Track-and-Trade: A liability approach to climate policy*

Etienne Billette de Villemeur† Justin Leroux‡

Abstract

We argue for the creation of a carbon liabilities market to address climate change. Each period, countries would be made liable for their share of responsibility in current climate damage. Because liabilities could be traded like financial debt, this decentralizes the choice of the rate by which countries discount future benefits and damage, thus yielding first-best emissions patterns. Rather than being based on an expected discounted sum of future marginal damage (as with a carbon tax or tradable emission permits) our proposal relies only on realized damage and on the well-documented emission history of countries.

Keywords: Carbon Liabilities, Climate Policy, Market Instruments.
JEL classification code: Q54, H23.

*The authors wish to thank Geir Asheim, Hassan Benchekroun, Yann Bramoullé, Alessio D'Amato, Julien Daubanes, Georges Dionne, Paul Ekins, Gérard Gaudet, André Grimaud, Christian Gollier, Jérémy Laurent-Luchetti, Peter Hartley, Marcus Hermann, Louis Hotte, Shi-Ling Hsu, Yann Kervinio, Sonja Köke, Stéphane Lambrecht, Derek Lemoine, Leslie Martin, Stephen Salant, Marc Santugini, Charles Séguin, Jens Steffek, Nicolas Treich, Edilio Valentini, Bruno Versaevel, Mirko Wiederholt, Ralph Winkler as well as participants of the 2012 Occasional California Workshop on Environmental and Resource Economics, the Montreal Resource and Environmental Economics Workshop, the 2012 Paris Environmental and Energy Economics Seminar (PEEES), and the 2014 GERAD Conference on Dynamic Games.

†LEM UMR 9221, Université de Lille, France.

‡HEC Montréal, CIRANO & CRÉ, Canada.

1 Introduction

The project of a worldwide harmonized optimal carbon tax policy is bound to fail. The reasons for that are at least twofold. First, the heterogeneities in beliefs about climate damage and in the weight given to future generations (through the discount factor) forbid any agreement on a single tax rate. Next, and perhaps more importantly, the very principle of the optimal carbon tax—or of the optimal cap-and-trade program, for that matter[1]—makes it politically unacceptable: it asks current emitters to pay for damage that may or may not happen to future generations. The willingness to pay simply isn't there.[2]

Arguments in favor of a carbon tax have been around for decades.[3] However, despite the significant achievements made over time in fine-tuning the inner workings of a well thought-out carbon tax and predicting its consequences (see, e.g., Metcalf and Weisbach, 2009; Fullerton et al., 2012, and Golosov et al., 2014, for state-of-the-art analyses), one cannot ignore the disheartening fact that very few countries, states, provinces or cities have thus far implemented a carbon tax. In addition, when a carbon tax does exist, the tax rate is often set below most reasonable estimates of the social cost of emissions, leading to suboptimal incentives to reduce emissions.[4] To this day, and despite its many desirable features (like efficiency, ease of implementation, revenue raising, and wholesome incentives for innovation, among others), enacting a carbon tax still requires much convincing.[5] Likewise, while some cap-and-trade programs exist (RGGI, WCI, EU-ETS), they are systematically weakened by the issuance of

[1] Like the optimal carbon tax, the optimal number of allowances is calculated on the basis of a present discounted value of future climate damage.

[2] Jenkins (2014) estimates at \$2-\$8 per ton of CO_2 the willingness to pay for the US population, whereas the full social cost of carbon emitted today is much higher (with estimates ranging from \$15 to \$150 per ton of CO_2).

[3] William Nordhaus was probably the first economist to put climate policy at the forefront of his research agenda (e.g., Nordhaus, 1977). Note, however, that a tax to address air pollution had already been proposed by Hardin (1968) in his celebrated contribution to the problem of externalities. The link between the concentration of gases in the atmosphere and the surface temperature of the earth had been established by Fourier (1824). Svante Arrhenius, who went on to obtain the Nobel Prize in Chemistry in 1904, made one of the first attempts to calculate the effect of a doubling of carbon dioxide concentration on temperature (Arrhenius, 1896). For more on the evolution of the awareness of mankind's impact on climate, see Kellogg (1987).

[4] According to the World Bank (2014), 14 countries and one Canadian province taxed CO2 emissions as of June 2014. Aside from Denmark, Finland, Sweden, Switzerland, and the province of British Columbia, carbon tax rates are significantly lower than the social cost of \$37/tCO2e estimated by the US government's Interagency Working Group on Social Cost of Carbon (US Government, 2013). These estimates are obtained from three established integrated assessment models: DICE, FUND and PAGE.

[5] The mere existence of Hsu's excellent book on the advantages of a carbon tax (Hsu, 2011) long after the first proposals is telling.

an excessive number of permits to be circulated (Hsu 2011).[6] We argue that the project of a worldwide carbon tax or cap-and-trade program should be abandoned in favor of an economically efficient policy that decentralizes beliefs and discount factors—thus allowing for disagreements in these dimensions—while exacting payments that are commensurate to damage that is *presently* occurring. We propose such a policy.

At the heart of our proposal lies the realization that a Pigovian climate policy need not exact the full payment of the social cost of carbon at the time of emission to yield first-best incentives. Consider a simple two-period illustration with a single decision maker. She obtains utility $u(e)$ from emitting e units of emissions in Period 1, which lead to $d \times e$ units of damage in Period 2. Hence, d can be thought of as the social cost of carbon per unit of emissions. Under an optimal carbon tax, the decision maker is required to pay (βd) per unit at the time of emission, where β is the discount factor. This leads her to choosing the socially optimal emissions level upon maximizing her total (period-one) payoff:

$$\max u(e) - (\beta d) e,$$

so that $u'(e^*) = \beta d$.

Suppose instead that the decision maker is only required to pay the (undiscounted) damage, (de), when it occurs in Period 2. Her objective is now to maximize the discounted sum of her payoffs, $u(e) - \beta(de)$, but her emission decision is the same as under the optimal carbon tax: $u'(e^*) = \beta d$. Notice, however, that this second approach allows for damage to be observed before payments are made, so that the decision maker never has to pay out-of-pocket amounts linked to damage that is still hypothetical. From the standpoint of political acceptability, this is a key difference.

Following this insight, we propose a new climate policy based on what we call *carbon liabilities*: Emitting CO_2 in the atmosphere[7] would be accompanied by the issuance of a carbon liability, a legal duty for the bearer to pay damages over time as climate damage occurs.[8] Simply put, our proposal consists in converting

[6] According to economic theory, the method of allocation of permits—whether they are given away or auctioned off—is immaterial for efficiency. Yet, governments are often compelled to give away permits, and to give away too many of them, due to political concerns. See Chapter 3 of Hsu (2011, chapter 3) for a recount of the many loopholes built in virtually all cap-and-trade programs, such as carbon offsets and a slew of exemptions, which severely limit the reach of cap-and-trade policies all on account of political acceptability.

[7] For expositional purposes, we shall speak only in terms of CO_2.

[8] Clearly, the issue of determining the magnitude of anthropogenic climate damage is a difficult one. Note, however, that this question is by no means resolved with a carbon tax.

CO_2 emissions into national (financial) debt. The debt would be owed to an international climate fund that could, for example, compensate participating countries to the tune of the climate damage incurred.[9] This requires nations to only pay for damage caused by their emissions 'upon delivery' rather than pay today for the infinite discounted sum of future marginal climate damage up front. Carbon liabilities would not expire, but would instead decay at the same rate as atmospheric CO_2, all the while holding bearers accountable for paying *carbon damages* over time.[10] By doing so, this avoids the principal-agent problem that arises when current generations pay the most while future generations benefit the most, as is the case for all carbon pricing policies proposed to date.

Despite their intention to decentralize the regulation of CO_2 emissions, existing carbon pricing policies are actually very centralized: they base the tax rate (or the cap) on a discount factor and on beliefs about future damage that are those of a planner. Pigovian taxation was intially developed in a single-period setting, where the externality was entirely characterized by the magnitude of its damage. The climate problem adds two dimensions to the externality: uncertainty and temporality. The Pigovian logic must therefore be revisited accordingly. If we take the 'Pigovian logic' to mean the internalization by agents of all aspects related to the externality, then our proposal applies this logic in full by letting countries internalize their own beliefs about future damage and their own view on discounting.

Note that our proposal is also very Coasian in spirit because it achieves efficiency via decentralization through trade, as we now illustrate. To build upon our previous example, consider now that carbon liabilities are tradable on a world market at price p and denote by X the net quantity of liability sold by the decision-making country. Thus, it maximizes $u(e) - \beta d(e - X) - pX - C[p(e - X)]$, where C is a convex cost function associated with holding climate

In fact it is made even worse, because predicting the damage caused by a ton of carbon emitted today decades or centuries into the future is a much more heroic feat than assessing the responsibility of past emissions in (observed) damage ocurring today.

[9] How revenues are used is beyond the scope of this paper. We shall simply assume here that revenues are allocated independently of emissions. We address this issue in a separate paper (Billette de Villemeur and Leroux, 2011).

[10] Throughout the paper, we use the word 'damage' to mean the harm caused by greenhouse gas emissions. By contrast, the plural 'damages' refers to the amount a nation is legally required to pay as the result of harm caused by its emissions.

debt.[11],[12] Optimizing in e and in X yields the following first-order conditions:

$$\begin{aligned} e: \quad & u'(e) && = \beta d + pC'\left[p\left(e - X\right)\right] \\ X: \quad & p && = \beta d + pC'\left[p\left(e - X\right)\right] \end{aligned} \tag{1}$$

Accordingly, countries who expect low discounted marginal damage—i.e., with low βd's, either because they are skeptical about climate change or because they have little concern for the future—will purchase liabilities. They will do so up to the point where the marginal cost of accumulating climate debt compensates the gap between its expectation of discounted marginal damage and the liability price. Likewise, countries with high expected discounted marginal damage will sell their liabilities and possibly accumulate 'carbon assets', should $e - X$ fall below zero. As a country accumulates assets, the marginal benefit of these assets falls (due to lack of trust that they will be repaid, or simply for reasons of portfolio diversification) until it closes the gap between p and its own βd. In equilibrium, the marginal utility of all countries is equal to the market-clearing price, p.

Relation to the literature

Our contribution is related to the literature on green accounting and the burgeoning literature on stakeholder value, which incorporate the externalities generated by economic agents into national accounting (see Weitzman 1976; Hartwick, 1990; or, more recently, Cairns, 2004, Cairns and Lasserre, 2006) or argue in favor of moving away from the conventional *share*holder value maximization (MaGill et al., 2013), respectively. In our proposal, the financiarization of carbon debt can be seen as an instrument to evaluate the 'climate debt' of a country to the world (i.e. to global stakeholders).

Many authors have proposed alternatives to the carbon tax and to cap-and-trade programs with the aim of facilitating global cooperation. Closely related to ours is the proposal of Gersbach and Winkler (2012) to revert the Pigovian logic effectively transforming the carbon tax into a carbon subsidy. Countries would initially contribute large lump sums to an international climate fund and

[11] Note the minus sign in front of pX. A seller of liabilities actually *pays* the buyer to hold them in its stead.

[12] In practice, economies often incur extra costs due to increased borrowing constraints as debt grows (Wachtel and Young, 1987; Engen and Hubbard, 2004; Laubach, 2009). From a technical standpoint, the introduction of a convex cost of holding debt guarantees an interior solution.

would be refunded over time proportionally to their abatement efforts relative to a business-as-usual scenario. While we retain their use of an international climate fund to collect and redistribute funds to countries, we view the large entry cost as a significant obstacle to their 'global refunding scheme'. In fact, the very aim of our proposal is to reduce the entry cost of countries as much as possible while retaining proper incentives.

The idea of using liabilities as a means to control externalities traces back to Calabresi (1970) and was recently compared to corrective taxation in Shavell (2011). On the one hand, regulation (i.e., taxation) is costly even in the absence of damage, whereas a liability approach only kicks in when harm actually occurs. On the other hand, a liability approach is typically more informationally demanding because it requires establishing tort (Kolstad et al, 1990; Shavell, 2011). Hence, a liability approach is likely to be more appropriate in situations where damage is highly uncertain but where its source can be easily established. This is precisely the case of climate change, where the magnitude of the damage is typically unknown ex ante but the responsibility of countries towards CO_2 concentration can be readily established thanks to available data on cumulated CO_2 emissions per country (e.g., from the World Resource Institute or the World Bank databases).[13] The general argument echoes that of Shleifer (2012) according to which the need for regulation arises where litigation is ineffective. Underlying this line of reasoning is the notion that turning to litigation (read liability) is a most natural reflex that should be left unhindered whenever it is an efficient option.

In a climate change context, liabilities have very recently been proposed as a means to make global cooperation more effective and less costly in the long run (Gampfer, Gsottbauer and Delas, 2014). Their merit lies in the fact that countries are more likely to adhere to an agreement on emissions reductions if they believe they will be compensated fairly for future climate damage incurred (Gampfer, 2014). The argument in these works is one of fairness. Ours is mainly one of economic efficiency.

The use of liabilities to address the climate problem is further supported by insights from the cost-sharing literature. An important lesson to be learned

[13] The liability approach is usually discussed in the context of tort law, involving private parties and legal costs attached to lawsuits, to establishing due care and negligence. By contrast, the liability approach we consider here is public, in the sense that it involves countries, and would consist in an automatic procedure where the negligence rule plays no role. Countries would be held responsible for climate damage according to their emissions emitted after an agreed-upon starting date.

from that literature is that the best properties of a payment scheme—whether in terms of efficiency, incentives, or even fairness—arise from mimicking the physical features of the externality to be managed (Moulin, 2002). Climate damage being a problem caused by the *stock* of CO_2 in the atmosphere, it is natural to condition payment on emission stocks rather than on emission flows as do the carbon tax and cap-and-trade programs. By making explicit—and, most importantly, financial—the somewhat intangible carbon debt that mankind accumulates along with atmospheric CO_2, carbon liabilities do just that.

The remainder of the paper is organized as follows. We first formally establish in Section 2 that turning carbon debt into financial sovereign debt leads to first-best incentives to reduce emissions without charging the full social cost of carbon up front. In Section 3 we introduce the instrument of carbon liabilities *per se*. Unlike carbon debt, which is based only on an estimation of marginal climate damage, payments are determined relative to realized damage. Section 4 shows that free trade leads to efficiency even in the presence of heterogeneous beliefs about future climate damage and heterogeneous discount factors. Hence, a track-and-trade policy allows for full decentralization along these dimensions, unlike a carbon tax or cap-and-trade programs. Section 5 concludes.

2 A carbon debt that yields correct incentives

Let $\{D_t\}_{t=0}^{+\infty} = \left\{\sum_i D_t^i\right\}_{t=0}^{+\infty}$ denote the flow of stochastic damage borne by all countries, indexed by i, as attached to anthropogenic climate change. At any period t, the occurrence and the magnitude of this damage is assumed to be an increasing function of Z_t, the current stock of anthropogenic CO_2 in the atmosphere. Our proposal consists in converting CO_2 emissions into financial debt. More precisely, in each period, all countries are required to contribute to an international climate fund to the tune of $\mu_t Z_t^j$ where $\mu_t = dD_t/dZ_t$ is the marginal climate damage due to anthropogenic emissions[14] and where $Z_t^j = \sum_{s=-\infty}^{t} \gamma^s X_s^j$ is the contribution of country j to the stock Z_t (it is the discounted sum of its past emissions X_s^j, for all $s \leq t$, accounting for their natural decay at rate $1-\gamma$).[15]

[14] Unlike in tort law, we do not aim for "full liability" because it is not optimal to cover the total cost.

[15] It is actually not required to trace back emissions to infinity. In fact, accounting only for, say, post-1990 emissions would result in the very same emissions pattern. The truncation simply amounts to lump-sum transfers to (developed) countries while preserving incentives at

Throughout, we shall work under the usual Pigovian assumption that no single agent has an impact on the marginal damage and that the impact of one's own emissions is negligible. The latter is perfectly consistent with the climate problem. The former implies that we ignore strategic issues associated with the relative size of countries. While not completely realistic, this assumption allows us to focus on the general features of the mechanism.

Proposition 1 *Such a carbon debt scheme yields first-best emission patterns.*

Proof. Under rational expectations, country i evaluates its present net benefit as:

$$PNB^i = \sum_{t=0}^{+\infty} \beta^t \left[B_t^i \left(X_t^i \right) - \mu_t Z_t^i \right], \tag{2}$$

where $B_t^i \left(X_t^i \right)$ is the per-period benefit of country i resulting from its emissions in the current period. Under the assumption that no single agent has an impact on marginal damage, country i chooses an emissions stream $\left\{ X_t^i \right\}_{t=0}^{+\infty}$ such that:

$$\frac{\partial B_t^i}{\partial X_t^i} = E_t \left[\sum_{s=t}^{+\infty} \beta^{s-t} \mu_s \frac{\partial Z_s^i}{\partial X_t^i} \right] = E_t \left[\sum_{s=t}^{+\infty} (\gamma\beta)^{s-t} \frac{\partial D_s}{\partial Z_s} \right]. \tag{3}$$

Each country equalizes its marginal benefit with the expected discounted value of marginal climate damage, thus achieving first-best efficiency.■

Notice that the only information required of the planner to implement our scheme, on top of the well-documented emission history of countries, is $\mu_t = \frac{dD_t}{dZ_t}$: the marginal impact of current *stock* of anthropogenic CO_2 in the atmosphere on the current *flow* of climate damage. While obtaining this information accurately may be no small task, it seems far less daunting to be working with observed data than with predictions over future decades or centuries. Indeed, the information required to implement an efficient carbon tax, τ, or the equivalent cap-and-trade program is the *expected*, *discounted* sum of the marginal impacts of current emissions on *future* climate damage:[16]

$$\tau_t = E_t \left[\sum_{s=t}^{+\infty} \beta^{s-t} \frac{\partial Z_s^i}{\partial X_t^i} \frac{\partial D_s}{\partial Z_s} \right] = E_t \left[\sum_{s=t}^{+\infty} (\gamma\beta)^{s-t} \frac{\partial D_s}{\partial Z_s} \right].$$

the margin.

[16] Using a dynamic stochastic general equilibrium model, Golosov et al. (2014) find a formula for the optimal carbon tax that does not rely on future values of output, consumption or stock of CO_2 in the atmosphere. While a remarkable finding, their formula cannot do away with the discount factor, however.

From a policy standpoint, implementing a carbon debt policy is simpler than implementing a cap-and-trade program. Indeed, carbon debt is issued and allocated systematically based on each country's observed emissions. By contrast, cap-and-trade schemes require a planner to issue and allocate permits with the obvious risks of miscalculation and misallocation, respectively.

The upshot of requiring less of the planner is that much more freedom is left to the countries, thus allowing for more decentralization than, say, a harmonized carbon tax policy or a global cap-and-trade program. Specifically, countries make their own predictions about future damage and work with their own discount factors. However, there is a limit to decentralization: making firms and consumers liable would increase default risk and lead to skewed incentives. By contrast, assigning liabilities at the country level significantly improves solvability and ensures correct incentives because nations are long-lived. In addition, this gives countries sovereignty on how to finance their carbon damages.

The following section presents a liability scheme where payments depend on the *realization* of climate damage, as opposed to having them depend on expectations.

3 From carbon debt to carbon liability

Although there is some evidence that climate change already has an impact on economic outcomes, climate damage remains highly uncertain and volatile. It follows that *ex ante* approaches to climate policy exhibit the unappealing feature of possibly requiring high payments when realized damage is low, or *vice versa*. A liability approach avoids this disconnectedness by linking payments to realized harm.[17]

In practice, while estimations of marginal climate damage recognize the large uncertainties surrounding the realization of damage (Tol, 2005), the notion of marginal damage only applies to the underlying climate patterns, stripped of all uncertainty. To clarify, if we model climate damage to be of the form $D(Z_t) = \tilde{D}(Z_t) + \varepsilon_t$ where all the uncertainty is contained in ε_t, the large body of work on the estimation of the marginal damage of emissions is aimed at obtaining the best estimates of $d\tilde{D}_t/dZ_t$. Accordingly, $\mu_t = d\tilde{D}_t/dZ_t$, so that the carbon

[17] In tort law, one aspect of the debate between the regulatory and the liability approaches we did not yet mention is the fact that payments reflect realized harm in the latter whereas they are based on the possibility of harm in the former. On this, see Shavell (1984, 2011) and Kolstad et al (1990).

debt payments considered up to now do not actually depend on realized harm.

We now turn to a formal presentation of our liability scheme. Assume that payments are adjusted according to realized damage, $D(Z_t)$. More precisely, assume that countries are actually required to pay $\mu_t Z_t^i \mathbb{I}_{D_t}$, where

$$\mathbb{I}_{D_t} = \frac{D(Z_t)}{\tilde{D}(Z_t)} \tag{4}$$

is the ratio of the realized over the predicted damage.

Proposition 2 *The liability rule $\mu_t Z_t^i \mathbb{I}_{D_t}$ is first-best efficient and yields payments proportional to realized climate damage.*

Proof. By definition, $E_s[\mathbb{I}_{D_t}] = 1$ for all $s \leq t$, so that expected payments are unchanged. Hence, from Proposition 1, the liability rule is first-best efficient. Furthermore, $\mu_t Z_t^i \mathbb{I}_{D_t}$ is indeed proportional to realized harm:

$$\mu_t Z_t^i \mathbb{I}_{D_t} = \frac{d\tilde{D}_t}{dZ_t} Z_t^i \frac{D(Z_t)}{\tilde{D}(Z_t)} \tag{5}$$

$$= \frac{\left(d\tilde{D}_t/dZ_t\right)}{\tilde{D}(Z_t)/Z_t} \frac{Z_t^i}{Z_t} D_t(Z_t). \blacksquare \tag{6}$$

Expression (6) displays the fact that liability payments are not only proportional to the realized harm, $D_t(Z_t)$, but also to one's relative contribution to the total stock of CO_2 in the atmosphere, Z_t^i/Z_t. This proportionality is modulated by the predicted ratio of marginal to average damage, which reflects the convexity of the damage function, D_t. If the damage function is linear, payments exactly cover total damage and countries pay in proportion to their emission contributions $\mu_t Z_t^i \mathbb{I}_{D_t} = \frac{Z_t^i}{Z_t} D_t(Z_t)$. If the damage function is convex, total payments add up to more than the realized damage because first-best efficiency requires going beyond full liability.

Remark 1 *Full liability—i.e., charging $\frac{Z_t^i}{Z_t} D_t(Z_t)$—constitutes a practical benchmark because payments rely only on realized damage, and do not require some estimate of the marginal damage function. However, full liability is only first-best efficient if the damage function, D_t, is linear.*[18]

[18] The superiority, in efficiency terms, of marginal-cost pricing over average-cost pricing is well-known. See, e.g., Friedmand and Moulin (1999).

Section 4 below addresses how trade can maintain efficiency if countries have different discount factors and different expectations about future anthropogenic climate damage.

4 Decentralization through trade

If liabilities can be traded, our approach is robust to heterogeneity in discount factors and to diverging forecasts. Indeed, if discount factors and forecasts are country-specific Expression (3) becomes:

$$\frac{\partial B_t^i}{\partial X_t^i} = E_t^i \left[\sum_{s=t}^{+\infty} (\gamma \beta_i)^{s-t} \mu_s \right]$$

where β_i and E_t^i are the discount factor and the expectations of country i, respectively.

Country heterogeneity yields trade opportunities: a market for liabilities leaves it to countries to determine how much liability they wish to hold based on their predictions of future climate damage. Should opinions differ on the likelihood or magnitude of future damage, or on the discount rate, efficiency is maintained through trade as we now show.

Specifically, given a competitive market price, p_t, countries may choose to buy carbon liabilities—and be *paid* to do so—or to sell them, by paying others to hold liabilities in their stead. We introduce a cost for countries of holding financial debt, c_t^i, which we interpret it as country default risk. We assume c_t^j to be increasing, strictly convex, and such that $c_t^i(0) = 0$.

Proposition 3 *Under heterogeneous preferences and beliefs, a tradeable carbon liabilities scheme—where installments are set to current marginal climate damage:* $\mu_t Z_t^i \mathbb{I}_{D_t} = \frac{d\tilde{D}_t}{dZ_t} Z_t^i \frac{D(Z_t)}{\tilde{D}(Z_t)}$*—is first-best efficient.*

Proof. We show that efficiency is robust to heterogeneity in countries' discount factors. The proof assuming heterogeneity in expectations about future damage, E_t^i, proceeds similarly.

Suppose countries have heterogeneous discount factors. Assume country j sells Y_t^j units of liability associated to its current emissons, $X_t^j + Y_t^j$. Its expected present net benefit writes as follows:

$$PNB^j = \sum_{t=0}^{+\infty} \beta_j^t E_0 \left[B_t^j \left(X_t^j + Y_t^j \right) - p_t Y_t^j - \left[\mu_t Z_t^j + c_t^j \left(p_t Z_t^j \right) \right] \right],$$

where,

$$Z_t^j = \gamma Z_{t-1}^j + X_t^j,$$

is the amount of carbon debt held by country j at date t.

Similarly, assume country i purchases Y_t^i units of liability, as measured in carbon stock units. Its expected present net benefit writes as follows:

$$PNB^i = \sum_{t=0}^{+\infty} \beta_i^t E_0 \left[B_t^i \left(X_t^i \right) + p_t Y_t^i - \left[\mu_t Z_t^i + c_t^i \left(p_t Z_t^i \right) \right] \right],$$

where the carbon stock for which country i is now considered to be responsible writes:

$$Z_t^i = \gamma Z_{t-1}^i + X_t^i + Y_t^i.$$

From the point of view of a net seller, the first-order conditions write as follows:

$$\begin{aligned} \frac{\partial B_t^j}{\partial X_t^j} &= E_t \left[\sum_{s=t}^{+\infty} \beta_j^{s-t} \frac{\partial Z_s^j}{\partial X_t^j} \left[\mu_s + p_s c'^j \left(p_s Z_s^i \right) \right] \right] \\ &= E_t \left[\sum_{s=t}^{+\infty} \left(\gamma \beta_j \right)^{s-t} \left[\frac{\partial D_s}{\partial Z_s} + p_s c'^j \left(p_s Z_s^j \right) \right] \right] \\ \frac{\partial B_t^j}{\partial Y_t^j} &= p_t \end{aligned}$$

From the point of view of a net buyer, the first-order conditions are the following:

$$\frac{\partial B_t^i}{\partial X_t^i} = E_t\left[\sum_{s=t}^{+\infty}\beta_i^{s-t}\frac{\partial Z_s^i}{\partial X_t^i}\left[\mu_s + p_s c'^i\left(p_s Z_s^i\right)\right]\right]$$
$$= E_t\left[\sum_{s=t}^{+\infty}(\gamma\beta_i)^{s-t}\left[\frac{\partial D_s}{\partial Z_s} + p_s c'^i\left(p_s Z_s^i\right)\right]\right]$$
$$p_t = E_t\left[\sum_{s=t}^{+\infty}\beta_i^{s-t}\frac{\partial Z_s^i}{\partial Y_t^i}\left[\mu_s + p_s c'^i\left(p_s Z_s^i\right)\right]\right]$$
$$= E_t\left[\sum_{s=t}^{+\infty}(\gamma\beta_i)^{s-t}\left[\frac{\partial D_s}{\partial Z_s} + p_s c'^i\left(p_s Z_s^i\right)\right]\right]$$

It follows that, for all i:

$$\frac{\partial B_t^i}{\partial X_t^i} = p_t,$$

and for all j

$$p_t = \frac{\partial B_t^j}{\partial Y_t^j} = \frac{\partial B_t^j}{\partial X_t^j},$$

yielding efficiency.■

Proposition 3 is in the vein of the First Welfare Theorem. Indeed, carbon liabilities act as tradable Arrow-Debreu-type securities that make markets complete, thus yielding allocative efficiency through decentralization.[19] Furthermore, another consequence of Proposition 3 is that the scheme allows for diverging opinions regarding climate change. It is also noteworthy that the mechanism is robust in the sense of being immune to strategic manipulation both in the discount factor and in the expectations because the final allocation of debt is a competitive market outcome. To sum up, the introduction of a market for carbon debt makes our mechanism robust to misrepresentation and to misreporting.

Morever:

Remark 2 *Because our scheme financiarizes the carbon debt, failure to honor the latter is now no different than a default in the repayment of financial debt.*[20]

[19] It would be interesting to explore whether an equivalent of the Second Welfare Theorem holds. For an attempt at conjugating allocative efficiency with the redistributive aspects of global warming, see Billette de Villemeur and Leroux (2011).

[20] However, there is a difference in the nature of these debts. In the unlikely case of default on financial debt, those affected are creditors who chose to expose themselves to financial risk. By contrast, those affected by default on carbon debt are likely to be residents of vulnerable countries that did not choose to expose themselves to climate risk. We thank Yann Kervinio

The previous remark highlights the seriousness of defaulting on carbon debt. Indeed, because the latter amounts to financial debt, failure to repay comes at a significant reputational cost. This is all the more the case as the agents involved are nations rather than individuals or firms.

Naturally, there is the question of whether countries will participate in the first place, leading to the general theme of coalition formation in international environmental agreements. Because the expected net discounted payoffs of countries are identical under our approach as with an optimal carbon tax, we do not expect to obtain in our setting markedly different conclusions from those of that litterature. Of course, those very important questions deserve to be addressed more carefully, which we leave to future work.

5 Concluding remarks

We propose a new approach to climate policy based on the trading of carbon liabilities between countries that yields the same emissions patterns as the optimal carbon tax. Moreover, we argued that such a mechanism would have significant advantages over the optimal carbon tax: it is less informationally demanding, it allows for disagreements in beliefs and preferences through decentralization, and it exacts payments commensurate to *actual* climate damage.

The key feature of the mechanism is that the prospect of being liable for future damage creates incentives to reduce emissions today. It is a general principle of justice that no party can be liable in the absence of 'constructive notice'.[21] In other words, a country should not be asked to pay for the climate consequences of emissions made before the discovery of the impact of greenhouse gases on the climate. As already mentioned, however, the mechanism does not require tracing emissions back to infinity to provide the right incentives, but can account for anthropogenic emissions starting at some agreed-upon reference date only. It thus does not violate this basic legal principle.[22]

There are at least two important tradeoffs to consider, however. One is that countries may find it tempting to abandon the program as climate damage becomes more and more pronounced in the future. We show in the Appendix that this temptation mirrors the current temptation, which is yielded to in

for this observation.

[21] We thank Shi-Ling Hsu for bringing this issue to our attention.

[22] Obviously, for incentives to exist, 'actual notice' is also needed; *i.e.*, countries must not only be aware of the fact that they are causing harm (constructive notice), but must also be informed that they will be considered liable for future climate damage.

practice, to continually delay the implementation of a carbon tax. We argue that this tradeoff between abandoning a carbon liability scheme in the future and postponing a carbon tax today leans towards a market for liabilities being the more effective policy. Indeed, before abandoning the program, at least some emissions reduction is being achieved.

The other tradeoff arises from the fact that revenues from a liabilities scheme would be transferred to an international climate fund rather than staying within the country. This may pose a significant political economy obstacle. However, a well-run climate fund would return revenues to countries commensurately with the climate damage suffered, say, or can also offer other redistributive possibilities. In turn, the existence of such a climate fund could go a long way towards improving the likelihood of reaching an international agreement.

Lastly, the liability approach reduces the number of dimensions of potential disagreement between would-be participants. Indeed, agreeing on an optimal carbon tax rate requires agreement on the discount factor and on the magnitude of yet unrealized climate events far off in the future; two highly debated issues that are unlikely to garner a consensus anytime soon. By constrast, an optimal liability scheme 'only' requires agreement on a starting date from which emissions should be counted. While such an agreement is by no means guaranteed, as there is a clear tension between developed and developing countries in this regard, reducing the dimensions of potential disagreement to one may significantly improve the likelihood of a joint international effort.

References

[1] Arrhenius, S. (1896) "On the Influence of Carbonic Acid in the Air upon the Temperature of the Ground," *Philosophical Magazine and Journal of Science*, **41**, 237-271.

[2] Billette de Villemeur, E. and J. Leroux (2011) "Sharing the Cost of Global Warming", *The Scandinavian Journal of Economics*, **113**(4), 758-783.

[3] Cairns, R. (2004). "Green Accounting for an Externality: Pollution at a Mine." *Environmental and Resource Economics* **27**: 409-427.

[4] Cairns, R. and P. Lasserre (2006) "Implementing Carbon Credits for Forests based on Green Accounting", *Ecological Economics* **56**, 610-621.

[5] Calabresi, G., *The Cost of Accidents: A Legal and Economic Analysis*, New Haven, CN.: Yale University Press, 1970.

[6] Engen, E.M. and R.G. Hubbard (2004) "Federal Government Debt and Interest Rates," in *NBER Macroeconomics Annual 2004, Vol. 19*, Mark Gertler and Kenneth Rogoff Eds., MIT Press.

[7] Fourier, J. (1824) "Remarques générales sur les températures du globe terrestre et des espaces planétaires", *Annales de Chimie et de Physique*, **2e série, XXVII**,136-167.

[8] Friedman, E. and H. Moulin (1999), "Three Methods to Share Joint Costs or Surplus," *Journal of Economic Theory*, **87**, 275-312.

[9] Fullerton, D., Heutel, G. and G.E. Metcalf (2012) "Does the Indexing of Government Transfers Make Carbon Pricing Progressive?" *American Journal of Agricultural Economics* **94**(2) 347-353.

[10] Laubach, T. (2009) "New Evidence on the Interest Rate Effects of Budget Deficits and Debt," *Journal of the European Economic Association*, **7,** 858-885.

[11] Metcalf, G.E. and D. Weisbach (2009) "The Design of a Carbon Tax" *Harvard Environmental Law Review* **33**(2), 499-506.

[12] Gampfer, R. (2014) "Do individuals care about fairness in burden sharing for climate change mitigation? Evidence from a lab experiment." *Climatic Change* **124(1)**, 65-77.

[13] Gampfer, R., E. Gsottbauer and A. Delas (2014) "Interactive experiment: compensation for loss and damage as a means to shore up climate ambition?" mimeo ETH-Zurich.

[14] Gersbach, H., and R. Winkler (2012) "Global Refunding and Climate Change," *Journal of Economic Dynamics and Control*, **36** 1775-1795,

[15] Gersbach, H., and R. Winkler (2011) "International emission permits markets with refunding," *European Economic Review* **55**, 759–773.

[16] Golosov, M., Hassler, J., Krusell, P. and A. Tsyvinski (2014) "Optimal taxes on fossil fuel in general equilibrium," *Econometrica*, **82**(1) 41–88.

[17] Hammond, Peter J., (1981) "Ex-Ante and Ex-Post Welfare Optimality under Uncertainty", *Economica*, **48** (n.191),235-250.

[18] Hardin, G. (1968) "The Tragedy of the Commons", *Science*, **162**, 1243-1248.

[19] Hartwick, J. M., (1990) "Natural resources, national accounting and economic depreciation," *Journal of Public Economics* **43**(3), 291-304.

[20] Hsu, S-L (2011) *The Case for a Carbon Tax*, Island Press eds.

[21] Jenkins, Jesse D. (2014) "Political economy constraints on carbon pricing policies: What are the implications for economic efficiency, environmental efficacy, and climate policy design?", *Energy Policy*, **69**, 467-477.

[22] Kellog, W. W. (1987) "Mankind's impact on climate: The evolution of an awareness", *Climatic Change*, **10**, 113-136.

[23] Kolstad, C. D., T.S. Ulen and G.V. Johnson (1990), "*Ex Post* Liability for Harm vs. *Ex Ante* Safety Regulation: Substitutes or Complements?", *American Economic Review*, **80(4)**, 888-901.

[24] Magill, M., Quinzii, M. and J.-C. Rochet (2013) "A critique of shareholder value maximization," mimeo, Swiss Financial Institute.

[25] Moulin, H. (2002) "*Axiomatic cost and surplus sharing*," in: K. J. Arrow & A. K. Sen & K. Suzumura (ed.), Handbook of Social Choice and Welfare, edition 1, volume 1, chapter 6, pages 289-357 Elsevier.

[26] Moulin, H. and S. Shenker, "Serial Cost Sharing", *Econometrica*, **60**, 1009-1037.

[27] Nordhaus, W. D. (1977), "Strategies for the control of carbon dioxide," Cowles discussion paper n. 443.

[28] US Goverment, Interagency Working Group on Social Cost of Carbon (2013) "Technical Support Document: Technical Update of the Social Cost of Carbon for Regulatory Impact Analysis" http://www.whitehouse.gov/sites/default/files/omb/assets/inforeg/technical-update-social-cost-of-carbon-for-regulator-impact-analysis.pdf (accessed Jan 7th 2015)

[29] Shavell, S. (1984) "Liability for Harm Versus Regulation of Safety", *Journal of Legal Studies*, June 1984, **13**, 357-374.

[30] Shavell, S. (2011), "Corrective taxation versus liability", *American Economic Review Papers and Proceedings*, **101**(3), 273-276.

[31] Shleifer, A., *The Failure of Judges and the Rise of Regulators.* MIT Press, 352 pages, 2012.

[32] Solow, R.M., 1986, "On the intergenerational allocation of natural resources," *Scandinavian Journal of Economics* **88**, no. 1, 141-149.

[33] Tol, R.S.J. (2005) "The marginal damage costs of carbon dioxide emissions: an assessment of the uncertainties," *Energy Policy*, **33**, 2064-2074.

[34] Wachtel, P. and J. Young (1987) "Deficit Annoucements and Interest Rates," *The American Economic Review*, **77**(5), 1007-1012.

[35] Weitzman, M. (1976), 'On the Welfare Significance of National Product in a Dynamic Economy', *Quarterly Journal of Economics* **90**, 156-162.

[36] World Bank (2014), "Putting a Price on Carbon with a Tax", http://www.worldbank.org/content/dam/Worldbank/document/SDN/background-note_carbon-tax.pdf (accessed Jan. 7th 2015)

A Appendix: Defaulting on carbon debt vs. delaying the carbon tax

This section compares the incentives to participate in a carbon tax scheme with the incentives to remain in the liability system over time. On the one hand, one drawback of the liability scheme is that countries face an increasing temptation to default on their accumulated carbon debt. On the other hand, if climate damage increases over time, there is also an increasing temptation to not participate in a carbon tax. Nevertheless, defaulting on the liability scheme is more tempting overall:

$$\Delta_{liability,T} - \Delta_{tax,T} = -\left(\gamma Z_{T-1}^{i}\right)\tau_T < 0,$$

where $\Delta_{liability,T}$ and $\Delta_{tax,T}$ are the costs of defaulting on the liability scheme and of not participating in the carbon tax from period T onward, respectively.

Proof. The expected cost of abandoning the carbon tax at date T is equal to:

$$\Delta_{tax,T} = E_T \left\{ \sum_{t=T}^{+\infty} \beta^{t-T} \left[B_t^i \left(X_t^i \right) - \tau_t X_t^i - D_t^i \left(Z_t \right) \right] - \sum_{t=T}^{+\infty} \beta^{t-T} \left[B_t^i \left(\tilde{X}_{t,T}^i \right) - D_t^i \left(\tilde{Z}_{t,T} \right) \right] \right\},$$

where $\tilde{X}_i$ and $\tilde{Z}_i$ refer to the carbon flow and stock in the case where country i is no longer taxed.

Likewise, the expected cost of defaulting on carbon debt from date T onward is equal to:

$$\Delta_{liability,T} = E_T \left\{ \sum_{t=T}^{+\infty} \beta^{t-T} \left[B_t^i \left(X_t^i \right) - \mu_t Z_t^i - D_t^i \left(Z_t \right) \right] - \sum_{t=T}^{+\infty} \beta^{t-T} \left[B_t^i \left(\tilde{X}_{t,T}^i \right) - D_t^i \left(\tilde{Z}_{t,T} \right) \right] \right\}.$$

Therefore,

$$\Delta_{liability,T} - \Delta_{tax,T} = E_T \left\{ \sum_{t=T}^{+\infty} \beta^{t-T} \left[\tau_t X_t^i - \mu_t Z_t^i \right] \right\},$$

with $\tau_t = E_t \left[\sum_{s=t}^{+\infty} \left(\gamma \beta \right)^{s-t} \frac{\partial D_s}{\partial Z_s} \right]$ and $\mu_t = \frac{\partial D_t}{\partial Z_t}$. Simple algebraic manipulations yield the result. ■

The above difference is negative, indicating that defaulting on the liability scheme is more tempting. Notice that the size of the difference is equal to the cost of the carbon stock inherited from the past, $\left(\gamma Z_{T-1}^i \right)$, priced at the carbon tax rate of the current period, τ_T. Notice that the difference does not account for the above-mentioned reputational costs associated with defaulting on (financialized) carbon debt. These reputational costs are relevant for the liability scheme but do not arise in the case of the carbon tax. Hence, the above difference constitutes an overestimate of the additional temptation to default when adopting a liability scheme rather than relying on a carbon tax.

On the other hand, a drawback of the carbon tax is that its adoption is costly up front. This is because it requires payments immediately for climate damage that may take decades or more to materialize. By contrast, a liability scheme asks for payments upon the realization of damage, effectively spreading payments over time. This makes the liability scheme more likely to be adopted in the short run, a considerable argument in the face of the current difficulties in reaching international climate agreements. To be precise, compare the net benefits of both schemes over the first $L+1$ periods:

$$\begin{aligned}
\Delta_{liability-tax,L} &= E_0 \left\{ \sum_{t=0}^{L} \beta^t \left[B_t^i \left(X_t^i \right) - \mu_t Z_t^i - D_t^i \left(Z_t \right) \right] - \sum_{t=0}^{L} \beta^t \left[B_t^i \left(X_t^i \right) - \tau_t X_t^i - D_t^i \left(Z_t \right) \right] \right\} \\
&= E_0 \left\{ \sum_{t=0}^{L} \beta^t \left[\tau_t X_t^i - \mu_t Z_t^i \right] \right\} = \beta^L E_0 \left[\tau_L Z_L^i \right] > 0.
\end{aligned}$$

The sign of the above difference is positive, implying that the liability scheme is strictly less costly over any finite horizon. Its magnitude is nothing but the (discounted) expected cost of the stock at date L.

Notice that despite the compounded discount factor, β^L, the above difference is not necessarily negligible, even if L is large. In fact, if damage is a convex function of total stock, and if stock increases over time, the tax rate $\tau_L = E_L \left[\sum_{s=L}^{+\infty} (\gamma\beta)^{s-L} \frac{\partial D_s}{\partial Z_s} \right]$ increases with L. Therefore, the size of the difference can even increase with L if $\tau_{L+1}/\tau_L > 1/\beta$. With discount factors close to one, this is a distinct possibility.