

Boon or Bane? Trade Sanctions and the Stability of International Environmental Agreements

Achim Hagen ^{*}, Jan Schneider [†]

Draft Version, December 5, 2017

Abstract

In spite of scientific agreement on the negative effects of anthropogenic climate change, efforts to find cooperative solutions on the international level have been unsatisfactory so far. Trade sanctions in the form of import tariffs are one principal measure discussed as a means to foster cooperation. Former studies have concluded that import tariffs are an effective mechanism to establish international cooperation. However, most of these studies rely on the assumption that outsiders are not able to retaliate, i.e. to implement import tariffs themselves. In this paper we use combined analytical and numerical analysis to investigate implications of retaliation. We find a threshold effect: below a certain coalition size the effect of retaliation predominates and decreases incentives to be a coalition member. In coalitions above the threshold size the effect of trade sanctions that stabilizes coalitions dominates and enables the formation of larger stable coalitions. Our analysis suggests that only after a sufficiently large climate coalition has already been formed, the threat of trade sanctions might be an effective stick to establish the grand coalition.

JEL-Classification: D58, Q54, Q58

Keywords: international environmental agreements; computable general equilibrium

Acknowledgements: We thank Christoph Böhringer, Thomas Rutherford and Roman Mendele-
vitch for valuable comments and conversations. We have further benefitted from discussions at the
Grantham Research Institute on Climate Change and the Environment, London School of Eco-
nomics and Political Science, the Resource Economics Group at Humboldt-Universität zu Berlin,
EAERE 2017, CESifo Area Conference on Energy and Climate Economics 2017 and IAEE Inter-
national Conference 2017.

^{*}Humboldt-Universität zu Berlin, Germany; email achim.hagen@hu-berlin.de

[†]Carl von Ossietzky University Oldenburg, Germany; email jan.schneider@uni-oldenburg.de

1 Introduction

In spite of scientific agreement on the negative effects of anthropogenic climate change, efforts to find cooperative solutions at the international UN climate negotiations have been unsatisfactory so far. Game theoretic strands of economic literature have studied the formation and stability of international environmental agreements and found that it is difficult to reach large stable climate coalitions due to strong freerider incentives (see Marrouch and Ray Chaudhuri, 2016, for a comprehensive overview of the literature).¹

Given the global public bad characteristics of greenhouse gas emissions, the absence of global cooperation gives rise to two important questions for countries intending to enhance global emission abatement: (i) How to design second-best unilateral policies and (ii) how to incentivize broader cooperation? A principal approach to both questions is to link climate and trade policy: the use or the threat of trade measures against countries without emission regulations.

As to (i), the policy debate is concerned with drawbacks of unilateral policies associated to carbon leakage, i.e. an emission increase in unconstrained regions triggered by domestic climate policy. Two interrelated mechanisms can lead to carbon leakage (Felder and Rutherford, 1993): a shift of emission-intensive production to competitors abroad due to cost-disadvantages from carbon pricing for domestic firms. Likewise, international prices for fossil fuels might drop due to climate policies, thereby incentivizing higher fossil fuel consumption in unregulated economies. Anti-leakage measures are discussed, among others, in the form of border carbon adjustments: countries with a domestic carbon price could tax carbon embodied in imports from unregulated regions, and likewise rebate carbon payments to exports to unregulated regions.² A major finding of applied studies on border carbon adjustments is that they shift substantial parts of the burden of emission reduction to developing countries. The burden shifting effect is related to the economic theory on optimal tariffs (Limão, 2008): countries are able to benefit from the introduction of import tariffs in terms of domestic welfare, while their trading partners suffer losses. The basic mechanism is a change in the ratio of export and import prices – the terms of trade – in favor of the tariff imposing country.

The rationale for trade sanctions to approach question (ii) builds on the aforementioned insight from trade theory that import tariffs can benefit the importer and hurt the exporting country, as this could make import tariffs a credible threat in the international game for greenhouse gas reductions. Tariffs in this case are imposed as sanctions and are meant to be punitive to non-

¹Recent extensions of the game theoretic literature include inter alia multiple agreements (Hagen and Eisenack, 2015) or minimum participation constraints (Weikard et al., 2015; Carraro et al., 2009).

²In policy proposals, border carbon adjustments are primarily considered for sectors that show high emission and trade intensities.

participants.³ The rationale for the use of trade sanctions is thus rather rooted in strategic considerations: cooperation on free trade is made conditional on cooperation on emission abatement.

To date, there are no trade measures in climate policy in place.⁴ Nonetheless, trade measures in the form of border carbon adjustments are mentioned as possible complementary measures in the EU Emissions Trading Directive (2009/29/EC) as well as in all the major US climate bills, i.e. the Waxman-Markey bill (US Congress, 2009), the Kerry-Boxer bill (Larsen et al., 2009), and the Cantwell-Collins bill (Larsen and Bradbury, 2010). The justification of trade measures in these climate policy legislations and bills has focused on carbon leakage and related concerns on competitiveness losses of emission-intensive and trade-exposed domestic industries.

In recent years, the focus of the debate on trade measures has shifted from tariffs as a second-best instrument combating carbon leakage to the appeal of tariffs as a strategic stick to foster cooperation. In particular, the announced withdrawal of the US from the Paris Agreement has opened a new debate on the role of punitive tariffs (Kemp, 2017).⁵

Previous studies on trade sanctions as a strategic means in climate policy have concluded that even low import tariffs are an effective tool to reach larger coalitions (Lessmann et al., 2009; Nordhaus, 2015). However, these studies rely on the assumption that outsiders of the coalition are not able to retaliate by imposing import tariffs themselves.⁶ This is a crucial assumption given that China has already threatened with trade war should it be subjected to border carbon adjustments (Voituriez and Wang, 2011) and given increased concerns over protectionism and trade war (Denyer, 2017).

In this paper, we take a more comprehensive approach and compare implications for coalition stability under three principal policy regimes: (i) A regime without trade sanctions; (ii) a regime in which coalition members use trade sanctions in the form of import tariffs against outsiders; and (iii) a regime in which coalition members use trade sanctions and outsiders retaliate with import tariffs.

We combine stylized theoretical analysis in a non-cooperative game theoretic model and numerical analysis in a static multi-region, multi-sector CGE model of global trade and energy use. Our theoretical findings suggest that trade sanctions increase the incentives to cooperate when retaliation is prohibited, which is in line with former findings in the literature. Considering retaliation

³Prominently, Joseph Stiglitz has explicitly argued in this vein, stating “Fortunately, we have an international trade framework that can be used to force states that inflict harm on others to behave in a better fashion.” (Stiglitz, 2006)

⁴In contrast, in the Montreal Protocol from 1987, which controls substances that deplete the ozone layer, the incorporation of trade measures is considered to have been successful as a threat to ensure full cooperation (Barrett, 2011).

⁵Former French President and then presidential candidate Nicolas Sarkozy: “And so I ask that Europe construct a carbon tax at Europe’s borders, a tax of one to three percent for all the products that come from the United States, if the United States exempts itself from the environmental regulations that we ourselves have imposed on our businesses.” (Harvey, 2016).

⁶An exception is Böhringer et al. (2016), who include retaliation but don’t use a concept of internal coalition stability.

by outsiders, however, leads to a “threshold”-effect: coalitions above a certain size are stabilized compared to the regime without trade sanctions, while coalitions below this threshold are destabilized. This leads to multiple equilibria. In particular, non-cooperation (the empty coalition) is always stable in the regime with retaliation. Our numerical analysis indicates that prospects for cooperation are reduced substantially when outsiders are able to retaliate; the size of the smallest internally stable coalition (other than the empty coalition) in scenarios with trade sanctions and retaliation is well above the size of the stable coalition in the absence of trade sanctions for most scenarios.

Our contribution to the existing literature is twofold. First, former analyses of trade sanctions as a means to foster cooperation in climate policy largely relies on the assumption that outsiders are not able to respond to trade sanctions with retaliation. We consider a regime in which outsiders do retaliate in a theoretical model of an international environmental agreement. Second, we use a multi-sector, multi-region CGE model with a full representation of international trade. This allows a quantification of our analytical findings in a setting with asymmetric regions and where regional and global welfare effects due to climate and trade policies are fully endogenized.⁷ This is of particular virtue as the basic mechanism affecting incentives for cooperation is welfare changes through trade policy.

The remainder of the paper is structured as follows: In section 2, we review related theoretical and applied literature. In section 3, we briefly discuss the economic rationale for the use of trade sanctions. After that, we lay out the assumptions in our three policy regimes in detail. In section 4, we formulate the theoretical model. We subsequently present results on coalition stability under the three regimes. Section 5 introduces the CGE model and data. The quantitative results are discussed in section 6 before we conclude in section 7.

2 Literature Review

Our paper relates both to the game theoretic literature on the nexus of coalition stability and trade and to the applied literature on trade measures in subglobal climate policies. We give an overview of these strands of literature where we focus on work that is more closely related to our considerations.

The game theoretic literature on international environmental agreements (IEAs) started to analyze the logic of coalition formation in the 1990s with seminal papers by Hoel (1992), Carraro

⁷Nordhaus (2015) argues that including a full international trade model is unnecessarily complex. Instead he uses what he calls “reduced-form tariff benefit functions” to represent regional welfare changes induced by tariffs in his C-DICE model. These functions are calibrated using a trade model by Ossa (2014).

and Siniscalco (1993), and Barrett (1994). Barrett (1997) shows in a model with symmetric countries that a trade ban accompanied by a minimum participation clause may help to sustain full cooperation on the provision of a global public good. In contrast, Dong and Zhao (2009) allow for endogenous tariffs which do not serve as a sanction to enforce higher cooperation in their model of an IEA. They find that the total effect of trade on IEA-participation can be positive or negative so that it is not clear if trade increases cooperation. Conconi and Perroni (2002) show in a cooperative game theoretical setting with three symmetric countries that linking decisions on trade and environment can have positive effects on cooperation if environmental costs and benefits are small compared to the costs and benefits of trade policies but can rather hinder cooperation for broader issues like climate change. Neumayer (2002) and Egger et al. (2011) empirically study determinants of cooperation in environmental agreements and find that trade openness has a positive influence on participation. Eichner and Pethig (2013) analyse IEA-formation in a model with consumption and production of fossil fuel and a composite consumer good and international trade. They find that with free trade larger coalitions may be sustained than under autarky, but achieve only slight emissions reductions. The model is extended in Eichner and Pethig (2014) where they show that the additional option of a fossil-fuel supply tax may increase global emissions reductions.

The applied literature has thoroughly studied border carbon adjustments as a measure to overcome drawbacks of unilateral carbon pricing associated with carbon leakage – for overview articles see Böhringer et al. (2012) and Branger and Quirion (2014). The focus of this literature has been on the ability of border measures as a means to combat carbon leakage, improve the global cost effectiveness of emission abatement, and reduce adverse impacts on domestic emission-intensive and trade-exposed industries being subjected to unilateral emission regulation. The main findings on border carbon adjustments are that they markedly reduce carbon leakage but their impact on global costs is only moderate. Their main effect is a strong burden shifting from abating to unregulated regions through changes in relative prices of traded goods (terms of trade).

Recently, the focus in applied research has shifted towards the idea of using trade measures not as means to improve sub-global policies given a certain climate coalition, but as a stick to incentivize cooperation and to enlarge the coalition. To our best knowledge, only two papers explicitly take into account possible retaliation by regions subjected to tariffs. Böhringer et al. (2016) use a static computable general equilibrium (CGE) model and set up a game between a coalition that is going forward with carbon pricing and non-coalition regions. The coalition can use carbon tariffs against outsiders. Non-coalition regions can either join the coalition, retaliate, or do nothing. They show that – even under the threat of retaliation – trade measures can spur prospects for cooperation. However, they do not use a concept of internal stability of the climate coalition. Böhringer and

Rutherford (2017) investigate prospects of trade sanctions against the US in order to coerce the US back into the Paris Agreement. They find that even trade war is no credible threat against the US, as the market power of the US on international markets is too large.

Most closely related to our analysis are the studies by Lessmann et al. (2009) and Nordhaus (2015). Lessmann et al. (2009) develop a dynamic model of cooperation and study the effect of trade sanctions in the form of import tariffs on participation in an IEA. They find that low tariff rates of 1.5 to 4% are sufficient to induce full cooperation. By assumption, however, outsiders are not able to retaliate.

In a more recent study, Nordhaus (2015) also suggests trade sanctions in his proposal for a mechanism that may help to stabilize an international environmental agreement that he calls 'climate club'.⁸ These sanctions – in the form of uniform import tariffs – are put in place against outsiders to increase cooperation and stabilize the climate agreement. From numerical simulations he concludes that prospects for international cooperation increase substantially when abating regions impose small trade sanctions against non-participants. As in Lessmann et al. (2009), his results rely on the assumption that outsiders are not able to respond to trade sanctions by members, assuming the treaty would prohibit retaliation.

3 Trade sanctions, policy regimes, and stability

In order to lay out policy regimes we need to be clear about the economic rationale behind import tariffs as a sanctioning mechanism, which traces back to Bickerdike (1906). We discuss this in a non-technical way and describe the details of the policy regimes we analyze in subsequent sections. This forms the basis for our analytical and numerical considerations.

Assume a large importing region in an undistorted equilibrium. In theory, the domestic welfare effect of imposing import tariffs is driven by two opposing factors: (i) The tariff reduces demand for the imported good, which puts a downward pressure on the respective import price. Consequently, the ratio of export and import prices changes in favor of the tariff imposing country, which can now pay for more physical units of imports with the same physical amount of exports. This is the terms-of-trade effect. In that sense, the tariff works as a substitute for the exertion of market power by consumers on the demand side. (ii) Starting from an undistorted equilibrium, the tariff creates a deadweight loss.

In a linear demand and supply structure, domestic welfare improvements due to the first effect are proportional to the tariff rate, while the deadweight loss is quadratic in the tariff rate. This

⁸Note that there exist other definitions of the concept of climate clubs – see e.g. Weischer et al. (2012), Widerberg and Stenson (2013), and Hagen and Eisenack (2015).

implies that for low tariff rates the former effect dominates the latter, i.e. there is scope for welfare improvements for tariff rates up to an “optimal tariff” for the taxing country. Global welfare, however, will unambiguously decline compared to the undistorted equilibrium due to the deadweight loss.

We represent these effects in a very stylized way in our analytical model. In the CGE framework, price changes – and thus terms-of-trade effects – as well as costs due to trade distortions are endogenous to the represented production, consumption, and trade activities.

We investigate three different policy regimes for international climate policy analytically and numerically. In all regimes, we assume a global social cost of carbon (SCC) as an exogenous parameter. The global SCC translates into regional SCCs via GDP-shares which can be interpreted as a regional constant marginal benefit of emission abatement. In a non-cooperative setting each region has an incentive to abate emissions up the point where marginal abatement cost equal marginal benefits, i.e. to introduce a CO₂ tax equal to its regional SCC. We refer to this reference situation as the empty coalition.

In our first policy regime – denoted *NTRF* – we allow cooperation but no trade sanctions. In this case, regions can form coalitions in order to jointly maximize their net payoffs. In the second regime – denoted *UTRF* – coalition members unilaterally impose trade sanctions on outsiders in the form of uniform import tariffs. Outsiders are prohibited from retaliating. Under the third regime – denoted *RTRF* – coalition members impose trade sanctions and outsiders retaliate by imposing uniform import tariffs themselves.

To investigate coalition stability under different regimes, we apply the concepts of internal and external stability (D’Aspremont et al., 1983) which are standard in the literature of IEAs (Hoel, 1992; Carraro and Siniscalco, 1993; Marchiori et al., 2017). Internal stability implies that no member country has an incentive to leave the agreement whereas the external stability condition is satisfied if no outsider has an incentive to join the existing coalition.

4 Analytical model

To study the qualitative effects of trade sanctions and retaliation on the stability of IEAs we setup a stylized non-cooperative game theoretical model which follows standard assumptions in the IEA-literature (Marrouch and Ray Chaudhuri, 2016; Finus, 2001; Hagen et al., 2017) before the simulations in the calibrated CGE model allow for a quantification of the impacts. The policy regimes without trade sanctions (*NTRF*) and with unilateral trade sanctions (*UTRF*) are completely in line with former work. We develop them here for the sake of completeness and in order to

set up definitions and function specifications to arrive at our main result on coalition stability under trade sanctions and retaliation (*RTRF*). We analyze stable coalitions in the *NTRF* regime and then study the impact of a policy regime shift to *UTRF* and *RTRF* on the stability of coalitions.

4.1 The two-stage game

The model is set up as a two-stage game (e.g. Barrett, 2001). At the first stage, countries decide if they join the climate coalition or not. At the second stage, the members of the coalition decide cooperatively about the amount of emission abatement in a simultaneous game between the coalition and the outsiders. Assume there are $i = 1, \dots, N$ symmetric regions with individual payoffs from emission abatement $\Pi(q_i, Q)$,

$$\Pi(q_i, Q) = bQ - \frac{1}{2}cq_i^2, \quad (1)$$

where q_i is the amount of abatement undertaken by region i , and $Q = \sum_i q_i$ is the global amount of abatement. Parameter $b \geq 0$ denotes the constant marginal benefit of emission abatement for an individual region. Hence, we can interpret b as the regional social cost of carbon, and bN as the global social cost of carbon, accordingly. Parameter $c \geq 0$ shapes the quadratic abatement cost function. The model is solved by backward induction, solving the abatement stage first. By assumption, trade sanctions don't affect abatement costs, thus results on optimal abatement are valid for all of our three policy regimes.

First, we determine the outsiders' abatement decisions. Each outsider country chooses q_{out} to maximize (1), so that

$$q_{out}^* = \frac{b}{c}. \quad (2)$$

It follows that the abatement decision of outsider countries is a dominant strategy and does not depend on other countries' decisions.⁹ To determine the member countries' abatement decision q_{coa}^* we consider the joint payoff for a coalition with k members, which is given by

$$k\Pi(q_{coa}, Q) = kbQ - k\frac{1}{2}cq_{coa}^2. \quad (3)$$

Note that at this stage of the game the number of coalition members k is taken as given and thus not considered as a decision variable. Maximization of (3) determines the amount of abatement

⁹Note that the dominant strategies of outsiders imply that the model results also hold for a sequential Stackelberg version of the game (as in e.g. Barrett, 1994; Diamantoudi and Sartzetakis, 2006; Rubio and Ulph, 2006) with the coalition deciding first and the outsiders behaving as Stackelberg-followers.

q_{coa}^* that is undertaken by any member country of the agreement as

$$q_{coa}^* = k \frac{b}{c}. \quad (4)$$

Comparing (2) and (4) shows that members of the coalition abate more emissions than outsiders do. A member country's emissions abatement depends on the size of the coalition but is independent of the outsiders' abatement decisions. Effectively, each member abates up to the point where marginal abatement cost equal joint marginal benefits in the coalition.¹⁰

Now we turn to the membership stage. Let $\Pi_{coa}(k)$ denote the payoff of an individual member in a coalition of k and $\Pi_{out}(k)$ the payoff for a outsider. Formally, the stability conditions for a coalition of k then read

$$\Pi_{coa}(k) \geq \Pi_{out}(k-1) \text{ for internal stability, and}$$

$$\Pi_{out}(k) > \Pi_{coa}(k+1) \text{ for external stability.}$$

Following Hoel and Schneider (1997) we define the stability function Φ in general as

$$\Phi(k) := \Pi_{coa}(k) - \Pi_{out}(k-1). \quad (5)$$

Thus, a coalition of size k is internally and externally stable if $\Phi(k) \geq 0$ and $\Phi(k+1) < 0$. In the following, we use superscripts to distinguish between the three policy regimes *NTRF*, *UTRF*, and *RTRF*.

4.2 Policy regime *NTRF*: Agreement stability without trade sanctions

In case of our functional specification, using (2) and (4), it is straightforward to see that for all coalitions $k > 2$ the stability function

$$\Phi^{NTRF}(k) = \Pi_{coa}^{NTRF}(k) - \Pi_{out}^{NTRF}(k-1) \quad (6)$$

is monotonically decreasing and thus an internally and externally stable coalition is characterized by the largest integer k that satisfies $\Phi \geq 0$, which is in our case $k = 3$. Note that this result hinges on the assumption of linear benefits of abatement (as in e.g. Fuentes-Albero and Rubio, 2010; Pavlova and de Zeeuw, 2013). Other specifications of the model, as in Barrett (1994), also

¹⁰We mirror this in our numerical analysis, where coalition members set a carbon tax equal to the sum of regional SCCs in the coalition, see section 5.

find the possibility for larger coalitions if benefits from cooperation are small. This effect is known as the paradox of cooperation. As we focus our analytical model on the effects of trade sanctions on the stability of IEAs we employ the simple linear-quadratic specification instead of emphasizing a more complex payoff structure.

4.3 Policy regime *UTRF*: Trade sanctions as a means for stability

Now assume that members of the coalition impose unilateral trade sanctions in the form of import tariffs at rate $\theta \geq 0$ on the outsiders (regime *UTRF*). We assume throughout this section that the imposed tariff rate is below the optimal tariff rate. In line with our considerations from section 3, each member will enjoy a welfare gain from the tariff $\beta(\theta, N - k) \geq 0$ while outsiders face a cost $\zeta(\theta, k) \geq 0$ and β and ζ are monotonically increasing in θ (up to the optimal tariff rate). The magnitude of gains and costs for a specific tariff rate depend on the coalition size. We can assume that for an individual coalition member, the welfare gain will be greater in smaller coalitions, as more import flows are taxed at the border. Likewise, losses for outsiders will increase in the coalition size.

Thus, the stability function becomes

$$\Phi^{UTRF}(k, \theta) = \Pi_{coa}^{NTRF}(k) + \beta(\theta, N - k) - \Pi_{out}^{NTRF}(k - 1) + \zeta(\theta, k - 1). \quad (7)$$

What can we say about moving from *NTRF* to *UTRF*, i.e. what is the effect of an imposition of unilateral trade sanctions on coalition stability? For a clear-cut comparison between *NTRF* and *UTRF*, we take a look at the net stability function, defined as:

$$\Delta\Phi^{UN}(\theta, k) := \Phi^{UTRF}(\theta, k) - \Phi^{NTRF}(k) = \beta(\theta, N - k) + \zeta(\theta, k - 1). \quad (8)$$

We can interpret the net stability function as the change in the basic incentive structure when moving from one regime to the other: positive values of the net stability function $\Delta\Phi^{UN}$ indicate that under the *UTRF* regime it is more attractive to be a member of a coalition of size k than under *NTRF*, thus the coalition would be stabilized through the regime shift; likewise, negative values indicate that the respective coalition would be destabilized.

The net stability function $\Delta\Phi^{UN}$ is unambiguously positive, which indicates that for every coalition size k the incentive to be member of the coalition is higher under *UTRF* than under *NTRF* for any tariff rate θ .¹¹

¹¹Recall that we are assuming tariffs rates below the optimal tariff, i.e. sufficiently low to imply welfare gains for the imposing region.

4.4 Policy regime *RTRF*: Agreement stability with trade sanctions and retaliation

The picture changes if outsiders react with retaliatory policies (regime *RTRF*). Assume that outsiders raise retaliatory trade sanctions at the same rate as the sanctioning tariff θ . These in turn benefit the outsiders who gain $\beta(\theta, k)$. The members of the coalition who are now targeted by trade sanctions themselves lose $\zeta(\theta, N - k)$. The loss for coalition members thus increases with the level of retaliatory trade sanctions as well as with the number of outsiders that apply those sanctions. Including trade sanctions and retaliation, the stability function now becomes

$$\Phi^{RTRF}(k, \theta) = \Pi_{coa}^{NTRF}(k) + \beta(\theta, N - k) - \zeta(\theta, N - k) - \Pi_{out}^{NTRF}(k - 1) + \zeta(\theta, k - 1) - \beta(\theta, k - 1). \quad (9)$$

It is apparent that retaliation has a destabilizing effect compared to unilateral trade sanctions only (*UTRF*), as the net stability function

$$\Delta\Phi^{RU} = \Phi^{RTRF} - \Phi^{UTRF} = -\beta(\theta, k - 1) - \zeta(\theta, N - k) \quad (10)$$

is unambiguously negative. The effect is the largest for small coalitions and decreases with the coalition size.

The more relevant question is: What can we say about a change from *NTRF* to *RTRF*? For simplicity, we make the assumption that a tariff-imposing region receives the same benefit $\beta(\theta)$ from each region it taxes, i.e. $\beta(\theta, N - k) = (N - k)\beta(\theta)$. A similar assumption about the loss of targeted countries implies $\zeta(\theta, k) = k\zeta(\theta)$. With these additional assumption, we can formulate our key analytical finding.

Proposition 1. *The effect of trade sanctions and retaliation on coalition stability compared to a no-tariff regime depends on the coalition size. Coalitions below a threshold size of $\bar{k} = (N + 1)/2$ are destabilized. For coalitions larger than size $\bar{k}$ the stabilizing effect of trade sanctions dominates.*

Proof. Appendix A. □

The stylized analytical representation of an IEA shows the main effects of the introduction of trade sanctions and the possibility of retaliation by outsiders on the agreement stability compared to the no-tariff regime. Trade sanctions serve as a means to stabilize coalitions if retaliation is prohibited. The stabilizing effect increases with the coalition size. However, retaliation has a destabilizing effect which is decreasing in the coalition size. The total effect of both sanctions and retaliation depends on the coalition size: for larger coalitions, the stabilizing effect of trade

sanctions dominates, while for small coalitions the destabilizing effect of retaliation dominates. These opposed effects are to be quantified in a more detailed modelling approach. Using a CGE model allows us to endogenize abatement costs as well as welfare effects triggered by trade measures and to quantify the effects of trade sanctions and retaliation.

5 Numerical model, data, and scenarios

5.1 Numerical model

For our quantitative assessment, we use a standard static multi-region, multi-sector CGE model of global trade and energy that was developed for numerical analyses on border carbon adjustments in sub-global climate policies. The CGE framework is particularly well suited to analyze quantitative implications of trade measures on coalition stability, as global and regional welfare implications of emission abatement and trade policies are fully endogenized. In this section, we provide a brief non-technical summary of the model. A detailed description including an algebraic formulation can be found in Böhringer et al. (2015).

Primary factors of production comprise labor, capital, and fossil resources. Labor and capital are assumed to be mobile across sectors within each region but not internationally mobile. Fossil resources (gas, crude oil, and coal) are sector-specific capital in fossil fuel production. Factor markets are perfectly competitive.

Final consumption in each region is represented through a representative agent who receives income from primary factors and maximizes welfare subject to a budget constraint and constant-elasticity-of-substitution (CES) utility.

The production of goods other than fossil resources is represented through a standard nested CES function, where at the top level a composite of value added, energy and material intermediate inputs trades off with a transport composite of international transport services. In fossil resource production, the specific resource factor trades off with a Leontief composite of all other inputs at a constant elasticity of substitution. Output of each production sector is allocated either to the domestic market or the export market according to a constant-elasticity-of-transformation function.

Government and investment demand are fixed at real benchmark levels. Investment is paid by savings of the representative agent while taxes pay for the provision of public goods and services.

International trade is modeled following Armington’s differentiated goods approach, where goods are distinguished by origin (Armington, 1969). The Armington composite for a traded good is a CES function of domestic production for that sector and an imported composite. The

import composite, in turn, is a CES function of production from all other countries. A balance of payment constraint incorporates the base-year trade deficit or surplus for each region.

CO₂ emissions are linked in fixed proportions to the use of fossil fuels, with region- and sector-specific CO₂ coefficients. Restrictions to the use of CO₂ emissions in production and consumption are implemented through exogenous CO₂ taxes. CO₂ emission abatement then takes place by fuel switching or energy savings – either by fuel-non-fuel substitution or by a scale reduction of production and final demand activities.

5.2 Data

For the calibration of model parameters, we use the latest version of the database from the Global Trade Analysis Project (GTAP version 9) with base-year 2011 (Aguiar et al., 2016). The GTAP database provides multi-region, multi-sector input-output tables, international trade flows on the sectoral level, sector- and fuel specific CO₂ data, as well as substitution elasticities for production and trade for 140 regions and 57 sectors. We aggregate the database according to our research questions, as summarized in Table 1. On the regional level, we follow Nordhaus (2015) and aggregate to the 15 major economic world regions including the USA, Europe, Japan, Russia, and China. On the sectoral level, we represent the most important sectors when it comes to the combination of carbon and trade policy: we individually represent the primary and secondary energy goods coal, crude oil, natural gas, and electricity. Additionally, we include aggregated sectors for energy- and trade- intensive industries, transport industries, and all other goods, respectively.

We calibrate the model to base-year input-output data, i.e. we determine the parameters such that the economic flows represented in the data are consistent with the optimizing behavior of the economic agents. The responses of agents to price changes are determined by a set of exogenous elasticities taken from the econometric literature. Elasticities in international trade (Armington elasticities) and substitution possibilities in production (between primary factor inputs) are directly provided by the GTAP 9 database. In fossil fuel production, elasticities of substitution between the resource and all other inputs are calibrated to match exogenous estimates of fossil-fuel supply elasticities (Graham et al., 1999; Krichene, 2002; Ringlund et al., 2008).

5.3 Scenarios

One particular scenario in our simulations is composed of assumptions along four dimensions: the global social cost of carbon (SCC), the climate coalition, the policy regime, and the tariff rate, as summarized in Table 2.

Table 1: Model sectors and regions

Sectors and commodities	Countries and regions
<i>Energy sectors</i>	United States of America
Coal	Europe
Crude oil	China
Natural gas	India
Refined oil products	Russia
Electricity	Japan
<i>Aggregated sectors</i>	Canada
EITE*	South Africa
Transport	Brasil
All other goods	Mideast and North Africa
	Eurasia
	Latin America
	Tropical Africa
	Middle-income Asia
	Rest of the World

* EITE – energy-intensive and trade-exposed sectors: chemical products; non-metallic minerals; iron and steel industry; non-ferrous metals; paper, pulp, and print.

5.3.1 Global social cost of carbon

We include four different assumptions about the global SCC – 12.5, 25, 50, and 100 USD per ton of CO₂.¹² These assumptions are in line with Nordhaus (2015) and with recent recommendations by the High-Level Commission on Carbon Prices chaired inter alia by Joseph Stiglitz and Lord Nicholas Stern (CPLC, 2017). We denote the assumptions with SCC-12.5, SCC-25, SCC-50, and SCC-100.

5.3.2 Coalitions

Our analysis comprises 15 regions, which implies 32767 possible coalitions. Each region has an assigned regional SCC, which is calculated as its base-year share of global GDP times the global SCC. In the same way as laid out in the analytical model in section 4, coalitions are specified as follows: within a coalition, individual members take the sum of their regional SCC’s into account, while outsiders act solely according to their own regional SCC. Effectively, coalition members apply a tax on CO₂ emissions equal to the sum of regional SCC’s of members, and outsiders apply a CO₂ tax equal to their own regional SCC. In the non-cooperative solution (equivalently: empty coalition) each region applies a CO₂ tax equal to their regional SCC.

¹²For a comprehensive overview of simulations and applications of the concept of the social cost of carbon, see Metcalf and Stock (2017).

5.3.3 Policy regimes and tariff rates

Our policy regimes in the numerical analysis exactly refer to those laid out in section 3: In the first regime *NTRF*, there are no trade sanctions, i.e. regions are solely applying CO₂ taxes, where the tax level depends on whether or not they are part of the coalition. In the second regime *UTRF*, we introduce trade sanctions in the form of import tariffs imposed by the climate coalition against outsiders. The import tariffs apply uniformly over all sectors of the economy. We include ad-valorem tariff rates from 1% to 10%. This range is in line with the political discussion, Nordhaus’ analysis and – as the presentation in the next section reveals – covers the spectrum of results conceivable from our analytical analysis.¹³ The third regime *RTRF* comprises trade sanctions as in *UTRF* and additionally retaliation by outsiders. In the case of retaliation (*RTRF*), our convention is that outsiders retaliate by applying the same uniform import tariff rate on imports from members that members apply for their part.

Table 2: Overview of simulation scenarios

Dimension	Description and specification
Global social cost of carbon	We include four different assumptions about the global social cost of carbon: SCC-12.5, SCC-25, SCC-50, SCC-100
Coalition	We include all possible 32767 coalitions
Policy regime	We include three different regimes: <i>NTRF</i> , <i>UTRF</i> , <i>RTRF</i>
Tariff rate	Under policy regimes <i>UTRF</i> and <i>RTRF</i> , we include ad-valorem import tariff rates of 1% to 10%

5.3.4 Business-as-usual

The non-cooperative solution serves as our business-as-usual for the assessment of welfare changes in the different scenarios.¹⁴ Obviously, we have a different business-as-usual for each assumption about the global SCC. Table 3 gives an overview of our business-as-usuals under the four different assumptions about the global SCC. The first column shows the GDP shares in our base-year data, which together with the global SCC define the regional SCC, and thus the regional CO₂ taxes in the business-as-usual. Consequently, the regional CO₂ taxes add up to the exogenous global SCC (row “Total”). Additionally, we report the global average CO₂ price, which is the emission-weighted sum of regional CO₂ taxes. We see that in the non-cooperative solution, the global average CO₂

¹³An interesting extension would be a setting in which regions apply the welfare maximizing – i.e. optimal – tariff rate as e.g. in (Böhringer and Rutherford, 2017). For our analysis, however, the simpler assumption of fixed tariff rates is more appropriate as it relates to the political discussion, the analytical part and to former analysis by Nordhaus (2015).

¹⁴We calculate welfare as Hicksian equivalent variation (HEV): the amount of USD that has to be added to the representatives agents’ business-as-usual income such that she enjoys the same utility level as in the counterfactual. In our case – as real government demand and real investment demand are fixed – regional HEV is the sum of the change in real private consumption and the change in global emissions valued at the regional SCC.

price is only slightly above one tenth of the assumed global SCC.

Table 3: Business-as-usual regional GDP shares and CO₂ taxes under different assumptions about the global social cost of carbon

	GDP share	SCC-12.5 CO ₂ tax	SCC-25 CO ₂ tax	SCC-50 CO ₂ tax	SCC-100 CO ₂ tax
Europe	26.4	3.30	6.60	13.21	26.41
United States of America	21.7	2.72	5.43	10.87	21.73
China	10.6	1.32	2.65	5.29	10.59
Japan	8.3	1.03	2.07	4.13	8.26
Mideast and North Africa	5.6	0.70	1.40	2.79	5.58
Latin America	4.8	0.61	1.21	2.42	4.85
Middle-income Asia	3.7	0.47	0.93	1.86	3.72
Brasil	3.5	0.43	0.87	1.73	3.47
Eurasia	3.3	0.42	0.83	1.66	3.32
Russia	2.7	0.33	0.67	1.33	2.67
India	2.6	0.33	0.66	1.32	2.63
Canada	2.5	0.31	0.62	1.24	2.49
Rest of the World	2.2	0.28	0.56	1.12	2.24
Tropical Africa	1.5	0.19	0.37	0.74	1.48
South Africa	0.6	0.07	0.14	0.28	0.57
Total	100	12.5	25	50	100
Global average		1.48	2.95	5.86	11.60

Note.— GDP share is given in % of world GDP; CO₂ tax is given in USD per ton of CO₂; Global average – emission-weighted average of regional CO₂ taxes.

Under each assumption about the global SCC, the policy regime, and the tariff rate, we calculate regional welfare changes for each of the 15 regions under all possible 32767 coalitions. In the next step, we check each coalition for internal and external stability, i.e. we compare for each member region whether it would be better off by leaving the coalition and for each outsiders region whether it would be better off by joining the coalition.

6 Results

In this section, again we discuss the results of our numerical analysis and relate them to insights from the analytical model in section 4. First we provide an insight to the basic incentive structure under the three considered policy regimes. In the next part, we analyze findings on stable coalitions, before we take a closer look at how the incentive structure changes when moving from *NTRF* to *RTRF*. Finally, we discuss implications for global welfare and emission levels in stable coalitions under the different policy regimes.

6.1 Basic incentive structures

Before we present quantitative results on stable coalitions under the three policy regimes *NTRF*, *UTRF*, and *RTRF* in detail, we present the basic incentive structure with regards to cooperation. Figure 1 shows indicators for these incentives for the four major model regions in terms of GDP – the United States (US), the European Union (EU), China (CHN), and Japan (JPN): the “Benefit of in” and the “Cost of out” at an assumed global social cost of carbon of 25 USD and a tariff rate of 2%. The “Benefit of in” is the welfare change of a region if it forms a coalition of 1. The “Cost of out” is the welfare change of a region if it leaves the grand coalition.¹⁵

In the absence of trade sanctions (regime *NTRF*), the “Benefit of in” is zero in all regions, as a coalition of 1 is technically identical to the non-cooperative equilibrium. The EU is the only region with a positive “Cost of out”, indicating they would even prefer to stay in the grand coalition. While the US and Japan would slightly benefit from leaving the grand coalition (in the form of negative cost), China has by far the strongest incentive to leave the grand coalition. They would save almost 20 billion USD when going from the grand coalition to the coalition of 14 without China. The underlying reason is the Armington structure of the model. Due to the differentiation of goods by origin, regions are able to pass costs due to carbon pricing through to trading partners. In that respect, carbon pricing can to some extent be a substitute for optimal tariffs and change the terms of trade in favor of the taxing region. Obviously, substitution possibilities for European goods are rather limited, thus Europe even prefers to remain in the grand coalition and apply higher carbon prices.¹⁶

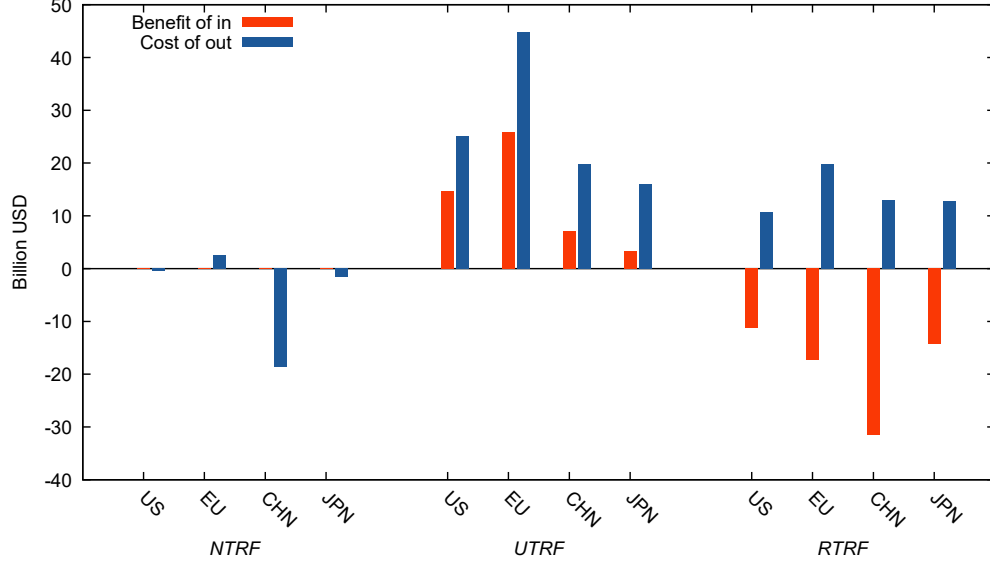
In case of unilateral tariffs by the coalition (regime *UTRF*), we clearly see the unambiguously positive effect on individual regions incentives. All regions would now prefer to form a coalition of 1, as they would enjoy welfare gains from taxing all there incoming imports. On the other hand, all regions would suffer dramatically under leaving the grand coalition and thus be subjected to tariffs by all 14 remaing coalition members.

Now we turn to the comparison between the unilateral tariff regime (*UTRF*) and the retaliation regime (*RTRF*), i.e., incentives that may be misrepresented when retaliation is precluded by assumption. Under *RTRF*, we find that the “Cost of out” is smaller compared to *UTRF*, yet still the represented regions clearly prefer to maintain the grand coalition. The incentive to starting coalitions, however, that was quite large under *UTRF*, has turned around and all the regions face substantial cost of between roughly 10 billion USD (US) and 30 billion USD (China) when forming

¹⁵These indicators where introduced in Nordhaus (2015).

¹⁶The possibility to pass through increased production costs might be exaggerated in the Armington structure compared to trade models relying on Melitz’ formulation of firm heterogeneity (Melitz, 2003). For the welfare assessment of tariffs at low rates, however, models employing the Armington structure work well. For comparisons of trade formulations see (Balistreri and Rutherford, 2012; Balistreri and Markusen, 2009).

Figure 1: “Benefit of in” and “Cost of out” at a global social cost of carbon of 25 USD and a tariff rate of 2% under *NTRF*, *UTRF*, and *RTRF*



Note.— US – United States; EU – Europe; CHN – China; JPN – Japan; Benefit of in – welfare change of a region if it forms a coalition of 1; Cost of out – welfare change of a region if it leaves the grand coalition.

a coalition of 1. In particular, this shows that the empty coalition is stable now.

This finding is exactly in line with our conclusion from the analytical part: Under a regime with unilateral trade sanctions, coalitions are stabilized unambiguously. But under a regime with trade sanctions and retaliation, larger coalitions tend to be stabilized, while smaller coalitions tend to be destabilized compared to regimes without trade sanctions.

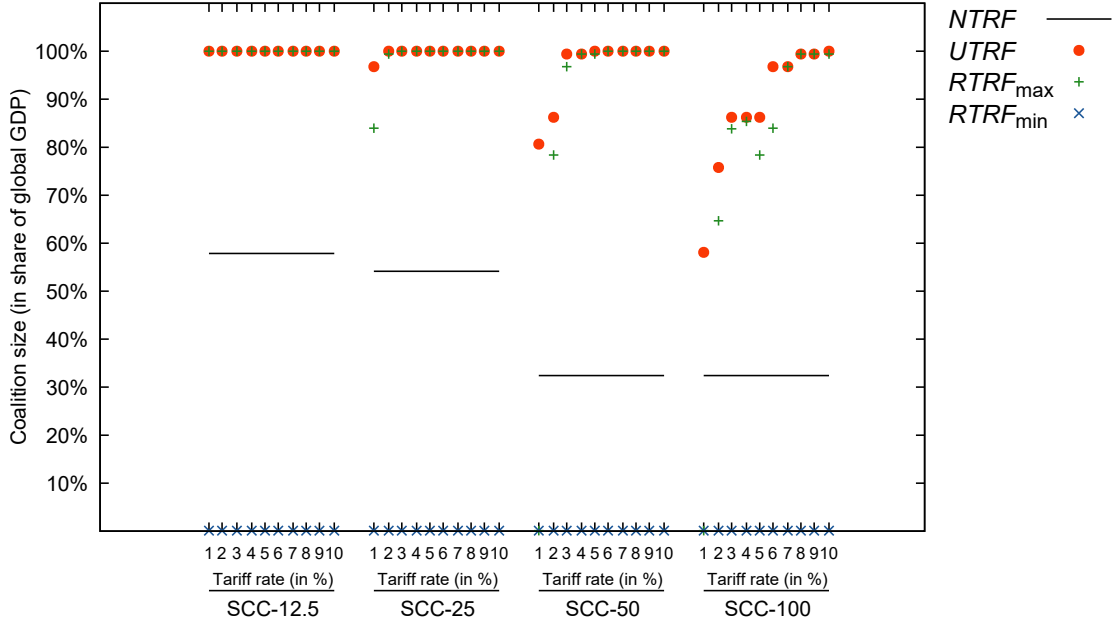
6.2 Stable coalitions

We now turn to stable coalition structures that result from the different policy regimes. Figure 2 shows the size of the stable coalitions in percentage share of world GDP in the policy regimes *NTRF*, *UTRF*, and *RTRF* for different assumptions about the global social cost of carbon and the different considered tariff rates from 1% up to 10% for the regimes *UTRF* and *RTRF*.

6.2.1 Policy regime *NTRF*: Stable coalitions in the absence of tariffs

In the absence of tariffs our simulations quantify the basic results from our analytical model in section 4. In contrast to the expository simple analytical model with symmetric regions the costs of abatement are fully endogenized in a model calibrated to real world data. Figure 2 shows the size of the stable coalition in terms of percentage of world GDP for the different scenarios of global social costs of carbon (SCC) in the policy regime without trade sanctions (*NTRF*). It indicates

Figure 2: Percentage share of world GDP of stable coalitions in the policy regimes *NTRF*, *UTRF*, and *RTRF* for different assumptions about the global social cost of carbon



Note.— $RTRF_{\max}$ refers to the largest stable coalition under *RTRF*; $RTRF_{\min}$ refers to the smallest stable coalition under *RTRF*.

that with rather low global SCC of 12.5 USD per ton of CO₂ a stable coalition covering 58% of world GDP can be maintained.¹⁷ With increasing social costs of carbon the size of the stable coalition decreases. For a global social cost of carbon of 50 USD only a coalition covering 32% of world GDP is stable. This coalition remains stable for a global social cost of carbon of 100 USD.

6.2.2 Policy regime *UTRF*: Stable coalitions with uniform tariffs

Introducing trade sanctions against outsiders of the coalition imposes a strong incentive to join. We find this effect, which has been extensively discussed by Nordhaus (2015), in the *UTRF* regime. Figure 2 shows the size of the stable coalition for the different tariff-scenarios from 1% up to 10% for the different assumptions about the global social cost of carbon. For low global social costs of carbon of 12.5 USD the threat of trade sanctions of 1% is already high enough to stabilize the grand coalition. Of course, if the grand coalition is stabilized by means of sanctions no region is actually subjected to the tariffs. All regions are members of the coalition while the trade sanctions serve as a threat for regions in case of leaving the coalition. With higher global social costs of carbon,

¹⁷We find a rather large stable coalition of 58% of world GDP under *NTRF*. This is due to the assumed Armington structure of international trade (see footnote 16). Furthermore, our discussion on costs and emissions in subsection 6.3 reveals that while the coalition seems quite large, effectively it only achieves an emission reduction of 1% compared to the business-as-usual (see Table 4).

the tariff rate which is necessary to stabilize the grand coalition increases so that in case of social costs of carbon of 100 USD only a substantial tariff of 10% could stabilize the grand coalition.

6.2.3 Policy regime *RTRF*: Stable coalitions with tariffs and retaliation

Allowing outsiders of the coalition to retaliate with the same tariff rate markedly changes the results (policy regime *RTRF*). Figure 2 shows the sizes of stable coalitions in percentage of world GDP for these scenarios. While the grand coalition can still be stabilized by sufficiently high trade sanctions, retaliation reduces the threat of being targeted. As a consequence, in case of social costs of carbon above 12.5 USD the size of stable coalitions that can be maintained with low trade sanctions decreases in comparison to the scenarios without retaliation. For social costs of carbon above 25 USD the effect of retaliation is sufficient to destabilize all possible coalitions for a tariff rate of 1%. More strikingly, in all scenarios with retaliation the non-cooperative outcome is an equilibrium: the non-cooperative solution. In the same way the threat of trade sanctions against outsiders can prevent members from leaving the coalition, retaliatory trade measures can prevent outsiders regions from joining small coalitions that impose trade sanctions on outsiders.

6.2.4 A closer look

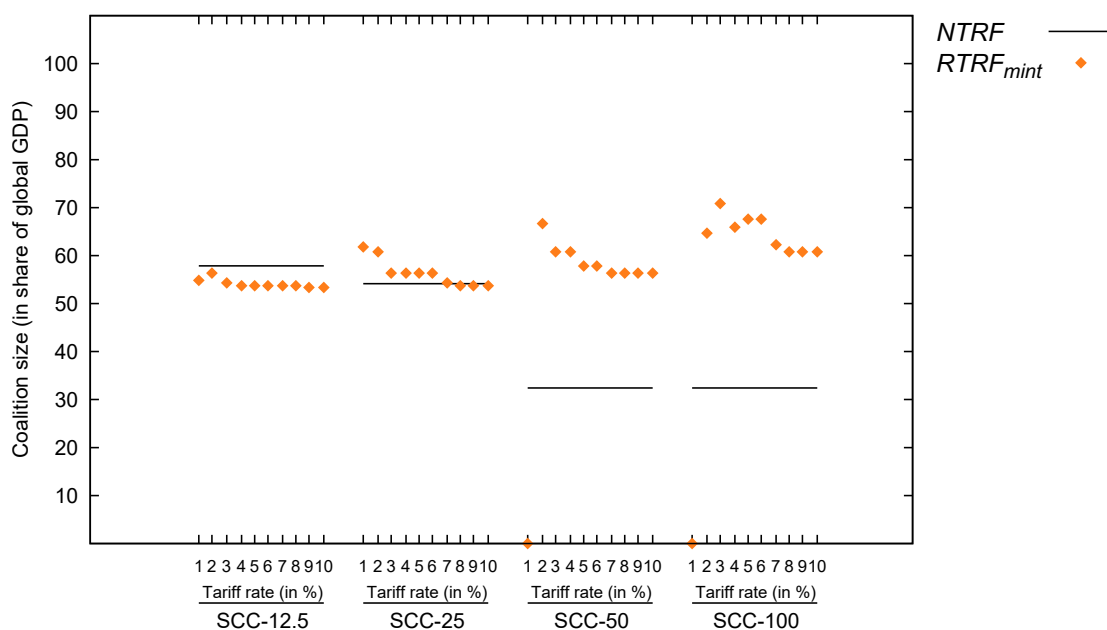
The results for the scenarios with retaliation call for a closer look. The existence of two stable equilibria for almost all of these scenarios raises the question which of the two equilibria might be reached. Although our analysis does not explicitly consider the dynamics of coalition formation, our results deliver some insights regarding this question. Our numerical results mirror our insight from Proposition 1.

The threat of being targeted by trade sanctions of coalition members and thus the stabilizing effect of tariffs increases with the size of the coalition as the volume of traded goods that is covered by tariffs increases for outsiders. By the same argument small coalitions are destabilized by retaliatory trade measures of outsiders: in small coalitions members have a high volume of trade with outsiders which makes them prone to retaliation. Retaliatory trade measures then have the potential to incentivize coalition members to leave the coalition to avoid retaliation. As a consequence, small coalitions are internally instable (and externally stable) which means that further regions want to leave the coalition while big coalitions are externally instable (and internally stable) so that further regions want to join the coalition. On the one hand, this means that if an already larger coalition exists, the introduction of trade sanctions could lead to an even bigger and possibly grand coalition, even in the presence of retaliation. On the other hand, starting from a small coalition the introduction of trade sanctions diminishes cooperation and the non-cooperative

equilibrium results. In other words: the introduction of both trade sanctions and retaliatory trade measures induce a threshold effect in the coalitional game. Below a certain coalition size the destabilizing effect of retaliation predominates and leads to the non-cooperative equilibrium whereas in coalitions above a certain size the stabilizing effect of trade sanctions enables the formation of even bigger stable coalitions.

Figure 3 shows the size of the smallest internally stable coalition (other than the empty coalition) in the scenarios with trade sanctions and retaliation (*RTRF*) and the size of the stable coalition in the absence sanctions (*NTRF*) in percentage of world GDP. These results show that the introduction of trade sanctions and retaliation with an already existing coalition that is stable without sanctions would lead to the non-cooperative equilibrium in most of the scenarios. Only for low global social costs of carbon of 12.5 USD per ton of CO₂ and for global social costs of carbon of 25 USD per ton of CO₂ with tariffs higher than 7% the introduction of trade sanctions with retaliation could help to stabilize a growing coalition that could ultimately become the grand coalition.

Figure 3: Percentage share of world GDP of stable coalitions in the policy regime *NTRF* and of the smallest internally stable coalitions under *RTRF* for different assumptions about the global social cost of carbon



Note.— *RTRF*_{mint} refers to the smallest internally stable coalition under *RTRF* other than the empty coalition.

6.3 Costs and emissions in stable coalitions

We report results on global costs and benefits, global emission changes, and global average CO₂ prices in all stable coalitions across our three policy regimes and four assumptions about the global social cost of carbon (SCC). Table 4 summarizes the results under *NTRF* and Table 5 shows the results for the regimes with uniform trade sanctions (*UTRF*) and with retaliation (*RTRF*).¹⁸ Global welfare is composed of two factors: (i) economic adjustment costs due to the CO₂ tax, i.e. the change in real consumption (item Cost CN in the tables); (ii) environmental benefits due to the change in global emissions, i.e. the global emission reduction times the respective global social cost of carbon (Benefit EM). Global welfare is the difference of environmental benefits and economic costs (Hicksian EV).¹⁹ Cost and welfare items are reported as percentage share of business-as-usual GDP. The global average CO₂ price is given in USD per ton of CO₂, and global emissions are given as percentage change from the business-as-usual.²⁰

Under *NTRF* (Table 4), we find that cooperation reduces the global emission level by 0.6% to 1.4% compared to the respective business-as-usual. The global average CO₂ price is about 30% of the efficient level for SCC-12.5 and decreases to roughly 15% of the efficient level for SCC-100 as the size of the stable coalition is smaller. If we focus on the economic adjustment costs, we find that they are slightly negative under SCC-50. This is again due to the terms of trade.²¹ As we move to SCC-100, where we find the same stable coalition as in SCC-50, economic costs have turned positive. The global welfare gain due to cooperation increases with the assumed SCC.

Table 4: Global cost and emission impacts in stable coalitions under *NTRF*

	SCC-12.5	SCC-25	SCC-50	SCC-100
Cost CN	0.005	0.010	-0.001	0.003
Benefit EM	0.005	0.013	0.011	0.039
Welfare	0.000	0.003	0.012	0.036
CO ₂ Price	3.86	6.42	7.61	15.08
Emissions	-1.0	-1.4	-0.6	-1.0

Note.— All costs are given as a percentage share of business-as-usual GDP; Cost CN – change in real consumption; Benefit EM – change in global emissions times the global social cost of carbon; Welfare – Hicksian equivalent variation (difference of Benefit EM and Cost CN); CO₂ Price – global average price on CO₂ emissions in USD/tCO₂; Emissions – percentage change in global emissions.

Now we turn to the comparison of effects under *UTRF* and *RTRF*. We find that in each constellation where the grand coalition is established, emission reductions are substantial compared to cooperative solutions under *NTRF*: it ranges from 6.8% for SCC-12.5 to 24.9% for SCC-100. In

¹⁸In the case of *RTRF*, we report outcomes for the largest stable coalitions.

¹⁹We take a utilitarian (Benthamite) perspective on global welfare accounting where welfare changes of individual regions are perfectly substitutable.

²⁰Be aware that each assumption about the global social cost of carbon constitutes a different business-as-usual.

²¹Recall the discussion of Figure 1.

these cases, global welfare improves compared to the business-as-usual, except for SCC-12.5. The reason is initial tax distortions in the business-as-usual: we set the CO₂ at the Pigouvian rate irrespective of interactions with pre-existing taxes. For lower assumptions about the externality (SCC), the externality is relatively small compared to initial tax distortions.

Let's turn to the cases where the grand coalition is not achieved. For assumptions about the global SCC of below 100 USD/t CO₂, *RTRF* unambiguously entails lower global emission reductions and lower global welfare levels than *UTRF*. Under SCC-100, however, the picture is mixed. For the cases with a tariff rate higher than 6%, where a very high level of cooperation is achieved, global welfare implications are rather similar. For tariff rates of 6% and lower, the comparison of welfare impacts between *UTRF* and *RTRF* hinges on the exact composition of the stable coalition. The coalition is larger under *UTRF* throughout these cases, which translates into a higher CO₂ tax inside the coalition. From a global efficiency perspective, this drives a larger wedge between marginal abatement cost inside and outside the coalition. Under *RTRF*, the CO₂ tax is lower inside the coalition, but additional trade distortions lead to more emission reductions outside the coalition. As a manifestation of this effect, the global average price for CO₂ is higher for a tariff rate of 4% under *RTRF* compared to *UTRF*, although the coalition is smaller. The overall effect plays out positive under *RTRF* for tariff rates between 2% and 4%, where the global welfare impact is greater than under *UTRF*.

Table 5: Global cost and emission impacts in stable coalitions under *UTRF* and *RTRF*

SCC-12.5										
	Trf 1%	Trf 2%	Trf 3%	Trf 4%	Trf 5%	Trf 6%	Trf 7%	Trf 8%	Trf 9%	Trf 10%
<i>UTRF</i>										
Cost CN	0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.041
Benefit EM	0.034	0.034	0.034	0.034	0.034	0.034	0.034	0.034	0.034	0.034
Welfare	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007
CO ₂ Price	12.50	12.50	12.50	12.50	12.50	12.50	12.50	12.50	12.50	12.50
Emissions	-6.8	-6.8	-6.8	-6.8	-6.8	-6.8	-6.8	-6.8	-6.8	-6.8
<i>RTRF</i>										
Cost CN	0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.041
Benefit EM	0.034	0.034	0.034	0.034	0.034	0.034	0.034	0.034	0.034	0.034
Welfare	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007	-0.007
CO ₂ Price	12.50	12.50	12.50	12.50	12.50	12.50	12.50	12.50	12.50	12.50
Emissions	-6.8	-6.8	-6.8	-6.8	-6.8	-6.8	-6.8	-6.8	-6.8	-6.8
SCC-25										
	Trf 1%	Trf 2%	Trf 3%	Trf 4%	Trf 5%	Trf 6%	Trf 7%	Trf 8%	Trf 9%	Trf 10%
<i>UTRF</i>										
Cost CN	0.091	0.103	0.103	0.103	0.103	0.103	0.103	0.103	0.103	0.103
Benefit EM	0.092	0.113	0.113	0.113	0.113	0.113	0.113	0.113	0.113	0.113
Welfare	0.001	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010
CO ₂ Price	22.20	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00
Emissions	-9.5	-11.6	-11.6	-11.6	-11.6	-11.6	-11.6	-11.6	-11.6	-11.6
<i>RTRF</i>										
Cost CN	0.060	0.101	0.103	0.103	0.103	0.103	0.103	0.103	0.103	0.103
Benefit EM	0.039	0.108	0.113	0.113	0.113	0.113	0.113	0.113	0.113	0.113
Welfare	-0.022	0.007	0.010	0.010	0.010	0.010	0.010	0.010	0.010	0.010
CO ₂ Price	14.36	24.50	25.00	25.00	25.00	25.00	25.00	25.00	25.00	25.00
Emissions	-4.0	-11.1	-11.6	-11.6	-11.6	-11.6	-11.6	-11.6	-11.6	-11.6
SCC-50										
	Trf 1%	Trf 2%	Trf 3%	Trf 4%	Trf 5%	Trf 6%	Trf 7%	Trf 8%	Trf 9%	Trf 10%
<i>UTRF</i>										
Cost CN	0.106	0.135	0.250	0.250	0.256	0.256	0.256	0.256	0.256	0.256
Benefit EM	0.116	0.138	0.332	0.332	0.346	0.346	0.346	0.346	0.346	0.346
Welfare	0.009	0.004	0.082	0.082	0.090	0.090	0.090	0.090	0.090	0.090
CO ₂ Price	24.99	29.74	48.90	48.90	50.00	50.00	50.00	50.00	50.00	50.00
Emissions	-6.0	-7.2	-17.3	-17.3	-18.0	-18.0	-18.0	-18.0	-18.0	-18.0
<i>RTRF</i>										
Cost CN	0	0.130	0.237	0.251	0.252	0.256	0.256	0.256	0.256	0.256
Benefit EM	0	0.111	0.288	0.332	0.332	0.346	0.346	0.346	0.346	0.346
Welfare	0	-0.020	0.052	0.081	0.080	0.090	0.090	0.090	0.090	0.090
CO ₂ Price	5.86	23.77	44.09	48.91	48.91	50.00	50.00	50.00	50.00	50.00
Emissions	0	-5.8	-15.0	-17.3	-17.3	-18.0	-18.0	-18.0	-18.0	-18.0
SCC-100										
	Trf 1%	Trf 2%	Trf 3%	Trf 4%	Trf 5%	Trf 6%	Trf 7%	Trf 8%	Trf 9%	Trf 10%
<i>UTRF</i>										
Cost CN	0.104	0.228	0.328	0.336	0.345	0.547	0.549	0.595	0.596	0.607
Benefit EM	0.172	0.312	0.432	0.434	0.437	0.808	0.809	0.925	0.925	0.962
Welfare	0.068	0.084	0.104	0.098	0.092	0.261	0.259	0.330	0.329	0.355
CO ₂ Price	27.49	43.24	58.51	58.56	58.61	87.09	87.12	97.49	97.49	100.00
Emissions	-4.6	-8.4	-11.6	-11.7	-11.7	-21.7	-21.7	-24.9	-24.9	-25.8
<i>RTRF</i>										
Cost CN	0	0.203	0.362	0.391	0.351	0.408	0.570	0.598	0.599	0.600
Benefit EM	0	0.328	0.504	0.523	0.351	0.423	0.814	0.926	0.926	0.927
Welfare	0	0.126	0.142	0.132	0.001	0.015	0.244	0.328	0.327	0.327
CO ₂ Price	11.60	36.09	57.50	62.27	46.49	55.94	87.25	97.53	97.54	97.55
Emissions	0	-8.8	-13.5	-14.1	-9.4	-11.4	-21.9	-24.9	-24.9	-24.9

Note.— All costs are given as a percentage share of business-as-usual GDP; Cost CN – change in real consumption; Benefit EM – change in global emissions times the global social cost of carbon; Welfare – Hicksian equivalent variation (difference of Benefit EM and Cost CN); CO₂ Price – global average price on CO₂ emissions in USD/tCO₂; Emissions – percentage change in global emissions; Trf – Tariff rate.

7 Conclusion

Former studies of import tariffs as a means to stabilize climate coalitions have concluded that they are an effective mechanism to foster international cooperation. However, most of these studies have relied on the assumption that outsiders are not able to retaliate, i.e. to use trade measures themselves. To close this research gap we use combined analytical and numerical analysis to investigate implications for internal and external stability under three policy regimes: (i) a regime without trade sanctions, (ii) a regime in which coalition members use trade sanctions in the form of uniform import tariffs against outsiders, and (iii) a regime in which coalition members use trade sanctions and outsiders retaliate with uniform import tariffs.

Our analytical model shows that while trade sanctions without the possibility to retaliate might stabilize coalitions, incorporating retaliation entails a “threshold” effect: Below a certain coalition size the destabilizing effect of retaliation predominates and leads to the non-cooperative equilibrium whereas in coalitions above a certain size the stabilizing effect of trade sanctions enables the formation of larger stable coalitions.

For our quantitative assessment we use a standard multi-sector, multi-region CGE model, where regional and global welfare effects due to policy interference is fully endogenized. Our assessment for scenarios with retaliation shows that while it is still possible to maintain larger stable coalitions, also the non-cooperative solution becomes stable. In our simulations, the “threshold” coalition size is well above 50% of world GDP across all assumptions about the global social cost of carbon.

We conclude that the consideration of retaliatory measures substantially decreases prospects for international cooperation through the threat of trade sanctions. Only after a sufficiently large climate coalition has already been formed, the threat of trade sanctions might be an effective stick to establish the grand coalition.

References

- Aguiar, A., B. Narayanan, and R. McDougall. (2016). An Overview of the GTAP 9 Data Base. *Journal of Global Economic Analysis* 1(1), 181–208.
- Balistreri, E. J. and J. R. Markusen (2009). Sub-national differentiation and the role of the firm in optimal international pricing. *Economic Modelling* 26(1), 47–62.
- Balistreri, E. J. and T. F. Rutherford (2012). Subglobal carbon policy and the competitive selection of heterogeneous firms. *Energy Economics* 34, Supple, S190–S197.
- Barrett, S. (1994). Self-Enforcing International Environmental Agreements. *Oxford Economic Papers* 46, 878–894.
- Barrett, S. (1997). The strategy of trade sanctions in international environmental agreements. *Resource and Energy Economics* 19(4), 345–361.
- Barrett, S. (2001). International cooperation for sale. *European Economic Review* 45(10), 1835–1850.
- Barrett, S. (2011). Rethinking Climate Change Governance and Its Relationship to the World Trading System. *The World Economy* 34(11), 1863–1882.
- Bickerdike, C. F. (1906). The Theory of Incipient Taxes. *The Economic Journal* 16(64), 529–535.
- Böhringer, C., E. J. Balistreri, and T. F. Rutherford (2012). The role of border carbon adjustment in unilateral climate policy: Overview of an Energy Modeling Forum study (EMF 29). *Energy Economics* 34(SUPPL.2), 97–110.
- Böhringer, C., J. C. Carbone, and T. F. Rutherford (2016). The Strategic Value of Carbon Tariffs. *American Economic Journal: Economic Policy* 8(1).
- Böhringer, C., A. Müller, and J. Schneider (2015). Carbon Tariffs Revisited. *Journal of the Association of Environmental and Resource Economists* 2(4), 629–672.
- Böhringer, C. and T. F. Rutherford (2017). Paris after Trump: An Inconvenient Insight. *Oldenburg Discussion Papers in Economics*, V – 400-17. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2988845.
- Branger, F. and P. Quirion (2014). Would border carbon adjustments prevent carbon leakage and heavy industry competitiveness losses? Insights from a meta-analysis of recent economic studies. *Ecological Economics* 99, 29–39.

- Carraro, C., C. Marchiori, and S. Orefice (2009). Endogenous Minimum Participation in International Environmental Treaties. *Environmental and Resource Economics* 42(3), 411–425.
- Carraro, C. and D. Siniscalco (1993). Strategies for the international protection of the environment. *Journal of Public Economics* 52(3), 309–328.
- Conconi, P. and C. Perroni (2002, aug). Issue linkage and issue tie-in in multilateral negotiations. *Journal of International Economics* 57(2), 423–447.
- CPLC (2017). Carbon Pricing Leadership Coalition (CPLC). Report of the High-Level Commission on Carbon Prices. Available at: <https://www.carbonpricingleadership.org/report-of-the-highlevel-commission-on-carbon-prices>.
- D’Aspremont, C., A. Jacquemin, J. J. Gabszewicz, and J. A. Weymark (1983). On the Stability of Collusive Price Leadership. *The Canadian Journal of Economics / Revue canadienne d’Econometrie* 16(1), 17–25.
- Denyer, S. (2017). China to Trump: We don’t want a trade war – but if there is one, you’d lose. *Washington Post*, March 15, 2017. Available at: https://www.washingtonpost.com/world/china-tells-trump-we-dont-want-a-trade-war-but-us-firms-would-suffer-most/2017/03/15/4e93c3da-0942-11e7-8884-96e6a6713f4b{}_story.htm.
- Diamantoudi, E. and E. Sartzetakis (2006). Stable International Environmental Agreements: An Analytical Approach. *Journal of Public Economic Theory* 8(2), 247–263.
- Egger, P., C. Jeßberger, and M. Larch (2011). Trade and investment liberalization as determinants of multilateral environmental agreement membership. *International Tax and Public Finance* 18(6), 605–633.
- Eichner, T. and R. Pethig (2013). Self-enforcing environmental agreements and international trade. *Journal of Public Economics* 102, 37–50.
- Eichner, T. and R. Pethig (2014). Self-Enforcing Environmental Agreements, Trade, and Demand- and Supply-Side Mitigation Policy. *Journal of the Association of Environmental and Resource Economists* 1(3), 419–450.
- Felder, S. and T. F. Rutherford (1993). Unilateral CO₂ Reductions and Carbon Leakage: The Consequences of International Trade in Oil and Basic Materials. *Journal of Environmental Economics and Management* 25(2), 162–176.

- Finus, M. (2001). Game theory and international environmental cooperation. *New Horizons in Environmental Economic Series*.
- Fuentes-Albero, C. and S. J. Rubio (2010). Can international environmental cooperation be bought? *European Journal of Operational Research* 202, 255–264.
- Graham, P., S. Thorpe, and L. Hogan (1999). Non-competitive market behaviour in the international coking coal market. *Energy Economics* 21(3), 195–212.
- Hagen, A. and K. Eisenack (2015). International Environmental Agreements with Asymmetric Countries: Climate Clubs vs. Global Cooperation. Technical report.
- Hagen, A., L. Kähler, and K. Eisenack (2017). Transnational environmental agreements with heterogeneous actors. *Economics of International Environmental Agreements: A Critical Approach*.
- Harvey, C. (2016). From France to Canada, countries are reeling from Trump’s climate plans.
- Hoel, M. (1992). International environment conventions: The case of uniform reductions of emissions. *Environmental & Resource Economics* 2(2), 141–159.
- Hoel, M. and K. Schneider (1997). Incentives to participate in an international environmental agreement. *Environmental & Resource Economics* 9(2), 153–170.
- Kemp, L. (2017). Better out than in. *Nature Climate Change* 7, 458–460.
- Krichene, N. (2002). World crude oil and natural gas: a demand and supply model. *Energy economics* 24(6), 557–576.
- Larsen, J. and J. Bradbury (2010). WRI Summary of the Carbon Limits and Energy for America’s Renewal Act.
- Larsen, J., A. Kelly, J. Bradbury, F. Litz, S. Forbes, and N. Bianco (2009). WRI Summary of S. 1733, the Clean Energy Jobs and American Power Act (Kerry-Boxer). *World Resources Institute, Washington, DC* 30.
- Lessmann, K., R. Marschinski, and O. Edenhofer (2009). The effects of tariffs on coalition formation in a dynamic global warming game. *Economic Modelling* 26(3), 641–649.
- Limão, N. (2008). optimal tariffs. In S. N. Durlauf and L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics*. Palgrave Macmillan.

- Marchiori, C., S. Dietz, and A. Tavoni (2017, jan). Domestic politics and the formation of international environmental agreements. *Journal of Environmental Economics and Management* 81, 115–131.
- Marrouch, W. and A. Ray Chaudhuri (2016). International Environmental Agreements: Doomed to Fail or Destined to Succeed? A Review of the Literature. *International Review of Environmental and Resource Economics* 9(3-4), 245–319.
- Melitz, M. J. (2003, nov). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica* 71(6), 1695–1725.
- Metcalf, G. E. and J. H. Stock (2017, jan). Integrated Assessment Models and the Social Cost of Carbon: A Review and Assessment of U.S. Experience. *Review of Environmental Economics and Policy* 11(1), 80–99.
- Neumayer, E. (2002). Does trade openness promote multilateral environmental cooperation? *World Economy* 25(6), 815–832.
- Nordhaus, W. (2015). Climate Clubs: Overcoming Free-riding in International Climate Policy. *The American Economic Review* 105(4).
- Ossa, R. (2014). Trade Wars and Trade Talks with Data. *The American Economic Review* 104(12).
- Pavlova, Y. and A. de Zeeuw (2013). Asymmetries in international environmental agreements. *Environment and Development Economics* 18, 51–68.
- Ringlund, G. B., K. E. Rosendahl, and T. Skjerpen (2008). Does oilrig activity react to oil price changes? An empirical investigation. *Energy Economics* 30(2), 371–396.
- Rubio, S. J. and A. Ulph (2006). Self-enforcing international environmental agreements revisited. *Oxford Economic Papers* 58(2), 233–263.
- Stiglitz, J. E. (2006). A New Agenda for Global Warming. *Economists’ Voice*.
- US Congress (2009). The American Clean Energy and Security Act of 2009 (HR 2454). *US House of Representative, Washington, DC*.
- Voituriez, T. and X. Wang (2011). Getting the carbon price right through climate border measures: a Chinese perspective. *Climate Policy* 11(5), 1257–1261.
- Weikard, H. P., L. Wangler, and A. Freytag (2015). Minimum Participation Rules with Heterogeneous Countries. *Environmental and Resource Economics* 62(4), 711–727.

Weischer, L., J. Morgan, and M. Patel (2012, nov). Climate Clubs: Can Small Groups of Countries make a Big Difference in Addressing Climate Change? *Review of European Community & International Environmental Law* 21(3), 177–192.

Widerberg, O. and D. E. Stenson (2013). Climate clubs and the UNFCCC. *Fores Study* 3, 1–52.

A Proof of Proposition 1

The net stability function reads as

$$\Delta\Phi^{RN}(\theta, k) = \Phi^{RTRF}(\theta, k) - \Phi^{NTRF}(k) = (N - 2k + 1)(\beta(\theta) - \zeta(\theta)). \quad (11)$$

Differentiating with respect to coalition size k gives us

$$\frac{\partial \Delta\Phi^{RN}(\theta, k)}{\partial k} = -2\beta(\theta) + 2\zeta(\theta). \quad (12)$$

We can assume that trade sanctions decrease global welfare due to trade distortions – recall the discussion in section 3. Thus, the total benefits from trade sanctions are lower than the total costs which, in our specification, boils down to the condition that benefits from putting tariffs on one trade-flow are lower than the costs, i.e. $\beta(\theta) < \zeta(\theta)$. Thus, we see that $\frac{\partial \Delta\Phi^{RN}(\theta, k)}{\partial k} > 0$ and that $\Delta\Phi^{RN}(\theta, k)$ has a single zero at $k = \frac{(N+1)}{2}$ for all $\theta > 0$. $\square$