

# Deforestation as an externality problem to be solved efficiently and fairly

Very preliminary version - please do not quote

Charles Figuières and Estelle Midler

August 5, 2011

## Abstract

The international community recently agreed on a mechanism called REDD+ to reduce deforestation in tropical countries. However the mechanism, by its very nature, has no reason to induce a Pareto optimal reduction of deforestation. The aim of this article is to propose an alternative class of mechanisms for negative externalities that implements Pareto optimal outcomes as Nash Subgame Perfect Equilibria, and that satisfies some fairness properties, in particular two original axioms of environmental responsibility. Outcomes are individually rational and the scheme does take into account environmental responsibility in the sense of our two axioms. However, envy freeness, even in a weak form adapted to the deforestation problem, turns out to be hard to achieve without dropping the other properties.

## 1 Introduction

The problem of deforestation has several dimensions. One of them takes on the appearance of negative externalities from forest-rich countries to the entire planet, affecting social efficiency.

Our goal in this paper is to imagine a solution to this externality aspect of the question, while  
20 taking into account of some particular equity issues that arise in this context.

Deforestation in tropical countries accounts for up to 20% of global emissions of CO<sub>2</sub>. It is  
the second most important source of Greenhouse Gas Emissions in the world and the first one  
in developing countries. It is also a leading cause of loss of global biodiversity. A new scheme  
called REDD, for Reduction of Emissions from Deforestation and Degradation of forests, has  
25 been agreed on at the COP16 of the UNFCCC in Cancun in 2010, to reward countries with low  
deforestation rates. The general principle is to compensate developing countries that reduce  
their deforestation with financial incentives. As far as details are concerned, there is still no  
consensus on the way such financial incentives should be calculated and allocated. The REDD  
transfers would be allocated per unit of real reduction of deforestation compared to a reference  
30 level, called the baseline (see for instance Parker et al, 2008). But the level of the baseline is an  
open issue.

Overall, the theoretical status of the REDD proposal is not entirely clear to us. Implicitly it  
looks like a cost-effectiveness tool: how to impose an exogenous limitation, or any limitation,  
of deforestation at the lowest cost for financing countries?

35 Now imagine that the issue be addressed from a different angle. Let the explicit goal be  
Pareto optimality, supplemented by additional criteria of equity and acceptability that seem  
relevant for an international externality problem like deforestation. Not surprisingly, the REDD  
program has no reason to induce a Pareto optimal reduction of deforestation (see Figuières et al,  
2010). Could we develop a class of proposals that will make it? We contend this is possible. At  
40 the conceptual level, the challenge is twofold. First, with the goal of Pareto optimality comes the  
issue of disclosure of private pieces of information about preferences, *i.e.* subjective attributes  
that are useful to calibrate the proposals but that countries possess and have an interest to hide.  
This is the heart of the literature on implementation. The second challenge is to retain, among  
the multiple Pareto optimal outcomes, only those that meet desirable principles of equity. There  
45 exists diverse criteria for equity and this paper is not an argument for one of them in the absolute.

Our aim is rather to make them explicit, to alter them in way that is relevant here, to draw a picture of their compatibility or incompatibility and, ultimately, to link them with some, not necessarily unique, solutions to the externality problem.

We write "solutions to the externality problem" rather than "solutions to the deforestation problem" on purpose. For it is clear that other aspects, ignored in this article, are also making up the issue of deforestation. For instance, monetary transfers should not pay inactivity. They are meant to be the financial counterpart of viable alternative activities to deforestation, may be intensive farming techniques, rural development or other local projects. Finding and supporting such parallel policies is part of the challenge. Also of crucial importance is the ability (or lack thereof) of southern countries to enforce measures or policies against deforestation. Some experts warn that most of the recipient countries for REDD are too fragile, facing institutional instability and even pervasive corruption. Therefore it is an error to incentivize them (Karsenty & Ongolo, 2011). They rightly argue that consolidating these states, expanding the coverage of property rights and justice, are necessary steps. But these steps are not sufficient, unless one believes heroically that those efforts will end up in a legal and institutional infrastructure not only strong enough to enforce local decisions, but also willing to internalize externalities worldwide. Put differently, as the title of the paper also underlines, we do not defend our analysis as a panacea but hopefully as a useful step, among others, in the way of a global and applicable knowledge to the deforestation issue.

In doing so, this paper also offers two by-products. Firstly, it contributes to the mechanism design literature by broadening the scope of the so-called "compensation mechanisms" (Danziger & Schnytzer, 1991, Varian, 1994). So far, compensation mechanisms apply to situations of positive externalities, under the guise of adequately designed subsidies, or to situations of negative externalities, with the use of particular taxes. But deforestation is a negative externality problem for which taxation is not possible, due to the lack of an international body with the power to impose it. We can shrewdly get around this problem by subsidizing avoided deforestation rather than taxing net deforestation (more explanations in Section 3). Secondly,

by introducing two new axioms of environmental responsibility in order to discriminate Pareto optima in a second best framework, the paper modestly contributes to the economic literature on equity.

The paper is organized as follows. In Section 2 we propose a simple static idealization of the North-South Deforestation problem. Section 3 introduces a class of incentive mechanisms - let us call it REDD\* - directly inspired from the compensation mechanisms, and analyses its efficiency, under different assumptions regarding the structure of information possessed by countries. Section 4 addresses the crucial questions of acceptability and equity of REDD\*. It proposes two complementary notions of environmental responsibility, and suggests possible formulas for baselines that comply with those notions. Section 5 concludes.

## 2 A north-south deforestation framework

Consider  $m$  countries in the developing South with a high endowment of tropical forests. Deforestation provides land and capital for development. Let  $d_i \in [0, \bar{d}_i[$  be the number of deforested hectares by country  $i$ , where  $\bar{d}_i$  is its total forest area.

Each country has a continuous increasing and concave technology that transforms deforestation into an index of composite economic goods and/or services<sup>1</sup>  $s_i(d_i)$ . This index then provides utility  $u_i(d_i) \equiv v_i(s_i(d_i))$  to country  $i$ . If it helps, one could think of  $v_i(\cdot)$  as a linear transformation of the services derived from deforestation, *i.e.*  $v_i(s_i) = \sigma_i s_i$  where  $\sigma_i \geq 0$  is a preference parameter. The functions  $u_i(\cdot) = v_i \circ s_i(\cdot)$  are assumed increasing and concave,  $u_i'' \leq 0 \leq u_i'$ . We also impose a usual Inada condition  $\lim_{d_i \rightarrow 0} u_i'(d_i) = +\infty$ .

Also, each country is endowed with an exogenous wealth  $y^i$ .

Country  $i$ 's preferences are defined over the pairs  $(d_i, y^i)$ , and represented by an additively

---

<sup>1</sup>There is an economic interest in deforestation that is not limited to timber exploitation. Forest also "compete", for instance, with agriculture and some form of tourism. Here  $s_i(d_i)$  captures all the opportunity costs of preserving forest.

separable total utility function:

$$U^i(d_i, y^i) = u_i(d_i) + y^i, \quad i = 1, \dots, m.$$

As regards deforestation there is a country-specific limit  $d_i^{bau}$ , beyond which nature cannot be  
 95 turned into arable lands within the time-scale captured by our static model; or put differently, for geographical, biophysical or economic reasons the marginal utility of deforestation is zero beyond those thresholds,  $u'_i(d_i) = 0, \forall d_i \geq d_i^{bau}$ . Therefore, on a non cooperative basis, southern countries push deforestation up to that threshold  $d_i^{bau}$ .

The north is a block that will be treated as a single country. It is also endowed with an exogenous wealth  $y^n$  and it is interested in aggregate tropical deforestation,  $D = \sum_i d_i$ , because it is linked with carbon emissions. Its preferences are captured by a utility function:

$$U^n(D, y^n) = u_n(D) + y^n,$$

which is strictly concave and decreasing with respect to the first argument,  $u''_n < 0, u'_n \leq 0$ .

This model is simple, yet it accounts for the asymmetric nature of the deforestation problem: at the business-as-usual, deforestation in the South fails to take into account of the negative externality it generates. Pareto optimal deforestation levels, denoted  $(d_1^*, \dots, d_m^*)$ , on the contrary, do not neglect those external effects. Discarding the possibilities of corner allocations for wealths, optimal issues equalize the marginal benefit for the south with the marginal cost for the North, *i.e.* they would solve the following system of equations (technical details are given in Appendix A):

$$u'_i = -u'_n, \quad i = 1, \dots, m. \quad (1)$$

100 Pareto optimality calls for different, actually lower, deforestation levels  $d_i^* \leq d_i^{bau}$ , because of their external negative effects<sup>2</sup>. But avoided deforestation represents an opportunity cost for southern countries.

---

<sup>2</sup>Pareto optimal levels are lower, as can be deduced from the properties of utility functions in the South. Their derivatives are positive only for levels of deforestation that fall in the range  $[0, d_i^{bau}]$ . And according to (1), at a Pareto optimal allocation, the marginal utilities in the South are positive.

### 3 A class of compensation mechanisms to curb deforestation

#### 3.1 The general design

105 There is a class of mechanisms, generically referred to as "compensation mechanisms", that rests on the following logic: agents involved in an economic environment with externalities solve the social dilemma by means of cross-subsidies (in case of positive externalities) or cross-taxes (in case of negative externalities) whose magnitude they decide by themselves. The role of the regulator, no matter who or what it may be, is simply to give effect to these decisions. 110 The classic reference is Varian (1994), but important predecessors are Guttman (1978, 1985 and 1987) and Danziger and Schnytzer (1991). These mechanisms implement first best allocations as subgame perfect Nash equilibria.

That kind of solution cannot be applied as it stands in our context of transnational negative externalities, because it would involve the developed North taxing the developing South! But a 115 trick can be found to retain the spirit of the mechanism, while turning taxes into subsidies. The description of what we call REDD\* is as follows. The North can now decide to subsidize developing countries who are willing to reduce their deforestation through a two-stage mechanism:

1. *Announcement stage*: in this first stage, countries choose subvention/tax rates simultaneously. The regulator asks the North to choose a vector of subsidy rates  $(t_1^n, \dots, t_m^n)$ , where 120  $t_i^n$  is the subsidy rate offered to developing country  $i$ , and asks developing country  $i$  to choose a tax rate  $t_i^s$ . The regulator also imposes that  $t_i^n \in [0, t_i^s]$ , in other words the span of the set of decisions of the North depends on the decisions taken in the South. As explained in Section 4.1, this first change in the initial mechanism is important to guarantee individual rationality in southern countries. Those announced rates are collected and end 125 up in the following formula for transfers: conditionally on the levels of deforestation to be decided at the next stage, the North would pay  $T^n = \sum_i T_i^n$ , with

$$T_i^n = \begin{cases} t_i^s(d_i^b - d_i) & \text{if } d_i < d_i^b, \\ 0 & \text{otherwise,} \end{cases}$$

and each southern country  $i$  would receive:

$$S_i = \begin{cases} t_i^n(d_i^b - d_i) - \varepsilon_i(t_i^n - t_i^s)^2 & \text{if } d_i < d_i^b, \\ 0 & \text{otherwise.} \end{cases}$$

The second change from the initial mechanism comes from the constants  $d_i^b, i = 1, \dots, m$ . These are the baselines, supposedly set by the regulator and through which price regulation obeys a logic of subsidy, not tax. This stage can be interpreted as a negotiation phase where countries discuss the correct price signal of avoided deforestation.

2. *Choice stage*: in this second stage, each southern country  $i$  determines its level of deforestation  $d_i$ . Transfers are then implemented.

So, under the mechanism, incomes become:

$$y^i = y_0^i + t_i^n(d_i^b - d_i) - \varepsilon_i(t_i^n - t_i^s)^2, \quad i = 1, \dots, m,$$

and:

$$y^n = y_0^n - \sum_i t_i^s(d_i^b - d_i).$$

Three important remarks about the specificities of this class of mechanisms are in order:

- Under Varian's mechanism, transfers are a linear function of the amount of negative externality produced. Here transfers are a linear function of  $(d_i^b - d_i)$ . Thereby REDD\* rewards the deforestation effort of the South as desired by the international community rather than taxing the net deforestation level.
- Under REDD+, each tropical country willing to reduce its deforestation level below its reference level would receive a transfer  $t(d_i^b - d_i)$  with  $t$  being the exogenous carbon price on the market. REDD\* differs because the price signal  $t$  is determined endogenously. This

will turn out to be important to ensure some attractive properties of the outcome, namely Pareto optimality (Section 3.2) and individual rationality (Section 4.1).

- in the end, the financing of the schemes does not rely on the carbon market. Hence one of the channels for the risk of carbon leakage<sup>3</sup> is eliminated.

## 145 3.2 Subgame perfect Nash equilibria

At a subgame perfect Nash equilibrium (SPNE), individual decisions are determined as usual by backward induction. In the last decision period, developing countries choose their optimal deforestation level  $d_i^*$  which maximizes their utility under the mechanism, knowing  $t_i^n$  and  $t_i^s$ . The first order condition for an interior optimal deforestation is:

$$\begin{aligned} \frac{\partial U^i}{\partial d_i} &= u'_i(\cdot) - t_i^n = 0. \\ \Leftrightarrow u'_i(\cdot) &= t_i^n. \end{aligned} \quad (2)$$

With the assumptions made so far,  $u'_i(\cdot)$  can be inverted, so  $d_i^*$  is a function of  $t_i^n$ , which we can write:

$$d_i^* = d_i^*(t_i^n).$$

Applying the implicit function theorem to (2), we can deduce that the larger the subsidy rate, the lower the deforestation:

$$d_i^{*'}(t_i^n) = \frac{1}{u''_i} < 0. \quad (3)$$

In the first period, countries choose the tax/subsidy levels. In the South, the first order condition for an interior solution is:

$$\frac{\partial}{\partial t_i^s} U^i = 2\varepsilon_i(t_i^n - t_i^s) = 0, \quad (4)$$

---

<sup>3</sup>Carbon leakage refers to an increase in carbon dioxide emissions in one country as a result of an emissions reduction by a second country..



therefore the best decision  $t_i^s$  is always to match the North's subsidy offer:

$$t_i^s = t_i^n \quad (5)$$

150 In the North, the marginal effect of each subsidy is:

$$\begin{aligned} \frac{\partial}{\partial t_i^n} U^n &= u'_n * d_i^{*'}(t_i^n) + \frac{\partial y^n}{\partial d_i^*} * d_i^{*'}(t_i^n) , \\ &= (u'_n + t_i^s) * d_i^{*'}(t_i^n) . \end{aligned} \quad (6)$$

It is the sum of two terms, the first one has an ambiguous sign whereas the second one is negative (see (3)).

At a SPNE, because of (5), necessarily  $t_i^s = t_i^n = t$ . It is easy to check that  $t = 0$  is not a SPNE. In such a case, (6) becomes:

$$\left. \frac{\partial}{\partial t_i^n} U^n \right|_{t_i^n=0} = u'_n * d_i^{*'}(0) > 0,$$

and the North has an interest to increase its subsidy rate. There are actually two kinds of SPNE, where the North always chooses the upper bound in its interval of possible decisions  $[0, t_i^s]$ . Those for which the North decision is constrained, when  $u'_n + t_i^s < 0$ . In that case  $\partial U^n / \partial t_i^n > 0$ , the marginal gain from reducing the deforestation is larger than the cost of the incentive to be offered to the South, but the choice of the tax  $t_i^s$  by southern country  $i$  does not allow to push further the subsidy  $t_i^n$ . And there are those for which  $u'_n + t_i^s = 0$ , in which the upper bound  $t_i^n = t_i^s$  is an unconstrained corner decision. We will restrict our attention to those SPNE, that are also Pareto optimal outcomes. Indeed, from (2), (5) and (6):

$$t_i^s = -u'_n = t_i^n = u'_i . \quad (7)$$

This last equation characterizes all the subgame perfect interior nash equilibria. Since an interior Pareto Optimum requires  $-u'_n = u'_i$ , one observes from (7) that it can be reached  
155 through the mechanism.

### 3.3 About the information structure

The solution concept used above to describe non cooperative decisions is indicative of the information structure under which the mechanism is supposed to work: the "regulator", whatever it may be, does not have any information about countries' preferences but countries themselves  
 160 know a great deal more. They know the utility function of each other; they know that they know that, and they know that they know that, and so on. In the terminology of game theory, there is complete information and common knowledge.

The assumption of complete information and common knowledge is sometimes justified as an approximation for situations where there exists a sufficient degree of familiarity among  
 165 agents. One may or may not subscribe to the view that this approximation is relevant for the deforestation problem. But is such an assumption really necessary? Or is it rather a convenience of presentation, a useful simplification? Would countries play the predicted Nash equilibrium under different, less demanding, information structures?

Empirical studies have found that supermodular (when agents best responses are upward  
 170 sloping) or near-supermodular games exhibit behavior of subjects that converges to the Nash equilibrium. Super-modularity is a technical property of games that ensures convergence to equilibrium under various learning dynamics, which include Bayesian learning, fictitious play, adaptive learning, and Cournot best reply (see Chen and Gazzale, 2004). This finding raises the important question of whether our class of compensation mechanisms is supermodular in the  
 175 subsidies?

By inspection of (4) and (6), one can deduce:

$$\frac{\partial^2}{\partial t_i^s \partial t_i^n} U^i = 2\varepsilon_i > 0, \quad \frac{\partial^2}{\partial t_i^n \partial t_i^s} U^n = 0.$$

So the game is super-modular.

To illustrate, Appendix B consider a case where complete information and common knowledge are ruled out. Countries take myopic decisions on the basis of their private information. Their interactions defines a sort of repeated international negotiation where countries proceed

by tatōnnement to find  $t_i^s$  and  $t_i^n$ . The important message is that, asymptotically, we get the same price signal  $t$  as before, when countries were supposed to have complete information and common knowledge. Therefore the efficiency of the mechanism does not necessarily disappears when countries do not have all the information on each other's preferences. The proposed class of mechanisms implement the optimum under less restrictive informational conditions than one may think at first sight. This property remains whatever the  $\varepsilon_i$  and the  $d_i^b$  chosen, allowing us to choose baselines which satisfy some fairness properties. Various types of baselines are discussed in the next section.

## 4 Baselines and equity

An important topic of the international debate about the financing of avoided deforestation in the South is the definition of the baselines. Several possibilities are under consideration. They could be based only on historical levels of deforestation but this would promote countries that have had "bad" past behavior. They could also take into account countries' development paths so that countries that have not cleared a lot of their forest until now would be favored. For more details on possible baseline definitions see Bush et al (2009). What is more likely to happen is a mix of those logics.

In addition, there exists an academic literature that addresses the question of equity from a more general perspective; it already gives a substantial and well organized bulk of knowledge (see for instance Roemer, 1998, Fleurbaey, 2008, or Clement et al, 2010) and we will borrow three important notions from it: individual rationality, no-envy and responsibility.

This section presents a series of round trips between the academic literature and the concerns currently expressed about the design of REDD. Equipped with qualified axioms that seem relevant for the deforestation problem, it is possible to suggest different formulas for baselines. The investigation will keep in mind the asymmetric nature of information. Thus, baselines should be designed without the recourse to pieces of information about utility functions not supposed to be publicly available.

## 4.1 Individual rationality

For international issues without a supranational authority, cooperation is problematic if the contemplated solution does not guaranty each country a level of national welfare at least equal to that they enjoyed under the business-as-usual scenario. The idea is already present in Steinhaus  
 210 (1948) and is sometimes called the *fair share guaranteed criterion*. Pareto optimal allocations that are individually rational prevent such kind of objections and can be viewed not only has an equity criterion but also, on more practical grounds, as a minimal condition for acceptability.

**Axiom 4.1 (IR).** *A Pareto optimal allocation  $(d_1^*, \dots, d_m^*, y^{1*}, \dots, y^{m*}, y^{n*})$  is individually rational (IR) if:*

$$\begin{aligned} u_i(d_i^{bau}) + y_0^i &\leq u_i(d_i^*) + y_0^i \\ &+ t^* (d_i^b - d_i^*) , \\ i &= 1, \dots, m, \end{aligned} \quad (8)$$

$$\begin{aligned} u_n(d_i^{bau}) + y_0^n &\leq u_n(d_i^*) + y_0^n \\ &- t^* \sum_{i=1}^m (d_i^b - d_i^*) . \end{aligned} \quad (9)$$

215 It will prove useful to break this requirement into two pieces, an individual rationality in the South when the first  $m$  inequalities hold, and an individual rationality in the North when the last inequality holds.

If all relevant pieces information about utility functions were available, setting baselines while respecting IR would simply amount to solve the system of linear inequalities (8) and (9). As explained in Appendix C, there exists, for each developing country, a set of acceptable baselines that takes the form of a neighborhood around its BAU, call it  $B_i(d_i^{bau})$ . If its baseline is chosen in this neighborhood, then this country does not reject the Pareto optimal allocation. Ignoring informational constraints the set of acceptable baselines for each country can be identified. In the South  $B_i(d_i^{bau}) = [\bar{d}_i^b, \bar{d}_i]$ , where:

$$\bar{d}_i^b = d_i^* + \frac{u_i(d_i^{bau}) - u_i(d_i^*)}{u_i'(d_i^*)} < d_i^{bau} , \quad i = 1, \dots, m.$$

The North also has such an interval, that contains the sum of the BAU deforestation levels. It is noted  $B_n \left( \sum_{i=1}^m d_i^{bau} \right) = [0, \widehat{D}]$ , where:

$$\widehat{D} = \sum_{i=1}^m d_i^* + \frac{u_n \left( \sum_{i=1}^m d_i^{bau} \right) - u_n (d_i^*)}{u'_n (d_i^*)} > \sum_{i=1}^m d_i^{bau}.$$

The set of acceptable baselines worldwide is then:

$$B^{IR} = \left\{ (d_1^b, \dots, d_m^b) / d_i^b \in B_i (d_i^{bau}), i = 1, \dots, m \text{ and } \sum_{i=1}^m d_i^b \in B_n \left( \sum_{i=1}^m d_i^{bau} \right) \right\}.$$

But our informational conditions are unfortunately quite hard. To compute the lower bound of  $B_i (\cdot)$  and the upper bound of  $B_n (\cdot)$ , one needs private pieces of information about countries' preferences. Yet, there clearly exists a vector of possible baselines that is common to all those sets and that is public knowledge: the vector  $(d_1^{bau}, \dots, d_m^{bau})$ . Let us emphasize this possibility:

**Proposition 4.1.** *Consider a Pareto optimal allocation. It is accepted worldwide as a REDD\* outcome when  $d_i^b = d_i^{bau}, \forall i = 1, \dots, m$ .*

It is worth noting that introducing the mechanism cannot be harmful for southern countries. Consider for instance a southern country. It could unilaterally secure the level of utility it enjoyed under the business-as-usual scenario. It suffices to set  $t_i^s = 0$ . Then, because  $t_i^n \in [0, t_i^s]$ , necessarily  $t_i^n = 0$  and  $d_i^* (t_i^n) = d_i^* (0) = d_i^{bau}$  while  $y^{i*} = y_0^i$ . If countries unilaterally settle for equilibrium tax rates that are not zero,  $t_i^{s*} \neq 0$ , then it must be the case that  $u_i (d_i^*) + y^{i*} \geq u_i (d_i^{bau}) + y_0^i$ ,  $i = 1, \dots, m$ . In other words, southern countries can unilaterally escape the mechanism. So if the mechanism does change behaviors, IR is guaranteed in the South. In this perspective, the choice of baselines is an issue, not to ensure IR in the South but to avoid rejection of the mechanism and to obtain IR in the North.

The above proposition identifies a sufficient condition,  $d_i^b = d_i^{bau}, \forall i = 1, \dots, m$ , to impose on baselines in order to ensure individual rationality of a REDD\* outcome to all countries, the North included. It does not necessarily mean that if baselines are larger or lower than the BAU levels, IR is violated. By inspection of the proof of Proposition 4.1, we can even say that

the stronger the concavity of utility functions, the more we may set baselines that depart from BAU levels. But, clearly, being too lax on baselines has the effect of increasing the volume of transfers, at the risk of transgressing individual rationality of the north. And being too strict  
 240 puts at risk acceptability by the South.

## 4.2 No-envy

Simply put, in our context an outcome is without envy if no country would prefer the deforestation-income bundle of another country<sup>4</sup>. This concept plays an important role in the economic analysis of equity (for seminal contributions, see Tinbergen, 1946, Foley, 1967, Kolm, 1971, Varian,  
 245 1974). It has often been discussed and criticized on several counts. It is well understood that no-envy is hard to achieve when agents have different and non transferable talents. And the ethical relevance of the notion has also been questioned. If envy can be considered a nasty feeling, why should it be used to elaborate a reflection on equity? Some argue however that no-envy may be proposed as a guide of justice in so far as it is indicative of social peace and, presumably,  
 250 stability of the proposed state of affairs. Because of those kind of objections and subtleties, many refinements or weakening of the no-envy criterion have been proposed, and we are no exception. In the deforestation context we shall introduce in this section three qualifications to the no-envy test.

First, because of the asymmetry between developed and developing countries, it makes sense to limit the use of this notion to southern countries. An allocation

$$(d_1^*, \dots, d_m^*, y^{1*}, \dots, y^{m*}, y^{n*})$$

has no-envy (NE) in the South if there exists no pair of developing countries  $i$  and  $j$  such that:

$$u_i(d_j^*) + y^{j*} > u_i(d_i^*) + y^{i*}.$$

---

<sup>4</sup>Envy is a social sentiment that is captured in a very particular way in much of the economic literature. We follow that tradition in this paper, but we refer to Kolm (1995) for an insightful discussion of the issue, and where envy is modelled as a negative consumption externality.

The above notion could still be criticized in our context, for it does not question the domain  
 255 over which it is reasonable to use the absence of envy as a guide for equity. Should the domain  
 incorporate the exogenous incomes  $y_0^i$ ? Those variables can be so dramatically different from  
 one developing country to another for reasons of size, history, geography... Although the issue  
 of justice along the dimension of incomes could be developed at length, one can admit that  
 redressing a feeling of envy grounded on income inequalities is far beyond the scope of REDD  
 260 transfers. This seems at best a welcome consequence of those transfers, at worst a requirement  
 not very realistic.

So the second refinement we propose is to discard from the domain of justice the exogenous  
 endowments of incomes. A modified and weaker condition of no-envy would then focus only  
 on deforestation decisions. It would just discard the possibility that:

$$u_i(d_j^*) + y_0^i + t^*(d_j^b - d_j^*) > u_i(d_i^*) + y_0^i + t^*(d_i^b - d_i^*) ,$$

or simply

$$u_i(d_j^*) + t^*(d_j^b - d_j^*) > u_i(d_i^*) + t^*(d_i^b - d_i^*) .$$

A last refinement is in order. Clearly, small countries may not be able to achieve the same  
 level of services derived from deforestation as those enjoyed by larger countries, for two rea-  
 sons. It might be because their forest endowment is (relatively) too small, or because their  
 technology to transform deforestation into services is (relatively) less efficient<sup>5</sup>. Formally, for  
 a particular level of service  $s_j^* = s_j(d_j^*)$  enjoyed by country  $j$ , there might be no admissible  
 value of deforestation in country  $i$  that would allow to achieve that level:

$$s_i(d_i) < s_j^* , \quad \forall d_i \in [0, \bar{d}_i] . \quad (10)$$

Then, how could country  $i$  has a claim against a particular allocation that would allow another

---

<sup>5</sup>By way of illustration, in 2005 the forest area of Solomon Islands was 18,770 km<sup>2</sup> (56th rank in the world),  
 to be compared with the 366,020 km<sup>2</sup> (15th in the world) for Argentina, or with the 4,502,770 km<sup>2</sup> (1st rank) of  
 Brazil. Source: FAO Global Forest Ressource Assessment 2005: Progress Towards Sustainable Forest Manage-  
 ment (Forestry Paper 147, Rome 2006).

country  $j$  a level of deforestation, and the corresponding services, which are beyond reach for country  $i$ ? Their respective situations are not commutable, for physical and/or technical reasons.

But in case another country's situation is technically within reach, i.e.  $\exists d_i \in [0, \bar{d}_i]$  such that  $s_i(d_i) = s_j^*$ , define the function that measures the number of deforested hectares necessary to produce a given service as

$$d_i = g_i(s) \quad , \quad g_i(\cdot) \equiv s_i^{-1}(\cdot) .$$

265 Finally, on that basis, a modified test for no-envy could be:

**Axiom 4.2** (NRES). *There is no restricted envy in the South (NRES) if there exists no pair of developing countries  $i$  and  $j$  such that:*

$$u_i(s_i^{-1} \circ s_j(d_j^*)) + t^*(d_j^b - s_i^{-1} \circ s_j(d_j^*)) > u_i(d_i^*) + t^*(d_i^b - d_i^*) .$$

In the particular case where countries have the same technologies and differs only with respect to their endowments of forests, then  $s_i^{-1} \circ s_j = 1$  and the above test becomes:

$$u_i(d_j^*) + t^*(d_j^b - d_j^*) > u_i(d_i^*) + t^*(d_i^b - d_i^*) .$$

If forest endowments are too different, so that inequality (10) holds, then the *NRES* test is somewhat satisfied by default.

**Proposition 4.2.** *Suppose that the set of acceptable baselines worldwide,  $B^{IR}$ , encompasses vectors of identical baselines for all southern countries. Whatever the differences in forest endowments, a Pareto optimal allocation implemented via the REDD\* mechanism where countries are offered the same baselines,  $d_i^b = d^b, \forall i$ , satisfies NRES and IR.*

*Proof.* First recall that for countries such that  $s_i^{-1} \circ s_j(d_j^*) \notin [0, \bar{d}_i]$ , then the NRES test for such countries is satisfied by default. Otherwise, the NRES test in the South requires that:

$$u_i(s_i^{-1} \circ s_j(d_j^*)) + t^*(d^b - s_i^{-1} \circ s_j(d_j^*)) \leq u_i(d_i^*) + t^*(d^b - d_i^*) \quad , \quad \forall i, j .$$

$$\Longleftrightarrow \quad u_i(s_i^{-1} \circ s_j(d_j^*)) \leq u_i(d_i^*) + t^*(s_i^{-1} \circ s_j(d_j^*) - d_i^*) \quad , \quad \forall i, j .$$



Now, because  $t^* = u'_i(d_i^*)$  the last expression is equivalent to:

$$u_i(s_i^{-1} \circ s_j(d_j^*)) \leq u_i(d_i^*) + u'_i(d_i^*)(s_i^{-1} \circ s_j(d_j^*) - d_i^*),$$

an inequality that is verified because the functions  $u_i(\cdot)$  are concave. ■

The above proposition leaves open the question of the existence of a Pareto optimal allocation sustained by the mechanism when baselines are identical. The above proposition just states that, if such an outcome exists, then it satisfies NRES (and IR by construction). But we know from the previous section that there is a risk that countries reject the mechanism, unless their baselines is identical to their BAU. Actually, the difficulty is whether the set  $B^{IR}$  of acceptable baselines defined in the previous section contains a common baseline. If all countries had very similar  $d_i^{bau}$ , it could be that  $d^b = \bar{d}^{bau} = \frac{\sum_i d_i^{bau}}{m}$  would do the job. But more realistically, countries have very different  $d_i^{bau}$ , and setting identical baselines may lead countries to fall back to the BAU scenario rather than to avoid envy.

### 4.3 Environmental responsibility

Getting back to propositions currently discussed at the UN, there is a concern that, based on observed current deforestation behaviors, countries have varying degrees of merit or accountability to the environment problem and, therefore, should be subject to differential treatments, particularly as far as baselines are concerned.

Until recently, the theme of responsibility was seldom analyzed in the economic academic literature (for a recent review see Fleurbaey and Maniquet, 2011). Much of the ethics of responsibility rests on the premise that agents have full control over some variables, and therefore should bear the consequence of the exercise of their control, without interference from society, whereas they should be compensated for the adverse influence of characteristics beyond their control (like handicaps for individuals)<sup>6</sup>.

---

<sup>6</sup>Not surprisingly, drawing a separating line between agents' characteristics that fall into the responsibility

In this section, this dichotomy between controlled and uncontrolled variables, serves as inspiration to develop an analysis of how baselines can be adjusted according to the responsibility of countries in terms of deforestation<sup>7</sup>. The first task is to give a precise content to the term "responsibility" that is used here.

Developing countries do have different characteristics, more or less beyond their control, like their natural land endowments and geoclimatic conditions. On the other hand, countries' preferences are usually considered sovereign. And, with regard to this question of responsibility, the status of their "technology"  $s_i(\cdot)$  that transforms deforestation into services is unclear. One may argue that, to some extent at least, current technologies reflect past choices. As a simplifying assumption, we will assume that countries can be considered responsible for their technology. In the end, countries' observable decisions taken in the past have resulted from a mixture of controlled and uncontrolled factors.

A possible measure of "environmental responsibility" could be the gap between the total possible deforestation and the BAU deforestation,  $\bar{d}_i - d_i^{bau}$ , that is, the contribution on a voluntary basis to pristine nature. However such a measure would attribute the same merit to countries with the same gap but with large differences in potential contributions, because some countries have much larger  $\bar{d}_i$  than others, a characteristic that can be considered outside the sphere of influence and those that do not, is no easy thing. Assuming this can be done in a satisfactory way, two main ethical principles have been studied. First, the *compensation principle* requires that some transfers be organized to neutralize the influence of factors that are not under the agents' responsibility. Second, according to the *natural reward principle* differences that are due to the exercise of variables under control should be respected.

<sup>7</sup>This existing literature on responsibility and compensation did not seem directly applicable to our subject for at least two reasons. Firstly, much of it is about pure redistribution problems, in microeconomic environments without externalities among agents. Our context, by contrast, is akin to a production environment - by turning to cooperation, countries "produce" a social surplus - with unilateral externalities (from the South to the North). Secondly, many analyses of responsibility are developed in a first-best context: all the necessary pieces of information required to draw the consequences of responsibility are supposed available, an assumption we have ruled out in the first place. Put differently, the deforestation problem has a special structure that raises specific moral issues and constraints.

responsibility. This objection is overcome if the responsibility is measured in relative terms, with the ratio:

$$M_i = \frac{\bar{d}_i - d_i^{bau}}{\bar{d}_i}.$$

305 Let us note  $\bar{M} = (1/n) \sum_i M_i$  the average relative responsibility and define  $\Delta M_i = M_i - \bar{M}$ . From the point of view of their contributions to the environment, countries can be partitioned into two subsets, those who are deserving ( $\Delta M_i > 0$ ) and those who are not ( $\Delta M_i \leq 0$ ).

Two possible requirements on transfers can be formulated, where each recognizes, in a particular way, the heterogeneous role played in the past by countries on the deforestation problem.

310 **Axiom 4.3 (d-ER).** *Let  $d$  be a reference vector of deforestation levels. A transfer scheme satisfies "d - Environmental Responsibility" (d-ER) if the baselines offered to "deserving" countries are at least equal to their deforestation level  $d_i$  indicated in  $d$ , whereas the baselines offered to "undeserving" countries are at most equal to  $d_i$ .*

The above requirement is *a priori* silent about the reference vector for baselines, but a natural candidate would grant each country its BAU level. This is the possibility carrying more weight currently; more precisely, the suggestion under scrutiny is to set up baselines on the basis of the average of historical national levels observed over a time period (usually 10 years)<sup>8</sup>. In a way, this suggestion assumes some libertarian perspectives (Nozick, 1974), according to which past practices are all vested rights. In addition, we saw in Section 4.1 that when baselines are exactly equal to BAU levels, IR is satisfied worldwide.

**Axiom 4.4 (I-ER).** *A transfer scheme satisfies "Incremental - Environmental Responsibility" (I-ER) if the baseline to country  $i$  is an increasing function of its departure from average relative environmental responsibility,  $\Delta M_i$ .*

Note that I-ER could admit a more demanding form, by imposing that the baseline offered to country  $i$  be a strictly increasing - instead of simply increasing - function of  $\Delta M_i$ .

---

<sup>8</sup>It is related to BAU levels if one believes that past behaviors are acceptable predictors for projected behaviors.

The two axioms can add up to embody the more general requirement to conceive baselines as deviations around a reference vector (for example the BAU), where the sign and the extent of the deviation by a particular country depends on its environmental responsibility.

#### 4.4 The $d$ -ER / NRES tension

330 If baselines comply with  $d$ -ER they generally propose a differential treatment to different countries, unless all the coordinates in  $d$  are identical. By contrast, offering identical baselines to all countries can avoid restricted envy in the South. Notice however that identical baselines are sufficient but not necessary to entail NRES. In general, no envy is closely related, though not identical, to equality. By and large, intuition suggests there is a difficulty to combine NRES  
335 and  $d$ -ER, but could this tension be ascertained? At least, this can indeed be proven when the reference deforestation vector  $d$  is given by the BAU.

**Theorem 4.5.** *Any transfer scheme that satisfies  $d$ -ER where  $d$  is fixed at the BAU,  $d = d^{bau}$ , does not respect NRES when countries are sufficiently heterogenous.*

*Proof.* Appendix D. ■

340 So, the goal of avoiding envy leads not only to disregard individual rationality, as we saw in section 4.2, but also to overlook a form of environmental responsibility.

#### 4.5 Some possible formulas for baselines

Cooperation produces a surplus of well-being. In a first best informational setting, it would be natural to set baselines in order to redistribute at least part of this surplus to the South. But this is  
345 not the path we take in our second best setting. Our ambition is simply to differentiate between countries according to their environmental responsibility, and the redistribution of welfare will be only an indirect consequence.

One can design at least three families of baselines that comply with the two axioms  $d^{bau}$ -ER and I-ER, and IR as well.

Under the first family, baselines are equal to the BAU levels plus a *bonus* for countries who are "deserving" ( $\Delta M_i > 0$ ):

$$d_i^{b1} = \alpha_i [\Delta M_i]^+ \sum_{h=1}^m (\bar{d}_h - d_h^{bau}) + d_i^{bau}, \quad \alpha_i \geq 0, \quad (11)$$

where the function  $[.]^+$  maps its argument onto itself when it is positive, and onto zero otherwise. The bonus is the total contribution to the environment by countries multiplied by country  $i$ 's "relative merit". The coefficients  $\alpha_i$  are used to adjust the weight of the bonus in relation to that of the BAU. By construction, transfers where baselines are given by (11) complies with  $d^{bau}$ -ER and I-ER. Clearly IR in the South also obtains (since  $d_i^{b1} \geq d_i^{bau}$ ), but it may be compromised in the North, except if each coefficient  $\alpha_i$  is sufficiently close to zero.

Under the second family, baselines are equal to the BAU levels possibly corrected by a *malus* for countries that are "undeserving" ( $\Delta M_i < 0$ ):

$$d_i^{b2} = \alpha_i [\Delta M_i]^- \sum_{h=1}^m (\bar{d}_h - d_h^{bau}) + d_i^{bau}, \quad \alpha_i \geq 0, \quad (12)$$

where the function  $[.]^-$  maps its argument onto itself when negative, and onto zero otherwise. By construction again,  $d^{bau}$ -ER and I-ER are satisfied. IR holds in the North (since  $\sum_i d_i^{b2} \leq \sum_i d_i^{bau}$ ), but undeserving countries in the South might reject the mechanism, unless their coefficients  $\alpha_i$  are sufficiently small. Those coefficients cannot be too large anyway, in order to ensure  $d_i^{b2} \geq 0$ .

The last family combines the bonus / malus of the two previous possibilities:

$$d_i^{b3} = \alpha_i \Delta M_i \sum_{h=1}^m (\bar{d}_h - d_h^{bau}) + d_i^{bau}, \quad \alpha_i \geq 0. \quad (13)$$

It complies with  $d^{bau}$ -ER and I-ER, and there exists "small" coefficients  $\alpha_i$  such that IR obtains. Note also that if  $\alpha_i = \alpha$ ,  $\forall i$ , then IR in the North is respected, even for a large value

of  $\alpha$ . Indeed in that case:

$$\begin{aligned}\sum_i d_i^{b3} &= \alpha \left[ \sum_h (\bar{d}_h - d_h^{bau}) \right] \sum_i \Delta M_i + \sum_i d_i^{bau} \\ &= \sum_i d_i^{bau},\end{aligned}$$

365 an equality that implies IR in the North (Proposition 4.1).

Under any of those formulas, when setting the baselines the regulator take the risk of rejection by the South or to violate IR in the North. This suggests a repeated “trial and error” process where baselines are first set at the BAU levels (with  $\alpha_i = 0$ ), and then parameters  $\alpha_i$  are progressively increased to meet the axioms  $d^{bau}$ -ER and I-ER and up to the point where a  
370 problem occurs. A fall back to the previous parameter values is then desirable.

And finally, a transfer scheme REDD\* where baselines are given by any of those three possibilities satisfies  $d^{bau}$ -ER and, therefore, may violate NRES if countries are too different (Theorem 4.5).

## 5 Summary

375 This article proposes a class of incentive mechanisms, called REDD\*, to curb deforestation efficiently in tropical countries. It is derived from the class of *compensation mechanisms* and adapted to the context of international negative externalities where no tax can be imposed on “polluters”. In summary, the proposed class of mechanisms allows one to choose some combinations of fairness axioms - individual rationality (IR), a form of no-envy (NRES), and two new  
380 axioms of environmental responsibility (d-ER and/or I-ER) - without losing Pareto optimality.

A first interesting conclusion is that IR, d-ER and I-ER can be compatible. There is no unavoidable and extreme trade-off between acceptability and environmental responsibility. Ultimately, such an arrangement could allay the fears of those who, perhaps rightly, warn that setting baselines equal to the business-as-usual produces perverse incentives overtime: “If I  
385 deforest more today, tomorrow my payments will automatically be greater”. But as soon as

baselines also depend on the environmental responsibility, such a calculation is no longer necessarily true. Less deforestation today will produce, *ceteris paribus*, a premium for tomorrow and may trigger a virtuous circle.

390 A tension exists however between no-envy and IR, and between no-envy and environmental responsibility. The first requirement tends to favor equal baselines for all, whereas the other two requirements calls for different baselines. Future research could explore more deeply the reasonable compromises between these two requirements.

## References and Notes

1. Busch, J., B. Strassburg, A. Cattaneo, R. Lubowski, A. Bruner, D. Rice, A. Creed, R. Ashton, and F. Boltz (2009). "Comparing climate and cost impacts of reference levels for reducing emissions from deforestation". *Environmental Research Letters*, 4 : 044006.
2. Chen, Y., Gazzale, R. (2004). "When Does Learning in Games Generate Convergence to Nash Equilibria? The Role of Supermodularity in an Experimental Setting". *American Economic Review*, American Economic Association, vol. 94(5), pages 1505-1535, December.
3. Danziger, L., Schnytzer, A., 1991. "Implementing the Lindahl voluntary-exchange mechanism". *European Journal of Political Economy*, 7 (1), 55-64.
4. Fleurbaey, M., 1994. "On fair compensation", *Theory and Decision* 36 : 277-307.
5. Fleurbaey, M., 2008. *Fairness, Responsibility, and Welfare*, Oxford University Press.
6. Fleurbaey M. and Maniquet F., 2011, Compensation and responsibility, in K. J. Arrow, A. K. Sen, K. Suzumura eds., *Handbook of Social Choice and Welfare*, vol. 2, North-Holland.
7. Figuières C., Leplay S., Midler E. and Thoyer S., "The REDD mechanism to curb deforestation : a well-defined system of incentives ?", Fourth World Congress on Environmental and Resource Economics, Montréal June-July 2010.
8. Foley, D., 1967, Resource allocation and the public sector. *Yale Economic Essays* 7, pp. 45-98.
9. Alain Karsenty, A., Ongolo, S., 2011, "Can 'fragile states' decide to reduce their deforestation? The inappropriate use of the theory of incentives with respect to the REDD mechanism", *Forest Policy and Economics*, In Press, Corrected Proof, Available online 16 June 2011, ISSN 1389-9341, DOI: 10.1016/j.forpol.2011.05.006. (<http://www.sciencedirect.com/science/article/pii/S1389934111000748>)



10. Kolm, S.-Ch., 1971, *Justice et Équité*, Paris: CEPREMAP (reprint, Paris: CNRS, 1972).
11. Kolm, S.-C., 1995, "The Economics of Social Sentiments: the Case of Envy". *The Japanese Economic Review*. Retrieved from <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1468-5876.1995.tb00005.x>
- 420 12. Maniquet F., 1999. "L'équité en environnement économique", *Revue Économique*, Programme National Persée, vol. 50(4), pages 787-810.
13. Nozick, R., 1974. *Anarchy, State, and Utopia*, New York: Basic Books.
14. Parker, C., Mitchell, A., Trivedi, M., Mardas, N., 2008. *The Little REDD Book: A Guide to Governmental and Non-governmental Proposals to Reducing Emissions from Deforestation and Degradation*. Global Canopy Programme, Oxford.
- 425 15. Roemer, J., 1985, "Equality of Talent", *Economics and Philosophy*, 1, p.151-87.
16. Roemer, J., 1996. "Theories of distributive justice", Cambridge, MA: Harvard University Press.
17. Steinhaus, H., 2008. The problem of fair division, *Econometrica* 16, 101-104.
- 430 18. Tinbergen, J., 1946, *Redelijke Inkomensverdeling*, Haarlem: De Gulden Pers.
19. Varian, Hal R., 1974. "Equity, envy, and efficiency," *Journal of Economic Theory*, Elsevier, vol. 9(1), pages 63-91, September.
20. Varian, H.R., 1994a. "A solution to the problem of externalities when agents are well-informed". *American Economic Review*, 84 (5), 1278-1293.

## Appendix

### A Pareto optimal allocations

Pareto optimal allocations can be found as a solution to the program:

$$\begin{aligned} & \max_{\{d_i\}_{i=1}^m, \{y^i\}_{i=1}^m, y^n} u_n \left( \sum_i d_i \right) + y^n \\ & \text{s.t.} \begin{cases} u_i(d_i) + y^i \geq \bar{U}^i, & i = 1, \dots, m, \\ y^n + \sum_i y^i = \Omega. \end{cases} \end{aligned}$$

The Lagrangian for this problem is:

$$\mathcal{L} = u_n \left( \sum_i d_i \right) + y^n + \sum_{i=1}^m \sigma_i [u_i(d_i) + y^i - \bar{U}^i] + \lambda \left( y^n + \sum_i y^i - \Omega \right)$$

The necessary conditions for optimality read as:

$$\frac{\partial \mathcal{L}}{\partial d_i} = u'_n + \sigma_i u'_i = 0, \quad i = 1, \dots, m, \quad (14)$$

$$\frac{\partial \mathcal{L}}{\partial y^i} = \sigma_i + \lambda \leq 0, \quad i = 1, \dots, m, \quad (15)$$

$$\frac{\partial \mathcal{L}}{\partial \sigma_i} = u_i(d_i) + y^i - \bar{U}^i = 0, \quad i = 1, \dots, m, \quad (16)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = y^n + \sum_i y^i - \Omega = 0, \quad (17)$$

$$\frac{\partial \mathcal{L}}{\partial y^n} = 1 + \lambda \leq 0. \quad (18)$$

$$\sigma_i [u_i(d_i) + y^i - \bar{U}^i] = 0, \quad i = 1, \dots, m. \quad (19)$$

In the sequel we focus on Pareto optimal allocations that involve strictly positive values for  $y^i, i = 1, \dots, m$  and  $y^n$ . Hence, conditions (15) and (18) must be satisfied as equalities. Then, from the resulting equations:

$$\sigma_i = 1, \quad i = 1, \dots, m.$$

Using this information in (14), one can deduce:

$$u'_i = -u'_n,$$

as indicated in the text by expression (1).

## 440 **B A myopic adjustment process**

Rule out complete information and common knowledge. Imagine that countries do not know each others' preferences; assume they are myopic and, at each announcement stage, they proceed by tâtonnement to find  $t^s_i$  and  $t^n_i$ . This kind of process could correspond to an international repeated negotiation, where, at each period, each and every country  $i$  in the South and the North can adjust their subsidy level as follows:

$$\begin{cases} t^s_{i,t+1} = t^n_{i,t}, \\ t^n_{i,t+1} = t^n_{i,t} - \gamma [U^n_1(D_t, y^n_t) + t^s_{i,t} U^n_2(D_t, y^n_t)] \end{cases}, \quad (20)$$

with  $\gamma > 0$  a parameter.

Along this myopic process, a southern country will match its level of transfer at  $t + 1$  with the one from the North at  $t$ . And the north will adjust its chosen level of transfer, if it sees that there is a marginal gain (respectively loss) from increasing (resp. decreasing)  $D_t$ . Then it will  
445 decrease (resp. increase)  $t^n_i$  proportionally.

**Proposition B.1.** *Assume that countries do not know each others' preferences and that each and every country behaves myopically as defined by the above adjustment process (20). Then if the mechanism is repeated over time, it converges asymptotically to a Pareto Optimum.*

*Proof.* System (20) can be written as a matrix equation:

$$\begin{bmatrix} t_{i,t+1}^s \\ t_{i,t+1}^n \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\gamma U_2^n & 1 \end{bmatrix} \begin{bmatrix} t_{i,t}^s \\ t_{i,t}^n \end{bmatrix} + \begin{bmatrix} 0 \\ -\gamma U_1^n \end{bmatrix} \quad (21)$$

To simplify notations, define:

$$\begin{aligned} t_{i,t+1} &= \begin{bmatrix} t_{i,t+1}^s \\ t_{i,t+1}^n \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 1 \\ -\gamma U_2^n & 1 \end{bmatrix}, \\ t_{i,t} &= \begin{bmatrix} t_{i,t}^s \\ t_{i,t}^n \end{bmatrix}, \quad \text{and } b = \begin{bmatrix} 0 \\ -\gamma U_1^n \end{bmatrix}. \end{aligned}$$

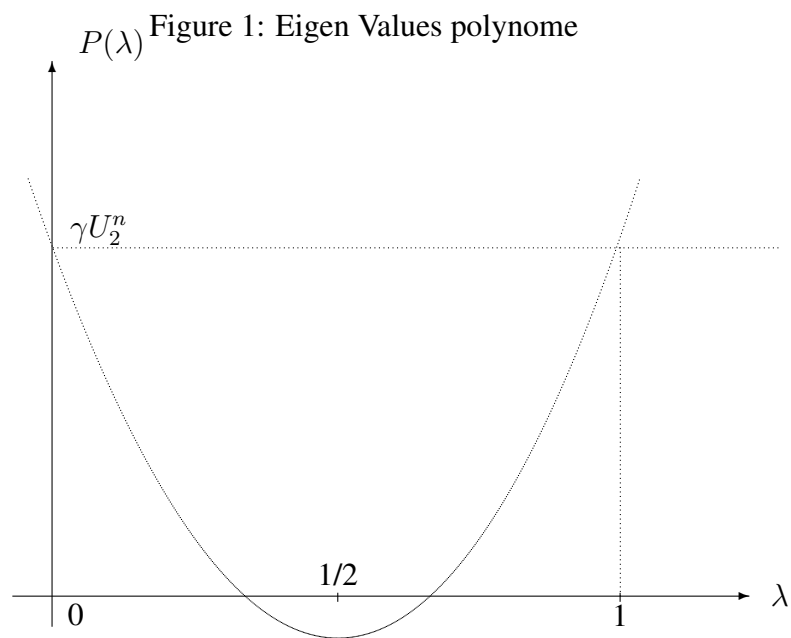
Then (21) becomes:

$$t_{i,t+1} = At_{i,t} + b \quad (22)$$

As one can check, a steady state, where  $t_{i,t+1}^n = t_{i,t}^n$ , is Pareto optimal. Indeed:

$$\begin{aligned} t &= t - \gamma(U_1^n + tU_2^n), \\ \iff t &= \frac{-U_1^n}{U_2^n}. \end{aligned} \quad (23)$$

450 It remains to notice that the last equation gives the expression for a Pareto optimal price  $t$ . We can infer the stability of the stationary states by studying the transition matrix  $A$ . The eigenvalues,  $\lambda_1$  and  $\lambda_2$ , of matrix  $A$  solve  $P(\lambda) = \lambda^2 - \lambda + \gamma U_2^n = 0$ . If  $\gamma < \frac{U_2^n}{4}$ , the shape of  $P$  is presented in figure 1. So  $(\lambda_1, \lambda_2) \in ]0, 1[ \times ]0, 1[$  and consequently a (Pareto optimal) steady state is asymptotically stable. ■



## C Individual rationality

Individual rationality for a southern countries requires:

$$u_i(d_i^{bau}) \leq u_i(d_i^*) + t^*(d_i^b - d_i^*).$$

Since  $u'_i(d_i^*) = t^*$  at a Pareto optimal allocation which is implemented via the mechanism, the requirement can be rewritten:

$$u_i(d_i^{bau}) \leq u_i(d_i^*) + u'_i(d_i^*)(d_i^b - d_i^*).$$

Because of concavity, when  $d_i^b \geq d_i^{bau}$ , the above inequality is guaranteed. By continuity, clearly there also exists a threshold lower bound baseline  $\bar{d}_i^b \leq d_i^{bau}$  for which this condition continues to be filled. It solves the equation:

$$u_i(d_i^{bau}) = u_i(d_i^*) + u'_i(d_i^*)(\bar{d}_i^b - d_i^*).$$

So it is:

$$\bar{d}_i^b = d_i^* + \frac{u_i(d_i^{bau}) - u_i(d_i^*)}{u'_i(d_i^*)}.$$

Note that  $d_i^* \leq \bar{d}_i^b \leq d_i^{bau}$ . To sum up, when the price signal is  $u'_i(d_i^*) = t^*$ , and because of concavity, there exists a range of baselines such that IR is obtained for country  $i$ . This range is called  $B_i = [\bar{d}_i^b, d_i^{bau}]$  in the text.

Using a similar reasoning, the range of acceptable baselines for the North is  $B_n = [0, \hat{D}]$ , where:

$$\hat{D} = \sum_{i=1}^m d_i^* + \frac{u_n(\sum_{i=1}^m d_i^{bau}) - u_n(\sum_{i=1}^m d_i^*)}{u'_n(\sum_{i=1}^m d_i^*)} > \sum_{i=1}^m d_i^{bau}.$$

## D Tension between $d^{bau}$ -ER and NRES

Remember that NRES requires

$$u_i(s_i^{-1} \circ s_j(d_j^*)) + t^*(d_j^b - s_i^{-1} \circ s_j(d_j^*)) \leq u_i(d_i^*) + t^*(d_i^b - d_i^*), \quad \forall i, j.$$

460 Rewrite this as:

$$\begin{aligned}
u_i(s_i^{-1} \circ s_j(d_j^*)) &\leq u_i(d_i^*) + t^*(s_i^{-1} \circ s_j(d_j^*) - d_i^*) + t^*(d_i^b - d_j^b), \quad \forall i, j. \\
u_i(s_i^{-1} \circ s_j(d_j^*)) &\leq u_i(d_i^*) + u'_i(d_i^*)(s_i^{-1} \circ s_j(d_j^*) - d_i^*) \\
&\quad + u'_i(d_i^*)(d_i^b - d_j^b), \quad \forall i, j.
\end{aligned} \tag{24}$$

We know that, because of concavity it is true that:

$$u_i(s_i^{-1} \circ s_j(d_j^*)) \leq u_i(d_i^*) + u'_i(d_i^*)(s_i^{-1} \circ s_j(d_j^*) - d_i^*),$$

But the last term of the inequality (24) is necessarily negative for some countries and may compromise the test for no-envy. Assume, without loss of generality, that country  $j$  is "deserving" ( $\Delta M_j > 0$ ) whereas country  $i$  is not ( $\Delta M_i < 0$ ). Then, if the baselines are chosen so as to meet  $d^{bau}$ -ER:

$$d_j^b \geq d_j^{bau} \text{ and } d_i^b \leq d_i^{bau}.$$

Assume also that  $d_i^{bau} < d_j^{bau}$ . So far, we can write:

$$|d_i^b - d_j^b| \geq |d_i^{bau} - d_j^{bau}|,$$

and

$$\begin{aligned}
&u_i(d_i^*) + u'_i(d_i^*)(s_i^{-1} \circ s_j(d_j^*) - d_i^*) + u'_i(d_i^*)(d_i^b - d_j^b) \\
&\leq u_i(d_i^*) + u'_i(d_i^*)(s_i^{-1} \circ s_j(d_j^*) - d_i^*) + u'_i(d_i^*)(d_i^{bau} - d_j^{bau}).
\end{aligned}$$

Now because the values  $d_i^{bau}$  and  $d_j^{bau}$  are deduced from the utility functions, they can be set arbitrarily so that:

$$u_i(d_i^*) + u'_i(d_i^*)(s_i^{-1} \circ s_j(d_j^*) - d_i^*) + u'_i(d_i^*)(d_i^{bau} - d_j^{bau}) < u_i(s_i^{-1} \circ s_j(d_j^*)),$$

an inequality that implies

$$u_i(d_i^*) + u'_i(d_i^*)(s_i^{-1} \circ s_j(d_j^*) - d_i^*) + u'_i(d_i^*)(d_i^b - d_j^b) < u_i(s_i^{-1} \circ s_j(d_j^*)),$$

in violation of NRES. QED.