Curbing fossil fuels: on the design of global reward payment funds to induce countries to reduce supply, reduce demand and expand substitutes

Lennart Stern *

May 20, 2021

The latest version can be found here.

Abstract

Existing international environmental institutions curb fossil fuels by paying countries to reduce demand and expand substitutes. This paper argues that it would be beneficial to create a new and separate institution that would pay countries to reduce their fossil fuel supply. In a model with endogenous funding I compare two architectures. In the first, these institutions would be separate so that donors could flexibly earmark their donations. Under a second architecture, there would be a unified institution with the mandate to split whatever funding it receives between the different approaches in the globally optimal way, treating the budget as if it was exogenous. The separated architecture always results in at least as much global welfare as the unified architecture. This is because it incentivizes fossil fuel exporters (importers) to donate to the institution paying countries to reduce fossil fuel supply (demand) since this raises (lowers) world market prices of fossil fuels. Using estimates of elasticities and the social cost of carbon and imposing several symmetry assumptions I find that emissions abatement is 1.32 times higher under the separated than under the unified architecture for the case of coal and 9.57 times for the case of oil.

1 Introduction

This project studies goods with global externalities. In applying the model I focus throughout the paper on the case of fossil fuels and particularly on coal and oil. However, the results of the static model that I analyze in section 3 are arguably relevant for most goods with global externalities.

*Paris School of Economics and EHESS, e-mail: lennart.stern@psemail.eu

Specifically, I study the problem faced by a global institution having an exogenous budget which it can split between the following three approaches to curbing coal: It can pay countries to reduce coal extraction (supply reduction), to reduce energy use (demand reduction) and to expand renewables (substitute expansion).

The question arises as to how to split the budget between these three approaches. An analogous question arises for any good with global externalities. The following diagram summarizes how the world is currently answering this question for several important goods with negative global externalities (left column in brown) and goods with positive global externalities (left column in green):

good with global externalities	**substitute**	**international institutions focused on supply**	**international institutions focused on demand**	**international institutions focused on substitute**
fossil fuels	renewables, nuclear		Clean Development Mechanism, Green Climate Fund, Global Environment Facility, Climate Investment Fund, UN Environment Program	Clean Development Mechanism, Green Climate Fund, Global Environment Facility, Climate Investment Fund, UN Environment Program, IAEA's Nuclear Fuel Bank
goods produced on previously forested land (soy, palm oil...)	same goods produced in Savannah	REDD+ UN-REDD	certification schemes	
fish	bivalve aquaculture	WTO negotiations on reducing fishery subsidies		
drugs and vaccines for infectious disease control	single product treatments increasing risk of drug resistance		Global Alliance for Vaccines and Immunization Global Fund for AIDS, TB and Malaria	
goods for pandemic surveillance			International Health Regulations	

To analyzes this question, this paper studies a static model with complete information. A global institution announces reward payment schemes for each country conditioning a positive transfer on the country's coal extraction, energy use and renewable energy production. Each country takes these reward payment schemes and world market prices as given. Assuming that all demand and supply elasticities are finite, I prove that the optimal amount of funding allocated to each of the three approaches is always strictly positive and an increasing function of the total available budget (see corollary 1).

For the case of coal, I find based on middle of the road elasticity estimates taken from the literature, that for an exogenous budget it is optimal for the global institution to spend 43% on paying countries to reduce coal supply. This contrasts with the current way that global institutions try to curb fossil fuels. So far, all of the money has been spent on demand reduction and substitute expansion. I find that for a given (not very large) available amount of money, 22% more welfare gains can be achieved if the money is split optimally between the three approaches than if the world deprives itself of the supply side approach. This provides a strong case for establishing a new global fund rewarding countries for reducing fossil fuel supply.

In practice, such a new global fund could define for each country reference levels of stocks of cumulative coal extraction based on business as usual scenario projections and then reward countries each year to the extent that their actual cumulative coal extraction is below the reference level for that year. This kind of scheme is already being used for rewarding countries for preserving tropical forests (Seymour and Busch (2016)).[1]

The question arises as to whether such a new global fund should be set up as a separate institution to which countries could donate directly. I refer to such an arrangement as "the separated architecture". In this case, donors could earmark their contributions to the fund that they prefer.

I compare this architecture to an alternative centralized architecture where a unified institution would be established that would split its budget according to a fixed rule between three funds corresponding to the different approaches (rewarding countries for supply reduction, rewarding countries for demand reduction and rewarding countries for substitute expansion). Specifically, I consider the rule that stipulates that whatever funding the unified institution receives, it splits it between the three funds so as to maximize global welfare (not taking into account that the rule itself might affect the funding it receives in the first place), which turns out to be equivalent to minimizing emissions[2]. For brevity, I refer to the latter architecture as "the unified architecture".

In section 5 I add a financing stage preceding the model from section 3 sketched above. I assume that the countries of the world are exogenously partitioned into players in a voluntary contributions game. All players decide simultaneously on their donations to the separate funds (under the separated

[1]The above diagram shows other goods with global externalities where supply-side approaches are currently absent, for example in the case of drugs for infectious disease control as displayed in the diagram. However, in that case, marginal costs of production are presumably approximately constant in the long run and so the price elasticity of supply is arguably very large in the long run. It turns out that this implies that the optimal amount of spending on rewarding supply reduction is very small. Thus the model can rationalize the fact that the world is focusing on demand side contracting in this case. Some of the other empty boxes in the above diagram appear to not be rationalizable within the model I will present. But I leave it for future work to analyze the specifics of these cases.

[2]This follows from the Constrained Efficiency Lemma 4 together with the fact that under the financing game sketched in the next paragraph funding will never exceed the amount required for achieving globally optimal emissions reductions.

architecture) or the unified institution (under the unified architecture). Players take into account how the funding translates into reductions in global emissions and in changes in the world market price for coal, in accordance with the model from section 3.

Under the unified architecture the world market price of coal is unaffected by the contributions (see the Price Preservation Lemma 3). However, under the separated architecture donations to the fund rewarding supply reduction raise the world market price of coal, whilst donations to the funds rewarding demand reduction and substitute expansion lowers it. As a result, coal exporters have an enhanced incentive to donate to the fund rewarding supply reduction and the coal importers have an enhanced incentive to donate to the funds rewarding demand reduction and substitute expansion.

For the comparison between the separated architecture and the unified architecture, there are two effects at work: One the one hand, the separated architecture leads to at least as much overall funding. On the other hand, under the unified architecture the available funding is by construction spent to maximize emissions reductions, whereas this generally not the case under the separated architecture.

I show that the second effect can never dominate the first: The unified architecture can never result in strictly more emissions reductions than the separated architecture. I show that there are generically at most two donors and that if there are exactly two then one contributes to the fund rewarding supply reduction and the other contributes to the funds rewarding demand reduction and substitute expansion.

In section 5.1.3 I specialize to the case where both the aggregate supply and demand functions for coal are linear. Moreover, I make several symmetry assumptions about the players (which, as indicated above, are subsets of countries that coalesce together into a unified actor in the financing game). Specifically, I assume that all coal exporters only differ in size but not in terms of what fraction of their coal production they export. Similarly, I assume that all coal importers only differ in size.

These assumptions allow me symbolically compute the unique Nash equilibrium in the financing game under each of the architectures, only as a function of the sizes of the largest players, the proportion of coal traded globally and the elasticities at the status quo. Calibrating the model with empirical estimates for the latter, I find that if the largest coal exporter is of the same size as the largest coal importer then the separated architecture leads to about 1.32 times as much emissions reductions as the unified architecture and about 1.76 times as much funding. Applying the model to oil instead yields the conclusion that the separated architecture leads to about 9.7 times as much emissions reductions as the unified architecture and 100 times as much funding.

These results weigh in favor of the separated architecture. Thus the policy conclusion arrived at in this paper is: It would be valuable to create a new global fund rewarding countries for reducing fossil fuel supply. Moreover, it would be best to keep this institution separate from the existing institutions that focus on reducing fossil fuel demand and to allow countries to donate to the institution that they prefer.

2 Related literature and contribution

2.1 Supply side vs demand side approaches to climate change mitigation

Harstad (2012) studies a model where the countries adversely affected by climate change act in a coordinated way. He finds that the coalition's best policy is to simply buy foreign deposits and conserve them.

The model presented in the current paper differs in that in it climate change mitigation happens only due to the countries' responses to the global institution's reward payment schemes. I find that for exogenous funding it is optimal for the global institution to use strictly positive amounts of money on contracts rewarding supply reduction, demand reduction and substitute expansion. The Coasian approach of simply buying up fossil fuel deposits is never optimal in the model.

The current study tries to complement the literature on carbon leakage by drawing out its implications for the design of global institutions. Fæhn et al. (2017) analyze the problem of a country that tries to cause a given reduction in global emissions at a minimal cost for itself. For the case of Norway they find that two thirds of the emissions reductions should optimally come from supply reduction. This result bears some similarity to my result that under exogenous funding a global institution should optimally use 43% of its budget on spending on rewarding supply reduction. Collier and Venables (2015) provide further considerations in favor of focusing on supply side approaches.

I also contribute to the literature on the optimal roles of deposit purchase contracts and leasing contracts as instruments of supply side climate policy. I find that restricting supply side approaches to deposit purchase comes at a welfare cost, which echoes the results from Eichner et al. (2020), despite the difference in the settings.

A major limitation of the current study is that it only models a single fossil fuel (interpreted to be coal) and a clean substitute. Daubanes et al. (2020) highlight the importance of taking into account the substitution between coal and gas. Extending the model presented here to simultaneously include coal, oil and gas is left for future research.

2.2 International Environmental Agreements

An assumption underlying this paper is that countries cannot reach the global optimum through Coasian bargaining. I do not provide an explanation for this. A large literature has provided explanations for the widespread observed failure to achieve efficiency. Dixit and Olson (2000) argue that the appropriate way to model Coasian bargaining is via a two stage game. In the first stage countries decide whether to participate and in the second stage they maximize their joint welfare. Since participation in the first stage is voluntary, there are strong free-riding incentives. The resulting pessimistic conclusions have also been found in the form of the Small Coalition Paradox in the literature on International Environmental Agreements (Barrett 2004).

An alternative explanation for the failure of the Coase Theorem is provided byMartimort and Sand-Zantman (2016) based on asymmetric information. They show that even if an agreement is such that upon withdrawal by one country all others will stop participating, the first best cannot be reached if countries' types are sufficiently heterogeneous.

The current study does not provide an explanation for the absence of Coasian bargaining. Instead, it takes it as given that the world will not achieve a full correction of global externalities.

2.3 The design of compensation funds and of mechanisms to fund them

Kornek & Edenhofer (2020) study different proposed designs for compensation funds for global public good provision. They endogenize the financing of such compensation funds and find that their design greatly affects the amount of funding that will be raised for them. Thus they highlight that the questions of how to best design compensation funds and how to best design mechanisms to fund them are inextricably linked. I arrive at a similar conclusion in section 5.

3 The model

The set of countries is denoted I. Each country is assumed to be of negligible size so that it acts as a price taker on the world market. z_i is the amount of energy from renewables that country i produces. x_i denotes the amount of coal that country i extracts. Coal is measured so that one unit of coal generates one unit energy via combustion. The energy generated from coal is assumed to be a perfect substitute to the energy generated from renewables. We denote by y_i the amount of energy that country i uses. All other energy sources are assumed away.

There is a common numeraire good. Its price is normalized to 1. There are no trade costs and there is a global market for coal. The world market price of coal is denoted p. Each country $i \in I$ takes world market prices as given. Each country i has energy x_i from coal and energy z_i from renewables. If the sum $x_i + z_i$ exceeds its energy use y_i then the country exports the excess amount of energy, $x_i - y_i + z_i$, in the form of coal. If the sum $x_i + z_i$ is less than its energy use y_i then the country imports the shortfall of energy, $x_i - y_i + z_i$, in the form of coal. In either case, the net revenue that the country gets is $p(x_i - y_i + z_i)$. [3]

[3] Currently, 20% of all coal is traded internationally. Consistent with our assumption of a globally integrated coal market, Steckel et al. (2015) find that "in the increasingly integrated global coal market the availability of a domestic coal resource does not have a statistically significant impact on the use of coal and related emissions".

Country i's utility is quasilinear in the numeraire:

$U_i(x_i, y_i, z_i, p) = B_i(y_i) - C_i(x_i) - G_i(z_i) + p(x_i - y_i + z_i) + f_i(x_i, y_i, z_i)$

Here $B_i(y_i)$ is the benefit that country i derives from energy use. [4] $C_i(x_i)$ is country i's cost of extracting x_i of coal. $G_i(z_i)$ is country i's cost of of producing z_i of energy from renewables. Moreover, $f_i(x_i, y_i, z_i)$ denotes the transfer that country i gets from the global institution, as explained further below.

It will be convenient to impose that costs are strictly convex and benefits strictly concave [5]s:

Assumption 1 (strict convexity of costs and strict concavity of benefits). $C_i'(x_i) > 0, C_i''(x_i) > 0 \forall i \forall x_i,\ G_i'(z_i) > 0, G_i''(z_i) > 0 \forall i \forall z_i,\ B_i'(y_i) > 0,\ B_i''(y_i) < 0 \forall i \forall y_i$.

There is a global institution which evaluates global welfare as follows:

$W = \sum_{i \in I} U_i - \eta(\sum_{j \in I} x_j)$

The interpretation is as follows: η is a positive and strictly increasing function. $\eta(\sum_{j \in I} x_j)$ is the aggregate value of the global climate change damages due to the aggregate amount $\sum_{j \in I} x_j$ of coal combusted. U_i is the utility that country i tries to optimize. Strictly, this should include the damage due to climate change that country i suffers due to its own coal use. However, we make the simplifying assumption that country i neglects this, in line with our assumption that all countries are of negligible size.

The global institution is endowed with an exogenous budget F. It offers reward payments to countries to induce them to reduce their coal supply, their coal demand and to expand their renewables supply. The timing is as follows: First, the global institution announces transfers that it will pay to countries conditional on their choices. $f_i(x_i, y_i, z_i)$ denotes the transfer that country i will receive if it chooses (x_i, y_i, z_i). Since countries are sovereign, the global institution cannot ask countries to pay it money, which means that the transfers $f_i(x_i, y_i, z_i)$ are constrained to be non-negative. Each country i takes the reward payment scheme $f_i(x_i, y_i, z_i)$ and the price p as given and chooses (x_i, y_i, z_i) so as to maximize its utility $U_i(x_i, y_i, z_i, p)$.

Definition 1. A **reward payment scheme** offered to country i is a map $f_i(x_i, y_i, z_i)$ assigning a nonnegative transfer to country i. A **world market equilibrium under a given set of reward payment schemes** $(f_i)_{i \in I}$ is a combination of an allocation $(x_i, y_i, z_i)_i$ and world market price p such that:

1) market clearing: $\sum_{i \in I} x_i - y_i + z_i = 0$

2) individual rationality: $(x_i, y_i, z_i) = argmax_{(x,y,z)} - C_i(x) + B_i(y) - G_i(z) + p(x - y + z) + f_i(x, y, z) \forall i \in I$

[4] This should be interpreted to be the entire surplus that the country reaps from energy use, both in the form of consumer surplus from end users and producer surplus from production using energy as an input.

[5] This assumption excludes the case of constant marginal cost, which is of interest for applications like the global health examples mentioned in the introduction. However, instead of treating this case separately, we will discuss this as a limiting case as price elasticities of supply go to infinity.

Definition 2. A set $(f_i)_{i\in I}$ of reward payment schemes implements the allocation-price pair $((x_i, y_i, z_i)_{i\in I}, p)$ with a budget F if $((x_i, y_i, z_i)_{i\in I}, p)$ is a world market equilibrium under $(f_i)_{i\in I}$ and $\sum_{i\in I} f_i(x_i, y_i, z_i) = F$.

Definition 3. An **reward payment scheme** $f_i(x_i, y_i, z_i)$ is called **additively separable** if it can be written as $f_i(x_i, y_i, z_i) = f_{ix}(x_i) + f_{iy}(y_i) + f_{iz}(z_i)$. A reward payment scheme is called a "positive affine linear scheme" if it can be written as $f_i(x_i, y_i, z_i) = max(0, \theta_{ixt}(\tilde{x_{it}} - x_{it})) + max(0, \theta_{ix}(\tilde{y_i} - y_i)) + max(0, \theta_{izt}(z_{it} - \tilde{z_{it}}))$. We will call the parameter θ_{ix} "the rate at which country i is rewarded for reducing coal extraction" and $\tilde{x_i}$ "the reference level relative to which country i is rewarded", and similarly for the other variables.

We will now prove that (under our assumption of convex cost functions and concave benefit functions) nothing is lost by restricting attention to the positive affine linear schemes for the reward payment schemes. We will also show that we can view the global institution as if it was choosing the allocation and the world market prices.

Lemma 1 (**The Surjectivity Lemma**). *Consider a combination of an allocation, $(x_i, y_i, z_i)_{i\in I}$ satisfying $\sum_{i\in I} x_i - y_i + z_i = 0$ and world market price p. There exists a set $(f_i)_{i\in I}$ of positive affine linear schemes implementing $((x_i, y_i, z_i)_{i\in I}, p)$. Moreover, the minimal transfers required to implement$((x_i, y_i, z_i)_{i\in I}, p)$ under affine linear schemes are F_{ix} for rewarding country i for supply reduction, F_{iy} for rewarding country i for demand reduction and F_{iz} for rewarding country i for substitute expansion with*

$$\begin{aligned} F_{ix} &:= \sup_x px - C_i(x) - (px_i - C_i(x_i)) \\ F_{iy} &:= \sup_y B_i(y) - py - (B_i(y) - py) \\ F_{iz} &:= \sup_z pz - G_i(z) - (pz - G_i(y)) \end{aligned}$$

Moreover, there does not exist any set of reward payment schemes implementing $((x_i, y_i, z_i)_{i\in I}, p)$ with a strictly smaller budget, i.e. with a budget strictly less than $\sum_{i\in I} F_{ix} + F_{iy} + F_{iz}$. Furthermore, these minimal required transfers are the same if we allow for any reward payment schemes (instead of restricting them to be positive affine linear).

Proof. See Appendix A.1 □

Let us now summarize the Lemmas using the diagram below. In practice, the global institution chooses reward payment schemes. The notion of market equilibrium defined above yields a mapping assigning a combination of world market price and an allocation to each set of reward payment schemes. By the surjectivity Lemma 1, this map is surjective.

The surjectivity of this market equilibrium map allows us to view the global institution as if it was choosing a combination of world market price and an allocation. Given world market price and an allocation, there are many reward payment schemes inducing them via the market equilibrium map. The transfers that end up being paid are $f_{xi}(x_i), f_{yi}(y_i), f_{zi}(z_i)$. The minimal required transfers F_{ix}, F_{iy}, F_{iz} are given by the formulae shown below:

$$\begin{tikzcd}
(f_{xi}, f_{yi}, f_{zi})_{i\in I} \arrow[r, "\text{market equilibrium (surjective map)}"] & (p, (x_i, y_i, z_i)_{i\in I}) \arrow[d, "\text{minimal required transfers}"] \\
 & (F_{ix}, F_{iy}, F_{iz})
\end{tikzcd}$$

$$F_{ix} = \sup_x px - C_i(x) - (px_i - C_i(x_i))$$
$$F_{iy} = \sup_y B_i(y) - py - (B_i(x_i) - px_i)$$
$$F_{iz} = \sup_z pz - G_i(z) - (pz_i - G_i(z_i))$$

In this diagram we are implicitly restricting attention to additively separable reward payment schemes. This is justified by the Surjectivity Lemma : This restriction does not affect the minimal transfers that are required.[6]

Hence the global institution's problem can be written as:

$max_{(p,(x_i,y_i,z_i)_{i\in I})} \sum_{j\in I} U_j - \eta(\sum_{j\in I} x_j)$ subject to the market clearing constraint, $\sum_{i\in I} x_i - y_i + z_i = 0$, and the budget balance constraint that $\sum_{j\in I} F_{jx} + F_{jy} + F_{jz} \leq F$, where F is the exogenous budget at the global institution's disposal.

Lemma 2 (The First Best). *For a sufficiently large budget F, the first best is characterized by the following conditions plus the feasibility condition $\sum_{i\in I}(x_i - y_i + z_i)$:*

$B_i'(y_i) = B_j'(y_j) = G_i'(z_i) = G_j'(z_j) = C_i'(x_i) + \eta = C_j'(x_j) + \eta \forall i, j \in I$

There exists a continuum of ways to split the budget between supply reduction reward payments schemes on the one hand and demand reduction and substitute expansion reward payment schemes on the other, all of which can achieve the optimal global welfare.

Proof. See appendix A.2. □

Lemma 2 asserts that if the global institution's budget constraint is not binding, then it does not matter how exactly the budget is split between the different approaches, as long as the resulting required budget does not exceed the available budget. However, from now on will will assume that the global institution's budget constraint *is* binding:

Assumption 2. *The global institution's budget constraint is binding. In other words: its budget is insufficient to fully correct the global externalities from coal.*

[6] Interestingly, this no longer holds if we were to depart from the assumption of complete information. In fact, one can deduce from the results in Armstrong and Rochet (1999) that even if types are independent across dimensions the optimal mechanism will not be additively separable.

This assumption will mean that it *will* matter for global welfare how the global institution's budget is split, thereby overturning the conclusion from Lemma 2. This is because the world market price of coal affects the sizes of the transfers required to make countries change their actions instead of just ignoring the reward payments. The world market price of coal, in turn, is affected by how countries are rewarded: The stronger the reward payments for supply reduction, the weaker the supply of coal on the world market and therefore the lower the resulting world market price. On the other hand, the stronger the reward payments for demand reduction and substitute expansion, the lower the demand for coal on the world market and thus the lower the resulting world market price for coal.

The following Lemma states that it is always optimal to choose a mixture of these two kinds of approaches, balanced precisely such that the net effect of the world market price of coal is neutral. It is important to emphasize that this "Price Preservation Lemma" refers to the *world market price* and not to the *net prices* that actors will face. Within a given country, the price that actors will face is the sum of the world market price and any taxes (or regulation-induced carbon prices, etc.) that the government will set. When the global institution rewards countries for supply reduction then this effectively means that it will pay countries for setting a carbon price on the coal extracted on its territory. When the global institution rewards countries for demand reduction then this effectively means that it will pay countries for taxing energy use (by households and firms). When the global institution rewards countries for expanding renewables it pays countries for subsidizing renewables.

The result of all this will always be that the *net price* of coal combustion (including taxes and other implicit or explicit carbon prices) will increase. What will optimally be preserved is the *world market price* of coal:

Lemma 3 (The Price Preservation Lemma). *"If a set of reward payment schemes achieves a given allocation with minimal aggregate transfer payments then it must preserve the world market price p of coal."*

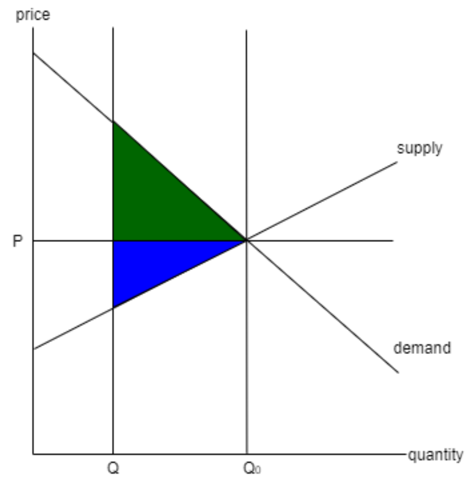
Formally: Consider a fixed allocation $(x_i, y_i, z_i)_{i \in I}$. Consider a set of reward payment schemes $(f_i(x_i, y_i, z_i))_{i \in I}$ implementing $((x_i, y_i, z_i)_{i \in I}, p)$ for some price p under a budget F. Then if there is no other set of reward payment schemes $(\tilde{f}_i(x_i, y_i, z_i))_{i \in I}$ implementing $((x_i, y_i, z_i)_{i \in I}, \tilde{p})$ for some price $\tilde{p}$ under a budget $\tilde{F}$ with $\tilde{F} < F$ then p must equal the price of energy in the absence of any reward payment schemes.

In particular, the optimal reward payment schemes must leave the world market price p of coal at the same level as when there are no reward payment schemes.[7]

[7] It is straightforward to generalize both the Surjectivity Lemma and the Price Preservation Lemma to the case where there are intermediate inputs used only for the good in question. This is relevant in other applications. For example, consider the problem of how to best cause the production of vaccines in normal times to increase so that the world is better prepared for the next pandemic. Consider all the intermediate inputs to vaccines that are only used for them. The Price Preservation Lemma implies that it is optimal for the global institution to use a part of its budget for paying countries to expand production of these intermediate inputs.

Proof. See Appendix A.3 □

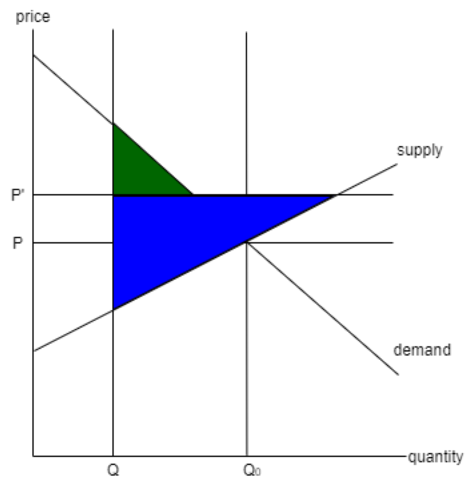
For the case where there is no substitute, we can illustrate the Price Preservation Lemma graphically. Suppose for simplicity that all countries have identical demand and supply functions shown in the following diagram:



In the absence of any reward payment schemes, the world market price for coal is p and the quantity of coal produced and used by each country is Q_0. Now let us compare different ways of reducing the quantity produced (and used) to $Q < Q_0$. Suppose first the global institution achieves this using a reward payment scheme that leaves the price unchanged. In that case, the minimal transfer that it has to pay each country for reducing their coal use from Q_0 to Q is given by the green area. The minimal transfer that it has to pay each country for reducing their coal extraction is given by the blue area.

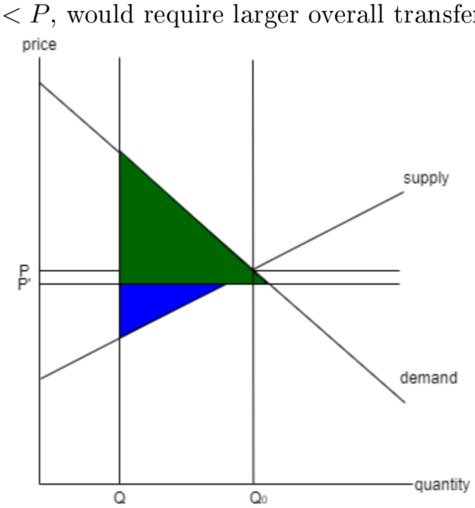
Now suppose the global institution were to implement Q with a higher world market price $P' > P$. This corresponds to a higher spending on rewarding supply reduction:

This is because if it does not then the prices for these intermediate inputs would increase as a result of the increased demand, in contradiction to the Price Preservation Lemma.



We see that the total size of the green and the blue areas together is larger now than when the price was preserved at P.

Similarly, greater demand side emphasis, corresponding to a smaller price $P' < P$, would require larger overall transfers:



Thus we have graphically recovered the Price Preservation Lemma: If a reward payment scheme is to achieve a given allocation with minimal aggregate transfers then the world market price p of coal must be the same as in the absence of any reward payment scheme.

Having established the Price Preservation Lemma by holding the allocation constant and finding the price that minimizes the required transfers, let us now hold the price constant and find the optimal allocation for the given price.

Lemma 4 (The Constrained Efficiency Lemma). *At the optimal reward payment scheme we have: The allocation* (x_i, y_i, z_i) *achieves maximal welfare among all allocations satisfying market clearing and having the same value of*

$\sum_{i \in I} x_i$. *Moreover, this constrained efficiency result even holds if we add the constraint that the world market price* p *of coal be any fixed value.*

Proof. See Appendix A.4 □

Given the Price Preservation Lemma and the Constrained Efficiency Lemma, it is intuitively clear that as the available budget F increases, so will the amounts spent on each of the three approaches at the optimal reward payment scheme. To see why, we note that if we were to only expand the budget for rewarding supply reduction, then the world market price p of coal would increase, in contradiction to the Price Preservation Lemma. Similarly, if we were to only expand the budget for rewarding demand reduction and the budget for rewarding substitute expansion, then the world market price of coal would fall. Moreover, from the Constrained Efficiency Lemma it is intuitively clear that the demand side budget and the substitute side budget must both expand: restricting marginal abatement to demand reduction or substitute expansion would come at an efficiency loss. I will formally validate this conclusion by proving the following:

Corollary 1 (The Interior Solution Corollary). *For the optimal reward payment scheme given the budget* F*, let* $F_x(F)$ *denote the amount optimally used for supply side payments and similarly* $F_y(F)$ *the optimal demand side budget and* $F_z(F)$ *the optimal substitute side budget. We have:* $\frac{dF_x}{dF} > 0, \frac{dF_y}{dF} > 0, \frac{dF_z}{dF} > 0 \forall F$.

Proof. See Appendix A.5 □

Corollary 2 (The Optimal Budget Split Corollary for Small Budgets). *For the optimal reward payment scheme given the budget* F*, let* $F_x(F)$ *denote the amount used for supply side payments and similarly* $F_y(F)$ *the demand side budget and* $F_z(F)$ *the substitute side budget. Moreover, let us denote by* $X(F)$ *the aggregate coal extraction and by* ϵ_x *the aggregate price elasticity of supply of coal,* $Y(F)$ *the aggregate energy consumption and and by* ϵ_y *the aggregate price elasticity of energy demand, by* $Z(F)$ *the aggregate renewable energy production and by* ϵ_z *the aggregate price elasticity of supply of renewable energy. We "generically" have:*

$$lim_{F \to 0} \frac{dF_x}{dF} = \frac{\epsilon_z \frac{Z(0)}{Y(0)} + \epsilon_y}{\epsilon_z \frac{Z(0)}{Y(0)} + \epsilon_x \frac{X(0)}{Y(0)} + \epsilon_y}$$

$$lim_{F \to 0} \frac{dF_y}{dF} = \frac{\epsilon_y}{\epsilon_z + \epsilon_y} \frac{\epsilon_x \frac{X(0)}{Y(0)}}{\epsilon_z \frac{Z(0)}{Y(0)} + \epsilon_x \frac{X(0)}{Y(0)} + \epsilon_y}$$

$$lim_{F \to 0} \frac{dF_z}{dF} = \frac{\epsilon_z}{\epsilon_z + \epsilon_y} \frac{\epsilon_x \frac{X(0)}{Y(0)}}{\epsilon_z \frac{Z(0)}{Y(0)} + \epsilon_x \frac{X(0)}{Y(0)} + \epsilon_y}$$

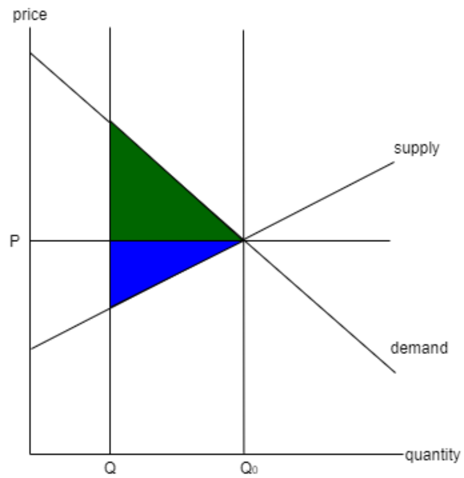
Proof. See Appendix A.5 □

We see that the more elastic the supply of coal, the smaller the proportion of money that will optimally be used to pay countries for reducing coal extraction. We even get the following:

Corollary 3. $lim_{\epsilon_x \to \infty} lim_{F \to 0} \frac{dF_x}{dF} = 0$

Proof. This follows directly from Corollary 2 □

For the case without any substitute we can again illustrate this result diagrammatically:



By the Price Preservation Lemma we know that the optimal reward payment scheme consists of spending the amount corresponding to the blue area on rewarding supply reduction. Making supply more and more elastic corresponds to making the supply curve flatter and flatter. This decreases the blue area, so the optimal proportion of spending on rewarding supply reduction decreases.

We conclude this section with some concavity results:

Lemma 5. *Let $W(F)$ be the maximal welfare achievable with a given budget F. Suppose that $\eta(\sum_i x_i)$ is linear. Then $W(F)$ is strictly concave.*

Proof. See Appendix A.6 □

Lemma 6. *Let $F(F_x, W)$ denote the budget required to achieve welfare W under the further constraint that the budget spent on reducing coal supply is F_x. Then $F(F_x, W)$ is convex in F_x.*

Proof. See Appendix A.7. □

Lemmas 5 and 6 make the following conjecture plausible:

Conjecture 1. *Let $W(F_x, F_y, F_z)$ be the maximal welfare achievable under the further constraint that the amount F_x be spent on coal supply reduction, F_y be spent on energy demand reduction and F_z be spent on renewable supply expansion. $W(F_x, F_y, F_z)$ is concave.*

In the special case where all supply and demand functions have constant elasticities, conjecture 1 does seem to hold, as suggested by the numerical results shown in the next section.

4 Numerical results for the model under constant elasticity specification

All the numerical calibrations whose results are summarized here are fully documented in the accompanying Mathematica notebook that can be downloaded here.

Drawing on the literature, I use the following middle-of-the-road for the parameters [8]:

$\epsilon_D = 0.85$ based on Freehan (2018) and Espey, J. A., & Espey, M. (2004).

$\epsilon_{S_G} = 2.7$ based on Johnson (2011)

$\epsilon_{S_C} = 1.3$ based on Dahl (2009)

$\eta = 0.4$ based on a social cost of carbon of \$36 per ton of CO2 (based on EPA (2015))

[9]

$\frac{X(0)}{Y(0)} = \frac{0.4}{0.26+0.4}$

$\frac{Z(0)}{Y(0)} = \frac{0.26}{0.26+0.4}$ [10]

By Lemma 2, the optimal budget split for an infinitesimally small budget only depends on the elasticities at the situation where no reward payment scheme is in place. Shown in the following figure is the optimal budget split as a function of the available budget:

[8] For the numerical calibrations that follow I use a slightly more complicated (but formally isomorphic model) that is detailed in the accompanying Mathematica notebook. The model takes into account that energy is required as an input to produce renewable energy.

[9] In the accompanying Mathematica notebook I take into account that there are costs for generating coal-powered electricity other than the coal itself. The model is arguably most relevantly applied to the non-Annex 1 countries, given that existing global environmental institutions limit their reward payments to these countries. This is why I take India for calibrating the cost parameters for coal powered electricity:

Electricity prices are around \$0.08 per kwh in India.

Per kwh of electricity from coal 900g of CO2 gets emitted.

Assume a social cost of carbon of \$36 per ton of CO2 (based on EPA (2015)).

Thus we have: $p = 0.08\, per\, kwh$ and $\eta = 0.9 \times 36/1000\, per\, kwh$

Hence $\frac{\eta}{p} = 0.4$. Thus with our normalization of $p = 1$ we get $\eta = 0.4$.

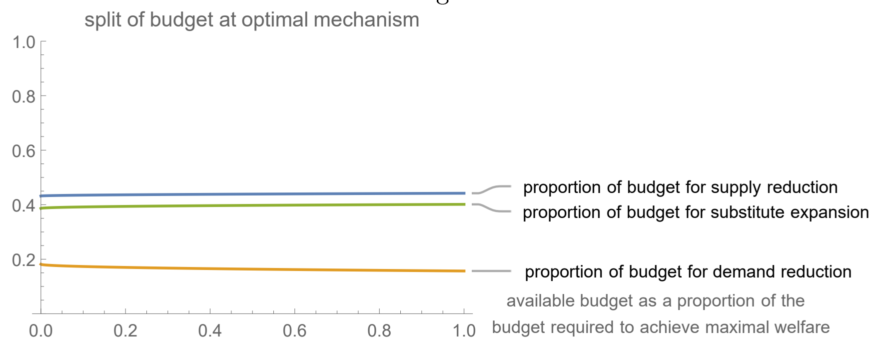
The amount of coal used to generate 1 kWh is 0.00052 short tons by https://www.doi.gov/sites/doi.gov/files/uploads/common_energy_units_conversion_other_commodities_review_final_1-30-17.pdf

price p=\$58.93 per short ton by https://www.eia.gov/energyexplained/coal/prices-and-outlook.php for bituminous coal

This gives a cost on coal inputs of \$0.0306436 per kWh of coal generated electricity. Assuming that the cost of generating coal-powered electricity equals its price in India (\$0.08 per kwh), this means that 38% of the cost is due to the coal itself. I use this this figure in the accompanying Mathematica notebook to commpute the plots that follow.

[10] As of 2018, renewables generates 26 % of global electricity (https : // www.iea.org/fuels - and - technologies/ renewables), whilst coal generates 40 % of global electricity \ (https : // data.worldbank.org/indicator/EG.ELC.COAL.ZS). Since the model does not take into account the other electricity sources, I ignore them for the purposes of this illustrative calibration.

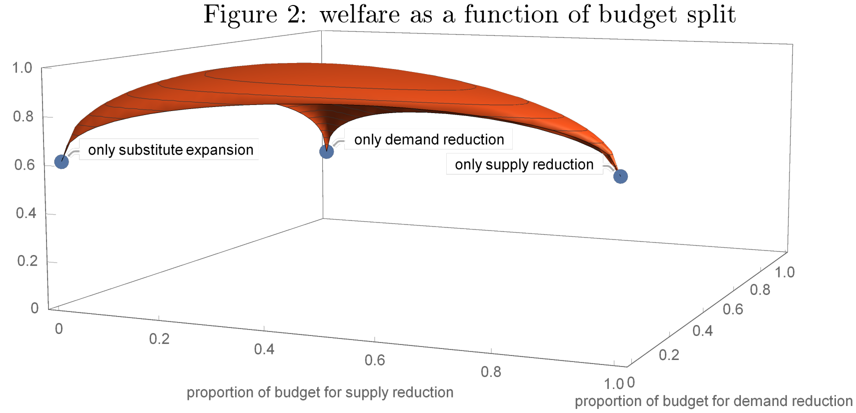
Figure 1:



The plots shows the results starting from an infinitesimal budget all the way to the minimal budget allowing the global institution to fully correct the global externality. Interestingly, the result for the optimal budget split hardly depends on the budget.

In the following figure I show how global welfare depends on how the budget is split between the three approaches:

Figure 2: welfare as a function of budget split



On the x-axis and the y-axis are the proportions of the budget used for rewarding supply reduction and rewarding demand reduction, respectively. By definition, the remaining proportion of the budget is used for rewarding substitute expansion. On the z-axis is the ratio of the global welfare gains achieved divided by the maximal global welfare gains achievable with the given budget. The figure depicts the case where the budget is small. It turns out that the results change very little when one chooses any other budget value between 0 and half the minimal value required to fully correct the global externality.

Around the optimum the surface is quite flat. In fact, as long as each of the three approaches to curbing coal is funded at at least 50% its optimal proportion, welfare losses relative to the optimal budget split are at most 10%. In D I show that this result is quite robust across the ranges of elasticity estimates found in the literature.

This result has important implications for the design of mechanisms to fund the global institution, as I will discuss in section 6. It suggests that it might not be so important to get the budget split exactly right and thus weighs in favor of decentralized funding mechanisms that have no guarantee for allocative efficiency but that create strong participation incentives by giving participating countries the opportunity to influence the allocation of funding across the different approaches to curbing fossil fuels.

5 Endogenous Funding

So far, the entire analysis has assumed that funding for the global institution is exogenous. Now I will endogenize this. From now on I will assume that there are only two global institutions, one institution rewarding countries for reducing coal extraction and one institution for rewarding countries for reducing coal combustion. The model discussed until now can be viewed as obtained by splitting the institution for rewarding countries for reducing coal combustion into two parts: one institution rewarding countries for reducing energy use and one institution for rewarding countries for expanding renewables. In fact, the results that I will present in this section will still hold under this differentiation.

I will compare endogenous funding under two different architectures:

architecture 1: *There is a separate global institution for rewarding countries for coal supply reduction and coal demand reduction. Donors can earmark their contributions to the specific institutions.*

architecture 2: *There is a single global institution that uses its available budget in the optimal way for rewarding countries for coal supply reduction and coal demand reduction. Donors cannot earmark their contributions.*

5.1 The financing game

5.1.1 Payoffs

The countries of the world are exogenously partitioned into a set J of players. There are two groups of players, the net coal exporters and the net coal importers: $J = J_{net\,coal\,exporters} \bigcup J_{net\,coal\,importers}$. The intended interpretation here is: a player is comprised of countries who decide individually on their respective coal extraction and coal combustion taking the reward payment schemes as given (in line with the preceding sections) but who act jointly as unitary players when it comes to deciding on how much money to give to the funds for rewarding supply reduction and demand reduction. The motivation for this interpretation is that it might be easier to cooperate on monetary contributions (which are directly comparable and that can be flexibly adjusted) than on individual domestic actions about fossil fuel extraction and combustion (which are harder to compare and that can only be adjusted with substantial time lags).

Player j chooses the monetary contribution t_{xj} that it makes to the supply reduction reward fund and the monetary contribution t_{yj} that it makes to the demand reduction reward fund. Thus the strategy profile in this financing game is $t := (t_{xj}, t_{yj})_{j\in J}$. The players anticipate how the aggregate funding for the two funds, $t_x := \sum_{j\in J} t_{xj}$ and $t_y := \sum_{j\in J} t_{yj}$ determine the world market price of coal, $p(t_x, t_y)$ and the individual countries' choices, $x_i(t_x, t_y), y_i(t_x, t_y)$. Player j's payoff in this financing game is thus:

$U_j(t) = -t_{xj} - t_{yj} + B_j(y_j(t_x, t_y)) - C_j(x_j(t_x, t_y)) + p(t_x, t_y)(x_j(t_x, t_y) - y_j(t_x, t_y)) + f_j(x_j(t_x, t_y), y_j(t_x, t_y), t_x, t_y) - \eta_j \sum_{k\in J} x_k(t_x, t_y)$

5.1.2 General results about the Nash equilibria

Lemma 7. *Under the unified architecture there exists generically a unique Nash equilibrium. Moreover, at this Nash equilibrium there is generically a single player making strictly positive contributions whilst all others contribute 0.*

Proof. See appendix B.1. □

Lemma 8. *Under the de-centralized architecture there exists generically a unique Nash equilibrium. Moreover, at this Nash equilibrium there are generically one or two players making positive contributions whilst all others contribute 0. Moreover, both the fund for rewarding supply reduction and the fund for rewarding demand reduction always receive positive contributions at the Nash equilibrium. Furthermore, if there are 2 players making strictly positive contributions then one of them is a net coal exporter only making contributions to the fund rewarding countries for supply reduction and the other is a net coal importer only making contributions to the fund rewarding countries for demand reduction.*

Proof. See appendix B.2. □

Lemma 9. *Let $x_{unified}$ denote the coal extraction that results under the unified architecture at the Nash equilibrium (which exists and is unique by Lemma 8). Let $x_{separated}$ denote the coal extraction that results under the separated architecture at the Nash equilibrium (which exists and is unique by Lemma 7). We always have: $x_{separated} \leq x_{unified}$. Moreover, if only one players makes contributions under the separated architecture then we have $x_{separated} = x_{unified}$ and if two players make contributions under the separated architecture then we have $x_{separated} < x_{unified}$.*

Proof. See appendix B.3. □

5.1.3 The Nash equilibria under linear demand and supply functions

In this section, I will assume the following specification:

$U_j(x_i, y_i, p) = \phi_i B(\frac{y_i}{\phi_i}) - \psi_i C(\frac{x_i}{\psi_j}) + p(x_i - y_i) + f_i(x_i, y_i) + \eta_j \sum_{k\in J} x_k$
with $\sum_{j\in J} \phi_j = \sum_{j\in J} \psi_j = 1$ and $\sum_{j\in J} \eta_j = \eta$.

Thus ϕ_j is the proportion of the coal combustion happening in j, ψ_j is the proportion of coal extraction happening in j and $\frac{\eta_j}{\eta}$ is the proportion of climate change damages accruing to j.

The set of players that are net coal exporters is denoted by $J_{net\,coal\,exporters}$. The set of players that are net coal importers is denoted by $J_{net\,coal\,importers}$. For ease of interpretation, I now make the following symmetry assumptions:

$\sum_{j \in J_{net\,coal\,exporters}} \phi_j + \psi_j = \sum_{j \in J_{net\,coal\,importers}} \phi_j + \psi_j$

$\sum_{j \in J_{net\,coal\,exporters}} \eta_j = \sum_{j \in J_{net\,coal\,importers}} \eta_j$

$\frac{\eta_j}{\psi_j} = \frac{\eta_k}{\psi_k}$ for $j, k \in J_{net\,coal\,exporters}$ or $j, k \in J_{net\,coal\,importers}$

$\frac{\phi_j}{\psi_j} = \frac{\phi_k}{\psi_k}$ for $j, k \in J_{net\,coal\,exporters}$ or $j, k \in J_{net\,coal\,importers}$

For $j \in J_{net\,coal\,importers}$, $\sigma := \frac{\phi_j}{\psi_j}$ is the ratio of coal combustion and coal extraction for each of the net coal importers. Let us denote by α the ratio of total global coal exports over total global coal production[11]. It follows from the above assumtions that $\sigma = \frac{1+\alpha}{1-\alpha}$.

Given the above assumptions, we can view $s_j := \frac{\eta_j}{\eta}$ as the "size of player j". By Lemma 8, we know that there will be at most two players who will make strictly positive contributions under the de-centralized architecture. Moreover, if there are two such players, then by Lemma 8, one of them is a net coal exporter only making contributions to the fund rewarding countries for supply reduction and the other is a net coal importer only making contributions to the fund rewarding countries for demand reduction. It follows that these players are the largest players (i.e. the one with the largest s_j) amongst the net coal importers and amongst the net coal exporters, respectively. We will denote these players by im and ex, respectively. Thus s_{im} denotes the size of the largest net coal importer and s_{ex} denote the size of the largest net coal exporter.

From now on I will assume a quadratic specification for the benefit and cost functions:

Assumption 3. *The coal extraction cost function is given by:*

$C(x) = \frac{x(2e_S + x - 2)}{2e_S}$

The coal combustion benefit function is given by:

$B(x) = \frac{x(2e_d - x + 2)}{2e_d}$

Lemma 10. *Aggregate global coal demand is given by $y(p) = 1 + e_D(1-p)$.*

Aggregate global coal supply is given by $x(p) = 1 - e_S(1-p)$.

Thus in the absence of any reward payment schemes we have $p = 1, x = y = 1$ and the price elasticity of supply at that point is e_S, whilst the price elasticity of demand at that point is e_D.

Lemma 11. *The aggregate contribution t_x to the fund paying countries for supply reduction and the aggregate contribution t_y to the fund paying countries for demand reduction result in the following cumulative coal extraction x and world market price for coal p:*

[11] We will later calibrate this for coal and oil. As of 2020, global coal exports make up 20% of global coal production (IAEA 2020). For oil we get from https://yearbook.enerdata.net/ that 50% of oil is exported.

$$x = 1 - \frac{\sqrt{2}e_D\sqrt{e_S}}{e_D+e_S}\sqrt{t_x} - \frac{\sqrt{2}e_S\sqrt{e_D}}{e_D+e_S}\sqrt{t_y}$$
$$p = 1 + \frac{\sqrt{2}\sqrt{e_S}}{e_D+e_S}\sqrt{t_x} - \frac{\sqrt{2}\sqrt{e_D}}{e_D+e_S}\sqrt{t_y}$$

Lemma 12. *Suppose $s_{im} \geq s_{ex}$. Switching from the unified architecture to the separated architecture multiplies the emissions reductions (resulting at the unique NEs of the financing games) by the following "amplification factor":*

$$a = \frac{\eta\big(s_{ex}s_{im}\alpha(e_d-e_s)^2+(e_d+e_s)(-e_d s_{ex}(1-s_{im})-e_s(1-s_{ex})s_{im})\big)+2\alpha(e_D+e_S)(s_{ex}(2s_{im}-1)-s_{im})}{\eta s_{im}(\alpha(e_d-e_s)(e_D s_{im}-e_s s_{ex})+(e_d+e_s)(e_d(s_{im}-1)+e_s(s_{ex}-1)))}$$

In particular, if the largest coal exporter is of equal size as the largest coal importer then this amplication factor is:

$$a = 1 + \frac{4(1-s)\alpha(e_d+e_s)}{\eta(s\alpha(e_d-e_s)^2-(1-s)(e_d+e_s)^2)}$$
$$a = 1 + \frac{4\frac{\alpha}{(e_d+e_s)}}{\eta}\frac{1}{1-\frac{s}{1-s}\alpha(\frac{e_d-e_s}{e_d+e_s})^2}$$

This expression can be explained as follows: The incentives to contribute funding can be decomposed into 2 effects: Firstly, donors cause reductions in the global externalities by contributing money to the reward payment funds. Let us call this the "welfare effect". Secondly, donors can lower the world market price of coal by giving money to the fund rewarding countries for demand reduction and raise the world market pricce of coal by giving money to the fund rewarding countries for supply reduction. Let us call this the "world market price effect".

The amplification factor a is determined by the relative strength of these two effects. The strength of the welfare effects is proportional to the strength of the externality, η. The world market price effect is increasing the fraction α of coal that is traded internationally. Moreover, it is decreasing in the sum of the elasticities, e_d+e_s, since the more elastic demand and supply then less the world market price will move as a result of the reward payment schemes. In fact, if $e_d = e_s$ then the world market price effect is exactly $\frac{4\alpha}{(e_d+e_s)}$. If the demand and supply elasticities are different, then there is another factor, $\frac{1}{1-\frac{s}{1-s}\alpha(\frac{e_d-e_s}{e_d+e_s})^2}$, that further increases the amplification factor a. This is because if $e_d > e_s$ then the fund rewarding demand reduction causes a relatively strong downward pressure on the world market price p, making it particularly attractive for the coal importer to fund it.

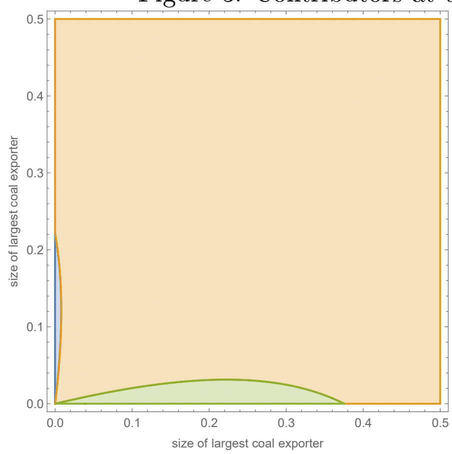
Example 1. Coal.

$\alpha = 0.2$, based on IAEA 2020 for the ratio of global coal exports to coal use
$e_d = 0.7$, based on Keen et al. (2019) for the price elasticity of coal demand
$e_s = 1.3$, based on Dahl (2009) for the price elasticity of coal supply
$\eta = 1.27$ based on a social cost of carbon of \$36 per ton of CO2 (based on EPA (2015)).[12]

[12] Recall that by Lemma 10 our demand and supply specifications have the following implicit normalisation: Price at the status quo is 1. Thus η is the ratio of the social cost of a unit of coal divided by its price. To compute this we use the following:
25 million Btu per short ton by: https://openei.org/wiki/Definition:Bituminous_coal#:~:text=Bulk%20density%20typically%20
79kgCO2 per GJ by https://www.engineeringtoolbox.com/co2-emission-fuels-d_1085.html
79/0.94791 kgCO2 per million Btu by http://convert-to.com/conversion/energy/convert-

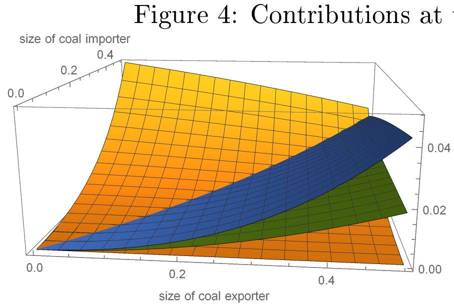
Lemma 14 tells us whether or not both players contribute at the Nash equilibrium. The following plot shows that this is the case for almost all combinations of the players' sizes:

Figure 3: Contributors at the Nash Equilibrium for coal



Here is the funding for the two funds as a function of the sizes of the two players:

Figure 4: Contributions at the Nash Equilibrium for coal



Now consider the special case where the largest coal exporter is of equal size, denoted s, as the largest coal importer. By Lemma 12 we have: The amplification factor a (i.e the factor by which emissions reductions are multiplied in the separate architecture relative to the centralised architecture) is:

$$a = 1 + \frac{4\frac{\alpha}{(e_d+e_s)}}{\eta} \frac{1}{1-\frac{s}{1-s}\alpha(\frac{e_d-e_s}{e_d+e_s})^2} \in [1.314, 1.319] \forall s \in [0, 0.5]$$

gj-to-btu.html

Hence 79/0.94791*25kgCO2 per short ton

price p=$58.93 per short ton by https://www.eia.gov/energyexplained/coal/prices-and-outlook.php

A social cost of carbon of 36$ per ton of CO2 leads to $\eta/p = 36 * 231.7/1000 * 25 = 1.27$

This corresponds to funding being between 1.73 and 1.78 times as high.

Example 2. Oil.

$\alpha = 0.425$, based on Enerdata (2018) for the ratio of global oil exports to oil use

$e_d = 0.5$, based on Keen et al. (2019) for the price elasticity of oil demand

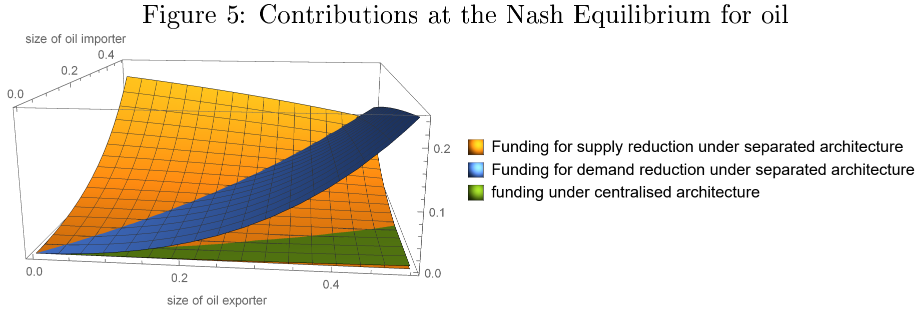
$e_s = 0.32$, based on Golombek et al. (2018) for the price elasticity of oil supply

$\eta = 0.24$ based on a social cost of carbon of \$36 per ton of CO2 (based on EPA (2015)).[13].

Evaluating the condition given in Lemma 14 reveals that for all combinations of sizes of the two plaers we have: at the unique Nash equilibrium both players contribute.

Here is the funding for the two funds as a function of the sizes of the two players:

Figure 5: Contributions at the Nash Equilibrium for oil



Now consider the special case where the largest coal exporter is of equal size, denoted s, as the largest coal importer. By Lemma 12 we have: The amplification factor a (i.e the factor by which emissions reductions are multiplied in the separate architecture relative to the centralised architecture) is:

$a = 1 + \frac{4\frac{\alpha}{(e_d+e_s)}}{\eta}\frac{1}{1-\frac{s}{1-s}\alpha(\frac{e_d-e_s}{e_d+e_s})^2} \in [9.58, 9.76] \forall s \in [0, 0.5]$

This corresponds to funding being between 95 and 110 times as high.

[13]Recall that by Lemma 10 our demand and supply specifications have the following implicit normalisation: Price at the status quo is 1. Thus η is the ratio of the social cost of a unit of coal divided by its price. To compute this we use the following: The EPA states: The average carbon dioxide coefficient of distillate fuel oil is 429.61 kg CO2 per 42-gallon barrel (EPA 2018). The fraction oxidized to CO2 is 100 percent (IPCC 2006). The eia states: The price of Brent crude oil, the international benchmark, averaged \$64 per barrel (b) in 2019. Assuming a social cost of carbon of \$36 per ton of CO2, we get: $\eta/p = 36 * 0.42961/64$=0.24.

6 Implications for the design of global public good institutions and mechanisms to fund them

6.1 Comparing the separated architecture with the unified architecture if the global institutions are financed via voluntary contributions

The primary way that global environmental institutions (and Global Public Good Institutions generally) are currently being financed is through assessed contributions and voluntary contributions by governments. Many global institutions have a wide range of different programs. For example, the Green Climate Fund finances both projects that can be viewed as rewarding countries for fossil fuel demand reduction and projects that can be viewed as rewarding countries for fossil fuel substitute expansion. An important design question is whether donors should be allowed to earmark their contributions to specific programs. This design question will be particularly important if a new fund for rewarding countries for reducing fossil fuel supply is established.

In the case of institutions for Global Health, countries have been enabled to earmark their contributions[14]. However, in the case of the Green Climate Fund donors can earmark only 20% of contributions.

A consideration weighing in favor of allowing countries to earmark their contributions is that this could increase the amounts that they will be willing to donate.[15] Statistical evidence suggests that this is true for international organizations generally (Bayram and Graham (2017)). One explanation for this could be that donors have different beliefs about the effectiveness of different programs. A further reason, highlighted in section 5, is that different approaches to curbing a good with global externalities will differently affect the world market price of this good. As a result, donors have self-interested reasons to prefer supply reduction approaches or demand reduction approaches, depending on whether they are net exporters or net importers of the good.

The results from section 4suggest that it would be valuable to create a new global institution rewarding countries for reducing fossil fuel supply. Moreover, the results from section 5 suggest that it would be best for such a supply-side fund to be separate and independent so that countries can donate money specifically for their preferred approach. I have referred to this way of separating out the different institutions as “the separated architecture”.

An alternative arrangement, which I have called “the unified architecture” would be to create a unified global institution that would commit to splitting its budget between the different approaches to curbing fossil fuels in the way that is ex ante optimal. For example, the Green Climate Fund could expand its scope to also reward countries for reducing fossil fuel supply and it could

[14]see Clinton and Sridhar (2019) for a critical take on this.

[15]For example, Raman (2014) reports on the negotiations about rules for earmarking for the Green Climate Fund: “Several interested contributors from developed countries led by the UK, US and Norway expressed the view that some targeting of funds should be allowed as this would enable more resources to come into the Fund.”.

adopt a rule specifying how to split the budget, depending only on the aggregate funding flows that it receives. The Green Climate Fund already splits each year the funding flows that it receives between mitigation and adaptation projects in equal parts. It could adopt a further such rule for splitting the mitigation spending between supply reduction, demand reduction and substitute expansion.

The static model with endogenous funding from section 5 can be applied each year. Thus by Lemma 9, we know that each year more funding will be raised under the separated architecture than under the unified architecture. Moreover this difference is typically a large factor, as we saw in calibrations 1 and 1.

In the static model the greatly larger funding under the separated architecture directly translates also into much larger emissions reductions. However, in a dynamic model along the lines just sketched this relationship is more complicated. This is because of the intertemporal nature of the fossil fuel market: Future spending on rewarding supply reduction will (if correctly anticipated) already reduce supply today. However, once the future arrives, forward looking donors will no longer take this into account. This gives rise to a time consistency problem.

This time consistency problem weighs potentially in favor the the unified architecture. This is because under this architecture the unified institution could commit to a path for how to split up the funding that it receives. Under the separated architecture, donors would by definition be able to decide at each point in time to which of the funds (for supply reduction, demand reduction and substitute expansion) to donate to. Thus the very attraction of the separated architecture (namely the ability for donors to always earmark all their donations) precludes the possibility for commitment.

In a companion paper (Stern 2021), I try to assess the extent of the time consistency problem in a dynamic model with three periods. For tractability, I suppose that the unified global institution receives an exogenous flow of funding that it has to use up in each period. I find that if the unified global institution cannot commit at all then there always arises a version of the Weak Green Paradox in the following sense: Increasing funding in the second period will increase coal extraction in the first period, thus accelerating initially global warming. However, I provide results suggesting that the Strong Green Paradox cannot arise under any plausible assumptions on the supply and demand functions: Additional funding in any period will always increase global welfare. These results suggest that taking into account the dynamic aspects and the associated time consistency problem is unlikely to overturn the conclusion of the current paper about the superiority of the separated architecture over the unified architecture.

6.2 Comparing the separated architecture with the unified architecture if the global institutions are financed via other mechanisms

The discussion in the section 6.1 has compared the separated and unified architectures (see the beginning of section 5for definitions), assuming that the different reward payment funds are all only financed through voluntary budgetary contributions. The model in section 5 aims to capture this form of financing.

However, it might be possible in the future to establish new mechanisms to fund these reward payment funds and other global public good institutions (GPGIs) more broadly. Such mechanisms could be purposefully designed in such a way that participants could through their contributions affect the overall allocation of funding in specific ways so that participants would have strong incentives to participate and contribute. An example of a proposal for such a mechanism is the so-called "MGF mechanism with Proportional Matching Funds" proposed in Stern (2020), where I find the following: Under the separated architecture the mechanism can raise twice as much aggregate funding as under the unified architecture[16]. This advantage of the separated architecture is likely to outweigh any loss in allocative efficiency, given the results from section 4: there we found that as long as each of the three reward payment funds gets at least half its optimal share of funding, the global welfare losses relative to the optimum split of the available money is at most 10%.

7 Conclusion and limitations

This paper has analyzed the problem faced by global institutions such as the Green Climate Fund, the Global Environment Facility and the Climate Investment Fund. With the part of the budgets allocated to climate change mitigation, these institutions can be viewed as being able to reward countries for reducing fossil fuel supply, reducing energy demand and expanding renewable energy. So far, these institutions only pursue the demand reduction and the substitute expansion approaches. The results in this paper suggest that it would be optimal for them to pursue a mixture of supply, demand and substitute based approaches to curbing fossil fuels.

Moreover, the results from the financing game presented here suggest that it would be best to enable countries to earmark their funding to a particular approach. This could be done by creating a separate institution for each approach. Net exporters of fossil fuels would then have a stronger incentive to give to the fund rewarding fossil fuel supply reduction, since this would increase the world market prices of fossil fuels. Net importers of fossil fuels would have a stronger incentive to give to the funds rewarding countries for fossil fuel demand reduction and substitute expansion, since this would decrease the world

[16]To be more precise, in Stern (2020) I consider a more general version of the architecture 1 and 2, applied to all GPGIs, not just the ones trying to curb fossil fuels that are the focus of the current study.

market prices of fossil fuels. An alternative centralized design with a unified institution, where all funding would be split between the different approaches so as to maximize global welfare would probably not raise as much funding. This would likely outweigh any efficiency gains it might achieve relative to a flexible decentralized architecture where countries can earmark their contributions to their preferred approaches.

An important limitation is that the model does not take into account the adverse effects of fossil fuel rents on global welfare via the natural resource curse (Ross (2015)). This consideration weighs in favor of focusing on demand reduction and substitute expansion instead of supply reduction. Possibly, it could rationalize the current absence of international institutions paying countries for reducing fossil fuel supply.

A further limitation is that the model abstracts away from informational asymmetries. Since in reality countries have private information about their costs and benefits of extracting coal, using energy and producing renewable energy, most of them will reap informational rents at the optimal mechanism. In Stern (2021) I develop a model taking this into account. Specifically, I study the optimal mechanism for reducing fossil fuel demand if the global institution can only condition its reward payments on each country's tax/subsidy rate on the fossil fuels. The model could be applied to the supply side and the substitute side and then integrated into the model presented in the current paper.

A Proofs for section 3

A.1 Proof of Surjectivity Lemma 1

Suppose first, hypothetically, the global institution could make countries pay transfers and suppose it were to impose the reward payment scheme $\theta_{ixt}(\tilde{x_{it}} - x_{it})$. Then since $x \mapsto px - C_i(x) + \theta_{ix}(\tilde{x_i} - x)$ is concave, its global optimum is x_i iff $\theta_{ix} = p - C_i'(x_i)$. Now suppose the global institution offers instead the reward payment scheme $f_{ix}(x_i) = max(0, \theta_{ixt}(\tilde{x_{it}} - x_{it}))$. x_{it} is still an optimal choice for i iff $px_i - C_i(x_i) + \theta_{ixt}(\tilde{x_{it}} - x_{it}) \geq sup_x px - C_i(x)$, or equivalently:

$\theta_{ixt}(\tilde{x_{it}} - x_{it}) \geq F_{ix} = sup_x px - C_i(x) - (px_i - C_i(x_i))$

Thus we have shown that the minimal transfer required to pay the country i to induce it to choose x_i under some affine linear reward payment scheme is indeed $F_{ix} = sup_x px - C_i(x) - (px_i - C_i(x_i))$. Analogously, one can prove the claims for the other variables.

What is left to be proved is that there does not exist another set $(f_i)_{i\in I}$ of reward payment schemes implementing $((x_i, y_i, z_i)_{i\in I}, p)$ with a strictly smaller budget. To establish this, we note that the individual rationality condition implies:

$B_i(y_i) - C_i(x_i) - G_i(z_i) + p(x_i - y_i + z_i) + f_i(x_i, y_i, z_i) \geq sup_{(x,y,z)} B_i(y) - C_i(x) - G_i(z) + p(x - y + z) = sup_y B_i(y) - py - sup_x(C_i(x) - px) - sup_z(G_i(z) - gz)$, so we have:

$f_i(x_i, y_i, z_i) \geq \sup_x px - C_i(x) - (px_i - C_i(x_i)) + \sup_y B_i(y) - py - (B_i(y_i) - py_i) + \sup_z pz - G_i(z) - (pz_i - G_i(z_i)) = F_{ix} + F_{iy} + F_{iz}$.

A.2 Proof of Lemma 2

Proof. If the budget is sufficiently large, the global institution's problem's Lagrangian becomes equal to:

$L = \sum_{j \in I} U_j - \eta(\sum_{j \in I} x_j) + \mu \sum_{i \in I}(x_i - y_i + z_i)$

where μ is the Lagrange multiplier associated with the feasibility constraint. The first order conditions are:

$\frac{\partial L}{\partial x_i} = \frac{\partial U_i}{\partial x_i} - \eta + \mu = -C_i'(x_i) - \eta + \mu = 0$

$\frac{\partial L}{\partial y_i} = \frac{\partial U_i}{\partial y_i} - \eta + \mu = B_i'(y_i) - \mu = 0$

$\frac{\partial L}{\partial z_i} = \frac{\partial U_i}{\partial z_i} - \eta + \mu = -G_i'(z_i) + \mu$

Thus the first best is characterized by the following conditions plus the feasibility condition $\sum_{i \in I}(x_i - y_i + z_i)$:

$B_i'(y_i) = B_j'(y_j) = G_i'(z_i) = G_j'(z_j) = C_i'(x_i) + \eta = C_j'(x_j) + \eta \forall i, j \in I$

In particular, the world market price p of energy does not appear in this characterization. The greater the emphasis on rewarding countries for supply reduction, the larger the resulting price p will be. But as long as the required overall budget does not exceed the available budget, this does not matter for global welfare. □

A.3 Proof of Price Preservation Lemma 3

Proof. Based on the Surjectivity Lemma (1), we view the global institution as choosing the price p and the allocation $(x_i, y_i, z_i)_{i \in I}$. Global welfare is determined by the allocation, $(x_i, y_i, z_i)_{i \in I}$, and the price p only is relevant because it affects the aggregate transfers required to get all countries to participate. The minimal required transfers $(F_{ix}, F_{iy}, F_{iz})_{i \in I}$ are given by the Surjectivity Lemma. In particular, the minimal aggregate required transfer is given by $\sum_{i \in I} F_{ix} + F_{iy} + F_{iz}$.

We now show that $\sum_{i \in I} F_{ix} + F_{iy} + F_{iz}$ is convex when viewed as a function of p. To do so, we use that $F_{ix} := \sup_x px - C_i(x) - (px_i - C_i(x_i))$, $F_{iy} := \sup_y B_i(y) - py - (B_i(y) - py)$, $F_{iz} := \sup_z pz - G_i(z) - (py - G_i(y))$ and we compute:

$\frac{d}{dp}(\sum_{i \in I} F_{ix} + F_{iy} + F_{iz}) = \sum_{i \in I} x_i^*(p) - x_i - (y_i^*(p) - y_i) + z_i^*(p) - z_i$

where $x_i^*(p)$ denotes country i's supply function in the absence of any mechanism, i.e. $px - C_i(x) = argmax_x px - C_i(x)$ and analogously for $y_i^*(p)$ and $z_i^*(p)$. By market clearing, we have:

$\frac{d}{dp}(\sum_{i \in I} F_{ix} + F_{iy} + F_{iz}) = \sum_{i \in I} x_i^*(p) - y_i^*(p) + z_i^*(p)$

But this is just the excess supply function, which is strictly increasing in p as can be deduced directly from assumption 1 . Thus the optimal p is characterized by the condition

$\sum_{i \in I} x_i^*(p) - y_i^*(p) + z_i^*(p) = 0$

The price obtaining at the market equilibrium in the absence of any mechanism satisfies this condition by market clearing. It is the unique price satisfying this condition. □

A.4 Proof of Constrained Efficiency Lemma 4

Proof. The basic reason for this result is as follows: The global institution has two considerations to take into account: it cares intrinsically about the countries' aggregate utility and it wants to minimize the required transfers. But the outside options are determined by the price and so the required transfers decrease in the countries' aggregate utility. Thus the two considerations perfectly align. I will now flesh out this argument in full formal detail.

Consider a fixed price p. Consider the set S of all allocations satisfying the market clearing condition and having a given value X for $\sum x_i$.

$S := \{(x_i, y_i, z_i) : \sum_{i\in I}(x_i - y_i + z_i) = 0, \sum_{i\in I} x_i = X\}$

Let μ denote the Lagrange multiplier associated with the market clearing constraint. Let β denote the Lagrange multiplier associated with the global institution's budget constraint. The global institution's Lagrangian is:

$L = \sum_{i\in I} U_i - \eta(\sum_{j\in I} x_j) + \mu \sum_{i\in I}(x_i - y_i + z_i) - \beta(\sum_{i\in I} F_{ix} + F_{iy} + F_{iz} - F)$

When choosing amongst allocations in this set S, there are only two terms in the Lagrangian that are affected, namely $\sum_{i\in I} U_i$ and $-\beta(\sum_{i\in I} F_{ix} + F_{iy} + F_{iz})$:

$L = \sum_{i\in I} U_i - \eta(\sum_{j\in I} x_j) + \mu \sum_{i\in I}(x_i - y_i + z_i) - \beta(\sum_{i\in I} F_{ix} + F_{iy} + F_{iz} - F)$

So we can write:

$L = \sum_{i\in I} U_i - \beta(\sum_{i\in I} F_{ix} + F_{iy} + F_{iz}) + \phi$

where ϕ does not depend on the allocation (as long as the allocation is chosen from the set S).

Using the expressions for the minimal transfers, we obtain:

$L = \sum_{i\in I} U_i - \beta(\sum_{i\in I} F_{ix} + F_{iy} + F_{iz}) + \phi$

We have:

$\sum_{i\in I} F_{ix} + F_{iy} + F_{iz} = \sup_x px - C_i(x) - (px_i - C_i(x_i)) + \sup_y B_i(y) - py - (B_i(y) - py) + \sup_z pz - G_i(z) - (py - G_i(y))$

$\sum_{i\in I} F_{ix} + F_{iy} + F_{iz} = \sum_{i\in I} \sup_{x,y,z} p(x - y + z) - C_i(x) + B_i(y) - G_i(z) + C_i(x_i) - B_i(y_i) + G_i(z_i))$

$\sum_{i\in I} F_{ix} + F_{iy} + F_{iz} = F + \sum_{i\in I} \sup_{x,y,z} p(x - y + z) - C_i(x) + B_i(y) - G_i(z) - U_i$

Now we can write the Lagrangian as follows:

$L = (1 + \beta) \sum_{i\in I} U_i + \phi^{\#}$

where $\phi^{\#}$ does not depend on the allocation (as long as the allocation is chosen from the set S). Thus for any given p and a given value X for $\sum x_i$, the global institution's optimization problem for the allocation $(x_i, y_i, z_i)_{i\in I}$ boils down to simply maximizing $\sum_{i\in I} U_i$ under the constraint that $\sum x_i = X$. □

A.5 Proof of the Interior Solution Corollary 1 and the Optimal Budget Split Corollary for Small Budgets 2

Proof. (of the Interior Solution Corollary and the Optimal Budget Split Corollary for Small Budgets) By the Price Preservation Lemma, the price at the optimal mechanism is the same as at the market equilibrium in the absence of any mechanism. We denote this price simply by p.

As before, let μ denote the Lagrange multiplier associated with the market clearing constraint. Let β denote the Lagrange multiplier associated with the global institution's budget constraint. The global institution's Lagrangian is:

$L = \sum_{i \in I} U_i - \eta(\sum_{j \in I} x_j) + \mu \sum_{i \in I}(x_i - y_i + z_i) - \beta(\sum_{i \in I} F_{ix} + F_{iy} + F_{iz} - F)$

By differentiating the Lagrangian with respect to the allocation, we obtain the following optimality conditions:

$$h_{xi}(x, y, z, \mu) := (1+\beta)(p - C_i'(x_i)) + \mu - \eta' = 0$$

$$h_{yi}(x, y, z, \mu) := (1+\beta)B_i'(y_i) - p - \mu = 0$$

$$h_{zi}(x, y, z, \mu) := (1+\beta)(p - G_i'(z_i)) - p + \mu = 0$$

We also have the market clearing condition:

$$f_\mu(x, y, z, \mu) := \sum_i x_i - y_i + z_i = 0 \tag{1}$$

Define $\sigma := \frac{1}{1+\beta}$, so $\sigma = 0$ corresponds to the case where the budget $F = 0$ and $\sigma = 1$ corresponds to the case where the budget F is the minimal amount sufficient to implement the global optimum. With this, we can rewrite the optimality conditions as follows:

$$h_{xi}(x, y, z, \mu) := p - C_i'(x_i) + \sigma(\mu - \eta') = 0 \tag{2}$$

$$h_{yi}(x, y, z, \mu) := B_i'(y_i) - p - \sigma\mu = 0 \tag{3}$$

$$h_{zi}(x, y, z, \mu) := (1+\beta)(p - G_i'(z_i)) - p + \sigma\mu = 0 \tag{4}$$

Now we can study what happens as we relax the budget constraint, which corresponds to increasing σ. Let us denote by $h := ((h_{xi})_{i \in I}, (h_{yi})_{i \in I}, (h_{zi})_{i \in I}, h_\mu)$, where h is a vector function whose components are defined in equations 2, 3, 4 and 1. The σ determines (x, y, z, μ) via the condition $h(x, y, z, \mu) = 0$.

For $\sigma \in (0, \infty)$ we know that this system has a unique solution. To see this, consider a fixed $\sigma > 0$. We can think of a choice of μ as determining the $(x_i, y_i, z_i)_{i \in I}$. Define $g(\mu) := h_\mu(x(\mu), y(\mu), z(\mu), \mu)$, where $x(\mu)$ denotes the vector of the x_i determined via equation 2 etc.. Each of the x_i and z_i are strictly increasing in μ, whilst all of the y_i are strictly decreasing in μ. Hence g is increasing. Moreover, $g(0) < 0$. To see why, we note that for $\mu = 0$ the y_i and the z_i are as in the absence of any mechanism whilst the x_i are strictly smaller. We also have that $g(\eta) > 0$. To see why, we note that for $\mu = \eta$ the x_i are as in the absence of any mechanism, whilst the y_i are strictly smaller and the z_i are strictly larger.

Given that thus $g(0) < 0$, $g(\eta) > 0$ and that $g(\mu)$ is increasing, we can apply the intermediate value theorem as long as g is continuous. But all the $x_i(\mu), y_i(\mu), z_i(\mu)$ are continuous which implies that g is continuous. Hence, by the intermediate value theorem, there exists a unique μ such that equation 1 holds and this μ is in $(0, \eta)$. This shows that there is a unique solution to equations 2, 2, 4 and 1. We now denote the unique solution by $(x, y, z, \mu)(\sigma)$.

Now we will show that $(x, y, z, \mu)(\sigma)$ is continuously differentiable at all $\sigma \in (0, 1)$. For this it suffices by the implicit function theorem to show that the Jacobian of h is nonsingular on $(0, 1)$, since f is continuously differentiable. The Jacobian J is as follows:

$$J = (\begin{array}{cccc} -C''(x) & 0 & 0 & \sigma \\ 0 & B''(y) & 0 & -\sigma \\ 0 & 0 & -G''(z) & \sigma \\ 1 & -1 & 1 & 0 \end{array})$$

where $C''(x)$ is a diagonal matrix with entries $C_i''(x_i)$ etc. and in a slight abuse of notation the 6 zeros in the upper left denote $|I|$ by $|I|$ matrices with all entries 0 and the σ denote column vectors of length $|I|$. For $\sigma > 0$, J is non-singular, since by assumption $C_i'' > 0, B_i'' < 0, G_i'' > 0$.

To see why, we note that if we want to write the bottom row vector as a linear combination of the other rows, then the weight given to each of the first $|I|$ rows (corresponding to the $(x_i)_{i \in I}$) must be strictly positive, whilst the weight given to the rows from row $|I| + 1$ to row $2|I|$ must be negative, whilst the weight given to the rows from row $2|I| + 1$ to row $3|I|$ must be negative. But all this together implies that the last component of this linear combination will be strictly positive, in contradiction to the fact that the last component of the last row is 0.

Hence we have established by the implicit function theorem that $(x, y, z, \mu)(\sigma)$ is continuously differentiable at all $\sigma \in (0, \infty)$. We also note that $lim_{\sigma \to 0^+} J$ is singular. This explains why we will now need to do some more work to establish a fact that we will later need, namely that "generically" $lim_{\sigma \to 0^+} \sigma \frac{d\mu}{d\sigma} = 0$.

Differentiation of the first order conditions (2,3,4) with respect to σ yields:

$$C_i''(x_i) \frac{dx_i}{d\sigma} = \mu - \eta + \sigma \frac{d\mu}{d\sigma} \tag{5}$$

$$B_i''(y_i) \frac{dy_i}{d\sigma} = \mu + \sigma \frac{d\mu}{d\sigma} \tag{6}$$

$$G_i''(z_i) \frac{dz_i}{d\sigma} = \mu + \sigma \frac{d\mu}{d\sigma} \tag{7}$$

Above we showed that $\mu(\sigma) \in (0, \eta) \forall \sigma > 0$. From this it follows that $\sigma \frac{d\mu}{d\sigma} \in (-\mu, \eta) \forall \sigma > 0$. This is because for σ such that $\sigma \frac{d\mu}{d\sigma} > \eta - \mu$ we would have $\frac{dx_i}{d\sigma} > 0 \forall i$ (by equation 5), $\frac{dy_i}{d\sigma} < 0 \forall i$ (by equation 6) and $\frac{dz_i}{d\sigma} > 0 \forall i$ (by equation 7), so that $\frac{d}{d\sigma}(x_i - y_i + z_i) > 0 \forall i$ which would contradict the market clearing condition.

Similarly, for $\sigma\frac{d\mu}{d\sigma} < -\mu$ we would have $\frac{dx_i}{d\sigma} < 0\forall i$ (by equation 5), $\frac{dy_i}{d\sigma} > 0\forall i$ (by equation 6) and $\frac{dz_i}{d\sigma} < 0\forall i$ (by equation 7) so that $\frac{d}{d\sigma}(x_i - y_i + z_i) > 0\forall i$, which would contradict the market clearing condition.

We are now ready to show that "generically" we must have $lim_{\sigma\to 0}\sigma\frac{d\mu}{d\sigma} = 0$. To establish this, let us first suppose that $lim_{\sigma\to 0}\sigma\frac{d\mu}{d\sigma}$ exists. Suppose we have $lim_{\sigma\to 0}\sigma\frac{d\mu}{d\sigma} = K \in (0,\infty]$. Then there exists some $\tilde{\sigma} > 0$ such that for all $\sigma \in (0,\tilde{\sigma}]$ we have $\frac{d\mu}{d\sigma} > \frac{K}{2\sigma}$. Integrating this yields for all $\sigma \in (0,\tilde{\sigma}]$: $\mu(\tilde{\sigma}) - \mu(\sigma) = \int_{s=\sigma}^{\tilde{\sigma}} \frac{d\mu}{d\sigma} > \int_{s=\sigma}^{\tilde{\sigma}} \frac{K}{s} ds = K(log(\tilde{\sigma}) - log(\sigma))$. Rearranging yields: $\mu(\sigma) < \mu(\tilde{\sigma}) - K(log(\tilde{\sigma}) - log(\sigma))$. But this would imply that $lim_{\sigma\to 0}\mu(\sigma) = -\infty$, in contradiction to the fact that, as shown above, we always have $\mu(\sigma) \in (0,\eta)\forall\sigma$. Similarly, we can show that $lim_{\sigma\to 0}\sigma\frac{d\mu}{d\sigma} = K < 0$ leads to a contradiction.

Now in our quest to show that $lim_{\sigma\to 0}\sigma\frac{d\mu}{d\sigma} = 0$ we only have one more case to show to be impossible, namely the case where $lim_{\sigma\to 0}\sigma\frac{d\mu}{d\sigma}$ does not exist. But in this case $\frac{d\mu}{d\sigma}$ has to fluctuate infinitely often by unbounded amounts as σ approaches 0. Intuitively, this will "generically" never happen.

Now using the surjectivity lemma 1yields:

$\frac{dF_x}{d\sigma} = \frac{d(\sum_i F_{ix})}{d\sigma} = \sum_i (C_i'(x_i) - p)\frac{\mu-\eta+\sigma\frac{d\mu}{d\sigma}}{C_i''(x_i)}$

$\frac{dF_y}{d\sigma} = \frac{d(\sum_i F_{iy})}{d\sigma} = \sum_i -(B_i'(y_i) - p)\frac{\mu+\sigma\frac{d\mu}{d\sigma}}{B_i''(x_i)}$

$\frac{dF_z}{d\sigma} = \frac{d(\sum_i F_{iz})}{d\sigma} = \sum_i (G_i'(x_i) - p)\frac{\mu+\sigma\frac{d\mu}{d\sigma}}{G_i''(x_i)}$

Using the optimality conditions yields:

$\frac{dF_x}{d\sigma} = \sum_i \sigma(\mu-\eta)\frac{\mu-\eta+\sigma\frac{d\mu}{d\sigma}}{C_i''(x_i)}$

$\frac{dF_y}{d\sigma} = \sum_i \sigma\mu\frac{\mu+\sigma\frac{d\mu}{d\sigma}}{-B_i''(x_i)}$

$\frac{dF_{i,z}}{d\sigma} = \sum_i \sigma\mu\frac{\mu+\sigma\frac{d\mu}{d\sigma}}{G_i''(x_i)}$

$$\frac{\frac{dF_x}{d\sigma}}{\frac{dF_x}{d\sigma}+\frac{dF_y}{d\sigma}+\frac{dF_z}{d\sigma}} = \frac{(\mu-\eta+\sigma\frac{d\mu}{d\sigma})(\mu-\eta)\sum\frac{1}{C_i''(x_i)}}{(\mu-\eta+\sigma\frac{d\mu}{d\sigma})(\mu-\eta)\sum\frac{1}{C_i''(x_i)}+(\mu+\sigma\frac{d\mu}{d\sigma})\mu\sum\frac{1}{-B_i''(y_i)}+(\mu+\sigma\frac{d\mu}{d\sigma})\mu\sum\frac{1}{G_i''(z_i)}}$$

Now we differentiate the market clearing condition, getting

$\sum\frac{dx_i}{d\sigma} + \frac{dz_i}{d\sigma} = \sum\frac{dy_i}{d\sigma}$

Substituting into this gives:

$(\mu-\eta+\sigma\frac{d\mu}{d\sigma})\sum\frac{1}{C_i''(x_i)} + (\mu+\sigma\frac{d\mu}{d\sigma})\sum\frac{1}{G_i''(x_i)} = (\mu+\sigma\frac{d\mu}{d\sigma})\sum\frac{1}{B_i''(y_i)}$

so $(\mu-\eta+\sigma\frac{d\mu}{d\sigma})\sum\frac{1}{C_i''(x_i)} = -(\mu+\sigma\frac{d\mu}{d\sigma})(\sum\frac{1}{-B_i''(y_i)} + \sum\frac{1}{G_i''(x_i)})$

which we can plug in to get:

$$
\begin{aligned}
\frac{\frac{dF_x}{d\sigma}}{\frac{dF_x}{d\sigma}+\frac{dF_y}{d\sigma}+\frac{dF_z}{d\sigma}} &= \frac{-(\mu+\sigma\frac{d\mu}{d\sigma})(\sum\frac{1}{-B_i''(y_i)}+\sum\frac{1}{G_i''(x_i)})(\mu-\eta)}{-(\mu+\sigma\frac{d\mu}{d\sigma})(\sum\frac{1}{-B_i''(y_i)}+\sum\frac{1}{G_i''(x_i)})(\mu-\eta)+(\mu+\sigma\frac{d\mu}{d\sigma})\mu\sum\frac{1}{-B_i''(y_i)}+(\mu+\sigma\frac{d\mu}{d\sigma})\mu\sum\frac{1}{G_i''(z_i)}}\\
&= \frac{-(\mu+\sigma\frac{d\mu}{d\sigma})(\sum\frac{1}{-B_i''(y_i)}+\sum\frac{1}{G_i''(x_i)})(\mu-\eta)}{-(\mu+\sigma\frac{d\mu}{d\sigma})(\sum\frac{1}{-B_i''(y_i)}+\sum\frac{1}{G_i''(x_i)})(\mu-\eta)+(\mu+\sigma\frac{d\mu}{d\sigma})\mu\sum\frac{1}{-B_i''(y_i)}+(\mu+\sigma\frac{d\mu}{d\sigma})\mu\sum\frac{1}{G_i''(z_i)}}\\
&= \frac{-(\sum\frac{1}{-B_i''(y_i)}+\sum\frac{1}{G_i''(x_i)})(\mu-\eta)}{-(\sum\frac{1}{-B_i''(y_i)}+\sum\frac{1}{G_i''(x_i)})(\mu-\eta)+\mu\sum\frac{1}{-B_i''(y_i)}+\mu\sum\frac{1}{G_i''(z_i)}}\\
&= \frac{-(\mu-\eta)}{-(\mu-\eta)+\mu}\\
&= \frac{\eta-\mu}{\eta}
\end{aligned}
$$

Increasing σ corresponds to increasing F. Therefore, we must have $\frac{dF_x}{d\sigma}+\frac{dF_y}{d\sigma}+\frac{dF_z}{d\sigma}>0$. But we already established that $\mu\in(0,\eta)$, so it follows that $\frac{dF_x}{d\sigma}>0$.

We now proceed analogously for F_y:

$$\frac{\frac{dF_y}{d\sigma}}{\frac{dF_x}{d\sigma}+\frac{dF_y}{d\sigma}+\frac{dF_z}{d\sigma}} = \frac{(\mu+\sigma\frac{d\mu}{d\sigma})\mu\sum\frac{1}{-B_i''(y_i)}}{(\mu+\sigma\frac{d\mu}{d\sigma})(\sum\frac{1}{-B_i''(y_i)}+\sum\frac{1}{G_i''(x_i)})(\eta-\mu)+(\mu+\sigma\frac{d\mu}{d\sigma})\mu\sum\frac{1}{-B_i''(y_i)}+(\mu+\sigma\frac{d\mu}{d\sigma})\mu\sum\frac{1}{G_i''(z_i)}}$$

$$\frac{\frac{dF_y}{d\sigma}}{\frac{dF_x}{d\sigma}+\frac{dF_y}{d\sigma}+\frac{dF_z}{d\sigma}} = \frac{\mu\sum\frac{1}{-B_i''(y_i)}}{\eta(\sum\frac{1}{-B_i''(y_i)}+\sum\frac{1}{G_i''(x_i)})}$$

Hence $\frac{dF_y}{d\sigma}>0$, since $\mu\in(0,\eta)$.q

Now we proceed analogously to F_z:

$$\frac{\frac{dF_z}{d\sigma}}{\frac{dF_x}{d\sigma}+\frac{dF_y}{d\sigma}+\frac{dF_z}{d\sigma}} = \frac{\mu\sum\frac{1}{G_i''(y_i)}}{\eta(\sum\frac{1}{-B_i''(y_i)}+\sum\frac{1}{G_i''(x_i)})}$$

Hence $\frac{dF_y}{d\sigma}>0$, since $\mu\in(0,\eta)$.

Rearranging the condition derived from market clearing yields: $(\mu-\eta+\sigma\frac{d\mu}{d\sigma})\sum\frac{1}{C_i''(x_i)} = -(\mu-\eta+\sigma\frac{d\mu}{d\sigma})(\sum\frac{1}{-B_i''(y_i)}+\sum\frac{1}{G_i''(x_i)})-\eta(\sum\frac{1}{-B_i''(y_i)}+\sum\frac{1}{G_i''(x_i)})$

$$\mu-\eta=-\eta\frac{\sum\frac{1}{-B_i''(y_i)}+\sum\frac{1}{G_i''(x_i)}}{\sum\frac{1}{C_i''(x_i)}+\sum\frac{1}{-B_i''(y_i)}+\sum\frac{1}{G_i''(x_i)}}-\sigma\frac{d\mu}{d\sigma}$$

$$\frac{\frac{dt_x}{d\sigma}}{\frac{dt_x}{d\sigma}+\frac{dt_y}{d\sigma}+\frac{dt_z}{d\sigma}} = \frac{\sum\frac{1}{-B_i''(y_i)}+\sum\frac{1}{G_i''(x_i)}}{\sum\frac{1}{C_i''(x_i)}+\sum\frac{1}{-B_i''(y_i)}+\sum\frac{1}{G_i''(x_i)}}+\frac{\sigma}{\eta}\frac{d\mu}{d\sigma}$$

using that $\mu_0 := lim_{\sigma\to0}\sigma\frac{d\mu}{d\sigma}=0$ we get:

$$lim_{\sigma\to0}\frac{\frac{dt_x}{d\sigma}}{\frac{dt_x}{d\sigma}+\frac{dt_y}{d\sigma}+\frac{dt_z}{d\sigma}} = \frac{\sum\frac{1}{G_i''(\hat{z_i})}+\sum\frac{1}{-B_i''(\hat{y_i})}}{\sum\frac{1}{G_i''(\hat{z_i})}+\sum\frac{1}{-B_i''(\hat{y_i})}+\sum\frac{1}{C_i''(\hat{x_i})}}$$

Letting $y_i^*(p)$ denote as before the energy demand function for country i in the absence of any mechanism, we have:

$B_i'(y_i^*(p))=p$, so $B_i''(p)\frac{dy_i^*}{dp}(p)=1$

Letting c denote the supply function for coal for country i in the absence of any mechanism, we have:

$C_i'(x_i^*(p))=p$, so $C_i''(p)\frac{dx_i^*}{dp}(p)=1$

Letting z_i^* denote the supply function for coal for country i in the absence of any mechanism, we have:

$G_i'(z_i^*(p))=p(1-q)$, so $G_i''(p)\frac{dz_i^*}{dp}(p)=1$

With this we obtain:

$\sum \frac{1}{-B_i''(\hat{y_i})} = \frac{dy_i^*}{dp} = \frac{y_i^*(p)}{p} \frac{p}{y_i^*(p)} \frac{dy_i^*}{dp} = \frac{y_i^*(p)}{p} \epsilon_y$, where ϵ_D denotes the price elasticity of demand for energy.

$\sum \frac{1}{C_i''(\hat{x_i})} = \frac{dx_i^*}{dp} = \frac{x_i^*(p)}{p} \frac{p}{x_i^*(p)} \frac{dx_i^*}{dp} = \frac{x_i^*(p)}{p} \epsilon_x$, where ϵ_x denotes the price elasticity of supply of coal.

$\sum \frac{1}{G_i''(\hat{z_i})} = \frac{dz_i^*}{dp} = \frac{z_i^*(p)}{p} \frac{p}{z_i^*(p)} \frac{dz_i^*}{dp} = \frac{z_i^*(p)}{p} \epsilon_{S_G}$, where ϵ_{S_G} denotes the price elasticity of supply of renewable energy.

Using these identities we get:

$lim_{\sigma \to 0} \frac{\frac{dF_x}{d\sigma}}{\frac{dF_x}{d\sigma} + \frac{dF_y}{d\sigma} + \frac{dF_z}{d\sigma}} = \frac{\epsilon_z \frac{Z(0)}{Y(0)} + \epsilon_D}{\epsilon_z \frac{Z(0)}{Y(0)} + \epsilon_x \frac{X(0)}{Y(0)} + \epsilon_y} = \frac{\epsilon_z \frac{Z(0)}{Y(0)} + \epsilon_y}{\epsilon_{S_G} \frac{Z(0)}{Y(0)} + \epsilon_x \frac{X(0)}{Y(0)} + \epsilon_y}$

where $X(0)$,$Y(0)$,$Z(0)$ denote the aggregate quantities for $F = 0$.

From this we deduce:

$lim_{\sigma \to 0} \frac{\frac{dF_y}{d\sigma}}{\frac{dF_x}{d\sigma} + \frac{dF_y}{d\sigma} + \frac{dF_z}{d\sigma}} = \frac{\epsilon_y}{\epsilon_z + \epsilon_y} \frac{\epsilon_x \frac{X(0)}{Y(0)}}{\epsilon_z \frac{Z(0)}{Y(0)} + \epsilon_x \frac{X(0)}{Y(0)} + \epsilon_y}$

$lim_{\sigma \to 0} \frac{\frac{dF_z}{d\sigma}}{\frac{dF_x}{d\sigma} + \frac{dF_y}{d\sigma} + \frac{dF_z}{d\sigma}} = \frac{\epsilon_z}{\epsilon_z + \epsilon_y} \frac{\epsilon_x \frac{X(0)}{Y(0)}}{\epsilon_z \frac{Z(0)}{Y(0)} + \epsilon_x \frac{X(0)}{Y(0)} + \epsilon_y}$

But we also have:

$\frac{dF_x}{dF} = \frac{\frac{dF_x}{d\sigma}}{\frac{dF}{d\sigma}} = \frac{\frac{dF_y}{d\sigma}}{\frac{dF_x}{d\sigma} + \frac{dF_y}{d\sigma} + \frac{dF_z}{d\sigma}}$ so the claimed result follows. □

A.6 Proof of Lemma 5

Proof. Let $(x_i(F), y_i(F), z_i(F))$ be the optimal allocation under the budget F. Given any $F_1, F_2 > 0$, suppose we have a budget of $\phi F_1 + (1-\phi)F_2$ with $\phi \in (0,1)$. Set p equal to the status quo value in the absence of any mechanism and set $(x_i, y_i, z_i) = (\phi x_i(F_1) + (1-\phi)x_i(F_2), \phi y_i(F_1) + (1-\phi)y_i(F_2), \phi z_i(F_1) + (1-\phi)z_i(F_2))$. The transfers required for this allocation to satisfy the participation constraints is lower than $\phi F_1 + (1-\phi)F_2$ by the assumed convexity of C_i and G_i and the assumed concavity of B_i. Suppose we set the transfer so that the participation constraints are satisfied with equality. Then, denoting by $\tilde{U}_i$ the values of U_i under the status quo, we have:

$W = \sum \tilde{U}_i - \eta(\phi x_i(F_1) + (1-\phi)x_i(F_2))$

$= \sum \phi(\tilde{U}_i - \eta x_i(F_1)) + (1-\phi)(\tilde{U}_i - \eta x_i(F_2)))$

$= \phi W(F_1) + (1-\phi)W(F_2)$s

But since we have not even used up all our budget, this shows that we can do strictly better than this. □

A.7 Proof of Lemma 6

Proof. The first order conditions for the x_i require that $C_i'(x_i) = C_j'(x_j) \forall i, j$. Thus once we stipulate a value for$\sum_i x_i$, all the x_i are determined via $C_i'(x_i) = C_j'(x_j) \forall i, j$. Then all the y_i, z_i are determined via $\sum_i x_i = \sum_i y_i - \sum_i z_i$, $B_i'(y_i) = B_j'(y_j) = G_i'(z_i) = G_i'(z_j)$. Thus in particular, once $\sum_i x_i$ is fixed, the welfare W is determined and the transfers only depend on p. Conversely, W determines all the x_i, y_i, z_i and the F_x then only depends on p.

Specifically, F_x corresponds to p via $F_x = \sum_i F_{ix} = \sum_i px_i^*(p) - C_i(x_i^*(p)) - (px_i - C_i(x_i))$

where, as usual, we denote by $x_i^*(p)$ i's coal supply in the absence of the mechanism.

$\frac{dF_x}{dp} = \sum_i x_i^* + (p-\alpha)\sum_i \frac{dx_i^*}{dp} - \sum_i \frac{dx_i^*}{dp} C_i'(x_i^*) = \sum_i x_i^*$, so $\frac{dp}{dF_x} = \frac{1}{\sum_i x_i^*}$

Similarly, we get:

$\frac{d\sum_i F_{iy}}{dF_x} = \frac{d\sum_i F_{iy}}{dp}\frac{dp}{dF_x} = \frac{-\sum_i y_i^*}{\sum_k x_k^*}, \frac{d\sum_i F_{iz}}{dF_x} = \frac{\sum_i z_i^*}{\sum_i x_i^*}$, so $\frac{dF}{dF_x} = \frac{\sum_i x_i^* + \sum_i z_i^* - \sum_i y_i^*}{\sum_k x_k^*}$

$\frac{d^2F}{dF_x^2} = \frac{d}{dp}\frac{\sum_i x_i^* + z_i^* - y_i^*}{\sum_k x_k^*}\frac{dp}{dF_x} = \frac{-\sum_k \frac{dx_k^*}{dp}\sum_i (x_i^* + z_i^* - y_i^*) + \sum_k x_k^* (\sum_i \frac{dx_i^*}{dp} + \frac{dz_i^*}{dp} - \frac{dy_i^*}{dp})}{(\sum_k x_k^*)^2}\frac{dp}{dF_x} = \frac{\sum_k \frac{dx_k^*}{dp}\sum_i (y_i^* - z_i^*) + \sum_k x_k^* (\frac{dz_i^*}{dp} - \frac{dy_i^*}{dp})}{(\sum_i x_i^*)^3}$

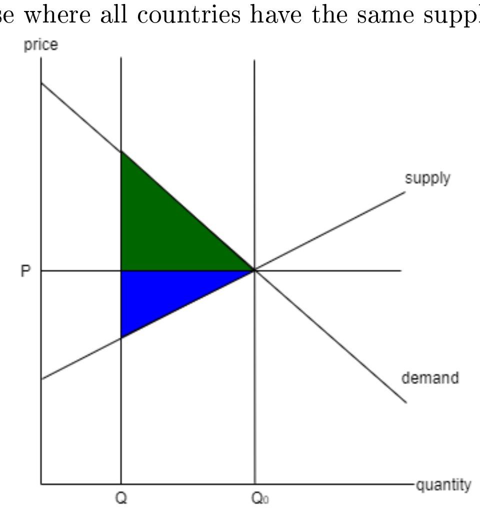
At the optimal mechanism, each country i will bee paid to lower their energy use relative to what it would individually choose were it to ignore the mechanism. Hence we must have $y_i \leq y_i^* \forall i$. Similarly, we must have $z_i \geq z_i^* \forall i$. Moreover, market clearing implies that $\sum_i y_i - z_i \geq 0$, so $\sum_i (y_i^* - z_i^*) \geq 0$.

Since $\frac{dz_i^*}{dp} \geq 0, \frac{dx_i^*}{dp} \geq 0 \forall i$ by the law of supply and $\frac{dy_i^*}{dp} \leq 0$ by the law of demand, it follows that $\frac{d^2F}{dF_x^2} \geq 0$ □

B Proofs and Additional Lemmas for section 5

B.1 Proof of Lemma 7

Proof. By proposition 1 from (Stern 2021) there generically is a unique NE and at most one player making positive contributions at this NE. What remains to be shown is that there is indeed one player making positive contributions. Diagramatically, it is clear that if no funding is provided for the centralised global institution then the marginal emissions reduction of an infinitesimal monetary contribution to it is infinite. To see why, consider the following diagram for the case where all countries have the same supply and demand functions:



Recall that by the Price Preservation Lemma the world market price p of coal will be preserved under the unified architecture. Now the ratio of $Q_0 - Q$ over the blue and green areas goes to ∞ as $Q \to Q_0$ from below.

Formally, recall that the minimal required transfers are

$F_{ix} := \sup_x px - C_i(x) - (px_i - C_i(x_i))$

$F_{iy} := \sup_y B_i(y) - py - (B_i(y) - py)$

But if no player makes any contributions then we have $p = C_i'(x_i) = B_i'(y_i)$, so $\frac{\partial F_{ix}}{\partial x_i} = \frac{\partial F_{iy}}{\partial y_i} = 0$. Thus for the first infinitestimal amount of emissions reductions, the marginal cost is 0. Hence all players have a profitable deviation consisting of making a small contribution to the global institution. □

B.2 Proof of Lemma 8

Proof. By proposition 1 from (Stern 2021) there generically is a unique NE and for each of the two funds there is at most one player making a positive contribution at this NE.

Recall that the minimal required transfers are

$F_{ix} := \sup_x px - C_i(x) - (px_i - C_i(x_i))$

$F_{iy} := \sup_y B_i(y) - py - (B_i(y) - py)$

Formally, recall that the minimal required transfers are

$F_{ix} := \sup_x px - C_i(x) - (px_i - C_i(x_i))$

$F_{iy} := \sup_y B_i(y) - py - (B_i(y) - py)$

Denoting as usual by $x_i^*(p)$ and $y_i^*(p)$ the supply and demand functions absent reward payment schemes, we get:

$\frac{\partial F_{ix}}{\partial p} := x_i^*(p) - x_i$

$\frac{\partial F_x}{\partial p} := x^*(p) - x$

$\frac{\partial F_{iy}}{\partial p} := -y_i^*(p) + y_i$

$\frac{\partial F_y}{\partial p} := -y^*(p) + y$

Also, by the Constrained Efficiency Lemma, we can think of each of the two reward payment fund as choosing the aggregate variables, x and y, it being understood that the reward payment schemes do so in the efficient way. Let us denote by C and B the aggregate cost and benefit functions.

$\frac{\partial F_x}{\partial x} := -p + C'(x)$

$\frac{\partial F_y}{\partial y} := p - B'(y)$

Using that $x = y$ we get the Jacobian matrix J for the mapping $(x, p) \mapsto (F_x, F_y)$:

$$J = \begin{pmatrix} \frac{\partial F_x}{\partial x} & \frac{\partial F_x}{\partial p} \\ \frac{\partial F_y}{\partial x} & \frac{\partial F_y}{\partial p} \end{pmatrix}$$

By the inverse function theorem, we can compute the Jacobian of the inverse by inverting the Jacobian. From this we obtain:

$\frac{\partial x}{\partial F_x} = \frac{y^*(p) - x}{(B'(x)-p)(x-x^*(p)) + (C'(x)-p)(y^*(p)-x)}$

$\frac{\partial p}{\partial F_x} = \frac{p - B'(x)}{(B'(x)-p)(x-x^*(p)) + (C'(x)-p)(y^*(p)-x)}$

Now let us show that none of the two funds can end up empty-handed at the Nash equilibrium. The argument provided in the proof of Lemma 7 shows that it cannot happen that both funds end up without any contribution.

Now suppose that $F_y > 0$. Then $y^*(p) - x > 0$ and $B'(x) - p > 0$. Since $lim_{F_x \to 0^+} - x + x^*(p) = 0^-$ and $lim_{F_x \to 0^+} C'(x) - p = 0^-$, we deduce:

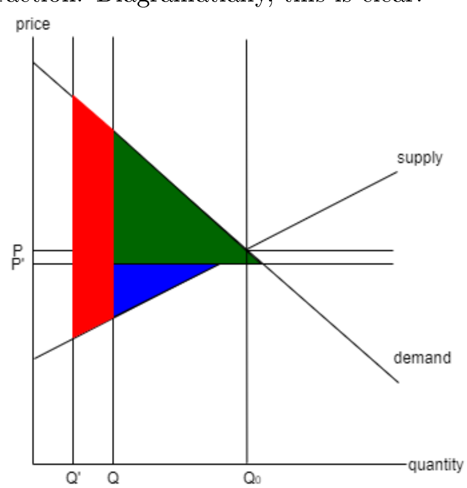
$lim_{F_x \to 0^+} \frac{\partial x}{\partial F_x} = -\infty$

$lim_{F_x \to 0^+} \frac{\partial p}{\partial F_x} = +\infty$

This implies that any coal importer will (by the envelope theorem) reap infinite marginal benefits from contributing an infinitesimal amount when $F_x = 0$. □

B.3 Proof of Lemma 9

Proof. First let us establish that the marginal money required to achieve a price-preserving reduction in coal extraction only depends on the level of coal extraction. Diagramatially, this is clear:



Here the red area corresponds to the additional transfers required for a price-preserving reduction of fossil fuel (production and use) from Q to Q'. This clearly does not depend on the price P'.

To verify the claim formally, recall that the minimal required transfers are

$F_{ix} := \sup_x px - C_i(x) - (px_i - C_i(x_i))$

$F_{iy} := \sup_y B_i(y) - py - (B_i(y) - py)$

Let $x_i(x)$ denote country i's coal extraction if the global institution implements an aggregate coal extraction of x. By the constrained efficiency lemma this does not depend on p. Now we can compute the change in the required aggregate budget for a price-preserving reduction in aggregate coal extraction:

$\frac{\partial F}{\partial x} = \sum_{i \in I} \frac{\partial F_{ix}}{\partial x} + \frac{\partial F_{iy}}{\partial x} = \sum_{i \in I} \frac{\partial F_{ix}}{\partial x_i} \frac{dx_i}{dx} + \frac{\partial F_{iy}}{\partial y_i} \frac{dy_i}{dx} = \sum_{i \in I} (-p + C_i'(x_i)) \frac{dx_i}{dx} + (p - B_i'(y_i)) \frac{dy_i}{dx}$

Now using that by market clearing we have $\sum_{i \in I} \frac{dy_i}{dx} - \frac{dx_i}{dx} = 0$, we get:

$\frac{\partial F}{\partial x} = \sum_{i \in I} C_i'(x_i) \frac{dx_i}{dx} - B_i'(y_i) \frac{dy_i}{dx}$

which does not depend on p. By the Constrained Efficiency Lemma, we know that all the $C_i'(x_i)$ are equal. Since each of the $C_i'(x_i)$ is strictly increasing, it now follows that $\frac{\partial F}{\partial x}$ is strictly increasing in x. Hence the marginal cost of buying price-preserving reduction in aggregate coal extraction, $-\frac{\partial F}{\partial x}$, is strictly decreasing in x

Now let us compare the Nash equilibria under the de-centralised and the centralised architectures. Consider the unique player j contributing at the NE under the centralized architecture. Under both architectures player j has the opportunity to buy an infinitesimal price-preserving reduction in aggregate coal extraction. At the Nash Equilibrium under the centralised architecture player j is indifferent between buying such a price-preserving reduction in aggregate coal extraction or not. At the Nash Equilibrium under the de-centralised architecture such a price-preserving reduction in aggregate coal extraction cannot be strictly cheaper, for otherwise player j would have a strictly profitable deviation consisting of buying it. But since we have shown that $\frac{\partial F}{\partial x}$ is strictly decreasing in x, we can now deduce that $x_{de-centralized} \leq x_{centralized}$.

Now suppose that only one player makes contributions under the de-centralised architecture. Then by 8, this player contributes to both institutions. This means that this player is indifferent between buying an infinitesimal price-preserving reduction in aggregate coal extraction or not. But since we showed above that the marginal cost of a price preserving reduction in coal, i.e.$-\frac{\partial F}{\partial x}$, is strictly decreasing in x, we know that there is a unique x such that player j is indifferent between buying an infinitesimal price-preserving reduction in aggregate coal extraction or not. This unique x must be $x_{centralized}$. □

Now come further intermediate Lemmas and proofs for the linear specification of demand and supply. All the results (and further ones) for the endogenous funding model are computed in a Mathematica notebook downloadable here.

Lemma 13. *The largest coal exporter, denoted by ex, has the following welfare:*

$$U_{ex} = s_{ex}\left((1-\theta)\left(\frac{(e_d(p-1)+x-1)^2}{2e_d} + \frac{x(2e_d-x+2)}{2e_d} - px\right) + (1+\theta)\left(\frac{((p-1)e_s-x+1)^2}{2e_s} + px - \frac{x(2e_s+x-2)}{2e_s}\right) - \eta x\right) - t_{x,ex} - t_{y,im}$$

where p and x are given by the expressionss from Lemma 11 with $t_x = t_{x,ex} + t_{x,im}$, $t_y = t_{y,ex} + t_{y,im}$. As defined above, θ denotes the ratio of total global coal exports over total global coal production.

Lemma 14. *With separate funds the emissions reductions resulting at there are 3 possible cases:*

Case 1:

$$(\theta(e_d-e_S)(e_d s_{im} - e_s s_{ex}) + (e_d+e_s)(e_d(s_{im}-1) + e_s(s_{ex}-1)))(e_d\eta(s_{ex} - s_{im})(s_{im}\theta(e_d-e_s) + (s_{im}-1)(e_d+e_s)) + 2\theta(e_d+e_s)(s_{ex}(2s_{im}-1) - s_{im})) \leq 0$$

In this case only the importer (i.e. player im) contributes at the Nash equilibrium.

Case 2:

$$(\theta(e_d-e_s)(e_d s_{im} - e_s s_{ex}) + (e_d+e_s)(e_d(s_{im}-1) + e_s(s_{ex}-1)))(\eta e_s(s_{ex} - s_{im})(e_d s_{ex}(\theta-1) + e_d - e_s s_{ex}(\theta+1) + e_s) + 2\theta(e_d+e_s)(s_{ex}(2s_{im}-1) - s_{im})) \leq 0$$

In this case only the exporter (i.e. player ex) contributes at the Nash equilibrium.

Case 3:Neither case 1 or case 2 obtains.

In this both im and ex contribute at the Nash equilibrium. im only contributes to the fund for demand reduction and ex only to the fund for supply reduction.

C Mathematica notebooks

A Mathematica notebook for the numerical computations under constant elasticity specifications can be downloaded here.

A Mathematica notebook computing the spending paths on the three approaches by the global institution at the optimal mechanism with full commitment and no borrowing or saving constraints can be downloaded here.

A Mathematica notebook computing the surfaces shown in section D about the loss from misallocation can be downloaded here.

D Robustness checks about the loss from misallocation

In section 4 I showed for a particular combination of elasticity estimates how welfare depends on the budget split. One takeaway was that the loss from misallocation is relatively small: as long as each of the three approaches gets at least 50% its optimal proportion of the budget, welfare losses are at most 10%. This particular way of summarizing the "flatness of the welfare surface around the optimum" is motivated by the discussion presented in section 6 concerning the pros and cons of flexible decentralized mechanisms: We should expect countries' allocation decisions in such mechanisms to be guided by a mixture of concern for global welfare and their own payoffs. Large fossil fuel exporters will strongly prefer money to go to the supply reduction approach since this will raise fossil fuel world market prices. Fossil fuel importers, on the other hand, will prefer money to go to the demand reduction and substitute expansion approaches. Those countries primarily concerned about climate change will prefer money to go at the margin to whatever of the three approaches is underfunded relative to the others. Thus we should expect the overall allocation to be somewhat responsive to what actually turns out to be good for global welfare. Based on this, I now assume for concreteness that for any given overall budget each of the three approaches gets at least half its optimal proportion.

Under this constraint, the worst outcome in terms of global welfare occurs when two of the three approaches each get only half their optimal proportion of the budget, with the third approach getting the rest. I refer to the three corresponding cases as "supply-side-heavy", "demand-side-heavy" and "substitute-side-heavy". I plot below the proportion of welfare realised under these three cases relative to the welfare that would be realised if the budget was split optimally across the three approaches. For these numerical simulations I assume constant-elasticity specifications for coal supply, energy demand and renewable energy supply. Throughout I assume that the overall budget is small. It turns out that all the results change little with the size of the budget. The plots explore the entire range of elasticity estimates that I have found in the literature, as I detail in the following subsections.

D.1 Estimates of long run price elasticities of demand for energy

Espey and Espey (2004) carried out a meta-analysis about residential electricity demand. of price and income elasticity estimates from 36 studies published over the period 1947 to 1997. The 125 estimates of long-run price elasticity fell in the range from -2.25 to -0.04 with a mean of -0.85. All the more recent studies that I have seen have estimates falling in this range[17] I thus consider the range -2.25 to -0.04 in the plots shown below.

D.2 Estimates of price elasticities of supply of renewable energy

I have only found a single study, namely Johnson (2011), which gives an estimate of 2.7. In the plots shown below I consider the range from 0.1 to 3 for the price elasticity of supply of renewables.

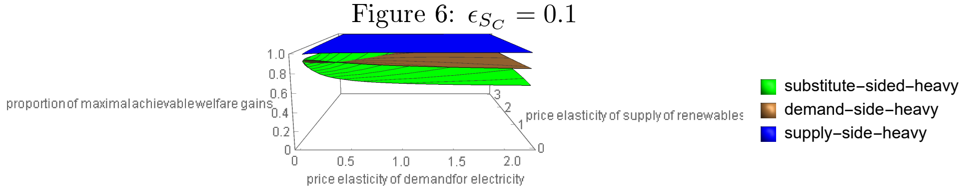
D.3 Estimates of the price elasticity of supply of coal

Daubanes, J., Henriet, F., & Schubert, K. (2020). note that the empirical literature on the price elasticity of coal supply—e.g., Labys et al. (1979), Beck et al. (1991), Light (1999), Light et al. (1999), and Dahl (2009)—finds estimates ranging from 0.1 and 1.9. Based on this, I consider the range from 0.1 to 1.9.

D.4 Results

Here is the case where the price elasticity of supply of coal is 0.1:

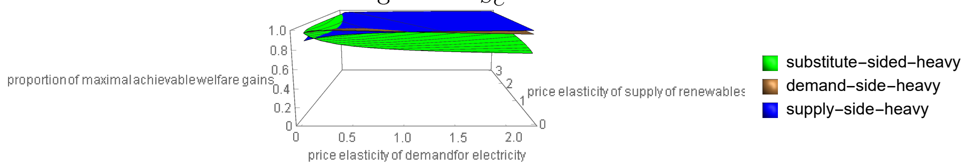
Figure 6: $\epsilon_{S_C} = 0.1$



Here is the case where the price elasticity of supply of coal is 1.9:

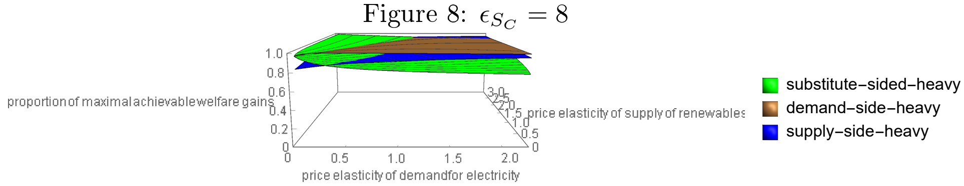
[17] E.g. Burke & Abayasekara (2018) find -1.

Figure 7: $\epsilon_{S_C} = 1.9$



Whilst the estimates for the price elasticity of supply of coal range from 0.1 to 1.9, we presumably cannot rule out potentially much large value for it in the long term. For illustration, consider the case where the price elasticity of supply of coal is 8:

Figure 8: $\epsilon_{S_C} = 8$



Overall, these results suggest that the conclusion that the welfare losses from misallocation are likely to be small is robust.

References

[1] Bayram, A. B., & Graham, E. R. (2017). Financing the United Nations: Explaining variation in how donors provide funding to the UN. The Review of International Organizations, 12(3), 421-459.

[2] Bretschger and Pattakou (2018), Bretschger, L., & Pattakou, A. (2019). As bad as it gets: how climate damage functions affect growth and the social cost of carbon. Environmental and resource economics, 72(1), 5-26.

[3] Burke, P. J., & Abayasekara, A. (2018). The price elasticity of electricity demand in the United States: A three-dimensional analysis. The Energy Journal, 39(2).

[4] Cramton, P., & Stoft, S. (2012). Global climate games: How pricing and a green fund foster cooperation. Economics of Energy & Environmental Policy, 1(2), 125-136.

[5] Collier, P., & Venables, A. J. (2014). Closing coal: economic and moral incentives. Oxford Review of Economic Policy, 30(3), 492-512.

[6] Dahl, C. (2009). Energy demand and supply elasticities. Energy Policy, 72.

[7] Daubanes, J., Henriet, F., & Schubert, K. (2020). Unilateral CO2 Reduction Policy with More Than One Carbon Energy Source.

[8] Edenhofer, O., & Kalkuhl, M. (2011). When do increasing carbon taxes accelerate global warming? A note on the green paradox. Energy Policy, 39(4), 2208-2212.

[9] Eichner, T., Kollenbach, G., & Schopf, M. (2020). Buying versus leasing fuel deposits for preservation. The Scandinavian Journal of Economics.

[10] Espey, J. A., & Espey, M. (2004). Turning on the lights: A meta-analysis of residential electricity demand elasticities. Journal of Agricultural and Applied Economics, 36(1379-2016-112600), 65-81.

[11] Fæhn, T., Hagem, C., Lindholt, L., Mæland, S., & Rosendahl, K. E. (2017). Climate policies in a fossil fuel producing country–demand versus supply side policies. The Energy Journal, 38(1).

[12] Harstad, B. (2012). Buy coal! A case for supply-side environmental policy. Journal of Political Economy, 120(1), 77-115.

[13] Hoel, M. (2014). Supply side climate policy and the green Paradox. Climate policy and nonrenewable resources: The green paradox and beyond, 21.

[14] IPPC (2014). Climate change, synthesis report, summary for policymakers

[15] Johnson, E. P. (2011). The price elasticity of supply of renewable electricity generation: Evidence from state renewable portfolio standards.

[16] Kornek, U., & Edenhofer, O. (2020). The strategic dimension of financing global public goods. European Economic Review, 127, 103423.

[17] Long, N. V. (2015).The green paradox in open economies: Lessons from static and dynamic models. Review of Environmental Economics and Policy, 9(2), 266-284.

[18] Martimort, D., & Sand-Zantman, W. (2016). A mechanism design approach to climate-change agreements. Journal of the European Economic Association, 14(3), 669-718.

[19] Rezai, A., & Van Der Ploeg, F. (2017). Second-best renewable subsidies to de-carbonize the economy: commitment and the Green Paradox. Environmental and Resource Economics, 66(3), 409-434.

[20] Ross, M. L. (2015). What have we learned about the resource curse?. Annual Review of Political Science, 18, 239-259.

[21] Seymour, F., & Busch, J. (2016). Why forests? Why now?: The science, economics, and politics of tropical forests and climate change. Brookings Institution Press.

[22] Steckel, J. C., Edenhofer, O., & Jakob, M. (2015). Drivers for the renaissance of coal. Proceedings of the National Academy of Sciences, 112(29), E3775-E3781.

[23] Teng, M., Burke, P. J., & Liao, H. (2019). The demand for coal among China's rural households: Estimates of price and income elasticities. Energy Economics, 80, 928-936.

[24] Van der Ploeg, F., & Withagen, C. (2014). Growth, renewables, and the optimal carbon tax. International Economic Review, 55(1), 283-311.

[25] Van der Ploeg, F., & Withagen, C. (2015). Global warming and the green paradox: A review of adverse effects of climate policies. Review of Environmental Economics and Policy, 9(2), 285-303.

[26] Van der Ploeg, F. (2016). Second-best carbon taxation in the global economy: the Green Paradox and carbon leakage revisited. Journal of Environmental Economics and Management, 78, 85-105.