$ and $t_{n^*i'}(j)$:

$$I_{n^*i'}(j) = \begin{cases} 1 & -(1-\alpha) \min\limits_{i^* \in W^*} \frac{d_{n^*i^*}}{A_{i^*}(j)} \frac{t^B}{(p^E)^\alpha} \geq \min\limits_{i^* \in W^*} \frac{d_{n^*i^*}}{A_{i^*}(j)}(p^E)^{1-\alpha} - \frac{d_{n^*i'}}{A_{i'}(j)}(p^E + t^B)^{1-\alpha} \\ & \text{and } \frac{d_{n^*i'}}{A_{i'}(j)} \leq \frac{d_{n^*w'}}{A_{w'}(j)} \text{ for all } w' \in W' \\ \\ 0 & \text{otherwise} \end{cases}$$

Rearranging this equation gives:

$$I_{n^*i'}(j) = \begin{cases} 1 & \big(\frac{p^E + t^B}{p^E}\big)^{1-\alpha} \big(1 + (1-\alpha)\frac{t^B}{p^E}\big)^{-1} \frac{d_{n^*i'}}{A_{i'}(j)} \leq \min_{i^* \in W^*} \frac{d_{n^*i^*}}{A_{i^*}} \\ & \text{and } \frac{d_{n^*i'}}{A_{i'}(j)} \leq \frac{d_{n^*w'}}{A_{w'}(j)} \text{ for all } w' \in W' \\ \\ 0 & \text{otherwise} \end{cases}$$

Written in this way, we can see that the export policy of the coalition effectively scales $d_{n^*i'}$ by a factor:

$$S(p^E, t^B) = \big(\frac{p^E + t^B}{p^E}\big)^{1-\alpha} \big(1 + (1-\alpha)\frac{t^B}{p^E}\big)^{-1}$$

This scaling factor passes through the trade share derivation, scaling the contribution of coalition countries to Φ_{n^*W} by a factor of $S(p^E, t^B)^{-\theta}$:

$$\pi_{n^*i'} = S(p^E, t^B)^{-\theta} \frac{T_{i'}(d_{n^*i'})^{-\theta}}{S(p^E, t^B)^{-\theta}\Phi_{n^*W'} + \Phi_{n^*W^*}}$$

$$\pi_{n^*i^*} = \frac{T_{i^*}(d_{n^*i^*})^{-\theta}}{S(p^E, t^B)^{-\theta}\Phi_{n^*W'} + \Phi_{n^*W^*}}$$

D.3 Country to Country Expenditure

In this section, we derive Y^G_{ni}, which is the amount consumers in n spend on goods that were originally produced by country n. We begin with the explicit formula:

$$\begin{aligned}
Y^G_{ni} &= \int_0^1 p_{ni}(j)c_{ni}(j)dj \\
&= \eta_n \int_0^1 I_{ni}(j)\big(p_{ni}(j)\big)^{1-\sigma_n} dj \\
&= \eta_n \int_0^\infty p^{1-\sigma_n}\mathbb{E}\big[I_{ni}(j) \mid p_{ni}(j) = p\big] Pr[p_{ni}(j) = p\big] dp \\
&= \eta_n \int_0^\infty p^{1-\sigma_n} \; Pr\Big[p_{ni}(j) \le \min_{w\in W\setminus i} p_{nw}(j) \mid p_{ni}(j) = p\Big] Pr\Big[p_{ni}(j) = p\Big] \; dp \\
&= \eta_n \int_0^\infty p^{1-\sigma_n} \; \Big(1 - Pr\big[\min_{w\in W\setminus i} p_{nw}(j) \le p\big]\Big) \; Pr\Big[p_{ni}(j) = p\Big] \; dp
\end{aligned}$$

In competitive equilibrium, we can derive probability distributions over prices as follows:

$$\begin{aligned}
Pr\big[p_{ni}(j) \le p\big] &= Pr\big[\frac{d_{ni}}{A_i(j)}(p^E)^{1-\alpha} \le p\big] \\
&= Pr\big[A_i(j) \ge \frac{d_{ni}}{p}(p^E)^{1-\alpha}\big] \\
&= 1 - Pr\big[A_i(j) \le \frac{d_{ni}}{p}(p^E)^{1-\alpha}\big] \\
&= 1 - e^{-T_i(\frac{d_{ni}}{p}(p^E)^{1-\alpha})^{-\theta}}
\end{aligned}$$

Taking the derivative gives:

$$Pr\big[p_{ni}(j) = p\big] = T_i(d_{ni})^{-\theta}(p^E)^{-\theta(1-\alpha)}\theta p^{\theta-1}e^{-T_i(\frac{d_{ni}}{p}(p^E)^{1-\alpha})^{-\theta}}$$

$$\begin{aligned} Pr\big[\min_{w\in W\setminus i} p_{nw}(j) \leq p\big] &= 1 - Pr\big[\min_{w\in W\setminus i} p_{nw}(j) \geq p\big] \\ &= 1 - \Pi_{w\in W\setminus i} Pr\big[p_{nw}(j) \geq p\big] \\ &= 1 - \Pi_{w\in W\setminus i} Pr\big[A_w(j) \leq \frac{d_{nw}}{p}(p^E)^{1-\alpha}\big] \\ &= 1 - \Pi_{w\in W\setminus i} e^{-T_w(\frac{d_{nw}}{p}(p^E)^{1-\alpha})^{-\theta}} \\ &= 1 - e^{-(p)^{\theta}(p^E)^{-\theta(1-\alpha)}\sum_{w\in W\setminus i} T_w(d_{nw})^{-\theta}} \end{aligned}$$

We can take the product of these two terms:

$$\Big(1 - Pr\big[\min_{w\in W\setminus i} p_{nw}(j) \leq p^*\big]\Big)\ Pr\big[p_{ni}(j) = p\big] = T_i(d_{ni})^{-\theta}(p^E)^{-\theta(1-\alpha)}\theta p^{\theta-1}e^{-p^{\theta}(p^E)^{-\theta(1-\alpha)}\Phi_{nW}}$$

We will make a u-substitution where $u = p^{\theta}(p^E)^{-\theta(1-\alpha)}\Phi_{nW}$. This implies:

$$du = \theta p^{\theta-1}(p^E)^{-\theta(1-\alpha)}\Phi_{nW}dp$$

$$p^{1-\sigma_n} = u^{(1-\sigma_n)/\theta}(\Phi_{nW})^{-(1-\sigma_n)/\theta}(p^E)^{1-\epsilon_n^D}$$

Plugging this into our integral:

$$
\begin{aligned}
Y_{ni}^{G} &= \eta_n \int_0^\infty p^{1-\sigma_n} \; \Big(1 - Pr\big[\min_{w\in W\setminus i} p_{nw}(j) \le p\big]\Big) \; Pr\Big[p_{ni}(j) = p\Big] \; dp \\
&= \eta_n \int_0^\infty u^{(1-\sigma_n)/\theta} (\Phi_{nW})^{-(1-\sigma_n)/\theta} (p^E)^{1-\epsilon_n^D} \frac{T_i(d_{ni})^{-\theta}}{\Phi_n W} e^{-u} du \\
&= \eta_n \pi_{ni} (\Phi_{nW})^{-(1-\sigma_n)/\theta} (p^E)^{1-\epsilon_n^D} \int_0^\infty u^{(1-\sigma_n)/\theta} e^{-u} du \\
&= \eta_n \pi_{ni} (\Phi_{nW})^{-(1-\sigma_n)/\theta} \; (p^E)^{1-\epsilon_n^D} \; \Gamma\Big[\frac{1-\sigma_n+\theta}{\theta}\Big] \\
&= \pi_{ni} Y_n^G
\end{aligned}
$$

Note that as a byproduct of this result, we find that the expected price a consumer in n pays for a good is independent of the country i that produced it.

D.4 Coalition to Noncoalition Energy Usage

Because of the export taxes the coalition imposes on its exports, there is no direct link between the energy coalition countries use when producing goods for noncoalition consumers and noncoalition expenditure on those goods. Because of this, we cannot derive a closed form for $C_{n^*i'}^E$ using $Y_{n^*}^G$ like we can for $C_{n^*i^*}^E$. Instead, we must evaluate $C_{n^*i'}^E$ directly. The amount of energy a coalition country i' uses to produce goods for noncoalition country n^* is given by:

$$
\begin{aligned}
C_{n^*i'}^E &= \int_0^1 e_{i'}\big(j, z_{n^*i'}(j)\big) d_{n^*i'} c_{n^*i'}(j) dj \\
&= \int_0^1 \frac{1-\alpha}{(p^E+t^B)^\alpha} \frac{d_{n^*i'}}{A_{i'}(j)} c_{n^*i'}(j) dj \\
&= \int_0^1 \frac{1-\alpha}{(p^E+t^B)^{-\alpha}} \frac{d_{n^*i'}}{A_{i'}(j)} I_{n^*i'}(j) \min_{i^*\in W^*} (p_{n^*i^*}(j))^{-\sigma_{n^*}} dj
\end{aligned}
$$

Amongst coalition producers, the $\mathbb{E}\Big[\frac{d_{n^*i'}}{A_{i'}(j)}\Big]$ is the same (the logic is similar to Appendix C.3). Because of this, we can rewrite this formula as:

$$C^E_{n^*i'} = \frac{\pi_{n^*i'}}{\pi_{n^*W'}} \int_0^\infty \frac{1-\alpha}{(p^E+t^B)^\alpha}\Big(\min_{w'\in W'} \frac{d_{n^*w'}}{A_{w'}(j)}\Big)\Big(\sum_{w'\in W'} I_{n^*w'}(j)\Big) \min_{i^*\in W^*} (p_{n^*i^*}(j))^{-\sigma_{n^*}} dj$$

$\sum_{w'\in W'} I_{n^*w'}(j) = 1$ if and only if $\min_{w'\in W'} \frac{d_{n^*i'}}{A_{i'}(j)} \leq S(p^E, t^B) \min_{i^*\in W^*} \frac{d_{n^*i^*}}{A_{i^*}(j)}$. Consequently, we can rewrite $\sum_{w'\in W'} I_{n^*w'}(j)$ into a second integral. To simplify our notation, we let $\delta_{ni}(j)$ denote $\frac{d_{ni}(j)}{A_i(j)}$:

$$\begin{aligned} C^E_{n^*i'} = \frac{\pi_{n^*i'}}{\pi_{n^*W'}} \frac{1-\alpha}{(p^E+t^B)^\alpha} \int_0^\infty \int_0^{\delta^*/S(p^E,t^B)} & \delta' \ Pr\Big[\min_{w'\in W'} \delta_{nw'}(j) = \delta'\Big] \ d\delta' \quad \times \\ & \eta_{n^*}(\delta^*(p^E)^{1-\alpha})^{-\sigma_{n^*}} \ P\Big[\min_{i^*\in W^*} \delta_{n^*w^*} = \delta^*\Big] \ d\delta^* \end{aligned}$$

We now derive the probability distribution over $\min_{w'\in W'} \delta'$. We have, for a single producer i':

$$\begin{aligned} Pr\big[\delta_{n^*i'} \geq \delta'\big] &= Pr\big[\frac{d_{n^*i'}}{A_{i'}(j)} \geq \delta'\big] \\ &= Pr\big[A_{i'}(j) \leq \frac{d_{n^*i'}}{\delta'}\big] \\ &= e^{-T_{i'}(d_{n^*i'}/\delta')^{-\theta}} \end{aligned}$$

Consequently:

$$\begin{aligned} Pr\big[\min_{i'\in W'} \delta_{n^*i'}(j) \leq \delta'\big] &= 1 - \Pi_{i'\in W} e^{-T_{i'}(d_{n^*i'}/\delta')^{-\theta}} \\ &= 1 - e^{-(\delta')^\theta \Phi_{n^*W'}} \end{aligned}$$

Taking the derivative:

$$Pr\Big[\min_{i' \in W'} \delta_{n^* i'}(j) \leq \delta'\Big] = \Phi_{n^* W'}\ \theta (\delta')^{\theta-1}\ e^{-(\delta')^\theta \Phi_{n^* W'}}$$

The derivation for $Pr\big[\min_{i^* \in W^*} \delta_{n^* i^*}(j) \leq \delta^*\big]$ is identical:

$$Pr\Big[\min_{i^* \in W^*} \delta_{n^* i^*}(j) \leq \delta^*\Big] = \Phi_{n^* W^*}\ \theta (\delta^*)^{\theta-1}\ e^{-(\delta^*)^\theta \Phi_{n^* W^*}}$$

We make two separate substitutions for our integral. We set $u = (\delta')^\theta \Phi_{n^* W'}$ and $v = (\delta^*)^\theta \Phi_{n^* W^*}$. We use the following relationships:

$$du = \Phi_{n^* W'} \theta\ (\delta')^{\theta-1} d\delta'$$

$$dv = \Phi_{n^* W^*} \theta\ (\delta^*)^{\theta-1} d\delta^*$$

$$\delta' = (u/\Phi_{n^* W'})^{1/\theta}$$

$$(\delta^*)^{-\sigma_{n^*}} = (v/\Phi_{n^* W^*})^{-\sigma_{n^*}/\theta}$$

Plugging our probability distributions and substitutions back into our double integral formula:

$$\begin{aligned} C^E_{n^* i'} = & \Bigg\{ \eta_{n^*} \frac{\pi_{n^* i'}}{\pi_{n^* W'}} \frac{1-\alpha}{(p^E + t^B)^\alpha} (p^E)^{-\sigma_{n^*}(1-\alpha)} \quad \times \\ & \int_0^\infty \int_0^{S(p^E, t^B)^{-\theta} \left(\frac{\Phi_{n^* W'}}{\Phi_{n^* W^*}} v\right.} (u/\Phi_{n^* W'})^{1/\theta} e^{-u}\ du\ (v/\Phi_{n^* W^*})^{-\sigma_{n^*}/\theta} e^{-v}\ dv \Bigg\} \\ = & \Bigg\{\ \eta_{n^*} \frac{T_{i'}(d_{n^* i'})^{-\theta}}{\Phi_{n^* W'}} \frac{1-\alpha}{(p^E + t^B)^\alpha} (p^E)^{-\sigma_{n^*}(1-\alpha)} \Big[\frac{(\Phi_{n^* W^*})^{\sigma_{n^*}}}{\Phi_{n^* W'}}\Big]^{1/\theta} \quad \times \\ & \int_0^\infty \gamma\Big(\frac{1+\theta}{\theta}, S(p^E, t^B)^{-\theta} (\frac{\Phi_{n^* W'}}{\Phi_{n^* W^*}}) v\Big) e^{-v} dv \Bigg\} \end{aligned}$$