

Free Trade Agreements with Environmental Standards

Hideo Konishi[†] Minoru Nakada[‡] Akihisa Shibata[‡]

November 23, 2018

Abstract

In this paper, we investigate the effects of a free trade agreement (FTA) with environmental standards between Northern and Southern countries, with explicit considerations for transferring clean technology and enforcing reduced emissions. Southern producers benefit greatly by having access to a Northern market without barriers, while they are reluctant to use new high-cost, clean technology provided by the North. Thus, environmentally conscious Northern countries should design an FTA where Southern countries provide sufficient benefits for the membership while imposing tighter enforcement requirements. Since including too many Southern countries dilutes the benefits of being a member of the FTA, it is in the best interest of the North to limit the number of Southern memberships while requiring strict enforcement of emissions reduction. This may result in unequal treatment among the Southern countries. We provide a quantitative evaluation of FTA policies by using a numerical example.

Keywords : Free trade agreements; Environmental standards.

JEL Classification Numbers : Q52; F18; F53.

1 Introduction

In the current era of rapid international integration of goods and financial markets, the environment of a country is significantly affected by other countries' economic activities. While various arguments have been raised about the relationship between free trade and the environment, one of the main issues is whether international trade between developed and developing countries affects positively or negatively the environment. There are a number of discussions on the above question among researchers both from theoretical and empirical points of view. Some researchers argue that trade liberalization may cause the relocation of pollution intensive firms from high-income countries with stringent pollution regulation (Northern countries) to low-income countries with weaker regulation (Southern countries). Although pollution in the North may be reduced, free trade is likely to have negative impacts on global pollution as well as pollution in the South, because dirtier firms are located in economies with laxer environmental regulation, this is called the Pollution Haven Hypothesis (Taylor, 2005). As Copeland and Taylor (1994) suggested, the role of income inequality between countries is important in determining the impact of trade on the environment; free trade can raise pollution when the degree of income inequality between countries is relatively high. Concerning the present income disparities between Northern and Southern countries, we could assume that trade liberalization affects the environment negatively if differences in incomes and the degree of environmental regulation among countries are significant.

However, empirical evidence demonstrates that this prediction may not necessarily be true. While Managi et al. (2009) displayed that trade has negative effects on the environment in non-OECD countries, Antweiler et al. (2001) showed that freer trade is in fact good for the environment. Cole (2004) demonstrated that the North American Free Trade Agreement (NAFTA) caused no pollution haven effect in Mexico. Gutierrez and Teshima (2018) pointed out the importance of technology upgrades induced by NAFTA for pollution reduction in Mexico. Such evidence highlights the importance of giving developing countries access to markets as their motivation to adopt cleaner technologies. This diffusion of such technologies via trade might be essential for developing countries to reduce pollution (Taylor 2004).

What would happen if a potential trade partner has a government that lacks the capacity for policy enforcement? In reality, there are some preferential trade agreements between developed and developing countries that contain provisions of environmental conservation. A good example is the European Union. The EU has enlarged member states ever since

participants increases (Proposition 2).

With Proposition 2, it is easy to see that there is a tradeoff between having more Southern countries in the FTA and the level of enforcement, but there are other tradeoffs as well. With more Southern memberships, a Northern country's consumer surplus increases, while its domestic firm's profit and its tariff revenue decrease. We also do not know how the total amount of emissions would be affected by an increase in the number of Southern countries in the FTA since the enforcement level for the FTA members goes down while the number of Southern countries goes up. Moreover, as the Southern membership goes up, the total transfers to them become more and more costly for the Northern countries. Since all of these factors are important and it is hard to get qualitative results, we will present an example with reasonable parameter values and observe the optimal FTA policy for the Northern country and its environmental implications.

In the numerical example, we confirm that these considerations play important roles in evaluating the FTA policies. Limiting Southern memberships is desirable for Northern countries, but it results in sizable inequality between the FTA members and nonmembers among Southern countries. Comparative static analyses of the numerical example demonstrate that, if the number of member states is kept constant, an increase in emissions from Southern countries raises the aggregate emissions. However, it also shows that, once the number of member states is endogenized, its overall effect on the aggregate emissions is negative, due to the subsequent increase in the number of Southern participants, which adopt clean technologies.

2 The Model

2.1 The basic structure of the model

There is one Northern country and m Southern countries in the world, and all Southern countries are identical ex ante. The set of Southern countries is denoted $S = \{1, \dots, m\}$. The Northern country (denoted by 0) has an inverse demand f_i^0 inverse

fixed at w_N and w_S , respectively ($w_N > w_S > 0$).³ There are two technologies that produce industrial goods, clean and dirty technologies. In order to produce one unit of an industrial good, the clean and dirty technologies (C and D , respectively) require c_C and c_D units of labor, respectively ($c_C > c_D > 0$). That is, the clean technology requires more labor input to produce one unit of output than the dirty technology. Initially, the Northern country has the clean technology C , while all Southern countries have the same dirty technology D . The amount of emissions by producing one unit of product with the clean and dirty technologies are denoted by e_C and e_D , respectively, with $e_D > e_C > 0$.

The Northern country applies a common specific tariff rate $\tau > 0$ on imports from Southern countries. Unless Southern country j has a free trade agreement with the Northern country, tariff rate τ applies. We fix throughout this paper (τ is not a policy variable). This is because the WTO prohibits increasing tariffs when countries form an FTA. A free trade agreement does not allow a country to export goods to a country indirectly via a third country. Each country $j \in \{1, \dots, S\}$ has a single firm (only Northern countries consume industrial good). Country j 's export quantity to Northern country 0 is denoted by Q_j , and country 0's domestic supply is denoted by Q_0 . Thus, the total supply in country 0 is $Q = \sum_{j \in S} Q_j + Q_0$.

We will also assume that Southern countries do not import industrial goods. This assumption is imposed for simplicity of the analysis, and gives the worst incentive for Southern countries to participate in the FTA.

2.2 Free trade agreement, environmental standard, and law enforcement

affects the total emissions in the world, we assume that for a Southern country to form an FTA with Northern country 0, the Southern country must accept an environmental standard set by the North with a required enforcement level. We denote FTA partners with Northern country 0 by set $A \subseteq S$.

This means that when Northern country 0 and country $j \in S$ form an FTA, country j must adopt clean technology C that requires c_c units of labor, and enforce the usage of the clean technology at least to some extent by spending a fixed cost to establish law enforcement. This is because the dirty technology has a lower marginal cost than the clean technology: $c_D < c_c$. Without an enforcement mechanism, producers are tempted to use the dirty technology, so law enforcement needs to randomly audit to check if the clean technology is being used. We will denote the level of enforcement of the clean technology implicitly by $\alpha_j \in [0, 1]$: country j 's firm produces only fraction α_j of its output with the clean technology and the rest of its output $(1 - \alpha_j)$ is produced with the dirty technology to save some money. Enforcing the usage of the clean technology can be costly, since it requires infrastructure such as an audit system and well-disciplined police, which in turn requires a fixed cost. Let $F_j(\alpha_j)$ be country j 's cost of introducing the clean technology together with the cost to establish law enforcement that achieve enforcement level $\alpha_j \in [0, 1]$. We assume $F_j(\alpha_j) = F + f_j(\alpha_j)$ with $F \geq 0$, $f_j(0) = 0$, $f_j'(\alpha_j) > 0$, and $f_j''(\alpha_j) > 0$. We assume that F_j 's are ordered by the efficiency of enforcement technology: i.e. for any $\alpha_j \in [0, 1]$, $f_1(\alpha_j) \leq f_2(\alpha_j) \leq \dots \leq f_m(\alpha_j)$ and $f_1'(\alpha_j) \leq f_2'(\alpha_j) \leq \dots \leq f_m'(\alpha_j)$ holds.

Let the total amount of pollutive emissions in the world be described by

$$E = e_c Q_0 + \sum_{j \in S} (\alpha_j e_c + (1 - \alpha_j) e_D)(Q_j + q_j);$$

where $\alpha_j e_c + (1 - \alpha_j) e_D$ is country j 's emission rate for $j \in A$, and $Q = (Q_0, \dots, Q_m)$ and $q = (q_1, \dots, q_m)$ denote supply vectors in the Northern and Southern countries, respectively. Northern and Southern countries receive negative externalities from pollutive emissions in an additive manner (global pollutive emissions) by $d_N E$ and $d_S E$, respectively, where $0 < d_S < d_N$.

3 Analysis

3.1 Northern market equilibrium allocation

We will analyze Northern country 0's market equilibrium. Firms in different countries have different effective marginal costs. The firm in country 0 has marginal cost $c_0 = w_N c$, the one in Southern country $j \in A$ has marginal cost $c_j = w_S c$ if $j \in A$, and the one in country $j \in S \setminus A$ has marginal cost $c_j = w_S D + c$ if $j \in S \setminus A$. When there are m countries that supply the product to country i , and they have heterogeneous costs (c_1, \dots, c_m) , the standard Cournot equilibrium solution can be obtained in the following manner: Country i 's best response to Q_j is a solution of

$$\max_{Q_j} P \left(\sum_{i=0}^m Q_i \right) - c_j Q_j;$$

i.e., the first order condition

$$P \left(\sum_{i=0}^m Q_i \right) - c_j + P' \left(\sum_{i=0}^m Q_i \right) Q_j = 0;$$

Summing them up, we have

$$(m+1) P \left(\sum_{i=0}^m Q_i \right) - \sum_{i=0}^m c_i + P' \left(\sum_{i=0}^m Q_i \right) \sum_{i=0}^m Q_i = 0; \tag{1}$$

This equation determines $Q = \sum_{j=0}^m Q_j$ and $P(Q)$ uniquely as long as the strategic substitute condition $(P'(Q) + P''(Q)Q_j) \leq 0$ for all Q and $Q_j < Q$ is satisfied. The equilibrium allocation is described only by Q : for all $j = 0, \dots, m$

$$Q_j(Q) = \frac{P(Q) - c_j}{P'(Q)}$$

and

$$Q_j(Q) = \frac{(P(Q) - c_j)^2}{P'(Q)^2}; \tag{2}$$

as long as $P(Q) - c_j \geq 0$ is satisfied (otherwise, $Q_j = 0$ holds and firm j becomes an inactive firm: i.e., the number of firms in the market shrinks, but all nice properties still hold even after some firms become inactive). We can show that under the strategic substitute condition,

Under this assumption, an FTA member country's monopoly output q_C is determined by $p - w_S c + p q_j = 0$. Its profit is denoted by $\pi_C = \frac{(p(q_C) - w_S c)^2}{p'(q_C)}$. Since $q_D < q_C$, $q_C < q_D$ and $c_C < c_D$ hold. The firm gets the exporting and domestic profits with the clean technology, and cheating workers get $(1 - \beta_j)(c_C - c_D)w_S(Q_j + q_j)$.

3.3 Global equilibrium allocation with an FTA

Suppose that Southern countries are in the FTA ($A_j = k$) and agree to use the clean technology: i.e., countries in $A \setminus \{0\}$ adopt the technology. Since Southern countries' marginal costs depend only on the (social) technologies they use, the equilibrium output allocation vector is solely determined by A (or k), too. Agreed enforcement level affects social welfare through the worldwide emission of pollutive substances E and Southern member countries' policy enforcement only.

Let $Q(k)$ be the solution of equation (1) for $c_0 = w_N c$, $q_j = w_S c$ for all $j \in A$, and $q_j = w_S c_D$ for all $j \notin A$. The Northern country's consumer surplus is described by $CS(k) = \int_0^{Q(k)} P(\sigma) - P(Q(k)) d\sigma$. Let $Q(k) = (Q_0(k); Q_1(k); \dots; Q_m(k))$ and $c(k) = (c_0(k); c_1(k); \dots; c_m(k))$ be such that $Q_j(k) = Q_j(Q(k))$ and $c_j(k) = c_j(Q(k))$ for the above $c = (c_0; c_1; \dots; c_m)$. Countries' supply and profit vectors in the Northern market are dependent on their technologies: $Q_j(k) = Q_C(k)$ and $c_j(k) = c_C(k)$ for $j \in A$, and $Q_j(k) = Q_D(k)$ and $c_j(k) = c_D(k)$ for $j \notin A$. Southern countries' domestic supply vector is simply determined as $q_j = q_C$ if $j \in A$, and $q_j = q_D$ otherwise.

The Northern country sets a clean-technology enforcement level $\beta \in [0, 1]$ and a sign-up subsidy $\tau \geq 0$ for its FTA member (Southern) countries, and the Northern country agrees to form a free trade agreement with Southern country j as long as country j is willing to adopt the clean technology by spending enforcement costs $\beta_j(\tau) \geq 0$ (open membership, or non-discrimination). The worldwide emission of pollutive substance under this free trade agreement is described by

$$\begin{aligned} E(k; \beta) &= e_C Q_0(k) + \sum_{j \in A} (e_C + (1 - \beta_j) e_D) (Q_j(k) + q_C) + \sum_{j \in S \setminus A} e_D (Q_j(k) + q_D) \\ &= e_C Q_0 + k (e_C + (1 - \beta) e_D) (Q_C + q_C) + (m - k) e_D (Q_D + q_D): \end{aligned}$$

firm's output decision is affected by β). In the former case, $\frac{dq_j}{d\beta} = 0$, while in the latter case, $\frac{dq_j}{d\beta} > 0$ holds. Despite the difference in the underlying assumption, the quantitative results are the same.

The Northern country's social welfare can be written as

$$SW(k; \tau) = CS(k) + \tau_0(k) - k - d_N E(k; \tau)$$

Southern countries' consumer surplus is described by $CS_j = CS_D \int_0^{R_{q_D}} (p(q) - p(q_D)) dq$ if $j \notin A$, and $CS_j = CS_C \int_0^{R_{q_C}} (p(q) - p(q_C)) dq$ if $j \in A$. Southern countries' social welfare can be written as

$$SW^{OUT}(k; \tau) = SW(k; \tau) - CS_D + \tau_D(k) + \tau_D - d_S E(k; \tau) \quad (5)$$

if $j \notin A$, and

$$SW^N(k; \tau) = SW(k; \tau) - CS_C + \tau_C(k) + \tau_C + F(\tau) + (1 - \tau)(\tau_C - \tau_D) w_S (Q_C + q_C) - d_S E(k; \tau) \quad (6)$$

if $j \in A$.

3.4 Participation decision in an FTA

Here, we consider an FTA between Northern country 0 and some Southern countries. We analyze the set of equilibrium participants in the free trade agreements with Northern country 0. Let $A \subseteq S$ be the set of Southern countries that participate in free trade agreements, and let its cardinality be $a = |A|$. Note that all countries j in A have marginal costs $\tau_j = w_S - c$ and countries j in $S \setminus A$ have marginal costs $\tau_j = w_S - D + \tau$. The equilibrium set A of the Southern FTA member countries k is described by the following two inequalities:

$$SW^N(k; \tau) - F - f_j(\tau) + SW^{OUT}(k - 1; \tau) \text{ for all } j \in A \text{ (internal stability)}$$

and

$$SW^N(k + 1; \tau) - F - f_j(\tau) + SW^{OUT}(k; \tau) \text{ for all } j \notin A \text{ (external stability).}$$

If a set of Southern country members satisfies both internal and external stability conditions then it is called a stable FTA. E7I013Tf FTe FTpr-27(ci)-304-27i 108.36569701 Tf 74=(hpremo)-303.

Proof. First note $f_1(\alpha) \geq f_2(\alpha) \geq \dots \geq f_m(\alpha)$ for all $\alpha \in [0, 1]$ by assumption. Thus, if $sw^N(k; \alpha) = F_k(\alpha) + sw^{OUT}(k-1; \alpha)$ holds then $sw^N(k; \alpha) = F_k(\alpha) + sw^{OUT}(k-1; \alpha)$ holds for all $\alpha^0 \leq \alpha$. And if $sw^N(k+1; \alpha) = F_k(\alpha) + sw^{OUT}(k; \alpha)$ holds then $sw^N(k+1; \alpha) = F_k(\alpha) + sw^{OUT}(k; \alpha)$ for all $\alpha^0 \leq \alpha$.

We will prove the statement by contradiction. Suppose that there is no stable FTA. We will use an induction argument.

1. Start with $k = 0$. If $sw^N(1; \alpha) = F_1(\alpha) + sw^{OUT}(0; \alpha)$, then $k = 0$ is a stable FTA. Since there is no stable FTA, we have $sw^N(1; \alpha) = F_1(\alpha) + sw^{OUT}(0; \alpha)$.
2. For $k \geq 1$, suppose that $sw^N(k^0; \alpha) = F_k(\alpha) + sw^{OUT}(k-1; \alpha)$ holds for all $\alpha^0 \leq \alpha$. This implies $sw^N(k; \alpha) = F_k(\alpha) + sw^{OUT}(k-1; \alpha)$. If $sw^N(k+1; \alpha)$

4.1 Linear Demand Functions

Here, we assume that the Northern country has the following inverse demand function: $P(Q) = a - bQ$, and each Southern country has $p(q) = a - bq$. We have the following basic results (the proof is in Appendix A).

Lemma 1. Suppose that there are k Southern countries in the FTA. The equilibrium total output in the Northern market, the Northern country's output, the Southern FTA and non-FTA country's export to the Northern market, and the Northern country's equilibrium consumer surplus CS are

$$Q(k) = \sum_{i=0}^n Q_i(k) = \frac{(m+1) (c_0 + kc_C + (m-k)(c_D + g))}{m+2};$$

$$Q_0(k) = \frac{1}{m+2} [1 + (kc_C + (m-k)(c_D + g)) - (m+1)c_0]g;$$

$$Q_C(k) = \frac{1 + c_0 - (m-k+2)c_C + (m-k)(c_D + g)}{m+2};$$

$$Q_D(k) = \frac{1 + c_0 + kc_C - (k+2)(c_D + g)}{m+2};$$

$$CS(k) = \frac{[(m+1) (c_0 + kc_C + (m-k)(c_D + g))]^2}{2(m+2)^2};$$

respectively. Profits from the Northern market earned by firms in the Northern country, Southern FTA country (with the clean technology), and Southern non-FTA country (with the dirty technology) are

$$\pi_0(k) = \frac{1}{m+2} [1 - (m+1)c_0 + kc_C + (m-k)(c_D + g)]^2;$$

$$\pi_C(k) = \frac{1}{m+2} [1 + c_0 - (m-k+2)c_C + (m-k)(c_D + g)]^2;$$

$$\pi_D(k) = \frac{1}{m+2} [1 + c_0 + kc_C - (k+2)(c_D + g)]^2;$$

respectively. Domestic outputs, profits, and consumer surpluses in FTA and non-FTA Southern countries are $q_C = \frac{a - c_C}{2b}$, $q_D = \frac{a - c_D}{2b}$, $CS_C = \frac{(a - c_C)^2}{8b}$, and $CS_D = \frac{(a - c_D)^2}{8b}$,

$c_{SD} = \frac{(a - c_D)^2}{8b}$, respectively. Finally, the amount of equilibrium total emissions is

$$E(k; \tau) = (2e_D - e_C) \frac{m+1}{m+2} \frac{c_0 + kc_C + (m-k)(c_D + \tau)}{m+2} \\
(e_D - e_C)(1 - c_C) + e_D k \frac{a - c_C}{2b} + (m-k) \frac{a - c_D}{2b} \\
(e_D - e_C) \frac{1 + c_0 + kc_C + (m-k)(c_D + \tau) - (m+2)c_0}{m+2} \\
(e_D - e_C) k \frac{1 + c_0 + kc_C + (m-k)c_D - (m+2)c_C}{m+2} + \frac{a - c_C}{2b} :$$

With these basic results, we can analyze the optimal FTA rule for the Northern country. The Northern country can choose a policy combination, the enforcement level $\tau \in [0, 1]$; and a sign-up subsidy τ_0 to the participants of the FTA from Southern countries in order to maximize its social welfare.

$$SW(k; \tau; \tau_0) = CS(k) + \tau_0(k) + (m-k)Q_D(k) - k - d_N E(k; \tau) \quad (7)$$

In order to find the optimal FTA policy for the Northern country, we can use the following two-step procedure. First, for each $k = 1, \dots, m$, find an optimal combination of policies $(\tau^k; \tau_0^k)$ by solving the following problem:

$$(\tau^k; \tau_0^k) \in \arg \max_{\tau, \tau_0} SW(k; \tau; \tau_0) \quad \text{s.t.} \quad sw^N(k; \tau) \geq F + f_k(\tau) + sw^{OUT}(k; \tau) \quad (8)$$

Second, choose the optimal size of an FTA:

$$k^* = \arg \max_k SW(k; \tau^k; \tau_0^k);$$

Then, $(\tau^k; \tau_0^k)$ is the optimal policy that implements a size k FTA. Recall that τ is an uncontrollable variable (see footnote 5). It is easy to see that a prohibitive tariff is optimal as long as there is at least one Southern FTA member, and it also minimizes non-FTA countries' emissions since it prohibits their access to the Northern market.

In the first step of the analysis, we rewrite the welfare maximization problem (8).

Lemma 2. The constraint of (8) with equality can be written as

$$\begin{aligned}
 (k; \cdot) = & \frac{3(a - c_C)^2}{8b} + \frac{3(a - c_D)^2}{8b} + F + f_k(\cdot) \\
 & \frac{1}{m+2} \left[(m-1)(c_C + (c_D + \cdot)) \right. \\
 & \left. f - 2(1 + c_0) - (m - 2k + 3)c_C + (m - 2k - 1)(c_D + \cdot) \right] g \\
 & + d_S (3e_D - 2e_C) \frac{c_C + (c_D + \cdot)}{m+2} - e_D \left[\frac{a - c_C}{2b} + \frac{a - c_D}{2b} \right. \\
 & \left. (e_D - e_C) \frac{1 + c_0 + kc_C + (m - k)(c_D + \cdot)}{m+2} - \frac{(m+2)c_C}{m+2} + \frac{a - c_C}{2b} \right. \\
 & \left. + (e_D - e_C)(k - 1) \frac{c_C + (c_D + \cdot)}{m+2} \right]
 \end{aligned}$$

This implies $\frac{\partial}{\partial k} > 0$ and the constraint gets tighter as k increases. Substituting this formula into (7), we can convert (8) into an unconstrained maximization problem.

Proposition 2. Under linear demand, we have $1 > k_1 > k_2 > \dots > k_m > 0$ with strict inequalities $k_{k-1} > k_k > k_{k+1}$ for all k s with an interior solution $1 > k > 0$.

Proof. Problem (8) can be written as

$$SW(k; \cdot; (k; \cdot)) = CS(k) + \pi_0(k) + (m - k)Q_D(k) - k(k; \cdot) - d_N E(k; \cdot)$$

Thus, given k , the social optimum k is characterized by

$$k \frac{\partial}{\partial k} + d_N \frac{\partial E}{\partial k} = 0:$$

Rewriting this, we obtain

$$\begin{aligned}
 f_k^0(k) = & (e_D - e_C) (d_N + d_S) \frac{1 + c_0 + kc_C + (m - k)(c_D + \cdot)}{m+2} - \frac{(m+2)c_C}{m+2} + \frac{a - c_C}{2b} \\
 & - (k - 1)d_S \frac{c_C + (c_D + \cdot)}{m+2}
 \end{aligned}$$

Since $(c_D + \cdot) > c_C$, the RHS is decreasing in k . Since $f_k^0(k) > 0$ and $f_k^0(k) = f_{k-1}^0(k)$ for all k , we conclude $k < k_{k-1}$ holds for all k as long as they are interior solutions.

This proposition shows that there is a tradeoff between the number of Southern participants and the level of enforcement. Although it is hard to analyze whether or not equilibrium

Table 1: A Numerical Example

k	0	1	2	3	4	5	6	7	8	9	10
Q	0.6875	0.69167	0.69583	0.7	0.70417	0.70833	0.7125	0.71667	0.72083	0.725	

become a member to avoid the considerably high tariff rate.

(3) If the clean technology is less costly (lower α_C), more states will join the FTA. Additionally, emissions decline because such reduction will be easier if it is less costly.

(4) An increase in the emission rate (higher α_D) in Southern countries raises the aggregate emissions as long as the number of member states is kept constant. However, these higher emissions induce the Northern country to persuade Southern countries to become members.

standards, for instance.). The number of Southern countries will be reduced as a byproduct, which also helps to pass the bill in Congress/Parliament. In such a case, it might also be interesting to analyze whether a political turnover would affect the number of Southern participants, as well as global emissions. These factors may require further investigation.

Appendix A: Linear Demand

Here, we assume that the Northern country has the following demand function $P(Q) =$

1. Firm j 's profit maximization problem is

$$\max_{q_j^D} \pi_j = 1 - \sum_{i=0}^n Q_i - c_j q_j$$

The first order condition is

$$1 - \sum_{i=0}^n Q_i - c_j = 0$$

Summing them up, we obtain

$$(m+1) - (m+2) \sum_{i=0}^n Q_i - \sum_{i=0}^n c_i = 0$$

and

$$Q_j = \sum_{i=0}^n Q_i = \frac{m+1}{m+2} - \frac{1}{m+2} \sum_{i=0}^n c_i$$

Let $c^W = c_0$, $c^C = c_C$, and $c^D = c_D$. We assume that in the presence of a tariff charged by the Northern country, the marginal cost of using the clean technology in the FTA is lower than the one of using the dirty technology outside of the FTA if they export $c^{OUT} = c_D + \tau > c^N = c_C$ naturally although $c_C > c_D$ holds. The equilibrium output by country j when k Southern countries participate in the FTA is

$$\begin{aligned} Q_j &= \frac{1}{m+2} + \frac{1}{m+2} \sum_{i=0}^n Q_i - c_j \\ &= \frac{1}{m+2} \left[1 + (c_0 + k c_C + (m-k)(c_D + \tau)) - (m+2) c_j \right] \end{aligned}$$

Thus, the Northern country's output and FTA and non-FTA Southern countries' exports are written as

$$\begin{aligned} Q_0(k) &= \frac{1}{m+2} \left[1 + (k c_C + (m-k)(c_D + \tau)) - (m+1) c_0 \right] \\ Q_C(k) &= \frac{1}{m+2} \left[1 + (k c_C + (m-k)(c_D + \tau)) - (m+1) c_C \right] \\ Q_D(k) &= \frac{1}{m+2} \left[1 + (k c_C + (m-k)(c_D + \tau)) - (m+1) c_D \right] \end{aligned}$$

respectively. Since $j = Q_j^2$, we have the following

$$c(k) = \frac{1}{m+2} [1 + c_0 + (m-k+2)c_c + (m-k)(c_D + \dots)]^2;$$

$$D(k) = \frac{1}{m+2} [1 + c_0 + kc_c + (k+2)(c_D + \dots)]^2$$

with clean technology in the Northern market. The fourth term represents an indirect effect of reductions in clean technology production in the existing $m - 1$ Southern member countries crowded out by the k th Southern country's participation.

Southern country j 's social welfare is written for two different cases: being a member or a nonmember of the FTA. Southern countries' social welfare can be written as

$$\begin{aligned} SW^{OUT}(k; \gamma) &= c_{SD} + c_D(k) + c_D + d_S E(k; \gamma) \\ &= \frac{(a - c_D)^2}{8b} + \frac{1}{m+2} \left[1 + c_0 + k c_C - (k+2)(c_D + \gamma) \right]^2 + \frac{(a - c_D)^2}{4b} \\ &\quad d_S E(k; \gamma) \end{aligned}$$

if $j \notin A$, and

$$\begin{aligned} SW^N(k; \gamma) &= c_{SC} + c_C(k) + c_C + (1 - \gamma)(c_C - c_D) + \gamma(Q_C + c_C) + d_S E(k; \gamma) \\ &= \frac{(a - c_C)^2}{8b} + \frac{1}{m+2} \left[1 + c_0 - (m - k + 2)c_C + (m - k)(c_D + \gamma) \right]^2 \\ &\quad + \frac{(a - c_C)^2}{4b} + d_S E(k; \gamma) \end{aligned}$$

if $j \in A$.

This implies

$$\begin{aligned} SW^N(k; \gamma) - SW^{OUT}(k - 1; \gamma) &= \frac{(a - c_C)^2}{8b} + \frac{1}{m+2} \left[1 + c_0 - (m - k + 2)c_C + (m - k)(c_D + \gamma) \right]^2 + \frac{(a - c_C)^2}{4b} \\ &\quad d_S E(k; \gamma) - \frac{(a - c_D)^2}{8b} - \frac{1}{m+2} \left[1 + c_0 + (k - 1)c_C - (k+1)(c_D + \gamma) \right]^2 \\ &\quad - \frac{(a - c_D)^2}{4b} + d_S E(k - 1; \gamma) \\ &= \frac{3(a - c_C)^2}{8b} - \frac{3(a - c_D)^2}{8b} \\ &\quad + \frac{1}{m+2} \left[(m - 1)(c_D + c_C) - 2(1 + c_0) - (m - 2k + 3)c_C + (m - 2k - 1)(c_D + \gamma) \right] \\ &\quad + d_S (E(k - 1; \gamma) - E(k; \gamma)): \end{aligned}$$

Northern country 0 can choose a policy combination: the enforcement level $\gamma \in [0, 1]$ and a

sign-up subsidy $\tau = 0$ to the participants of FTA from Southern countries. In order to find the optimal FTA policy for the Northern country, we can use the following procedure. First for each $k = 1, \dots, m$, find an optimal combination of policies $(\tau; k)$ by solving the problem:

$$(\tau; k) \in \arg \max_{\tau} SW(\tau; k) \quad \text{s.t.:} \quad SW^N(\tau; k) = F + f_k(\tau) + SW^{OUT}(k-1; \tau);$$

For describing the binding constraint of the above problem, we express the subsidy amount as a function of τ and k :

$$\begin{aligned} &= \tau(k) \\ &= SW^N(\tau; k) + F(0) + f(\tau) + SW^{OUT}(k-1; \tau) \\ &= \frac{3(2a - c_C - c_D)(c_C - c_D)}{8b} \\ &\quad (1 - \tau)(c_C - c_D) \frac{1 + c_0 + (m - k + 2)c_C + (m - k)(c_D + \tau)}{m + 2} + \frac{a - c_C}{2b} \\ &\quad c_D \frac{1}{m + 2} \left[(m - 1)(c_D + c_C) + f + 2(1 + c_0) + (m - 2k + 1)(c_D + c_C) \right] \\ &\quad + F + f_k(\tau) + c_D E(k; \tau); \end{aligned}$$

Problem (8) can be written as

$$SW(\tau; k; (k; \tau)) = CS(\tau) + \tau_0(k) + (m - k)Q_D(k) - k(k; \tau) - c_D E(k; \tau);$$

Thus, given k , the social optimum τ_k is characterized by

$$k \frac{\partial}{\partial \tau} + c_D \frac{\partial E}{\partial \tau} = 0;$$

Thus, we have

$$\begin{aligned} k f_k'(\tau_k) - c_D (e_D - e_C) k &\frac{1 + c_0 + k c_C + (m - k)(c_D + \tau)}{m + 2} - \frac{(m + 2)c_C}{m + 2} + \frac{a - c_C}{2b} \\ &+ c_D (e_D - e_C) k (k - 1) \frac{c_C + (c_D + \tau)}{m + 2} \\ &- c_D (e_D - e_C) k \frac{1 + c_0 + k c_C + (m - k)(c_D + \tau)}{m + 2} - \frac{(m + 2)c_C}{m + 2} + \frac{a - c_C}{2b} \\ &= 0; \end{aligned}$$

Rewriting this, we obtain

$$f_k^Q(k) = (e_D - e_C) (d_N + d_S) \frac{1 + c_0 + k c_C + (m - k)(c_D + c_C)}{m + 2} + \frac{a - c_C}{2b} - (k - 1) d_S \frac{c_C + (c_D + c_C)}{m + 2} :$$

Since $(c_D + c_C) > c_C$, the RHS is decreasing in k . Since $f_k^{Q0} > 0$ and $f_k^Q(k) < f_k^{Q1}(k)$

Table A2: Higher Tari Rate: = 0:15

k	0	1	2	3	4	5	6	7	8	9	10
Q	0.64583	0.65417	0.6625	0.67083	0.67917	0.6875					

Table A4: Higher Emission Rate: $e_D = 0:5$

k	0
---	---

- [8] Eichner, Thomas and Rüdiger Pethig (2013a). Self-enforcing environmental agreements and international trade. *Journal of Public Economics* 102, 37-50.
- [9] Eichner, Thomas and Rüdiger Pethig (2013b). Trade tariffs and self-enforcing environmental agreements. CESifo Working Paper No. 4464.
- [10] Eichner, Thomas and Rüdiger Pethig (2015). Is trade liberalization conducive to the formation of climate coalitions? *International Tax and Public Finance* 22, 932-955.
- [11] EPA (1994). Environmental Protection Along the U.S.-Mexico Border.
- [12] EU (2014). Candidate and Potential Candidate Countries, The European Union (http://ec.europa.eu/environment/archives/enlarg/enlargement_en.htm).
- [13] Gutiérrez, Emilio, and Kensuke Teshima (2018). Abatement expenditures, technology choice, and environmental performance: Evidence from firm responses to import competition in Mexico. *Journal of Development Economics* 133, 264-274.
- [14] Kuhn, Thomas, Radomir Pestow, and Anja Zenker (2015). Self-enforcing climate coalitions and preferential free trade areas. *WWDP* 123/2015.
- [15] Managi, Shunsuke, Akira Hibiki, and Tetsuya Tsurumi (2009). Does trade openness improve environmental quality? *Journal of Environmental Economics and Management* 58(3), 346-363.
- [16] Taylor, M. Scott (2005). Unbundling the pollution haven hypothesis. *The BE Journal of Economic Analysis & Policy* 3(2), 1-28.
- [17] WB (2007) *Journey to a Cleaner Future*, The World Bank (<http://www.worldbank.org>).