



Contents lists available at ScienceDirect

## Journal of Public Economics

journal homepage: [www.elsevier.com/locate/jpube](http://www.elsevier.com/locate/jpube)

# Mirage on the horizon: Geoengineering and carbon taxation without commitment <sup>☆</sup>

Daron Acemoglu <sup>a,\*</sup>, Will Rafey <sup>b</sup><sup>a</sup> MIT Department of Economics, Cambridge, MA, United States<sup>b</sup> UCLA Department of Economics, Los Angeles, CA, United States

## ARTICLE INFO

## Article history:

Received 7 August 2020

Revised 19 December 2022

Accepted 20 December 2022

## JEL Classification:

Q01

Q4

Q54

Q55

Q58

O30

O31

O33

C65

## Keywords:

Carbon tax

Geoengineering

Environment

Innovation

Time inconsistency

## ABSTRACT

We show that, in a model without commitment to future policies, geoengineering breakthroughs can have adverse environmental and welfare effects because they change the (equilibrium) carbon taxes. In our model, energy producers emit carbon, which creates a negative environmental externality, and may decide to switch to cleaner technology. A benevolent social planner sets carbon taxes without commitment. Higher future carbon taxes both reduce emissions given technology and encourage energy producers to switch to cleaner technology. Geoengineering advances, which reduce the negative environmental effects of the existing stock of carbon, decrease future carbon taxes and thus discourage private investments in conventional clean technology. We characterize the conditions under which these advances diminish—rather than improve—environmental quality and welfare, and show that given current estimates of costs and environmental damages, these conditions are likely to be satisfied in our model.

© 2022 Elsevier B.V. All rights reserved.

<sup>☆</sup> We thank Marco Battaglini, Patrick Behrer, Steve Cicala, Bård Harstad, Gustavo Joaquim, Antoine Levy, Torben Mideksa, William Nordhaus, Alp Simsek, James Stock, Ludwig Straub, Massimo Tavoni, Gernot Wagner, Martin Weitzman, Iván Werning, and seminar participants at the University of Oslo Political Economy of the Environment Workshop, the U.C. Berkeley Advanced Workshop in Climate Economics, the Harvard Environmental Economics Lunch, the Harvard Solar Geoengineering Research Program, the MIT Joint Program on the Science and Policy of Global Change, and the MIT Macro Lunch for comments and useful suggestions. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 1122374. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Rafey also acknowledges financial support from the Becker Friedman Institute at the University of Chicago and the MIT Environmental Solutions Initiative.

\* Corresponding author.

E-mail address: [daron@mit.edu](mailto:daron@mit.edu) (D. Acemoglu).

## 1. Introduction

There is increasing recognition that a transition to cleaner technology has to be the bedrock of future reductions in carbon emissions.<sup>1</sup> While there have been important advances in “conventional” clean technologies, such as wind and solar, and economic research supports the notion that carbon taxes and other

<sup>1</sup> On anthropogenic climate change, see, e.g., [Mann et al. \(2017\)](#) on the contribution of human activity to heat waves, [Yan et al. \(2016\)](#) and [Gertler and O’Gorman \(2019\)](#) on tropical and extratropical storms, [Kopp et al. \(2016\)](#) on sea level rise, and [Cook et al. \(2016\)](#) on scientific consensus. On economic costs of climate change, see [Dell et al. \(2012\)](#) on growth, [Hsiang et al. \(2013\)](#) on conflict, and [Greenstone et al. \(2013\)](#) on calculating the social cost of carbon. For recent macroeconomic analyses of climate change, see [Hassler et al. \(2016\)](#), [Nordhaus \(2008, 2017\)](#), [Desmet and Rossi-Hansberg \(2015\)](#), and [Golosov et al. \(2014\)](#).

subsidies can contribute to the adoption of these technologies,<sup>2</sup> some experts and policymakers instead pin their hopes on geoengineering breakthroughs, such as large-scale carbon sequestration, ocean fertilization, and solar radiation management (for example, Keith, 2013; Flannery, 2015; and Morton, 2015). Although such breakthroughs, if realized, could enable the global economy to achieve lower environmental damages without high carbon taxes, there are concerns that the prospect of geoengineering may delay or undermine other policy responses to climate change. As the Intergovernmental Panel on Climate Change (IPCC) argues,

One of the most prominent arguments against geoengineering suggests that geoengineering research activities might hamper mitigation efforts...which presumes that geoengineering should not be considered an acceptable substitute for mitigation. The central idea is that research increases the prospect of geoengineering being regarded as a serious alternative to emission reduction. (IPCC 2014, p. 219).

In this paper, we provide a new and complementary reason why the prospect of geoengineering may, paradoxically, lead to worse environmental outcomes. In addition to incorporating geoengineering, our model features two plausible modifications relative to the simplest model of Pigovian carbon taxation. First, we introduce a conventional clean technology, which firms can adopt in order to reduce emissions when faced with a future carbon tax. Consistent with much of the evidence in the area of innovation, we assume that the development or adoption of cleaner conventional technologies today will make it cheaper to adopt them in the future.<sup>3</sup> Second, we assume that policy is chosen by a social planner without the ability to commit to future policies. That policy-making is potentially “time inconsistent” —both because future decision-makers may be different than the current one and because even the same decision-maker may wish to revise policy plans and deviate from promises made in the past—has long been emphasized in many areas of economics.<sup>4</sup> In environmental economics, several policy reversals have illustrated the relevance of the commitment problem for climate change.<sup>5</sup>

The core of our argument is that time-inconsistency—beyond its general import—qualitatively changes the positive and normative implications of new technologies. A natural reference point for the carbon tax in a model with harmful carbon emissions is the

<sup>2</sup> We refer to wind, solar and geothermal technologies and to energy-saving incremental improvements, which firms themselves develop or invest in, as “conventional” technologies to distinguish them from the less-tested geoengineering technologies (which are likely to be developed by other entities).

<sup>3</sup> This type of externality arises naturally in almost all models of endogenous technology, including the quality ladder models of Aghion and Howitt (1992), Grossman and Helpman (1991) and Klette and Kortum (2004), as well as many of the horizontal innovation models, such as Jones (1995). See Acemoglu (2007) for a discussion. Aghion and Griffith (2005) and Akcigit and Kerr (2018) provide evidence consistent with this type of externality in general, while Aghion et al. (2016) provide evidence for it in the context of the adoption of cleaner technologies in the automobile industry and Bollinger and Gillingham (2014) provide evidence in the context of solar installations.

<sup>4</sup> Throughout, we use the term “time-inconsistency” in the spirit of Kydland and Prescott (1977) and Calvo (1978) to signify that Bellman’s (1957) theorem of sequential optimality fails even for a standard, additively separable exponentially-discounted objective function because the constraint facing the decision-maker changes over time (here, due to decisions made by other agents).

<sup>5</sup> Major political revisions include Canada’s 2011 withdrawal from Kyoto, Australia’s repeal of its carbon tax in 2014, and the U.S.’s repudiation of the Paris Accord in 2017. Examples of energy incentives revised in response to technological change include the Spanish solar feed-in-tariff, where the government reneged on solar subsidies after production costs fell unexpectedly, and the U.K.’s decision in late 2011 to cut solar subsidies under the 2008 Energy Act by 55%. Optimal climate policies are often computed for the next 100–200 years (e.g., Golosov et al., 2014), while the U.S. legislature commits to the wind power production tax credit for three years on average (from 1992–2015) and caps under the E.U. Emissions Trading Scheme are renegotiated between each phase (every four to eight years).

Pigovian benchmark (where the carbon tax equals the marginal damage from one more unit of carbon). This carbon price can be implemented via a price instrument (a tax) or a quantity instrument (a “cap-and-trade”). However, when the (social) planner would also like to encourage a transition to cleaner (conventional) technology, she would like to deviate from the Pigovian benchmark and set a higher tax rate to encourage more rapid technology adoption. But in a world without commitment to future policies, firms will anticipate that any promised taxes above the Pigovian level will be revised, and underinvest in clean technology.

It is into this setting that we introduce the prospect of geoengineering. For clarity, and in line with the IPCC’s own taxonomy, we distinguish between two different types of geoengineering technologies (recognizing that some real-world technologies are a mixture of these two types): type I technologies, *carbon removal*, which correspond to a rightward shift of the environmental damage function (or, equivalently, reduce the effective stock of carbon that enters the damage function by a constant amount), and type II technologies, *climate adaptation* or *solar radiation management*, which reduce marginal damages from carbon in the atmosphere.<sup>6</sup> Examples of type I technologies include all forms of large-scale carbon dioxide removal, like mass afforestation, biochar, ambient air capture, and ocean fertilization (Lenton and Vaughan, 2009). Examples of type II technologies include solar radiation management, such as albedo enhancement, space reflectors, or stratospheric aerosols (National Research Council, 2015). While some type II geoengineering technologies appear to be the most empirically relevant, due to their low predicted costs, type I technologies may also experience breakthroughs.

Though the two types of geoengineering technologies have somewhat different implications, they both work in a similar manner in an environment without commitment. With clean technology held fixed, geoengineering breakthroughs of either type reduce future damages, and thus future Pigovian carbon taxes. But because firms investing in clean technology anticipate that such breakthroughs will lower the profitability of their investments, fewer such firms invest. For this reason, geoengineering *increases* underinvestment in these socially valuable technologies.

More specifically, we demonstrate that type I geoengineering technologies reduce investment in conventional clean technology so much that overall damages remain at the same level as before geoengineering. Intuitively, in our model to restore incentives for the adoption of conventional clean technologies, the Pigovian tax (marginal damage of carbon) needs to be at a certain level. With type I technologies, when the marginal value of damages remains the same so does the level of overall damages. Interestingly, even though overall damages remain constant, welfare may decline because the problem of underinvestment in cleaner technologies becomes more severe with the geoengineering advances. More ominously, we show that type II geoengineering technologies may actually lead to greater damages (depending on an elasticity condition for the damage function) and are more likely to reduce welfare.

The reason why geoengineering technologies backfire in our model is very different from those emphasized in previous discussions, which focus on potential downsides of the prospect of geoengineering because major geoengineering breakthroughs may not be realized or may create new, unrelated environmental risks. Instead, our framework identifies potential inefficiencies from geo-

<sup>6</sup> See IPCC (2014, pp. 484–489): “Two categories of geoengineering are generally distinguished. Removal of GHGs, in particular carbon dioxide termed ‘carbon dioxide removal’ or CDR, would reduce atmospheric GHG concentrations... ‘Solar radiation management’ or SRM technologies aim to increase the reflection of sunlight to cool the planet and do not fall within the usual definitions of mitigation and adaptation.”

engineering that arise precisely because the breakthroughs *will be* realized.

We first develop these ideas in the simplest setting, which is a static world with *ex ante* identical firms. Each firm first undertakes a costly investment to switch to a cleaner production technology anticipating the future carbon tax and any geoeengineering breakthroughs. A benevolent planner sets the carbon tax after these conventional clean technology investments are made, but before production decisions. Production decisions create emissions, which contribute to the stock of carbon in the atmosphere, and a convex (social) damage function determines the welfare costs from this stock of carbon. The key technological externality—that clean technology investments make future clean technology cheaper—arises from a simple premise: a fraction of firms are replaced by new entrants, and if they have invested in the clean technology, the new entrant can inherit this improvement. This externality implies that the planner would like to choose a carbon tax rate above the Pigovian level, but the aforementioned time-inconsistency problem means that she cannot deviate from Pigovian taxes, leading to underinvestment in the conventional clean technology.

What simplifies the analysis of this model is that there exists a unique level of the carbon tax that satisfies the technology incentive-compatibility constraint (IC), which makes the *ex ante* identical firms indifferent between investing in the clean technology and not. Provided that it is optimal to have some firms invest in the clean technology, the stock of carbon in the atmosphere has to adjust in order to satisfy the technology IC. In this light, the implications of various different types of geoeengineering technologies become straightforward. A type I geoeengineering technology, for example, shifts the damage function to reduce the level of the Pigovian carbon tax at a given stock of carbon in the atmosphere. But at this lower level of carbon tax, the technology IC is violated. To restore IC, the stock of carbon in the atmosphere must increase to offset the benefits from geoeengineering. The logic for type II technologies is similar, except that in this case following geoeengineering, the overall level of damages increase not to their original level but to restore the original marginal value of damages. Depending on the elasticity of the damage function, this may involve an increase in the level of damages relative to the benchmark without geoeengineering.

Our baseline results are greatly simplified thanks to our focus on *ex ante* identical firms. We establish that our general conclusions are robust to introducing firm heterogeneity, so that some firms are more likely to switch to clean technology at any given level of carbon tax. Another important simplifying assumption is that policymakers have to rely on the carbon tax in order to encourage the development of the clean technology. However, as emphasized in Acemoglu et al. (2012, 2016), optimal policy in this class of models involves both carbon taxes and direct subsidies to clean technology, and we also show that our results are robust to allowing for such subsidies. We discuss various other extensions of our basic framework as well.

After expositing our main ideas in a transparent manner in a static environment, we move to a continuous-time model of endogenous technological change with quality ladders. This model is useful for micro-founding the technological externality introduced above and demonstrating that our results do not depend on a static setting or on having no technological advances in dirty technologies. We also use the dynamic model for a simple quantitative exercise to investigate whether geoeengineering breakthroughs could reduce welfare in practice.

In our dynamic model, each active firm operates the best available technology in a given energy-related activity, and is stochastically replaced by a new entrant that builds and improves upon its productivity. The key technology externality emerges from the assumption that firms face nonzero probabilities of replacement.

We characterize the dynamic Markov equilibrium with a time-inconsistent planner in this setting. To do so, we show that (1) the interior equilibrium level of clean technology in the dynamic economy converges uniquely and in finite time, and (2) the equilibrium tax trajectory without commitment after the clean technology transition must be Pigovian. Though Pigovian taxes become more complicated (because they take into account future damages), we show that the results in the unique balanced growth path (BGP) are qualitatively identical to those we obtained in the static setting (and that the dynamic equilibrium converges to the BGP).

In our quantitative exercise, we calibrate the BGP of the dynamic model to world emissions and use estimates of environmental damages from Nordhaus (2017) to investigate the implications of geoeengineering for carbon emissions and welfare. We find that type I geoeengineering technologies are likely to improve welfare, but type II geoeengineering advances will lead to a significant decline in investment in clean technology, increase emissions and reduce welfare. Though primarily illustrative, this quantitative exercise suggests that unless appropriate policy responses can be designed (for example, in the form of credible commitment to future carbon taxes), the opportunities presented by geoeengineering technologies may be squandered and even backfire.

Our work is related to several literatures. First, we provide the first analysis of geoeengineering breakthroughs in an environment where the planner cannot commit and firms undertake clean technology investments. Our work consequently builds on and contributes to a growing literature on clean technology investments and innovations. In addition to Acemoglu et al. (2012, 2016) and Aghion et al. (2016), which have been mentioned above, Bovenberg and Smulders (1995), Goulder and Mathai (2000), Grimaud et al. (2011), Hartley et al. (2016), Hassler et al. (2012), Newell et al. (1999), and Popp (2002, 2019) also discuss endogenous technology in the context of environmental policy and climate change. Particularly relevant is Battaglini and Harstad (2016), who study incentives to invest in clean technologies in multilateral settings.

Second, several recent papers analyze the economics of geoeengineering technologies. Barrett (2008), Weitzman (2015), Moreno-Cruz (2015), and Meier and Traeger (2020) focus on the international political economy dimensions of geoeengineering technologies to study the risks of unilateral geoeengineering when the technology imposes externalities on other countries. Heutel et al. (2018) and Emmerling and Tavoni (2018) explore the optimal combination of conventional technology and geoeengineering under full commitment. Some authors have also argued that geoeengineering technologies may be harmful because of moral hazard—it is giving an additional option to policymakers who may not fully internalize social objectives (Lawrence and Crutzen, 2017). For instance, Morrow (2014) argues that if policymakers form false beliefs or are morally corrupt or unethical, then geoeengineering research can lead to worse outcomes. Similarly, Quaas et al. (2017) show that when the planner uses hyperbolic discounting, geoeengineering may be harmful. Our paper contributes to this literature by developing a new perspective on how geoeengineering might affect mitigation based on the interplay between the development of conventional clean technologies and time inconsistency of policymaking.

Third, existing work has emphasized the relevance of the commitment problem for exhaustible-resource management (Karp and Newbery, 1993), environmental regulation (Laffont and Tirole, 1996, 2000, 2002), climate policy (Helm et al., 2003; Brunner et al., 2012), and firms' incentives to adopt abatement technology (Requate and Unold, 2001; Requate and Unold, 2003). However, we are not aware of any papers that model or note how time-inconsistency distorts policies so much that technological improvements lead to lower welfare.

Fourth, our results are also related to the literature on second-best (environmental) policy. In the classic work by Peltzman (1975), a regulator sets safety technology standards but cannot directly control agents' risk-taking activities. Also related is Sinn's (2012) "green paradox," whereby future taxes on resources or resource use can lead to faster extraction of exhaustible supplies by firms today (see Jensen et al., 2015 and Lemoine, 2017). We contribute to this literature by showing the powerful effects that follow from the interplay of time-inconsistent policymaking and new (geoengineering) breakthroughs.

To the best of our knowledge, none of the literatures mentioned above contain comparative statics similar to the ones we emphasize: geoengineering reducing aggregate welfare and this paradoxical result becoming more powerful when there are stronger innovation externalities and reductions in the marginal damages from carbon. We finally note that these results also apply beyond the context of geoengineering, for example, when there are interactions between different jurisdictions or other changes that reduce marginal damages.

The remainder of the paper is organized as follows. Section 2 introduces our model and characterizes the equilibrium. Section 3 shows the robustness of our main results to alternative assumptions that incorporate richer distributions of firm types, more policy instruments, additional aspects of geoengineering, or political economy concerns. Section 4 then extends our baseline results to an infinite-horizon setting in continuous time. Section 5 presents our quantitative exercise. Section 6 concludes. Appendix A contains proofs omitted from the text, online Appendix B contains proofs for extensions and the infinite-horizon model, and online Appendix C contains details on the quantitative evaluation.

## 2. Baseline model

In this section, we introduce our simplest baseline static model. In the next two sections, we show its robustness to various extensions, including a dynamic model which provides a clearer micro-foundation for some of the baseline model's assumptions and fits more naturally with existing economic models of climate change, but still delivers essentially identical results.

### 2.1. Production and environmental damages

We consider an economy consisting of a range of energy-related activities, represented by the continuum  $[0, 1]$ , which are used to produce a consumption good. For simplicity, we take these activities to be perfectly substitutable. Initially, firm  $i$  controls the production technology for activity  $i \in [0, 1]$ , and by using  $k_i$  units of the consumption good as inputs, it can produce

$$f_d(k_i)$$

units of the consumption good. The production function  $f_d$  is assumed to be twice continuously differentiable, increasing and concave with the usual Inada conditions to ensure interior solutions (i.e.,  $\lim_{k \rightarrow 0} f'_d(k) = \infty$  and  $\lim_{k \rightarrow \infty} f'_d(k) = 0$ ). Since all activities are perfectly substitutable, energy firms will act competitively. We choose the consumption good as numeraire (normalizing its price to 1).

As indicated by the subscript "d," the initial production technology is "dirty," and generates  $k_i$  units of carbon emissions. By incurring a cost  $\Gamma > 0$ , each firm can upgrade to a (conventional) cleaner technology that produces

$$f_c(k_i)$$

units of the consumption good (where  $f_c$  is also twice continuously differentiable, increasing and concave, and satisfies

$\lim_{k \rightarrow 0} f'_c(k) = \infty$  and  $\lim_{k \rightarrow \infty} f'_c(k) = 0$ ), but only  $\gamma k_i$  units of carbon (where  $\gamma < 1$ ). We think of clean firms as switching to a technology using clean sources, such as wind or solar energy, or upgrading their existing plant's efficiency to reduce emissions. Note that  $f_c$  is allowed to differ arbitrarily from  $f_d$ . In particular, a clean firm's productivity may differ from that of a dirty firm (though our results also apply to the case in which both types have identical production functions). In general, when  $f_c \neq f_d$ , clean and dirty firms will use different levels of inputs even without a carbon price. The assumption that firms innovate only in clean technology is relaxed in the dynamic model of Section 4.

We also assume that in each activity  $i \in [0, 1]$ , a new entrant replaces the incumbent firm with probability  $\lambda \in [0, 1]$ . If the incumbent has already transitioned to clean technology, the entrant inherits it.<sup>7</sup> The entrant has access to the same production technology as the incumbent ( $f_d$  if there has not been a transition to clean technology, and  $f_c$  if there has been such a transition).<sup>8</sup>

Given these assumptions, denoting the fraction of activities that have switched to clean technology by  $q$ , total emissions in the economy can be written as

$$E = q\gamma k_c + (1 - q)k_d, \tag{1}$$

where  $k_c$  is the equilibrium input level of clean technology and  $k_d$  is the equilibrium input level of dirty technology (here we are using the fact that both entrants and incumbents will choose the same level of investment given their technology). The presence of the term  $\gamma < 1$  captures the fact that input usage by clean firms creates lower emissions.

Finally, we assume that the stock of carbon in the atmosphere is given by

$$S = (1 - \delta)S_0 + E, \tag{2}$$

where  $S_0 \geq 0$  is the initial level of carbon, and  $\delta \in [0, 1]$  denotes "depreciation" of this stock of carbon (for example, by absorption by oceans and forest cover). We choose this formulation to create continuity with the dynamic model in the next section. The damages from carbon in the atmosphere are denoted by

$$D(S; \xi, v), \tag{3}$$

where  $D$  is an increasing, twice continuously differentiable and strictly convex function, and enters utility additively.<sup>9</sup> The parameters  $\xi$  and  $v$  will be used to model the effects of other types of geo-engineering advances on environmental damages; for now, we

<sup>7</sup> Nothing in our qualitative results below change if we instead assume that the entrant could use the clean technology at some cost  $\Gamma_{\text{entrant}} < \Gamma$ . Our specification can be viewed as the special case with  $\Gamma_{\text{entrant}} = 0$ , adopted for simplicity. An incumbent's clean investment may create lower costs for an entrant if, for example, there exists any learning-by-doing in that activity, or if imperfect patent protection (or, in the dynamic model, quality-ladder-location-specific investment) prevents an exiting incumbent from recouping the entire value of their clean investment. For evidence on these types of spillovers, see footnote 3.

<sup>8</sup> This structure of entrants replacing incumbents will be further micro-founded in the context of the dynamic model in the next section.

<sup>9</sup> Welfare is usually assumed to be convex in the range of predicted temperature increases (Nordhaus and Moffat, 2017, p. 16), but differing assumptions give various specifications for the map from carbon to temperature. Some argue for a linear relationship based on historical climate records (Hassler et al., 2016, §3.2.6); others suggest concavity based on principles of climate physics (Golosov et al., 2014, §4.2); still others note that a convex relationship may arise in the presence of "tipping points" (see, e.g., Lemoine and Traeger, 2014). Hassler et al. (2016, §4.7) review the geophysics and its implications for the damage function. Our assumption that  $D$  is strictly convex rules out cases where the map from the carbon stock to global temperature is more concave than welfare losses are convex in temperature, which is unlikely. (An exception is the knife-edge case of Golosov et al., 2014, whose exponential temperature damage function and logarithmic carbon-temperature make marginal damages exactly linear). Finally, the additive formulation of environmental damages simplifies our analysis, but is not essential. In Appendix B, we show that our qualitative results are unaffected if damages affect productivity as in Nordhaus (1991, 2008) and Golosov et al. (2014), or affect utility in a non-additively separable manner.

suppress these parameters, writing environmental damages simply as  $D(S)$ .

### 2.2. Carbon tax and production decisions

Firms pay a carbon tax of  $\tau$  per unit of their emissions. Thus the profit maximization problems of the two types of firms can be written as

$$\begin{aligned} \pi_d(\tau) &= \max_{k \geq 0} f_d(k) - (1 + \tau)k \\ &= f_d(k_d(\tau)) - (1 + \tau)k_d(\tau), \end{aligned}$$

and

$$\begin{aligned} \pi_c(\tau) &= \max_{k \geq 0} f_c(k) - (1 + \gamma\tau)k \\ &= f_c(k_c(\tau)) - (1 + \gamma\tau)k_c(\tau), \end{aligned}$$

where  $k_d(\tau)$  is defined as the profit-maximizing level of input choice for a dirty firm, and  $k_c(\tau)$  is the profit-maximizing level of input choice for a clean firm.

The difference between the profit-maximization problem of the two types of firms stems from the difference in their production functions and—more crucially for our focus—from the fact that clean firms pollute less per unit of input (i.e.,  $\gamma < 1$ ). That clean firms pollute less per unit of input does not, however, guarantee that their overall emissions are less than that of dirty firms, since they may choose higher levels of input usage. This possibility, first noted by [Jevons \(1866\)](#), may lead to greater overall emissions by clean firms. Our next assumption ensures that this is not the case.

**Assumption 1** (No Jevons). For all  $\tau \geq 0$ , we have

$$\Lambda(\tau) \equiv k_d(\tau) - \gamma k_c(\tau) > 0.$$

This assumption is not restrictive and is automatically satisfied when  $\gamma = 0$  and  $f_c = f_d$ .

### 2.3. Clean technology decisions

The difference in profits between a clean and a dirty firm can be written as

$$\begin{aligned} \Psi(\tau) &= \pi_c(\tau) - \pi_d(\tau) \\ &= [f_c(k_c(\tau)) - (1 + \gamma\tau)k_c(\tau)] - [f_d(k_d(\tau)) - (1 + \tau)k_d(\tau)] \\ &= [f_c(k_c(\tau)) - k_c(\tau)] - [f_d(k_d(\tau)) - k_d(\tau)] + \tau\Lambda(\tau), \end{aligned} \tag{4}$$

where  $\Lambda(\tau)$  is the change in emissions from switching to a clean technology defined in [Assumption 1](#).

Recall that firms make their investment to switch to clean technology before they know whether they will be replaced by a new firm, and enjoy the additional profits from clean technology,  $\Psi(\tau)$ , only if they are not thus replaced (an event of probability  $1 - \lambda$ ). Consequently, a firm will find it (privately) optimal to switch to clean technology only if the condition

$$(1 - \lambda)\Psi(\tau) \geq \Gamma$$

is satisfied. Our key results will follow from the interplay between the effect of various geoengineering technologies and incentives for investment in (traditional) clean technology implied by this constraint.

In what follows, we denote the fraction of firms that switch to clean technology by  $q$ . The following lemma is immediate (proof omitted):

**Lemma 1** (Incentive Compatible Technology Choice).

$$\begin{cases} \Psi(\tau) > \frac{\Gamma}{1-\lambda} \Rightarrow q = 1 \\ \Psi(\tau) = \frac{\Gamma}{1-\lambda} \Rightarrow q \in [0, 1] \\ \Psi(\tau) < \frac{\Gamma}{1-\lambda} \Rightarrow q = 0. \end{cases} \quad (\text{Technology IC})$$

Note that when (Technology IC) holds exactly, i.e.,

$$\Psi(\tau) = \frac{\Gamma}{1-\lambda}, \tag{5}$$

any fraction of firms switching to clean technology is privately optimal. Conversely, when this equality does not hold, either all firms or no firm will make the switch to clean technology. Since we show that  $\Psi(\tau)$  is increasing in the next lemma, (5) defines a unique carbon tax rate, which we denote by  $\hat{\tau}$ .

The following lemma shows that higher taxes increase the incentives to switch to clean technology (proof omitted).

**Lemma 2** (Carbon Tax and Technology IC). Suppose [Assumption 1](#) holds. Then

$$\frac{d\Psi(\tau)}{d\tau} = \Lambda(\tau) > 0.$$

The result that a small increase in the carbon tax affects (Technology IC) only through  $\Lambda(\tau)$  follows from the Envelope Theorem, or simply from using the fact that both clean and dirty firms are choosing profit-maximizing input levels. That this effect is positive is a consequence of [Assumption 1](#). This result greatly simplifies our analysis by ensuring that the function  $\Psi$  is monotone.

### 2.4. The planner's problem

The (social) planner maximizes utilitarian welfare. Imposing, without loss of any generality, that all dirty (clean) firms choose the same level of inputs, welfare can be written as

$$\begin{aligned} W &= (1 - q)[f_d(k_d) - k_d] + q[f_c(k_c) - k_c] - q\Gamma - D(S) \\ &= (1 - q)\pi_d + q\pi_c + (k_d - q\Lambda)\tau - q\Gamma - D(S), \end{aligned} \tag{6}$$

where, as in [Assumption 1](#), we write  $\Lambda = k_d - \gamma k_c > 0$ .

Three important observations are in order. First, differently from private firms, the planner cares about the actual cost of inputs, and not about the taxes; this can be seen by the presence of the term  $\Lambda$ . Second, she also cares about the externality from emissions, as captured by the term  $D(S)$ . Third, the probability that a current producer is replaced by a new entrant,  $\lambda$ , which was important for private decisions to invest in clean technology, does not feature in this objective function because the new entrant will be able to produce with the same technology.

Until Section 3, we assume that the planner has access to a single instrument—a carbon tax,  $\tau$ . As mentioned in the Introduction, we can equivalently interpret the carbon tax instrument in our model as a quantity instrument or a cap-and-trade policy where the planner issues a number of tradable permits (and the carbon tax  $\tau$  in this case is given by the market-clearing permit price).

### 2.5. Timing of events

The key assumption, already highlighted in the Introduction, is that of *the lack of commitment* to future policies, which induces *time-inconsistency*. Namely, the planner is not able to choose, and

commit to, the carbon tax sequence ahead of all other decisions. In the static model, we incorporate this feature with the following timing of events:

- All firms simultaneously make their technology decisions.
- Firms that will be replaced by new entrants are revealed.
- The planner chooses the carbon tax,  $\tau$ .
- Given the carbon tax  $\tau$ , all firms simultaneously choose their input levels.

### 2.6. Equilibrium

Given the above description, a (subgame perfect) equilibrium can be defined as tuple  $(q^*, \tau^*, k_d^*, k_c^*)$  such that.

- Given  $q^*$ ,  $\tau^*$  maximizes  $W$  as in (6);
- $q^*$  satisfies (Technology IC);
- Given  $\tau^*$ ,  $k_d^*$  and  $k_c^*$  maximize, respectively,  $\pi_d$  and  $\pi_c$ .

Since the maximization problem of both clean and dirty firms is strictly concave, the equilibrium will always feature the same level of inputs for a given type of firm, denoted respectively by  $k_d(\tau)$  and  $k_c(\tau)$  as defined above. Then, once  $q^*$  and  $\tau^*$  are determined, the level of emissions can be computed from Eq. (1) as  $E(\tau^*, q^*)$ , and the level of stock of carbon in the atmosphere from Eq. (2) as  $S(\tau^*, q^*)$ . In view of this, we summarize the equilibrium simply by  $(\tau^*, q^*)$ , corresponding to the level of carbon tax and fraction of firms switching to clean technology.

### 2.7. Pigovian carbon taxes

A first implication of the timing of events adopted here (which incorporates the time-inconsistency feature mentioned above) is that the carbon tax will always be Pigovian—it will equal the marginal damage created by one more unit of emissions. This structure of Pigovian taxation contrasts with the case in which the planner can commit to carbon taxes, as we will see later.

More formally, we have:

**Proposition 1.** In equilibrium, the carbon tax is given as

$$\tau^* = D'(S(\tau^*, q^*)).$$

This result follows straightforwardly by differentiating the planner's objective function, (6). The Pigovian tax given in (7) will play a central role throughout the paper.

### 2.8. Characterization of equilibrium

In the rest of the analysis, we impose the following assumption, which ensures the existence of an *interior equilibrium*, meaning one in which some firms switch to clean technology, while others do not.<sup>10</sup>

**Assumption 2** (Conditions for Interior Equilibrium). We have

$$\frac{\Gamma}{1 - \lambda} \in (\Psi(\underline{\tau}), \Psi(\bar{\tau}))$$

where  $\underline{\tau} = D'((1 - \delta)S_0 + \gamma k_c(\underline{\tau}))$  and  $\bar{\tau} = D'((1 - \delta)S_0 + k_d(\bar{\tau}))$ .

<sup>10</sup> This equilibrium can also be labeled “asymmetric” because some ex-ante identical firms switch to clean technology, while others do not. We show in Section 3.1, however, that asymmetry is not the important feature, and similar results obtain when firms are heterogeneous ex ante in terms of their cost of switching to clean technology. The important feature is that the transition to clean technology is not complete.

This assumption ensures that condition (5) holds and the equilibrium is interior. It implies that when all firms make the switch to clean technology, the stock of carbon is low enough that the planner chooses a relatively low level of carbon tax (the one given by  $\underline{\tau}$  in this assumption), and when no firm makes the switch, the stock of carbon is high enough that the planner chooses a relatively high level of carbon tax (the one given by  $\bar{\tau}$  in this assumption).

When this assumption does not hold, there exists a unique equilibrium in which all firms switch to the clean technology or no firm switches to the clean technology, and in neither case do we have interesting comparative statics of investment in clean technology (small changes in parameters will not impact clean technology decisions). Thus Assumption 2 restricts the analysis to the interesting subset of the parameter space, where the equilibrium is interior. This is also empirically reasonable—in practice, only a limited fraction of energy producers have made the transition to clean technology, and there exist marginal clean investment decisions that will be impacted by future carbon taxes.

The next proposition characterizes the unique interior equilibrium.

**Proposition 2** (Interior Equilibrium). Suppose Assumptions 1 and 2 hold. Then there exists a unique equilibrium given by  $(\tau^*, q^*) = (\hat{\tau}, \hat{q})$ , where  $(\hat{\tau}, \hat{q}) = (D'(S(\hat{\tau}, \hat{q})), \hat{q})$ . This equilibrium is interior in the sense that the fraction of firms switching to clean technology  $\hat{q}$  is strictly between 0 and 1.

**Proof.** See Appendix A.

The first noteworthy result in this proposition is the uniqueness of an interior equilibrium. The reason why the equilibrium is interior and only a fraction of firms switch to the clean technology relates to the main economic force in our model. Firms, at the margin, switch to clean technology because of the carbon tax. The higher the carbon tax, the more inclined they are to make this transition. However, the carbon tax is determined by the planner after the technology decisions are made and will be lower when more firms have made the switch to clean technology—and herein lies the time-inconsistency problem. In particular, as already emphasized, in an interior equilibrium (5) needs to hold as equality. This implies that the carbon tax needs to take a specific value,  $\hat{\tau}$ , as given in Proposition 2 and represented by the horizontal line in Fig. 1. Given the convexity of damages in (3), the Pigovian carbon tax the social planner will set is increasing in the stock of carbon and hence decreasing in  $\hat{q}$ , guaranteeing uniqueness. But, for  $\hat{\tau}$  to emerge as the planner's choice, the stock of carbon in the atmosphere needs to take a specific value,  $S(\hat{\tau}, \hat{q})$ , and exactly  $\hat{q}$  fraction of firms need to switch to clean technology. If more firms than  $\hat{q}$  were to switch to clean technology, there would be less carbon in the atmosphere than  $S(\hat{\tau}, \hat{q})$ , and consequently, the planner would choose a lower carbon tax than  $\hat{\tau}$ , violating (5). Likewise, if fewer firms than  $\hat{q}$  made the switch, the carbon tax rate would be higher than  $\hat{\tau}$ , once again violating (5).

In addition to the existence of a unique interior equilibrium, the most important conclusion of Proposition 2 is that the level of carbon taxes will be Pigovian. This is dictated by the timing of events. At the time the planner sets the tax rate, technology decisions have already been made—in view of the fact that the planner cannot commit to carbon taxes ex ante. Without an influence on technology decisions, there is no reason for the planner to deviate from the Pigovian benchmark.

This contrasts with what the planner would have preferred if she could commit to the carbon tax, as we show next.

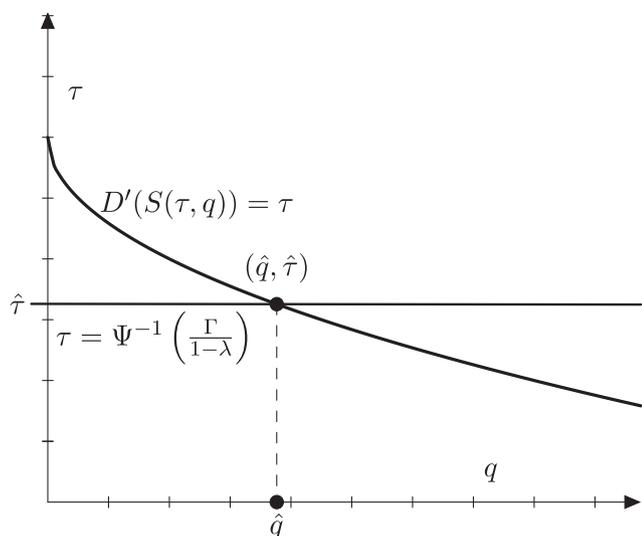


Fig. 1. Unique interior equilibrium  $(\hat{\tau}, \hat{q})$ .

2.9. Second-best

In this subsection, we briefly contrast the equilibrium with the “second-best” allocation where the planner can commit to carbon taxes in advance of the technology decisions of energy firms (but still cannot dictate input choices and technology decisions, hence the label “second-best”). This comparison will highlight the implications of time-inconsistency in our model.

Suppose that the planner sets a carbon tax rate  $\tau$ , and commits to it, before the technology decisions of firms.<sup>11</sup> The next proposition shows that as long as  $\lambda > 0$  the planner deviates from the Pigovian tax and induces more firms to switch to the clean technology than in the case without commitment.

**Proposition 3** (Second-best).

1. Suppose  $\lambda > 0$ . Then the planner commits to a carbon tax  $\tau^{SB} = \hat{\tau} > D'(S^{SB})$ , and the equilibrium fraction of firms that switch to clean technology is  $q^{SB} > \hat{q}$ , where  $S^{SB}$  is the stock of carbon in the second-best allocation (with commitment).
2. Suppose  $\lambda = 0$ . Then the planner commits to a carbon tax  $\tau^{SB} = \hat{\tau} = D'(S^{SB})$ , and the equilibrium fraction of firms that switch to clean technology is  $q^{SB} = \hat{q}$ .

**Proof.** See Appendix A.

The first part of this proposition shows that, provided that  $\lambda > 0$ , the planner would like to deviate from Pigovian taxation. Recall that Pigovian taxation implies  $\tau^{SB} = D'(S^{SB})$ , whereas the planner would like to commit to a tax  $\tau^{SB} > D'(S^{SB})$ .<sup>12</sup> This is because when

<sup>11</sup> When we endow the planner with commitment power, if the planner commits to the (unique) incentive-compatible tax  $\hat{\tau}$ , then any  $q \in [0, 1]$  may be an equilibrium. In the spirit of general mechanism design or principal-agent problems, we impose the incentive-compatibility constraints and let the planner choose her favorite allocation consistent with incentive-compatibility. An alternative approach to eliminating this multiplicity is to introduce heterogeneity in costs across firms, as in Section 6 of Requate and Unold (2001, p. 550).

<sup>12</sup> Note, however, that we still have  $\tau^{SB} = \hat{\tau}$ , since the planner cannot control investments in clean technology and thus has to satisfy (Technology IC).

$\lambda > 0$ , there is underinvestment in clean technology, because firms do not take into account the benefit they create for others who will build on their clean technology investments. As a result, in the second-best allocation where she cannot directly control technology investments but still can commit to a tax, the planner would like to encourage greater investment in clean technology by setting higher carbon taxes than the Pigovian benchmark, in order to induce more firms to switch to the clean technology. However, without commitment, the planner cannot achieve a non-Pigovian carbon tax, and the equilibrium always involves too little investment in clean technology, i.e.,  $\hat{q} < q^{SB}$ .

The second part of the proposition highlights the role of  $\lambda > 0$ . When  $\lambda = 0$ , firms fully internalize the benefits from a switch to clean technology. In this case, our model delivers a familiar result from the literature on commitment and technology adoption (e.g., Phaneuf and Requate, 2017, Proposition 11.9), where setting the right price of carbon—i.e., the Pigovian tax—induces the optimal level of technological investments, and thus the planner has no reason to resort to a non-Pigovian tax.

One consequence of Proposition 3 is that, when  $\lambda > 0$  as we assume to be the case throughout the rest of the analysis, there is too little investment in clean technology and too much carbon in the atmosphere. Any further increase in the stock of carbon reduces welfare.

2.10. The effects of geoengineering

We next study the implications of geoengineering technologies on equilibrium carbon taxes, investment in clean technologies, environmental damages and welfare when the planner lacks commitment. By geoengineering technologies, we refer to technological advances that reduce the damages from a given stock of carbon and are operated by the government or some other entity (but not the firms themselves and are thus distinct from the traditional clean technologies studied above). We distinguish between two different types of geoengineering technologies, which we first enumerate and motivate. We then analyze their implications separately. Actual geoengineering breakthroughs may combine features from these two types, but it is useful for our purposes to exposit their implications separately.<sup>13</sup>

To incorporate each type of geoengineering, let us make the role of the different parameters explicit as follows, writing

$$D(S; \xi, v) = (1 - v)\tilde{D}((1 - \delta)S_0 - \xi + E),$$

where  $\tilde{D}$  is a base damage function, and changes in the parameters  $\xi \geq 0$  and  $v \in [0, 1]$  each shift the environmental damage function.

More specifically, the first type of geoengineering technology, which we refer to as *carbon removal* or *geoengineering technology of type I*, corresponds to an increase in  $\xi$ , and thus leads to a parallel rightward shift of the environmental damage function as shown in the left panel of Fig. 2. In practice, this corresponds to large-scale carbon sequestration schemes that capture carbon from the air, such as permanent afforestation or algae blooms.

The second type of geoengineering technology, *climate adaptation*, *solar radiation management*, or *geoengineering technology of type II*, corresponds to an increase in  $v$ , a proportional rightward shift or rotation of the environmental damage function as illus-

<sup>13</sup> The distinction is perhaps best understood as one between the functional forms each type takes within our simplified setting, rather than fundamental attributes of the underlying technologies. As discussed in footnote 9, our simple damage function abstracts from complicated geophysical and economic relationships between the atmospheric carbon stock, global temperatures, and welfare. In practice, the extent to which actual geoengineering technologies correspond to type I, type II, or a mixture of the two will depend on the specification of the carbon cycle and damage function within a given integrated assessment model.

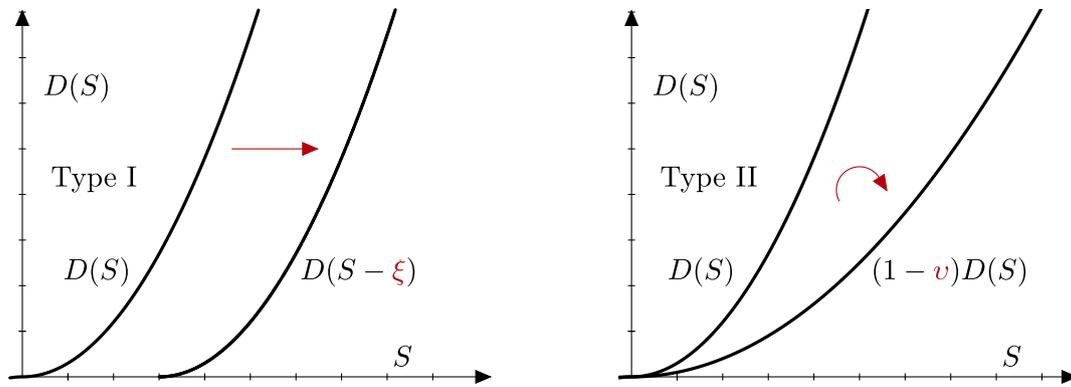


Fig. 2. Geoengineering of type I (“carbon removal”) and type II (“climate adaptation” or “solar radiation management”).

trated in the right panel of Fig. 2. We interpret this class as representing a range of technologies related to solar radiation management, aiming to slow temperature rise at a given emissions level. The most well-known example is the injection of sulfur dioxide into the stratosphere—suggested most notably by the Nobel prize-winning chemist Paul Crutzen—in order to reduce surface temperatures. Less esoteric geoengineering solutions that reduce marginal damages via various technological adaptations fall within this category as well. We take these geoengineering technologies to be exogenous (see below).

We now show that, in our framework, both types of geoengineering technologies do, to some extent, backfire, and they may increase emissions and even reduce welfare.

**Proposition 4** (*Geoengineering Technologies of Type I*). Suppose that Assumptions 1 and 2 hold. Consider a geoengineering technology improvement of type I that increases  $\xi$  by a small amount  $d\xi$ . Then we have

- $d\hat{\tau}/d\xi = 0$  (there is no effect on the equilibrium carbon tax).
- $d\hat{q} = -\frac{1}{\lambda}d\xi < 0$  (investment in clean technology declines).
- $dE = d\xi > 0$  (emissions increase, through lower  $\hat{q}$ ).
- $dD/d\xi = 0$  (environmental damages remain constant).
- $dW/d\xi < 0$  if and only if  $\lambda(\pi_c - \pi_d) > \Lambda\tau$  (welfare may decline).

**Proof.** See Appendix A.

The key economic force driving the result in Proposition 4 is that even after the geoengineering advances, the Technology IC (5) still pins down the carbon tax rate at  $\hat{\tau}$ . This is because with a small change in  $\xi$ , Assumption 2 will continue to hold and the equilibrium has to be interior. This in turn requires that the social planner still prefers to set the carbon tax at  $\hat{\tau}$ , which is only possible if the marginal environmental damage remains constant. Since geoengineering shifts the damage function rightward by  $d\xi$ , the total stock of carbon must increase by  $d\xi$ . This happens by fewer firms making the switch to clean technology. This is visually illustrated in Fig. 3. Geoengineering shifts the curve representing the level set of marginal damages rightwards as shown by the red curve. If there was no change in investment in conventional clean technology, marginal and overall damages would both decline. But in equilibrium, marginal damages have to remain constant, and as shown by the arrows along the red curve, the adjustment involves a reduction in investment in conventional clean technology, which

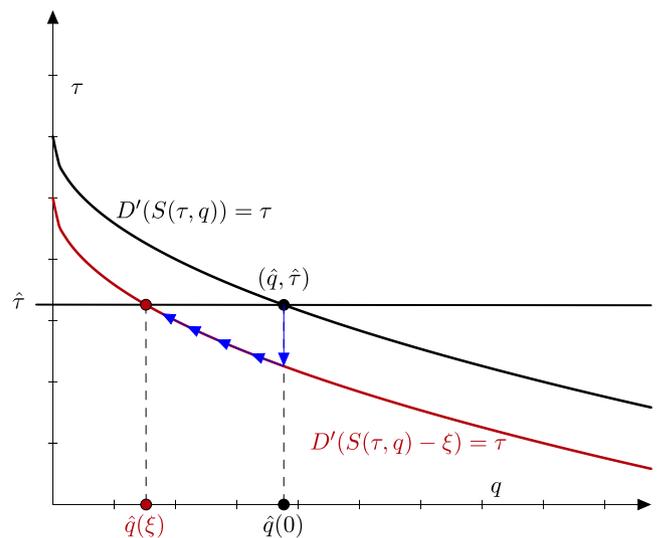


Fig. 3. Equilibrium response of clean technology after type I geoengineering.

increases emissions and the total stock of carbon, and restores marginal damages to the same level that prevailed before geoengineering.

The effects of this type of geoengineering advance on welfare are ambiguous because of two competing forces. On the one hand, since  $S$  remains constant and  $q$  declines, society saves the costs of switching to clean technology. If investment in clean technology were optimal (which happens when  $\lambda = 0$ ), this would be its sole impact because reductions in investment in clean technology would only have second-order welfare costs. Thus in this case, despite the increase in emissions, welfare would go up. However, because  $\lambda > 0$ , investments in clean technology are distorted, and a further reduction in the fraction of firms making the switch to clean technology creates a first-order welfare loss. Put differently, the benefit from investment in clean technology is not only the reduction in emissions, but also the fact that  $\pi_c = f_c(k_c(\hat{\tau})) - (1 + \gamma\hat{\tau})k_c(\hat{\tau})$  may be greater than  $\pi_d = f_d(k_d(\hat{\tau})) - (1 + \hat{\tau})k_d(\hat{\tau})$ . A reduction in  $q$  implies that this gain is forgone, which can outweigh the cost savings from lower investments. The condition for welfare to diminish as a result of a geoengineering advance of type I in the last part of the proposition indeed requires that  $\lambda$  and  $\pi_c - \pi_d$  are sufficiently large to compensate for the fixed cost savings. In fact, a large value of  $\lambda$ , by creating a larger wedge between the planner’s objective function and private incentives to switch to clean technology, is sufficient to ensure

that welfare declines as a result of this type of geoengineering advance.

Fig. 4 illustrates the intuition for our welfare result diagrammatically. The planner equates the downwards-sloping marginal value of consuming an additional unit of carbon with the upward-sloping marginal environmental damage. The first effect of geoengineering is to lower marginal environmental damages, as shown by the downward shift of the marginal damages curve from  $D'(\hat{S}(0))$  to the blue line  $D'(\hat{S}(\xi) - \xi)$ . This shift leads to social gains depicted by region A (in blue), which are approximately  $D'(S)$ . If the level of investment in clean technology were first best, this would be the only effect of geoengineering on welfare. However, when  $\lambda > 0$ , clean technologies were already suboptimal, so further discouraging clean technology with geoengineering also has welfare consequences. In the figure, the distortion in clean technology investments leads to a twist of the marginal product of an additional unit of carbon, from  $Y'(S, q(0)) = \hat{q}(0) \frac{\partial k_c}{\partial S} \left[ \frac{\partial f_c}{\partial k_c} - 1 \right] + (1 - \hat{q}(0)) \frac{\partial k_d}{\partial S} \left[ \frac{\partial f_d}{\partial k_d} - 1 \right]$  to the red curve,  $Y'(S, q(\xi)) = \hat{q}(\xi) \frac{\partial k_c}{\partial S} \left[ \frac{\partial f_c}{\partial k_c} - 1 \right] + (1 - \hat{q}(\xi)) \frac{\partial k_d}{\partial S} \left[ \frac{\partial f_d}{\partial k_d} - 1 \right]$ . The decline in clean technology compresses the marginal product curve towards the  $x$ -axis because dirty firms use carbon less efficiently, while stretching the curve away from the  $y$ -axis because dirty firms use fewer inputs of the consumption good at each level of emissions. This twist leads to a loss of welfare, represented by region B (in red), which is approximately equal to  $|d\hat{q}/d\xi| \cdot \lambda[\pi_c - \pi_d]$ . Overall welfare falls when region B is greater than region A—i.e., when  $\lambda(\pi_c - \pi_d)/\Lambda > \tau = D'(S)$  as in Proposition 4.

If instead of a small increase in  $\xi$  there is a large increase, Assumption 2 may be violated. In this case, the planner may wish to deviate from (5), forgoing any investment in clean technology. A similar caveat applies to the type II geoengineering technology discussed next.

**Remark 1 (Carbon Leakage).** Though our focus is on geoengineering technologies, Proposition 4 holds identically in a different setting. Suppose that our model applies to a specific country (say the United States) and another country (say China) reduces its emissions by an amount  $d\xi > 0$ . his reduction in the global carbon stock would reduce

the Pigovian tax of the domestic government, violating (5). To restore this constraint, emissions by domestic firms increase, again through reduced investments in clean technology.

The implications of geoengineering technologies of type II are broadly similar but slightly more involved because they can also lead to greater overall damages.

**Proposition 5 (Geoengineering Technologies of Type II).** Suppose that Assumptions 1 and 2 hold. Consider a geoengineering technology improvement of type II that increases  $v$  by a small amount  $dv > 0$ , and let  $\eta = \hat{S}D''(\hat{S})/D'(\hat{S})$  be the elasticity of the marginal damage function (where  $\hat{S} = S(\hat{\tau}, \hat{q})$ ). Then we have

- $d\hat{\tau}/dv = 0$  (there is no effect on the equilibrium carbon tax).
- $dS/dv > 0$  (the total stock of carbon increases).
- $d\hat{q}/dv < 0$  (investment in clean technology declines).
- $dE/dv > 0$  (emissions increase, through lower  $\hat{q}$ ).
- $dD/dv > 0$  if and only if  $\eta \leq \eta^*$  (environmental damage increases if the damage function is not too convex), where  $\eta^* \geq 1$ .
- $dW/dv < 0$  if and only if

$$\eta \leq \eta^{II}(\lambda) \equiv a\lambda \left( \frac{\pi_c - \pi_d}{\Lambda \tau} \right)$$

(welfare declines if the damage function is not too convex), where  $a \equiv SD'(S)/D(S) > 1$ .

**Proof.** See Appendix A.

As in Proposition 4, the results of Proposition 5 are a consequence of the fact that to sustain an interior clean technology adoption rate, the carbon tax needs to remain at  $\hat{\tau}$ , and this necessitates an increase in emissions. In the case of a type I geoengineering improvement, emissions increased in such a way as to keep the total stock of carbon in the atmosphere and overall environmental damages constant. With a type II advance, emissions must again increase to keep the marginal damage constant, but this might involve a higher level of overall damages. In particular, if the elasticity of the marginal damage function,  $\eta$ , is high, marginal dam-

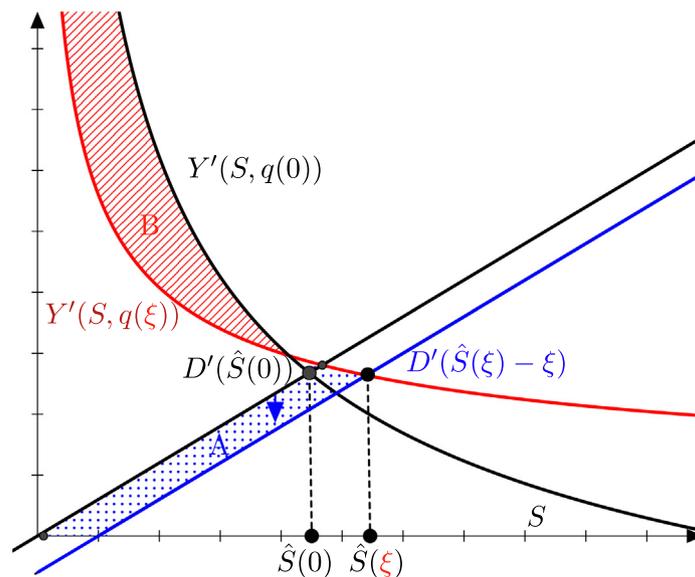


Fig. 4. Competing welfare effects of type I geoengineering.

ages can change significantly without a large change in the level of the stock of carbon. In this case, the direct environmental benefit from geoengineering dominates the equilibrium decline in clean technology, and environmental damages fall. Conversely, if  $\eta$  is low (in particular, less than some  $\eta^*$ ), to restore marginal damages to their initial value and thereby sustain the Pigovian tax at  $\hat{\tau}$ , the stock of carbon needs to change by a large amount, which translates into an increase in overall environmental damages. This is the case illustrated in the right panel of Fig. 5.<sup>14</sup>

The effects on welfare are once again ambiguous for similar reasons to those discussed above. But provided that the elasticity of the marginal damage function  $\eta$  is sufficiently low (in this case less than  $\eta^*$ ) and  $\lambda > 0$ , the negative effect of distorting investment in clean technology dominates savings from the transition costs  $q\Gamma$ , and overall welfare declines (in fact, as in the previous case, a sufficiently large  $\lambda$  ensures that welfare always declines). Conversely, a sufficiently elastic marginal damage function or a sufficiently low  $\lambda$  will make welfare increase as a result of an improvement in geoengineering technologies of type II. But the reason why these two conditions rule out negative welfare effects are somewhat different: when  $\eta$  is sufficiently high, overall environmental damages decline by a large amount; when  $\lambda$  is very small, private investment in clean technology is nearly optimal, and a further reduction in these investments only has second-order welfare costs.

Together, Propositions 4 and 5 are the main results of our static model: the negative equilibrium response of clean technology entirely offsets the environmental benefits of geoengineering improvements. In all interior equilibria, geoengineering technologies that remove carbon directly from the atmosphere do not affect overall environmental damages (Proposition 4), while geoengineering technologies that flatten the damage function sometimes increase environmental damages (Proposition 5). When  $\lambda$  is sufficiently large, both geoengineering improvements will reduce welfare.

### 3. Robustness and other extensions

Many of the assumptions adopted so far are for simplicity and transparency. In this section, we relax several of them. First and most importantly, we show that our results do not require firms to be identical and extend our results to an environment in which firms are heterogeneous in terms of their costs of switching to clean technology. Second, we show that if the planner can partially commit to subsidizing clean entrants directly, such instruments will prevent the adverse equilibrium outcomes to the extent that this commitment is possible. Third, we show that our results do not change when geoengineering outcomes are uncertain, endogenous, or involve additional harmful side effects. Finally, we show that when the planner does not internalize the full social cost of carbon, this will exacerbate the commitment problem, leading welfare to decline for a broader range of parameters. Throughout, we assume that slight variations of Assumptions 1 and 2 (adapted to the extended environment) continue to hold, but do not state them formally to conserve space.

#### 3.1. Heterogeneous costs of clean investment

In our analysis above, the (Technology IC) constraint holds with equality both before and after the arrival of geoengineering technologies. This sharp result simplifies our analysis but relies on mar-

<sup>14</sup> The figure is drawn for a quadratic  $D$  function. We show in Appendix A that when  $D$  is quadratic, the condition  $\eta \leq \eta^*$  is always satisfied, and thus environmental damages always increase as a result of a geoengineering technology improvement of type II. In this case, the condition  $\lambda \geq \lambda^* \equiv \frac{1}{2} \left( \frac{\lambda \tau}{\pi_c - \pi_d} \right)$  is necessary and sufficient for welfare to decline (note that  $\lambda^* < 1/2$ ).

ginal firms having identical switching costs. Here, we show that our welfare results do not rely on the assumption that firms are the same. Whether or not firms differ, geoengineering will always save society some fixed costs by offsetting some of the clean technology transition. When firms differ in adoption cost, these cost savings increase because geoengineering displaces the most costly marginal clean firms. The resulting lower marginal cost of clean investment makes society prefer slightly more environmental protection, lowering marginal environmental damages and the equilibrium tax. This effect undoes the knife-edge result that environmental damage does not change after a Type I improvement. Nevertheless, geoengineering still incurs social costs proportional to the value of that transition (in terms of spillovers) and welfare may fall for similar reasons as before. The primary difference is that the local curvature of the aggregate investment cost function becomes an additional primitive needed to determine welfare outcomes.

More formally, suppose that each firm  $i$ 's cost of switching to clean technology is now  $\Gamma_i = \Gamma + \chi_i$ , with  $H(x) \equiv \mathbb{P}(\chi_i \leq x)$ . While before, equilibrium technology adoption was the jump-discontinuous function

$$\hat{q}(\tau) = \hat{q} \mathbf{1}_{\{(1-\lambda)\Psi(\tau)=\Gamma\}} + \mathbf{1}_{\{(1-\lambda)\Psi(\tau)>\Gamma\}}, \tag{8}$$

the effect of heterogeneity is to smooth equilibrium clean technology adoption,

$$\hat{q}(\tau) = H((1-\lambda)\Psi(\tau) - \Gamma).$$

Heterogeneity also alters the aggregate cost of the clean transition,

$$\tilde{\Gamma}(q) = q\Gamma + \int_{-\infty}^{H^{-1}(q)} x dH(x),$$

and the costs faced by the marginal adopter,  $\tilde{\Gamma}'(q) = \Gamma + H^{-1}(q)$ . We can build some intuition for our more general results in the presence of heterogeneity, by totally differentiating (8) with respect to  $\xi$  (Type I):

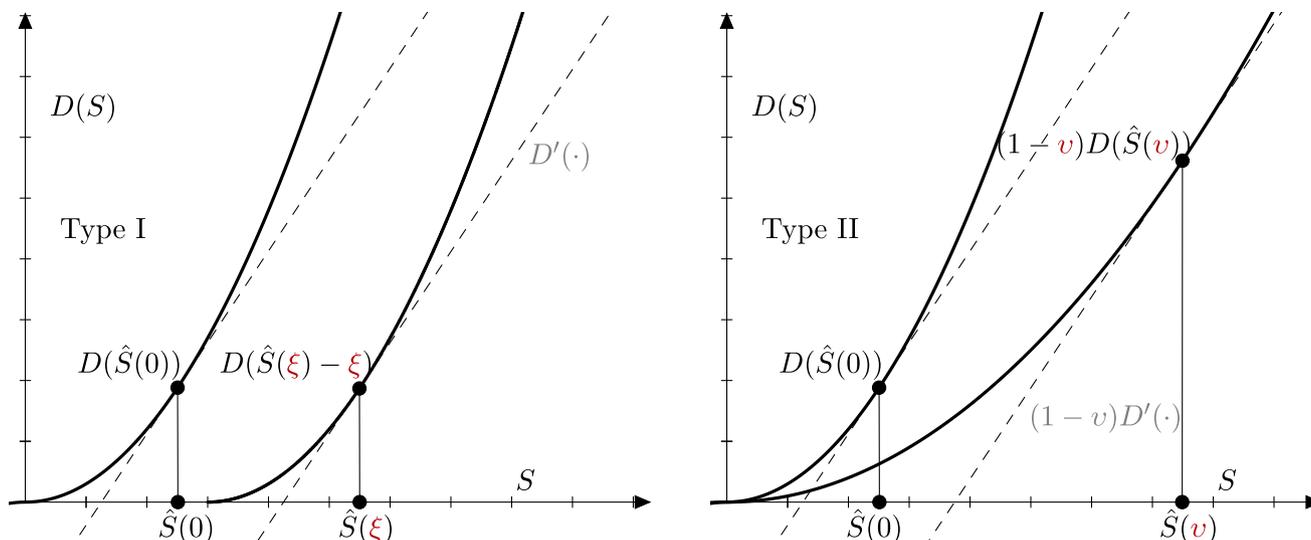
$$\frac{d\hat{\tau}}{d\xi} = \frac{1/h(0)}{\Lambda(1-\lambda)} \frac{d\hat{q}}{d\xi}. \tag{9}$$

This expression shows that taxes will decline after geoengineering in proportion to the convexity of the aggregate cost function  $\tilde{\Gamma}(q)$ . This convexity is related to firm heterogeneity. In the edge case where all firms are concentrated around the same level of cost  $\Gamma$ , we have  $\tilde{\Gamma}''(q) = 1/h(0) \Rightarrow 0$ . Then the right-hand side of (9) vanishes and  $d\hat{\tau}/d\xi = 0$  for all  $\lambda > 0$  as in our benchmark case, and we recover the same results as before. Away from this edge case, an increase in  $\xi$  reduces the carbon tax, which contributes to gains from geoengineering. Nevertheless, provided that  $\tilde{\Gamma}''(q) = 1/h(0)$  is not very large relative to the other parameters, welfare continues to decline because the lower marginal environmental damages continue to distort clean technology investments. The next proposition provides the exact conditions for this generalization. Like the proofs of all remaining results in the paper, the proof of this proposition is provided in the online Appendix B.

**Proposition 6 (Heterogeneous Firms).** Suppose that firms differ by fixed costs,  $\Gamma_i = \Gamma + \chi_i$ , with  $\chi_i \sim H$ . Then

- $dW/d\xi < 0 \Rightarrow \lambda(\pi_c - \pi_d) > [\Lambda + \Theta]\tau$ , where  $\Theta = \left( \frac{1}{D''(S)} - \frac{1-q}{f''_d(k_d)} - \frac{\gamma^2 q}{f''_c(k_c)} \frac{1/h(0)}{\Lambda(1-\lambda)} \right) > 0$ .
- $dW/dv < 0 \Rightarrow \eta < a\lambda \frac{\pi_c - \pi_d}{(\Lambda + \Theta)\tau}$ , where  $a > 1$  as before.

Proposition 6 provides necessary and sufficient conditions for geoengineering breakthroughs to reduce welfare. As the discussion before the proposition illustrated, these conditions depend on the



**Fig. 5.** Environmental damage before and after type I (“carbon removal”) and type II (“climate adaptation” or “solar radiation management”) geoengineering. The left (right) panel depicts the increase in  $\hat{S}$  necessary to keep marginal damages constant with a type I (type II) technology.

extent of heterogeneity in the cost of switching to clean technology, as captured by  $\Theta$ . In particular, if  $h(0)$  is very small, indicating that there is a very large amount of heterogeneity around the current equilibrium, then  $\Theta$  will be very high and the reduction in future carbon taxes will not trigger a large decline in socially valuable clean investments. Conversely, when  $h(0)$  is not very small or when  $\lambda(\pi_c - \pi_d)$  is large, geoengineering advances continue to reduce welfare.

### 3.2. Direct subsidies for clean technology

We have so far assumed that the social planner has access to a single policy instrument, the carbon tax. This is not central to our results as long as the social planner cannot perfectly control investments in clean technology. Suppose, for illustration, that she can commit to subsidize a proportion  $\sigma \in [0, 1]$  of each firm’s investments (fixed costs  $\Gamma$ ) in clean technology. We capture the fact that the planner cannot perfectly control these clean technology investments by assuming that these subsidies are not pure transfers and there is a social cost of  $\sigma\omega\Gamma$  in terms of the consumption good, where  $\omega > 0$  parameterizes the extent of inefficiency from the clean technology subsidies. In this case the qualitative results in Propositions 3–5 continue to apply. In particular, we can summarize the results in this case with the following proposition.

**Proposition 7 (Direct Subsidies for Clean Technology).** Suppose that in addition to the carbon tax, the social planner has access to a clean technology subsidy whereby a fraction  $\sigma \in [0, 1]$  of clean investment costs is subsidized at a social cost  $\sigma\omega\Gamma$  in terms of the consumption good. Then

- If  $\omega > 0$ , the planner still prefers to commit to a tax above the Pigovian benchmark when  $\lambda > 0$ .
- $dW/d\xi < 0 \Rightarrow \tilde{\lambda}(\pi_c - \pi_d) > \Lambda\tau$ .
- $dW/dv < 0 \Rightarrow \eta \leq \eta^{\text{II}}(\tilde{\lambda})$ , where  $\tilde{\lambda} \equiv \frac{\lambda - \sigma - (1-\lambda)\sigma\omega}{1-\sigma}$ .

This result is intuitive. As long as the planner cannot perfectly control clean technology decisions—in particular because clean technology subsidies are socially costly (the case where  $\omega > 0$ )—she would prefer a carbon tax above the Pigovian level to encourage additional clean technology investments. The commitment

problem, however, prevents this, and the logic of our results above apply and yield the same insights. However, our results do not extend to the case in which  $\omega = 0$ , where the planner can perfectly control clean technology investments without any social distortions. In this case, she has no reason to deviate from Pigovian taxation, the technology externality emphasized above is no longer present, and geoengineering improves welfare because the equilibrium is constrained efficient.

### 3.3. Other features of geoengineering

Our analysis was simplified by assuming that a geoengineering advance is anticipated to arrive with certainty and does so. In practice, there is considerable uncertainty about whether and when large-scale geoengineering will be feasible. We can incorporate this feature by assuming that both the agents in the economy and the social planner expect the geoengineering technology to arrive with some probability  $\kappa \in (0, 1)$ , realized after clean investments and carbon taxation.<sup>15</sup> Then the following hold:

**Proposition 8 (Stochastic Geoengineering).** Suppose an increment  $d\xi$  of type I geoengineering arrives with probability  $\kappa \in (0, 1)$ . Then expected welfare declines if and only if

$$\lambda(\pi_c - \pi_d) > \Lambda \frac{\mathbb{E}_\kappa[D''(S)]}{D''(S - d\xi)} (\tau - (1 - \kappa)b)$$

where  $\mathbb{E}_\kappa[D''(S)] = \kappa D''(S - d\xi) + (1 - \kappa)D''(S)$  and  $b \equiv D'(S) - D'(S - d\xi) > 0$ .

Suppose an increment  $dv$  of type II geoengineering arrives with probability  $\kappa \in (0, 1)$ . Then expected welfare declines if and only if

$$\eta < \frac{a}{\Lambda} \left( \frac{1 - dv}{1 - \kappa \cdot dv} \right) \left( \frac{\lambda(\pi_c - \pi_d)}{\tau - (1 - \kappa)b'} \right)$$

where  $b' \equiv dv \cdot D'(S) > 0$ .

Proposition 8 shows how our results generalize immediately. There is also an additional insight. The worst outcomes in terms of welfare and environmental damages are realized when geoen-

<sup>15</sup> The case in which the event of geoengineering is realized before carbon taxation is essentially identical, since clean investments are still made with reference to the expected tax, though the resulting calculations involve more expectation operators, given the additional uncertainty over the realized carbon tax.

neering is expected to succeed with high probability—substantially lowering investment in clean technology—but then fails. In this case, our model delivers a specific channel for the general concerns noted by the IPCC in the Introduction.

It is also straightforward to see that all of our results apply with endogenous geoen지니어ing—in particular, when the social planner has access to a technology to generate possibly stochastic geoen지니어ing advances and cannot commit to not deploying this technology. In the no-commitment equilibrium, the social planner chooses her optimal geoen지니어ing investment after clean technology investments are made, and then the equilibrium is very similar to the one with stochastic geoen지니어ing technologies given the resulting equilibrium probability of geoen지니어ing success.<sup>16</sup>

### 3.4. Political economy considerations

Our model takes a charitable view of policymakers, assuming that they fully internalize environmental externalities. In practice, many regulators and politicians appear to be far from this ideal benchmark. For example, they may be captured by special interest groups or receive campaign contributions that influence their policy agendas. In several simple political economy settings, such behavior can be modeled by assuming that the policymaker maximizes a weighted social welfare function, with greater weights on groups capable of lobbying or making campaign contributions. In our setting, this corresponds to the policymaker having an objective function that assigns a lower weight to environmental damage. The next proposition shows that our qualitative results remain unchanged in this case, except that negative welfare outcomes become more likely.

**Proposition 9** (*β-Benevolence*). Suppose that the policymaker values only  $\beta \in (0, 1)$  of environmental damages and thus maximizes

$$W(\beta) = (1 - q)[f_d(k_d) - k_d] + q[f_c(k_c) - k_c - \Gamma] - \beta D(S) \quad (6')$$

rather than (6)—but that true welfare is still given by (6). Then Proposition 4 and Proposition 5 hold, except that in the latter case welfare declines if

$$\eta \leq a \left( \beta \lambda \frac{\pi_c - \pi_d}{\Lambda \tau} + 1 - \beta \right) < \eta^H. \quad (10)$$

Intuitively, because the policymaker undervalues environmental damages, the equilibrium is further away from the second-best, making it more likely that the decline in clean investment resulting from geoen지니어ing reduces welfare.

## 4. Dynamic model

We now extend our static model to a dynamic economy where production decisions are made continuously, firms enter and exit,

<sup>16</sup> Many also fear that geoen지니어ing interventions may lead to unanticipated side effects. Our model abstracts from the damages and operational costs of geoen지니어ing to present the most favorable case for these advances and emphasize that the commitment problem is sufficient for welfare to decline. Including harmful side effects in our model will only make geoen지니어ing less likely to improve welfare. To see this, suppose geoen지니어ing of either type entails additively separable damages. Then, welfare falls after a breakthrough for parameters that satisfy Propositions 4 and 5 as before. However, welfare will also fall even if the conditions of Propositions 4 and 5 are not satisfied, as long as geoen지니어ing's benefits from the baseline model fall below its (expected) side effects. Identical reasoning applies to costly geoen지니어ing. Similar results also obtain when geoen지니어ing's harmful side effects are uncertain. For example, suppose geoen지니어ing is either (a) harmless or (b) sufficiently dangerous as to be optimal to prohibit. Uncertain side effects that take this form give the same results as in Proposition 8, where geoen지니어ing succeeds with some nonzero probability and fails otherwise.

and technological quality and the stock of carbon accumulate over time. Our model is constructed to mimic both the structure of our static setup and the quality-ladder models of Aghion and Howitt (1992) and Grossman and Helpman (1991) as closely as possible. The quality-ladder structure enables us to endogenize the replacement probability  $\lambda$  as the flow rate of creative destruction. After deriving the unique balanced growth path (BGP) and characterizing the structure of the dynamic equilibrium, we show that the effects of geoen지니어ing technologies on the BGP are essentially identical to those derived in the static model.

### 4.1. Production, entry and environmental damages

As in the static model, we consider an economy with a unique consumption good, produced by a continuum of perfectly substitutable energy activities indexed by  $[0, 1]$ . We think of each activity  $i \in [0, 1]$  as being produced at a site  $i \in [0, 1]$  dedicated to that activity and each site can house only one firm (so that the number of active firms will be constant as in the static model). Time  $t$  is infinite and discrete, of length  $\Delta > 0$ . In what follows, we simplify the exposition by taking  $\Delta \rightarrow 0$  to work directly with differential equations. The production technology differs from the static model only in that the productivity of each activity depends on where it is located on a quality ladder, denoted by  $n_{it} \in \mathbb{N}$  for activity  $i$  at time  $t$ . This productivity applies both to dirty and clean technologies. If there has not been a switch to clean technology in activity  $i$ , then the firm with the best technology in this line at time  $t$  will be active in site  $i$  and produce

$$A^{n_{it}} f_d(k_{it}), \quad (11)$$

where  $k_{it} \equiv K_{it}/A^{n_{it}}$  is “normalized investment,”  $K_{it}$  is investment (again in terms of the consumption good),  $A = 1 + \alpha > 1$  so that each higher rung on the quality ladder secures a proportional improvement in productivity, and we continue to make the same assumptions on  $f_d$  ( $f'_d > 0$ ,  $f''_d < 0$ , and the Inada conditions).<sup>17</sup> We also assume that the dirty production technology emits  $k_{it}$  units of carbon given investment  $K_{it}$ .

If, on the other hand, activity  $i$  has switched to clean technology, the firm with the best technology for this activity at time  $t$  has access to the production technology

$$A^{n_{it}} f_c(k_{it}), \quad (12)$$

where again the same assumptions as in the static model apply to  $f_c$ , and as before, clean technology emits  $\gamma k_{it}$  units of carbon when the level of investment is  $K_{it}$ , where  $\gamma < 1$ . Consequently, total carbon-intensive investment at time  $t$  is

$$E_t = \int_0^1 [\mathbf{1}_{\{i \text{ is dirty}\}} K_{it} + \gamma \mathbf{1}_{\{i \text{ is clean}\}} K_{it}] di. \quad (13)$$

An important feature of our formulation is that even though productivity varies across activities, normalized investment will only differ between dirty and clean activities, and we thus denote these normalized values by  $k_{dt}$  and  $k_{ct}$  respectively for dirty and clean technologies at time  $t$ . Consequently, total carbon-intensive investment can also be expressed as

$$E_t = (1 - q_t) k_{dt} \mathbb{E}[A^{n_{it}} | i \text{ is dirty}] + q_t \gamma k_{ct} \mathbb{E}[A^{n_{it}} | i \text{ is clean}]$$

where  $q_t$  denotes the aggregate fraction of clean firms at time  $t$ .

<sup>17</sup> As in the static model, the Inada conditions imply that, despite productivity differences across activities, all activities will produce positive output. Moreover, since each site can house a single firm, only the firm with the best technology in that activity will produce.

The dynamics of the stock of carbon in the atmosphere, which we write directly in differential form since we focus on  $\Delta \Rightarrow 0$ , are given as

$$\dot{S}_t = \frac{E_t}{A_t} - \delta S_t, \tag{14}$$

where  $S_0 \geq 0$ ,  $\delta > 0$  is the environmental regeneration rate, and environmental damages are

$$A_t D(S_t; \zeta, v),$$

where

$$A_t \equiv \int_0^1 A^{n_{it}} di \tag{15}$$

is the average productivity of the economy at time  $t$ ,

$$D(S; \zeta, v) \equiv (1 - v)\tilde{D}(S - \zeta) \tag{16}$$

as in the static model, and  $\tilde{D}(\cdot)$  is increasing, strictly convex, and twice continuously differentiable in the stock of carbon  $S$ . We set the geoengineering parameters as  $(\zeta, v) = 0$  and omit them from our notation until the final subsection of this section. Note that damages are multiplied by average productivity, while carbon-intensive investment is divided by average productivity to obtain emissions. This formulation captures the fact that when the productivity level of the economy is higher, a given stock of carbon in the atmosphere will have more negative productivity or disutility implications, while ensuring that damages grow at the same rate as the economy.

Finally, we assume that the economy is inhabited by a representative household, discounting the future at the exponential rate  $\rho > 0$ . In the text we simplify the analysis (and keep it as close as possible to the static model) by assuming that this household obtains linear flow utility (more general utility functions are discussed in Appendix B). Thus the objective function of the household at time  $t$  is

$$\sum_{s=t}^{\infty} [C_{t+\Delta(s-t)} - A_{t+\Delta(s-t)} D(S_{t+\Delta(s-t)}; \zeta, v)] e^{-\rho \Delta(s-t)},$$

where  $C_s$  is consumption at time  $s$ . Once again, taking the limit  $\Delta \Rightarrow 0$ , we work with the continuous-time equivalent,

$$\int_t^{\infty} [C_s - A_s D(S_s; \zeta, v)] e^{-\rho(s-t)} ds. \tag{17}$$

The switch from dirty to clean technology has a fixed cost of  $A^{n_{it}} \Gamma > 0$  in terms of the consumption good for activity  $i$  with productivity  $A^{n_{it}}$ , incurred as a flow cost over an interval of positive measure, and incurred only once for each activity (because once an activity switches to clean technology, all future productivity improvements build on the existing clean technology in that activity or site). This formulation, which makes the cost of switching to clean technology proportional to productivity, ensures that the incentives to switch to clean technology remain independent of an activity's productivity.

Productivity improvements take place in a manner analogous to the standard quality-ladder models. Specifically, potential entrants invest in research and development (R&D) in order to improve over existing products. R&D uses a scarce input, say scientists, which has an inelastic supply of  $Z > 0$ .<sup>18</sup> We also assume that R&D is undirected, meaning that entrants decide their R&D effort, but cannot choose which activity they are researching and are randomly

<sup>18</sup> This formulation with an inelastic supply of scientists ensures that the overall growth rate of the economy will be insensitive to the rate of carbon taxation. We view this as a desirable benchmark property, since otherwise the planner would have an incentive to manipulate carbon taxes in order to affect the long-run growth rate.

matched to one of the activities in  $[0, 1]$ . A successful innovation for activity  $i$  currently with productivity  $A^{n_{it}}$  enables the entrant to replace the incumbent producer of this activity with a new technology with productivity  $A^{n_{it}+1}$ . Let us denote R&D effort (scientists hired) at time  $t$  by  $z_t$ . Then the (Poisson) arrival rate of a successful innovation is

$$\lambda_t = \varphi z_t, \tag{18}$$

where  $\varphi > 0$ . The cost of R&D effort of  $z_t$  is  $z_t w_t$ , where  $w_t$  denotes the equilibrium wage for scientists. This wage is determined from the market-clearing condition for scientists given by

$$z_t = Z \quad \text{for all } t \geq 0. \tag{19}$$

This naturally ensures that in equilibrium

$$\lambda_t = \lambda \equiv \varphi Z.$$

Taking into account the expenditures on switching to clean technology, the resource constraint of the economy implies that consumption at time  $t$  is given as

$$C_t = \int_0^1 A^{n_{it}} [f_i(k_{it}) - k_{it} - \mathbf{1}(t = \inf\{t \geq 0 : q_{it} = 1\}) \Gamma] di,$$

which integrates over the output levels of different activities and then subtracts the costs of investment in clean technology (where  $\mathbf{1}(t = \inf\{t \geq 0 : q_{it} = 1\})$  is the indicator function for the time at which activity  $i$  switches to clean technology and incurs the fixed cost  $A^{n_{it}} \Gamma$ ).

#### 4.2. Carbon tax and production decisions

As in the static model, there is a carbon tax of  $\tau_t$  at time  $t$ . Profits of dirty and clean firms can be written, respectively, as

$$\begin{aligned} \Pi_{idt} &= \max_{k \geq 0} A^{n_{it}} [f_d(k) - (1 + \tau_t)k] \\ &= A^{n_{it}} [f_d(k_d(\tau_t)) - (1 + \tau_t)k_d(\tau_t)] \end{aligned} \tag{20}$$

and

$$\begin{aligned} \Pi_{ict} &= \max_{k \geq 0} A^{n_{it}} [f_c(k) - (1 + \gamma \tau_t)k] \\ &= A^{n_{it}} [f_c(k_c(\tau_t)) - (1 + \gamma \tau_t)k_c(\tau_t)], \end{aligned} \tag{21}$$

where  $k_c(\tau_t)$  and  $k_d(\tau_t)$  are then defined as the optimal input decisions for dirty and clean firms respectively. We use  $\pi_j(\tau_t) \equiv \Pi_{ijt}/A^{n_{it}}$  to denote normalized profits of activity  $j \in \{c, d\}$  at time  $t$ .

We next write the value functions of firms with clean and dirty technologies as a function of their productivity. At time  $t$ , a clean incumbent with productivity  $A^n$  has (expected) net present discounted value given by the usual dynamic programming recursion (provided that this value is a differentiable function of time):

$$r_t V_{ct}(n) = A^n \pi_c(\tau_t) + \dot{V}_{ct}(n) - \lambda V_{ct}(n).$$

Intuitively, the firm receives a "dividend" of  $A^n \pi_c(\tau_t)$  on its asset of  $V_{ct}(n)$ , but also recognizes that this asset may change value, captured by the term  $\dot{V}_{ct}(n)$ , and may entirely disappear because of creative destruction coming from improvements by entrants, which takes place at the Poisson rate  $\lambda_t$  and will make the incumbent lose the asset entirely. This stream of profits is then discounted at the interest rate  $r_t$ . Because the household's preferences are linear, the interest rate is always equal to the discount rate, i.e.,

$$r_t = \rho,$$

and the current owner obtains the flow of profits associated with  $A^n$  only until an entrant supplants them, this expected net present discounted value can be expressed as

$$V_{ct}(n) = A^n \int_t^\infty \pi_c(\tau_s) e^{-(\rho+\lambda)(s-t)} ds, \tag{22}$$

which is just the discounted integral of flow profits  $\pi_c(\tau_s)$  over time, adjusted for the baseline productivity of the firm and the Poisson rate  $\lambda$  of arrival of creative destruction.

The expected net present discounted value of dirty firms is similar, except that they can choose whether to switch to clean technology at a cost  $A^{n_c}\Gamma$ , so that

$$V_{dt}(n) = \max \left\{ V_{ct}(n) - A^n \Gamma, \frac{A^n \pi_d(\tau_t) + \dot{V}_{dt}(n)}{\rho + \lambda} \right\}. \tag{23}$$

The max operator takes care of the choice to switch to clean technology, while the second part is the dynamic programming recursion rearranged (with  $r_t = \rho$  imposed).

Eqs. (23) and (24) show that  $V_{jt}(n)/A^n$  is independent of  $n$  for  $j \in \{c, d\}$ , and we thus define  $v_{jt} \equiv V_{jt}(n)/A^n$  as the normalized value function.

### 4.3. Clean technology and R&D decisions

Eq. (23) immediately gives us the equivalent of (Technology IC) in the static model. Firms are happy to switch to clean technology if the maximization operator in this expression picks the first term, which, put in terms of normalized value functions, holds if

$$v_{dt} = v_{ct} - \Gamma. \tag{24}$$

This binding constraint will play an analogous role to (5) in the static model, and implies the following form for incentive-compatible technology choice  $q_t$ :

$$\begin{cases} v_{dt} = v_{ct} - \Gamma \Rightarrow q_t \in [0, 1] \\ v_{dt} > v_{ct} - \Gamma \Rightarrow q_t \leq q_s \text{ for all } s < t. \end{cases} \text{ (Dynamic Technology IC)}$$

which closely resembles its analogue in the static model.<sup>19</sup>

Next, using the characterization of the value functions in the previous subsection, we derive equilibrium R&D decisions. Since potential entrants have access to the R&D technology given by (19), equilibrium requires the following free-entry condition to hold with complementary slackness

$$\varphi \int_0^1 [q_t V_{ct}(n_{it} + 1) + (1 - q_t) V_{dt}(n_{it} + 1)] di - w_t = 0,$$

where  $V_{jt}(n)$  for  $j \in \{c, d\}$  are the expected value functions defined in (23) and (24),  $w_t$  is the equilibrium wage for scientists, and the integral reflects the fact that R&D is undirected and may lead to an improvement over a clean or dirty technology. Using the definition of normalized value functions, the free-entry condition can be simplified to the following form

$$q_t v_{ct} + (1 - q_t) v_{dt} = \frac{w_t}{\varphi A_t}. \tag{25}$$

At each  $t$ , the wage for scientists,  $w_t$ , adjusts to satisfy (25) (so  $z_t = Z$ ).

### 4.4. Planner's problem

As in the static model, the (social) planner is benevolent, and therefore maximizes the same objective as the representative household, (17). She will seek to achieve this objective by choosing

<sup>19</sup> Unlike the static condition (Technology IC), however, there is no case in which  $v_{dt} < v_{ct} - \Gamma$ , since  $v_{dt} = \max \left\{ v_{ct} - \Gamma, (\rho + \lambda)^{-1} (\pi_{dt} + \dot{v}_{dt}) \right\}$  implies that  $v_{dt} \geq v_{ct} - \Gamma$  for all  $t \geq 0$ . Naturally, the equilibrium involves  $q_t = 1$  when the max operator always strictly picks the first term in (23). We provide conditions for this not to be the case in equilibrium in Assumption 2' below.

a sequence of carbon taxes,  $(\tau_t)_{t \geq 0}$ . We also continue to assume that the planner does not have access to a commitment technology, so the sequence of carbon taxes can be revised at any  $t$ . As in the static model, the planner's preferred allocation differs from that of the firms in two ways. First, firms do not internalize the environmental damage they create (except through the carbon taxes that the planner imposes). Second, they fail to internalize the positive externality that they create for future producers of the same activity when they switch to clean technology. This externality is again proportional to the likelihood of replacement,  $\lambda$ .

### 4.5. Definition of equilibrium

We focus on (subgame perfect) Markovian equilibria where agents condition their strategies at  $t$  on the state  $\Omega_t \equiv (S_t, q_t, \tau_t, \{n_{it}\}_{i \in [0,1]})$ , and form rational expectations about the evolution of future states. The state  $\Omega_t$  contains all current payoff-relevant variables. In a Markov equilibrium, therefore, at each  $t$ , both the government's tax policy  $\mathcal{T}(\Omega_t)$  and the clean firms' switching decision  $\mathcal{Q}(\Omega_t)$  are functions of the current carbon stock  $S_t$ , clean technology level  $q_t$ , carbon tax  $\tau_t$ , and technological progress  $\{n_{it}\}_{i \in [0,1]}$ . This focus on Markovian equilibria is motivated by our main interest, which is to understand the implications of lack of commitment to future carbon taxes. In an infinite-horizon setup, non-Markovian equilibria may sometimes mimic commitment policies.<sup>20</sup> Our Markov concept precludes the possibility of these reputation-based commitment devices by ruling out strategies at  $t$  that condition on (payoff-irrelevant) aspects of the history of play not captured by  $(S_t, q_t, \{n_{it}\}_{i \in [0,1]})$ .

A *dynamic (Markov) equilibrium*, or an equilibrium for short, is given by a path of technology choices, taxes, input decisions, wages for scientists, and stock of carbon  $\{(q_t^*)_{t \geq 0}, (\tau_t^*)_{t \geq 0}, (k_{dt}^*)_{t \geq 0}, (k_{ct}^*)_{t \geq 0}, (w_t^*)_{t \geq 0}, (S_t^*)_{t \geq 0}\}$ ,

- Given  $(q_t^*)_{t \geq 0}$ , carbon taxes  $\tau_t^* = \mathcal{T}(\Omega_t)$  maximize household utility (18) at each  $t \geq 0$ ,
- Given  $(\tau_t^*)_{t \geq 0}$ , clean technology decisions  $q_t^* = \mathcal{Q}(\Omega_t)$  satisfy (Dynamic Technology IC) for all  $t \geq 0$ ,
- Given  $\tau_t^*$ , input choices  $k_{dt}^*$  and  $k_{ct}^*$  maximize, respectively,  $\pi_{dt}$  and  $\pi_{ct}$  in (20) and (21), for all  $t \geq 0$ ,
- Given  $(\tau_t^*)_{t \geq 0}$  and  $(q_t^*)_{t \geq 0}$ , the equilibrium R&D intensity  $z_t$  and wages  $w_t$  satisfy labor market clearing (19) and free entry (25) for each  $t \geq 0$ .
- The carbon stock  $(S_t^*)_{t \geq 0}$  evolves according to the law (14).

Our definition embeds the Markov feature discussed above in the policy functions  $\mathcal{T}$  and  $\mathcal{Q}$ , which preclude carbon tax or clean technology strategies that depend on any features of the history of play prior to  $t$  not contained in the state  $\Omega_t$ . The equilibrium has a block recursive structure whereby the remaining variables can be determined from  $(\tau_t^*)_{t \geq 0}$  and  $(q_t^*)_{t \geq 0}$ . In view of this, we use the shorthand of referring to an equilibrium as  $(\tau_t^*, q_t^*)_{t \geq 0}$ .

We also define a *Balanced Growth Path Equilibrium (BGP)* as an equilibrium in which  $(\tau_t^*, q_t^*) = (\hat{\tau}, \hat{q})$  for all  $t$ , so that aggregate output  $A_t$  grows at a constant rate given by

<sup>20</sup> In our setup, this would take the form of trigger strategies where the social planner expects worse actions from the firms following a lower-than-promised carbon tax. Though such schemes are not always feasible, they nevertheless complicate the analysis. See, for example, Phelan and Stacchetti (2001), who characterize the equilibrium value set for a class of dynamic policy games between a taxation authority that cannot commit and a continuum of households.

$$g \equiv \alpha\lambda = \alpha\varphi Z,$$

where the presence of the term  $\alpha = A - 1$  follows from the properties of the Poisson process.<sup>21</sup> We will also see that in a BGP,  $S_t = \widehat{S}$  for all  $t$ . When this causes no confusion, we will also include  $\widehat{S}$  in the definition of a BGP (or  $S_t^*$  in the definition of an equilibrium).

#### 4.6. Farsighted Pigovian taxes

To characterize the equilibrium tax sequence, we start by determining the evolution of marginal environmental damages or, equivalently, the shadow price of carbon emissions, which will give us the dynamic equivalent of Pigovian taxation (or what we will call “farsighted Pigovian taxes”). Consider the Hamiltonian corresponding to the planner’s maximization problem, in (17), subject to the evolution of the stock of carbon given in (14),

$$H_t(K_t, S_t) = C_t - A_t D(S_t) - \mu_t [E_t/A_t - \delta S_t], \tag{26}$$

where  $\mu_t$  is the costate variable associated with the stock of carbon in the atmosphere.<sup>22</sup> Since emissions are divided by average productivity,  $A_t$ , the shadow value of carbon emissions is given by

$$p_t = \mu_t/A_t. \tag{27}$$

Since the Hamiltonian is concave, the necessary and (with the usual transversality condition) sufficient first-order condition for optimality is

$$\partial H_t / \partial S = \dot{\mu}_t - \rho \mu_t,$$

which yields a simple form for the shadow price of carbon emissions provided that the planner’s maximization problem in (26) is well-behaved (in particular has a finite value). The next assumption ensures this:

#### Assumption 3 (Growth).

$$g \equiv \alpha\varphi Z < \rho + \delta.$$

Under this assumption, we have:

**Lemma 3 (Shadow Cost of Carbon).** Suppose Assumption 3 holds. Then, along any optimal path,

$$\dot{p}_t = -D'(S_t) + (\delta + \rho - \alpha\lambda)p_t \tag{28}$$

and thus

$$p_t = \int_t^\infty D'(S_s) e^{-(\delta + \rho - \alpha\lambda)(s-t)} ds, \tag{29}$$

for all  $t \geq 0$ .

We refer to the tax trajectory implied by (29) as “farsighted Pigovian.” This terminology emphasizes that this tax sequence is a direct generalization of our static Pigovian tax. The generalization accounts for the fact that emissions create damages not only today but at all future dates, which means that the shadow price of carbon emissions must incorporate the discounted cost of these future damages.

Our next result shows that equilibrium taxes—due to the lack of commitment of the planner—must equal the farsighted Pigovian

<sup>21</sup> Each  $n_{it}$  is a sample path of a Poisson process with intensity  $\lambda t$ , so  $\int_0^1 A^{n_{it}} di$  corresponds to the expectation of  $A^{N_t} = \exp(N_t \log A)$ , where  $N_t \sim \text{Pois}(\lambda t)$  so that  $E[e^{\phi N_t}] = \exp(\lambda t(e^\phi - 1))$  for any  $\phi \in \mathbb{R}$ .

<sup>22</sup> The full maximization problem would also need to impose constraints for the evolution of the states of clean technology,  $q_t$ , and average productivity in the economy,  $A_t$ , but as these constraints do not change the expression for the shadow price of carbon, we omit them from our exposition in the text.

<sup>23</sup> The main technical detail, showing that equilibrium clean technology indeed always converges in finite time, is stated and proven as Lemma B1 in the Appendix.

taxes characterized in (29), at least once clean technology converges.<sup>23</sup>

**Proposition 10 (Pigovian Best-response).** There exists  $T < \infty$  such that equilibrium taxes are given by

$$\tau_t = p_t \tag{30}$$

for all  $t \geq T$ .

Proposition 10 shows that, despite the complicated dependence of clean technology and R&D decisions on the entire tax trajectory, equilibrium carbon taxes take a simple form. In fact, in (29), these taxes only depend on the evolution of the stock of carbon in the atmosphere  $(S_t)_{t \geq 0}$ . The key to understanding this result is that absent technology choices, the (farsighted) Pigovian taxes are optimal (with or without commitment), and the lack of commitment, combined with the Markovian restriction, precludes the planner from choosing a tax sequence that is ex post distortionary (different from Pigou), once the transition to cleaner technology is complete (either with  $q_t = 1$  or  $q_t = \hat{q} < 1$ ). This transition is completed within some finite time  $T$ , enabling us to use backward induction to prove the proposition.<sup>24</sup>

**Remark 2 (Counterexample to pure Pigovian taxes).** Proposition 10 establishes that  $\tau_t = p_t$  for all  $t \geq T$ . In addition, we can prove that  $\tau_t \leq p_t$  for all  $t \geq 0$ . But there might be some circumstances in which the social planner prefers to set a tax rate strictly less than the Pigovian one in the interval  $[0, T]$  in order to increase future Pigovian taxes and encourage a faster switch to clean technology. We analyze the conditions under which this possibility could arise in Appendix B, but also prove that such a counterexample is possible only if  $\lambda$  is very high (in fact, so high that all geoengineering technologies are strictly welfare reducing.)

We will see later that, if she could commit, the planner would prefer to deviate from this Pigovian tax scheme.

#### 4.7. Characterization of equilibrium

To characterize the dynamic equilibrium, we impose dynamic analogues of Assumptions 1 and 2, which will again rule out Jevons’ paradox and guarantee an “interior” equilibrium.

**Assumption 1’ (Dynamic No Jevons)** For all  $t \geq 0$  and all  $\tau \geq 0$ , we have

$$\Lambda(\tau_t) \equiv k_d(\tau_t) - \gamma k_c(\tau_t) > 0.$$

This assumption enables us to develop another parallel with the static model. Analogously with (4), let us define

$$\Psi(\tau_t) \equiv \pi_c(\tau_t) - \pi_d(\tau_t) = f_c(k_c(\tau_t)) - k_c(\tau_t) - (f_d(k_d(\tau_t)) - k_d(\tau_t)) + \Lambda(\tau_t)\tau_t \tag{31}$$

as the difference in (normalized) profits between clean and dirty technologies at carbon tax  $\tau_t$ . Recall that in the static model, Lemma 2 ensured that  $\Psi'(\tau) = \Lambda > 0$ . Here, we similarly have  $\Psi'(\tau_t) = \Lambda(\tau_t) > 0$  by Assumption 1’. Moreover, in an interior BGP where  $(\tau_t, q_t) = (\hat{\tau}, \hat{q})$  with  $\hat{q} \in (0, 1)$ , we obtain a simplified form of (Dynamic Technology IC),

$$v_{dt} - v_{ct} = \frac{\Psi(\hat{\tau})}{\rho + \lambda} = \Gamma. \tag{32}$$

<sup>24</sup> This result is reminiscent of the generic time-inconsistency result of Calvo (1978). It is also simplified since  $\Delta \rightarrow 0$ , which removes the possibility of choosing a distortionary tax today in order to affect behavior until the taxes are adjusted tomorrow.

Here the first equality exploits the fact that when  $\hat{q} < 1$ ,  $v_{dt}$  is equal to the discounted stream of profits from dirty technology, and that profits, taxes and the creative destruction rate,  $\lambda$ , are constant, and the second equality follows from (24).

**Assumption 2' (Conditions for Dynamic Interior Equilibrium)** Let the initial carbon stock be  $S_0$ . Then for all  $t \geq 0$ ,

$$\Gamma \in \left( \int_t^\infty \Psi(\underline{\tau}_s) e^{-(\rho+\lambda)(s-t)} ds, \int_t^\infty \Psi(\bar{\tau}_s) e^{-(\rho+\lambda)(s-t)} ds \right)$$

where

$$\bar{\tau}_t = \int_t^\infty D' \left( S_0 e^{-\delta s} + \int_0^s k_d(\bar{\tau}_v) e^{-\delta(s-v)} dv \right) e^{-(\delta+\rho-\alpha\lambda)(s-t)} ds$$

and

$$\underline{\tau}_t = \int_t^\infty D' \left( S_0 e^{-\delta s} + \int_0^t k_d(\underline{\tau}_v) e^{-\delta(t-v)} dv + \gamma \int_t^s k_c(\underline{\tau}_v) e^{-\delta(s-v)} dv \right) e^{-(\delta+\rho-\alpha\lambda)(s-t)} ds.$$

Although notationally cumbersome, this assumption has an identical interpretation as its static counterpart, Assumption 2. Specifically, it ensures that the cost of switching to clean technology is neither too high nor too low—and the relevant thresholds depend on the farsighted Pigovian taxes and R&D intensities that will prevail when no firm ever switches to clean technology,  $(\bar{\tau}_t)_{t \geq 0}$ , or all firms switch to clean technology,  $(\underline{\tau}_t)_{s \geq t}$ . As in its static analogue, Assumption 2, the conditions in Assumption 2' depend on the initial stock of carbon, because this determines the entire path of Pigovian taxes.

We start by characterizing the BGP in which  $(\tau_t, q_t) = (\hat{\tau}, \hat{q})$  for all  $t$ , which also ensures that the stock of carbon in the atmosphere converges to some finite  $\hat{S}$ . From (14), this limiting value of the stock of carbon must satisfy

$$\hat{q} \gamma k_c(\hat{\tau}) + (1 - \hat{q}) k_d(\hat{\tau}) = \delta \hat{S}. \tag{33}$$

Using (29) and (30), the stationary Pigovian tax  $\hat{\tau}$  is given by

$$\hat{\tau} = \frac{D'(\hat{S})}{\delta + \rho - \alpha\lambda}. \tag{34}$$

These two equations together with (32) determine  $(\hat{S}, \hat{\tau}, \hat{q})$ . The next proposition establishes that such a BGP exists and is unique.

**Proposition 11 (Existence, Uniqueness of the Balanced Growth Path).** Suppose Assumptions 1', 2', and 3 hold. Then there exists a unique BGP where  $(S_t, \tau_t, q_t) = (\hat{S}, \hat{\tau}, \hat{q})$ , and  $(\hat{S}, \hat{\tau}, \hat{q})$  is the unique solution to equations (32)–(34).

The existence of a BGP  $(\hat{S}, \hat{\tau}, \hat{q})$  follows from the equations and arguments proceeding the proposition. The uniqueness of this BGP is a consequence of the fact that the BGP farsighted Pigovian tax  $\hat{\tau}$  is a decreasing function of  $\hat{q}$ . Once the incentive-compatible carbon tax,  $\hat{\tau}$ , is pinned down by (32), there exists a unique  $\hat{q}$  that solves (34). These two variables then yield a unique value of  $\hat{S}$ .

A noteworthy feature of the unique BGP is that, as in our static model,  $\hat{q} \in (0, 1)$  and the equilibrium is “interior.” This, in particular, ensures that in the BGP, (32) holds, which restricts the value of the BGP carbon tax to  $\hat{\tau}$ . The next proposition shows that every equilibrium converges to the BGP equilibrium in Proposition 11, and does so by some  $T < \infty$ .

**Proposition 12 (Interior Dynamic Equilibrium).** Suppose Assumptions 1', 2', and 3 hold. Then the unique dynamic equilibrium takes the following form. There exists a  $T < \infty$  such that:

- for all  $t \in [0, T)$ ,  $\tau_t$  and  $S_t$  grow continuously and  $q_t = 0$ .
- for all  $t \geq T$ ,  $(S_t, q_t, \tau_t) = (\hat{S}, \hat{q}, \hat{\tau})$ , where  $(\hat{S}, \hat{q}, \hat{\tau})$  is given in Proposition 11.

Fig. 6 illustrates the shape of the dynamic equilibrium. The stock of carbon is always nondecreasing, and smoothly increasing until the economy reaches the BGP. Therefore marginal environmental damages and Pigovian taxes also increase until they reach their constant BGP level  $\hat{\tau}$ . As the Pigovian tax grows, clean technology incentives also increase—eventually (by monotonicity of  $\Psi(\tau_t)$ ) reaching the value for which (32) holds, at which point clean technology leaps from zero to  $\hat{q}$  along a most rapid approach path (Spence and Starrett, 1975).

The proof of Proposition 12 is provided in Appendix B. Here we give some intuition. Proposition 11 established that the BGP has to be “interior”—if all activities eventually switched to clean technology, the subsequent carbon taxes would be too low to make such a switch optimal, whereas if no activity switches to clean technology, the stock of carbon and thus future carbon taxes would be sufficiently high to incentivize investment in clean technology. Proposition 12 then shows how we get to this BGP. Initially, with a lower stock of carbon in the atmosphere than the BGP value, the marginal damage of carbon emissions is low, so Pigovian taxes are also low, and consequently the transition path involves faster growth of emissions than in the BGP. When the stock of carbon reaches  $\hat{S}$ , the fraction of firms that have already transitioned to clean technology must be exactly the BGP value,  $\hat{q}$ , to sustain the (stationary) Pigovian tax sequence that maintains the dynamic technology IC, (32), so that we have  $\tau_t = \hat{\tau}$  for all  $t \geq T$ .

#### 4.8. Second-best

We noted above that, as in the static model, if she could commit, the planner would set a carbon tax sequence different than the Pigovian one. In this subsection, we prove this claim. As in Proposition 3 in our static analysis, the next result shows that whenever  $\lambda > 0$ , the second-best deviates from Pigovian taxation. The main differences are that the condition that  $\lambda > 0$  is now automatically satisfied in any BGP with productivity growth (provided that  $Z > 0$ ). Second-best carbon taxes,  $\tau_t^{SB}$ , exceed Pigovian ones (are greater than the shadow price of carbon emissions,  $p_t$ ), and induce more firms to switch to clean technology. In contrast, if  $\lambda = 0$  so that there is no growth in productivity in this economy, second-best and Pigovian taxes coincide.

**Proposition 13 (Dynamic Second-best).**

- Suppose that  $Z > 0$  (which ensures that  $\lambda > 0$ ). Then the planner commits to a carbon tax  $\tau_t^{SB} \geq p_t^{SB}$  for all  $t \geq 0$ , with  $\tau_t^{SB} > p_t^{SB}$  for some  $t \geq 0$ , and the equilibrium fraction of firms that switch to clean technology converges to  $q^{SB} > \hat{q}$ .
- Suppose that  $Z = 0$  (so that  $\lambda = 0$ ). Then for all  $t \geq 0$ ,  $\tau_t^{SB} = p_t$  and the equilibrium fraction of firms that switch to clean technology converges to  $q_t^{SB} = \hat{q}$ .

#### 4.9. Geoengineering

We next consider the implications of geoengineering breakthroughs on dynamic carbon taxation, environmental damages and welfare. We focus on the BGP derived in Proposition 11, and show that the results are essentially identical to the effects of geoengineering in the static model, derived in Section 2.10. We again

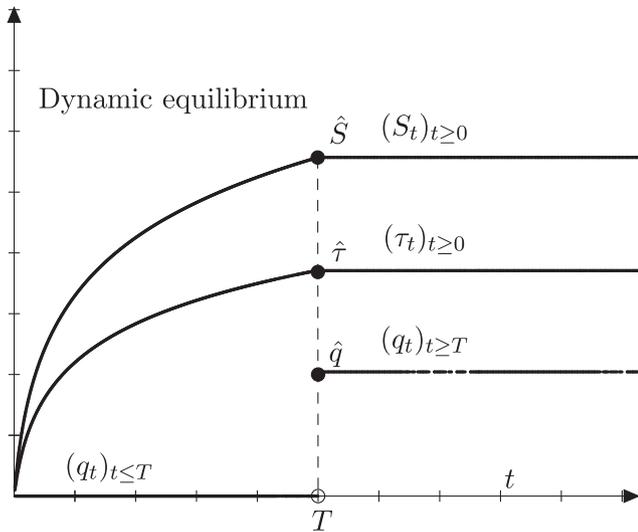


Fig. 6. Time paths of the carbon stock  $(S_t)_{t \geq 0}$ , optimal taxes  $(\tau_t)_{t \geq 0}$ , and clean technology  $(q_t)_{t \geq 0}$  in the dynamic equilibrium.

distinguish between the two types of geoengineering advances, captured by the parameters  $\xi$  and  $v$  in the general damage function  $(1 - v)D(S_t - \xi)$ .

**Proposition 14** (*Dynamic Implications of Type I Geoengineering Technologies*). Suppose that Assumptions 1', 2', and 3 hold, and the economy's unique BGP is given by  $(\hat{S}, \hat{q}, \hat{\tau})$ . Consider a geoengineering technology improvement of type I that increases  $\xi$  by a small amount  $d\xi > 0$ . Then:

- $d\hat{\tau}/d\xi = 0$  (taxes do not change).
- $d\hat{S}/d\xi = 1$  (the stock of carbon increases).
- $d\hat{q}/d\xi = -\delta/\Lambda < 0$  (clean technology falls).
- $dW/d\xi < 0 \Rightarrow \frac{\lambda}{\rho + \lambda}(\hat{\pi}_c - \hat{\pi}_d) > \Lambda \hat{\tau}$  (welfare may decline).

This proposition shows that any geoengineering advance of type I results in conclusions similar to Proposition 4—in the BGP, the stock of carbon in the atmosphere increases and welfare (in the BGP) may even decline if there is a sufficiently strong response of investment in clean technology.

The next proposition gives the dynamic analogues of Proposition 5. Geoengineering lowers equilibrium clean technology adoption, and welfare may decline.

**Proposition 15** (*Dynamic Implications of Type II Geoengineering Technologies*). Suppose that Assumptions 1', 2', and 3 hold, and the economy's unique BGP is given by  $(\hat{S}, \hat{q}, \hat{\tau})$ . Consider a geoengineering technology improvement of type II that increases  $v$  by a small amount  $dv > 0$ , and let  $\eta = \hat{S}D''(\hat{S})/D'(\hat{S})$  be the elasticity of the marginal damage function. Then

- $d\hat{\tau}/dv = 0$  (taxes do not change).
- $d\hat{S}/dv = \frac{D'(\hat{S})}{(1-v)D''(\hat{S})} > 0$  (the stock of carbon increases).
- $d\hat{q}/dv = -\frac{\delta S}{(1-v)\Lambda} \frac{1}{\eta} < 0$  (clean technology declines).
- $dW/dv < 0 \Rightarrow \eta < \eta^{\text{II}}(\lambda)$ , where

$$\eta^{\text{II}}(\lambda) \equiv a \left( \delta(\rho - g + \delta) \frac{\lambda}{\rho + \lambda} \frac{\hat{\pi}_c - \hat{\pi}_d}{\Lambda \hat{\tau}} + \frac{\delta}{\rho - g + \delta} \right)$$

and  $a \equiv \hat{S}D'(\hat{S})/D(\hat{S}) > 1$  (welfare may decline).

We note in addition that the conditions for welfare to decline as a result of a geoengineering advance of type II are again very similar to those we have obtained in the static model in Proposition 5. In particular, as in the static model, if  $\lambda$  is sufficiently large, welfare declines following both types of geoengineering.

### 5. Quantitative evaluation

We now undertake an illustrative quantitative exercise to investigate whether, for environmental damages, emissions, entry and exit rates, and clean technology investment costs consistent with available evidence, geoengineering advances could indeed reduce welfare. For this exercise, we focus on the BGP of the dynamic model, and abstract from various other important factors, such as political economy issues, uncertainty and transitional dynamics.

#### 5.1. Production

We assume that firm-level dirty and clean production technologies are identical and Cobb-Douglas,

$$f_d(k) = f_c(k) = \vartheta k^v,$$

which implies that the only benefit from switching to clean technology is reducing emissions. The parameter  $v$  is the elasticity of output with respect to energy use, and we set  $v = 0.046$  to match the total cost share of energy.<sup>25</sup> The parameter  $\vartheta$  is then chosen so that the model matches world output  $Y_t$  in 2016 (from World Bank, 2019).

Energy inputs  $k_{dt}$  and  $k_{ct}$  are denominated in gigatonnes of carbon (GtC), and emissions are given by (13). We set  $\gamma = 0$  since emissions from non-fossil fuel energy are negligible and normalize  $A_{2016} = 1$ . Since the dirty energy input and emissions are equal in our model, we set  $(1 - q_t)k_{dt} = \sum_j \epsilon^j K_{dt}^E$ , where  $K_{dt}^E$  is the global energy consumption of type  $d_j$  (in particular,  $d_j \in \{\text{coal, oil, natural gas}\}$ ) and  $\epsilon^j$  is the average carbon intensity of energy source  $d_j$ , obtained from EPA (2018).

Given the choice of units in our model,  $q_t k_{ct}$  should be measured in “carbon-equivalent” units, so that we capture the amount of carbon emissions that would have been generated from the same consumption of energy had it not been generated by clean technologies. To do this, we scale total clean energy consumption by the average carbon intensity of dirty inputs,  $\epsilon_d = \frac{\sum_j \epsilon^j K_{dt}^E}{\sum_j K_{dt}^E}$ , and thus

$q_t k_{ct} = \epsilon_d K_{ct}^E$ . We normalize the price of the inputs  $k_{dt}$  and  $k_{ct}$  to 1 using a carbon price index  $P_k$ .<sup>26</sup> Table 1 summarizes our parameter choices.

<sup>25</sup> This follows Hassler et al. (2016), who argue that Cobb-Douglas is a reasonable long-run approximation in view of the fact that the cost share of energy does not have a trend. In computing  $v$  from the cost share of energy, we take into account the effect of carbon taxes. In particular, with the variables defined below, we use the equation  $v = P_k [\epsilon_d K_{ct}^E + (1 + \tau_t) \sum_j \epsilon^j K_{dt}^E] / Y_t = 0.046$ , which follows from firms' cost minimization. This yields  $v = 0.046$ .

<sup>26</sup> Using commodity price data averaged over 2000–2016 and the carbon intensities mentioned above, the prices per carbon tonne are \$410.38/tC for oil, \$336.12/tC for natural gas, and \$89.88/tC for coal, or an average carbon price index of  $P_k = \$284.11/tC$ .

**Table 1**  
Production technology parameters.

| Parameter                   |                                    | Value  | Source   |
|-----------------------------|------------------------------------|--------|--|
| $\nu$                       | energy-output elasticity           | 0.0460 | energy services as a fraction of global GDP  |
| $\gamma$                    | clean tech emissions rate          | 0      | n.a.   |
| $Y_t$                       | world output                       | 77.797 | trillion 2010 USD, 2016, market exchange rate (World Bank, 2019)   |
| $\vartheta$                 | production function TFP            | 69.843 | trillion 2010 USD (World Bank, 2019)   |
| $g$                         | rate of emissions growth           | 0.0199 | $(K_{2016}^E - K_{2000}^E)^{1/18} - 1$   |
| $\lambda$                   | replacement probability            | 0.0175 | U.S. firm-level exit rate (Acemoglu et al., 2016)  |
| $\{K_{d_j}^E\}$             | coal consumption                   | 3.731  | gigatonnes of oil equivalent (Gtoe) (IEA, 2018)  |
|                             | oil consumption                    | 4.390  | Gtoe (IEA, 2018)   |
|                             | natural gas consumption            | 3.035  | Gtoe (IEA, 2018)   |
| $\{\vartheta^j K_{d_j}^E\}$ | coal emissions <sup>a</sup>        | 2.931  | $3.731 \text{ Gtoe} \cdot 0.786 \frac{\text{tC}}{\text{toe coal}} = 2.931 \text{ GtC}$   |
|                             | oil emissions <sup>b</sup>         | 3.791  | $4.390 \text{ Gtoe} \cdot 0.863 \frac{\text{tC}}{\text{tonne oil}} = 3.791 \text{ GtC}$  |
| $(1 - \hat{q})\hat{k}_d$    | natural gas emissions <sup>c</sup> | 1.742  | $3.035 \text{ Gtoe} \cdot 1.58 \frac{\text{tonnes natural gas}}{\text{tonne oil}} \cdot 0.574 \frac{\text{tC}}{\text{tonnes natural gas}} = 1.742 \text{ GtC}$ |
|                             | total dirty inputs                 | 8.463  | $\sum_j \vartheta^j K_{d_j}^E \text{ GtC}$   |
| $\hat{K}_c^E$               | clean energy consumption           | 2.604  | Gtoe (IEA, 2018)   |
| $\hat{q}\hat{k}_c$          | total clean inputs                 | 1.975  | $\frac{1}{1.318} \frac{\text{GtC}}{\text{Gtoe}} \cdot 2.604 \text{ Gtoe} = 1.975 \text{ GtC}$  |
| $\{p_d\}$                   | coal price                         | 89.88  | USD/tC, Australian world coal price (2000–2017) <sup>d</sup>   |
|                             | oil price                          | 410.38 | USD/tC, Brent crude price (2000–2017) <sup>e</sup>   |
| $P_k$                       | natural gas price                  | 336.12 | USD/tC, U.S. Henry Hub Gas world price (2000–2017) <sup>f</sup>  |
|                             | carbon price index                 | 284.11 | USD/tC, $P_k = (\sum_j p_d k_{d_j}) / (\sum_j \hat{k}_{d_j})$  |

<sup>a</sup> From EPA, 2018. Coal:  $21.11 \frac{\text{mmbtu}}{\text{tonne coal}} \cdot 26.05 \frac{\text{kgC}}{\text{mmbtu}} \cdot 10^{-3} \frac{\text{tonne}}{\text{kg}} \cdot \frac{1 \text{ tC}}{0.7 \text{ toe}} = 0.786 \frac{\text{tC}}{\text{toe coal}}$ .  
<sup>b</sup> Oil:  $\frac{1}{7.33} \frac{\text{toe}}{\text{barrel}} \cdot 5.80 \frac{\text{mmbtu}}{\text{barrel}} \cdot 20.31 \frac{\text{kgC}}{\text{mmbtu}} \cdot 10^{-3} \frac{\text{tonne}}{\text{kg}} \cdot 10^{-3} \frac{\text{tC}}{\text{toe}} = 0.863 \frac{\text{tC}}{\text{toe oil}}$ .  
<sup>c</sup> Natural gas:  $0.1 \frac{\text{mmbtu}}{\text{thermal natural gas}} \cdot 14.46 \frac{\text{kgC}}{\text{mmbtu}} \cdot 10^{-3} \frac{\text{tonne}}{\text{kg}} \cdot 396.8321 \frac{\text{therm}}{\text{toe}} = 0.574 \frac{\text{tC}}{\text{toe natural gas}}$ .  
<sup>d</sup> Average Australian world coal price 1 January 2000 to 1 July 2017, of \$70.61/tonne, and  $0.786 \frac{\text{tC}}{\text{toe coal}}$ .  
<sup>e</sup> Average price of Brent crude, \$64.87/barrel from 1 January 2000 to 1 January 2017. The calculation is  $64.87 \frac{\text{USD}}{\text{barrel}} \cdot 7.33 \frac{\text{barrel}}{\text{toe}} \cdot 0.863 \frac{\text{tC}}{\text{toe}} = 410.38 \frac{\text{USD}}{\text{tC}}$ . Golosov et al. (2014), use \$70/barrel, from 2005–2009.  
<sup>f</sup> Average world price of U.S. Henry Hub Gas, 1 January 2000 to 1 July 2017. The calculation is  $4.86 \frac{\text{USD}}{\text{natural gas mmbtu}} \cdot \frac{1}{14.46} \frac{\text{mmbtu}}{\text{kgC}} \cdot 10^3 \frac{\text{kgC}}{\text{tC}} = 336.12 \frac{\text{USD}}{\text{tC}}$ .

We identify the equilibrium fraction of clean firms  $\hat{q}$  in the BGP from the fraction of global clean energy use.<sup>27</sup> In Appendix C we use the first-order conditions of energy firms to derive the following relationship between the measure of clean firms and the observed level of clean energy production,  $K_c^E$ :

$$q_t = \frac{\vartheta_d K_{ct}^E}{\vartheta_d K_{ct}^E + (1 + \tau_t)^{1-\nu} \sum_j \vartheta^j K_{d_j}^E} \tag{35}$$

Evaluating (35) with data from IEA (2018) yields  $\hat{q} = q_t = 0.158$  for  $t = 2016$ , where  $\tau_t = \hat{\tau} = 72.82$  USD/tC is the BGP Pigovian tax derived below. Given the values for  $q_t, \nu$ , and  $Y_t/P_k$ , the parameter  $\vartheta$  of the firm-level production technologies is chosen as

$$\begin{aligned} \vartheta &= \frac{Y_t/P_k}{(1 - q_t)^{1-\nu} (K_{dt}^E)^\nu + q_t^{1-\nu} (K_{ct}^E)^\nu} \\ &= 69.843 \frac{\text{trillion USD}}{284.11 \text{ USD/tC}} \text{ for } t = 2016. \end{aligned}$$

The growth rate of average productivity,  $g$ , is identified with the average growth rate 0.0199 of energy consumption  $K_{dt}^E + K_{ct}^E$  between 2000 and 2016. We set the parameter  $\lambda = 0.0175$ , which is the exit rate of US energy firms between 1975 and 2004 (Acemoglu et al., 2016). Finally, we calibrate the cost to switch to clean technology,  $\Gamma$ , from the Technology IC Constraint in (32), with

<sup>27</sup> The clean technologies are hydroelectric, nuclear, renewables, geothermal, and biofuels. The unadjusted fraction of clean energy in 2016 (i.e., (35) evaluated at  $\tau_t = 0$ ) is 0.189. Another option would be to directly use  $q = 0.11$ , the fraction of U.S. firms classified as clean based on the energy technology patents they hold (Acemoglu et al., 2016, p. 76), which leads to very similar results. Note that this measure of clean technology excludes potential improvements in energy efficiency.

$\hat{\Psi} = \hat{\pi}_c - \hat{\pi}_d = \vartheta \hat{k}_c^d - \hat{k}_c - [\vartheta \hat{k}_d^d - \hat{k}_d] + \hat{k}_d \hat{\tau}$ , which also ensures that (A2') holds. This yields  $\hat{q}(\rho + \lambda)\Gamma = 132.34$  billion USD of fixed costs incurred per year.<sup>28</sup>

### 5.2. Damage function

We set the discount rate  $\rho = 0.0425$  as in Nordhaus (2017). We also follow Nordhaus and adopt the environmental damage function

$$\mathcal{D}(T_t) = a_2 T_t^2, \tag{36}$$

where  $T_t$  is the difference of the current global average surface temperature and the long-run historical average; the coefficient  $a_2 = 0.00236$  is calibrated such that 3°C of warming reduces GDP by 2.1%. Mean global surface temperature in Nordhaus (2017) is given by

$$T_t(S_t; \xi, \nu) = bF_t(S_t; \xi, \nu) + c_t \tag{37}$$

where  $F_t(S_t, \xi, \nu)$  is the total radiative forcing from the anthropogenic carbon stock  $S_t$ .<sup>29</sup> We introduce geengineering by parameterizing this term as

<sup>28</sup> This is in the ballpark of the total cost of investment in clean energy in 2016, \$287.5 billion; see BNEF (2018).

<sup>29</sup> Nordhaus's Nordhaus's (2017) specification includes a feedback loop between surface and deep ocean temperatures as well as exogenous radiative forcing (e.g., from non-CO<sub>2</sub> emissions and land use changes). Specifically,  $c_t = T_{t-1} - b(b_2 + b_3)T_{t-1} + bb_3 T_{t-1}^{\text{ocean}} + bF_t^{\text{exog}}$ , with mean deep ocean temperature evolving as  $T_{t-1}^{\text{ocean}} = (1 - b_4)T_{t-1}^{\text{ocean}} + b_4 T_{t-1}$  (Nordhaus, 2017, eq. 6–8). These terms do not have natural counterparts in BGP and tend to be less important than the direct effect of atmospheric carbon on temperature, so like Golosov et al. (2014) we ignore them and set  $c_t = 0$ .

$$F_t(S_t; \xi, v) = \frac{\theta(1-v)}{\ln 2} \ln \left( \frac{S_t - \xi}{\underline{S}} \right) \quad (38)$$

where  $\theta$  is the “equilibrium climate sensitivity,” calibrated to generate 3.1°C of global warming under a CO<sub>2</sub> doubling from pre-industrial levels and  $\underline{S} = 581$  GtC is the pre-industrial atmospheric stock (Golosov et al., 2014). Type I geoengineering,  $\xi$ , reduces the effective atmospheric carbon stock. Type II,  $v$ , directly affects global temperature through total radiative forcing (see Moreno-Cruz and Keith, 2013; Emmerling and Tavoni, 2018; and Heutel et al., 2018). Specifically, (39) uses the multiplicative form of Heutel et al. (2018).

We set the constant rate of carbon depreciation,  $\delta$ , as  $\delta = (1 - 0.5^{1/30}) \cdot (1 - 0.2) = 0.0183$ , in order to match the IPCC estimates that one-half of new carbon emissions persist in the atmosphere after 30 years and that one-fifth persists indefinitely (Golosov et al., 2014).<sup>30</sup> Finally, we set the BGP level of carbon stock  $\hat{S}$  to equal the total amount of carbon in the atmosphere in 2016. Table 2 summarizes these parameters.

With these assumptions, equations (36)–(38) imply the following form for the environmental damage function:

$$D(\hat{S}; \xi, v) = \mathcal{D} \circ T \circ F = a_2 \left( \frac{\theta(1-v)}{\ln 2} \ln \left( \frac{\hat{S} - \xi}{\underline{S}} \right) \right)^2.$$

Note that even though (36) is quadratic in temperature, our damage function is not quadratic in the carbon stock, nor does it have a constant elasticity. Given this damage function, Eq. (34) yields a BGP

Pigovian carbon tax of  $\hat{\tau} = \frac{A_t \theta D'(\hat{S}; \xi, v)}{\rho + \delta - g}$ . With our other parameter choices, this implies a carbon tax of  $\hat{\tau} = 72.82$  USD/tC<sup>31</sup>, as reported in Table 3. The calculation of this carbon tax completes the description of the BGP equilibrium in our calibrated economy.

### 5.3. Results

We now calculate versions of the key statistics discussed Propositions 14 and 15. The geoengineering parameters are  $\xi = v = 0$  in the baseline. We start with small changes in these geoengineering parameters, and then turn to larger variations.

We begin with type I geoengineering. Proposition 14 established that geoengineering advances of type I improve welfare provided that  $\frac{\lambda}{\rho + \lambda}(\hat{\pi}_c - \hat{\pi}_d) < \Lambda \hat{\tau}$ . In our calibrated economy, we have

$$\frac{\lambda}{\rho + \lambda}(\hat{\pi}_c - \hat{\pi}_d) = 0.820 < 2.517 = \Lambda \hat{\tau},$$

which implies that  $dW/d\xi > 0$ . That is, this type of geoengineering always improves welfare. The intuition for this result is that our estimates imply relatively small values for  $\hat{\pi}_c - \hat{\pi}_d$  and  $\lambda$ , and thus relatively small indirect costs due to lower investments in clean technology. The lower carbon emissions resulting from geoengineering of type I,  $\Lambda \hat{\tau}$ , comfortably exceed these indirect costs.

The implications of geoengineering of type II are very different. In this case, Proposition 15 shows that welfare effects depend on the curvature of the damage function. In particular, recall that

<sup>30</sup> In Appendix C we discuss how our results would be affected by alternative models of carbon depreciation and different specifications of the effects of geoengineering technologies on climate.

<sup>31</sup> The average E.U. ETS price between 2009 and 2018, 10.76 euro/tCO<sub>2</sub>, translates into 44.21 USD/tC, using 1.12 USD/euro and 0.273 tC/tCO<sub>2</sub> (3.66 tCO<sub>2</sub>/tC). Nordhaus's Nordhaus's (2017) calculation, which takes into account the full set of temperature and carbon dynamics and future technological improvements, is \$114.19/tC (\$31.23/tCO<sub>2</sub>), with a 10–90%-tile range of [\$26, \$282]. Golosov et al.'s (2014) benchmark estimate is \$60/tC.

**Table 2**  
Damage function parameters.

| Parameter | Value                          | Source  |
|-----------|--------------------------------|---|
| $\theta$  | climate sensitivity            | 3.1 Nordhaus, 2017                                      |
| $a_2$     | economic damage of temperature | 0.00236 Nordhaus, 2017                                  |
| $\hat{S}$ | carbon stock                   | 854.06 gigatonnes of carbon (GtC) in 2016 (NOAA, 2019)  |
| $\hat{E}$ | emissions                      | 8.463 gigatonnes of carbon (GtC) (IEA, 2018; EPA, 2018) |
| $\rho$    | discount rate                  | 0.0425 Nordhaus, 2017                                   |
| $\delta$  | carbon depreciation            | 0.0183 See text; based on IPCC (2007)                   |

**Table 3**  
Calibrated equilibrium quantities.

| Parameter       | Value                   | Source   |
|-----------------|-------------------------|--|
| $\hat{\tau}$    | carbon tax              | 72.82 2010 USD/tC                                    |
| $\hat{q}$       | fraction of clean firms | 0.158 IEA, 2018                                      |
| $\hat{q}\Gamma$ | switching fixed cost    | $\frac{132.34}{\rho + \delta}$ billion 2010 USD/year |

$dW/dv < 0$  if and only if  $\eta < \eta^{\text{II}}(\lambda)$ . In our calibrated economy, we have

$$\eta = \frac{1 - \ln \left( \frac{\hat{S} - \xi}{\underline{S}} \right)}{\ln \left( \frac{\hat{S} - \xi}{\underline{S}} \right)} = 1.596 < 2.324 = \eta^{\text{II}}(\lambda).$$

therefore, for this type of geoengineering technologies, we are comfortably in the region where  $dW/dv < 0$ , and so marginal advances in geoengineering technology of type II will reduce welfare. The intuition for this result is that given our damage function, environmental damages are relatively insensitive to the carbon stock around the BGP level  $\hat{S}$ , and thus restoring equilibrium investment incentives following geoengineering advances requires a very large increase in carbon emissions.

The next question concerns the size of welfare losses due to geoengineering advances. In Appendix C, we show that an improvement that neutralizes 10% of all climate damage ( $0.1 = 1 - (1 - dv)^2$ , or  $dv = 0.051$ ) reduces investment in conventional clean technology by 78.8%, increases carbon emissions by 14.8%, and reduces global welfare by 0.113% (83.7 billion USD/year). In contrast, if investment in clean technology were kept constant (for example, by committing to a future carbon tax schedule), the same technology would have improved welfare by 0.066%. This underscores the importance of considering new policy tools in response to potential geoengineering advances.<sup>32</sup>

## 6. Conclusions

Many scientists and policymakers are pinning their hopes on major geoengineering advances to stem damages from the rapidly-rising concentration of carbon in the atmosphere. Others, on the other hand, have worried that the prospect of geoengineering advances may jeopardize more conventional solutions to our environmental maladies, most notably the necessary increases in carbon taxes. Many of these concerns center around the possibility that the promise of geoengineering solutions may not materialize, or that geoengineering may have harmful side effects. In this paper, we have proposed an alternative perspective on the possible dark

<sup>32</sup> For slightly larger (smaller) improvements, welfare losses are larger (smaller). For much larger improvements, our Assumption 2' would be violated, all clean technology investment would be displaced and the economy would move out of the interior equilibrium. This does not by itself guarantee that welfare increases, but with even further geoengineering advances welfare would eventually rise. See Appendix C for the numerical results.

side of geoengineering. We have argued that geoengineering may damage the environment and welfare precisely when it is expected to materialize (or at least do so with a high probability). At the center of our argument is the possibility that the expectation of geoengineering makes future carbon taxes non-credible (because once geoengineering advances have been made, the damage from carbon emissions is reduced), which will discourage current investments in conventional cleaner technology, increase emissions and perhaps even reduce welfare.

We see this paper as a first step in the investigation of corrective policies without commitment and in the context of technological breakthroughs like geoengineering. In addition to considering richer menus of different technologies for reducing carbon emissions and combating climate change, future theoretical work could consider direct competition between firms using clean and dirty technologies (see Acemoglu et al., 2016 for one attempt in this direction). A major element missing from our analysis is the interaction between different countries and jurisdictions, which would require political economy considerations in addition to the issues of policy-making without commitment. Potentially more important is to provide empirical evidence on the two-way interactions between technology and policy—how current and future policy affects investments in clean technology, and how new technologies impact future policies.<sup>33</sup>

How empirically relevant is the possibility that new geoengineering breakthroughs affect clean technology investments through the carbon price? Geoengineering breakthroughs have not yet occurred, and in this sense our specific predictions are untestable. In other historical settings, however, exogenous technological breakthroughs appear to have dampened the incentives to adopt clean technology. Consider the celebrated sulfur dioxide (SO<sub>2</sub>) trading market introduced by the U.S. Clean Air Act of 1990. Following the act, unexpected declines in the transportation cost of low-sulfur Appalachian coal reduced sulfur emissions for many regulated firms and depressed sulfur permit prices (Schmalensee and Stavins, 2013, pp. 11–12) and depressed sulfur permit prices. As in our theory, these developments discouraged the adoption of conventional cleaner technologies, such as fuel scrubbers, to one-third of the government target (Burtraw, 1996, pp. 20–21).

Finally, while the paper's analysis has been positive, its results have normative implications. First, much of the conversation over climate engineering thus far has centered on scientific assessments of the probabilities that geoengineering will succeed or create adverse environmental risks. We suggest that the prospect of this research may itself affect economic equilibria by impacting investments in conventional clean technology, and in this sense our model provides a note of caution for geoengineering policymakers. Second, by stressing the costs of the policymakers' inability to commit to future carbon taxes, our results highlight that there are additional benefits from efficient subsidies for clean technology (which would remove the excessive reliance on carbon taxes to incentivize innovation) and from new commitment devices in the context of environmental policy.<sup>34</sup>

<sup>33</sup> For example, Lemoine (2017) provides evidence from the U.S. coal futures market that the anticipation of a possible federal carbon price meaningfully altered firms' ex-ante coal storage decisions.

<sup>34</sup> For example, enforceable carbon price floors, such as the auction reserve price in the California cap-and-trade mechanism under AB 32 (Borenstein, et al., 2019) may help to provide some medium-term commitment in the context of carbon markets. Laffont and Tirole (1996) also discuss a number of more complicated options contracts that can be used to mimic commitment in dynamic pollution permit markets.

### Data availability

Data will be made available on request.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Appendix A. Omitted Proofs

**Proof (Proof of Proposition 2 (Interior Equilibrium)).** (Interiority). From Proposition 1,  $\hat{\tau} = D'(\cdot; \hat{q})$ . Assumption 2, which imposes that  $(1 - \lambda)\Psi(\underline{\tau}) < \Gamma$  and  $(1 - \lambda)\Psi(\bar{\tau}) > \Gamma$ , then implies that neither  $q = 0$  nor  $q = 1$  are subgame perfect equilibria. (Existence and uniqueness). The private gain from switching to clean technology,  $\Psi(\tau)$ , is continuous in  $\tau$ , so the intermediate value theorem gives existence of a point  $\hat{\tau}$  such that  $(1 - \lambda)\Psi(\hat{\tau}) = \Gamma$ . Since  $\Psi(\tau)$  is increasing (from Lemma 2),  $\hat{\tau}$  is unique. Moreover, because  $D'' > 0$ , the Pigovian tax,

$$\tau = D'((1 - \delta)S_0 - q\Lambda(\tau) + k_d(\tau)),$$

is decreasing in  $q$ . Consequently  $\hat{q}$  is also unique.

**Proof (Proof of Proposition 3 (Second-best)).** The derivative of welfare with respect to  $q$  is

$$\frac{\partial W}{\partial q} = f_c(k_c) - k_c - \Gamma - (f_d(k_d) - k_d) - \frac{\partial S}{\partial q} D'(S)$$

which, using  $\partial S/\partial q = \Lambda$  and the fact that in the interior,  $f_c(k_c) - k_c - (f_d(k_d) - k_d) + \Lambda\tau = \frac{\Gamma}{1-\lambda} = \Gamma - \frac{\lambda}{1-\lambda}\Gamma$ , becomes

$$\frac{\partial W}{\partial q} = \frac{\lambda}{1-\lambda}\Gamma - \Lambda\tau + \Lambda D'(S). \tag{A1}$$

At  $\tau = D'(S)$ , (A1) is positive, implying that  $\tau^{SB} > D'(S)$  yields strictly higher welfare than  $\tau = D'(S)$  if and only if  $\lambda > 0$ .

**Proof (Proof of Proposition 4 (Type I Geoengineering)).** (Taxes, damages do not change). In an interior equilibrium,

$$\hat{\tau} = \frac{1}{\Lambda} \left[ \frac{\Gamma}{1-\lambda} - f_c(k_c) + k_c + f_d(k_d) - k_d \right]$$

and the RHS is invariant to a level shift in  $S_0$ , so  $d\hat{\tau} = 0$ . If  $\hat{\tau} = D'(S)$ , then  $dS = 0$ , which implies that  $-\Lambda d\hat{q} = d\zeta$ . (Welfare). We can calculate the total derivative of welfare,  $W = q(f_c(k_c) - k_c - \Gamma) + (1 - q)(f_d(k_d) - k_d) - D((1 - \delta)S_0 - \zeta + E)$ , with respect to  $\zeta$  as

$$\begin{aligned} \frac{dW}{d\zeta} &= \left[ q(f'_c(k_c) - 1) \frac{dk_c}{d\zeta} + (1 - q)(f'_d(k_d) - 1) \frac{dk_d}{d\zeta} \right] \frac{d\tau}{d\zeta} \\ &\quad + [f_c(k_c) - k_c - \Gamma - (f_d(k_d) - k_d)] \frac{dq}{d\zeta} + D' - \frac{dE}{d\zeta} D' \\ &= [f_c(k_c) - k_c - \Gamma - (f_d(k_d) - k_d)] \frac{dq}{d\zeta} \end{aligned} \tag{A2}$$

where the second line uses  $d\hat{\tau}/d\zeta = 0$  and  $dE/d\zeta = 1$ . (And  $dE/d\zeta = 1$  confirms  $dD/d\zeta = 0$ ). Using (5), (A2) simplifies to

$$\frac{dW}{d\zeta} = [\lambda(f_c(k_c) - k_c - (f_d(k_d) - k_d)) - (1 - \lambda)\Lambda\tau] \frac{d\hat{q}}{d\zeta}.$$

Using  $d\hat{q}/d\zeta = -1/\Lambda$ , and  $\pi_c - \pi_d - \Lambda\tau = f_c(k_c) - k_c - (f_d(k_d) - k_d)$ , we conclude that  $\frac{dW}{d\zeta} < 0 \Rightarrow \lambda(\pi_c - \pi_d) > \Lambda\tau$ .

**Proof** (Proof of Proposition 5 (Type II Geoengineering)). (I. Taxes). As in the proof of Proposition 4, only  $d\hat{\tau}/dv = 0$  sustains IC. (II. Environmental damage). Differentiating total environmental damage,  $(1 - v)D(S)$ , with respect to  $v$ , we obtain

$$\frac{dD(\cdot)}{dv} = -D(S) + (1 - v)D'(S) \frac{dS}{dv}. \tag{A3}$$

To calculate  $dS/dv$ , note that because  $d\hat{\tau}/dv = 0$ , we can differentiate

$$(1 - v)D'(S) = \hat{\tau}$$

with respect to  $v$  to obtain

$$-D'(S) + \frac{dS}{dv}(1 - v)D''(\cdot) = 0 \Rightarrow \frac{dS}{dv} = \frac{1}{1 - v} \frac{D'(S)}{D''(\cdot)}.$$

The total effect in (A3) then becomes

$$\begin{aligned} \frac{dD(\cdot)}{dv} &= -D(S) + \frac{1}{1 - v} \frac{D'(S)}{D''(\cdot)} (1 - v)D'(S) \\ &= -D(S) + \frac{1}{\eta} \cdot SD'(S), \end{aligned} \tag{A4}$$

where  $\eta \equiv SD''(S)/D'(S)$  is the relative curvature of  $D(\cdot)$  at  $S$ . By convexity ( $D'' \geq 0$ ), the quantity  $D(S)$  is bounded above by  $SD'(S)$ , so letting  $\eta^* \equiv SD'(S)/D(S) > 1$  we have

$$\eta < \eta^* \Rightarrow dD/dv > 0.$$

(III. Welfare). Aggregate welfare changes with  $v$  according to

$$\begin{aligned} \frac{dW}{dv} &= \frac{\partial}{\partial q} [q[f_c(k_c) - k_c] + (1 - q)[f_d(k_d) - k_d] - q\Gamma] \frac{dq}{dv} - \frac{dD(\cdot)}{dv} \\ &= [f_c(k_c) - k_c - [f_d(k_d) - k_d] - \Gamma] \frac{dq}{dv} - \frac{dD(\cdot)}{dv} \\ &= \left[ \frac{\lambda}{1 - \lambda} \Gamma - \Lambda\tau \right] \frac{dq}{dv} - \frac{dD(\cdot)}{dv}, \end{aligned}$$

where the last substitution follows from (5).

Differentiating the tax invariance condition  $D'(S) = (1 - v)D'(S)$  as before, noting that  $S$  can adjust only through  $q$ , and that  $\partial E/\partial q = -\Lambda$ , we obtain

$$\frac{dq}{dv} = \left[ \frac{\partial E}{\partial q} \right]^{-1} \frac{dS}{dv} = -\frac{1}{\Lambda(1 - v)} \frac{D'(S)}{D''(\cdot)}, \tag{A5}$$

or equivalently,

$$\frac{dq}{dv} = -\frac{1}{\Lambda(1 - v)} \frac{S}{\eta}.$$

As  $\tau = (1 - v)D'(S)$ , using this expression for  $dq/dv$  above gives

$$\left[ \frac{\lambda}{1 - \lambda} \Gamma - \Lambda\tau \right] \frac{dq}{dv} = -\frac{\lambda}{1 - \lambda} \Gamma \cdot \frac{1}{\Lambda(1 - v)} \frac{S}{\eta} + \frac{1}{\eta} SD'(S). \tag{A6}$$

From above, the total effect on environmental damage is

$$-\frac{dD(\cdot)}{dv} = D(S) - \frac{1}{\eta} SD'(S). \tag{A7}$$

The last term in each of the previous two expressions cancels when summed, and we obtain

$$\frac{dW}{dv} = D(S) - \frac{\lambda}{1 - \lambda} \Gamma \cdot \frac{1}{\Lambda(1 - v)} \frac{S}{\eta}. \tag{A8}$$

From (5), we have

$$\Gamma = (1 - \lambda)(\pi_c - \pi_d), \tag{A9}$$

and multiplying both sides by  $\Lambda(1 - v)D'(S)/D(S)$ , we obtain

$$\frac{dW}{dv} < 0 \Rightarrow \Lambda\tau - \lambda(\pi_c - \pi_d) \frac{1}{\eta} \frac{SD'(S)}{D(S)} < 0$$

and letting  $a \equiv SD'(S)/D(S) > 1$  (where the inequality follows from the strict convexity of  $D(\cdot)$ ), we conclude that

$$\eta < \eta^{\text{II}}(\lambda) \equiv a\lambda \left( \frac{\pi_c - \pi_d}{\Lambda\tau} \right)$$

characterizes the family of damage functions for which  $dW/dv < 0$ .

**Proof** (Proof of Claim in Footnote 14 (Quadratic Damages)). (Damages always increase). If  $D$  is quadratic, then the approximation

$$D(S) \approx SD' - \frac{1}{2} S^2 D''$$

is exact, so that  $D/SD' = 1 - \eta/2$ . By (A4),  $dD/dv > 0 \Rightarrow -D(S) + \eta^{-1}SD'(S) > 0$ , so

$$dD/dv > 0 \Rightarrow -1 + \eta/2 + 1/\eta > 0,$$

or  $dD/dv > 0 \Rightarrow \eta^2/2 - \eta + 1 > 0$ . But  $\eta^2/2 - \eta + 1$  is a polynomial with only imaginary roots, and is thus always positive. (Welfare). Under the assumption that  $D$  is quadratic,  $a = 1 - \eta/2$  and from (A4), we conclude that the condition

$$\eta(1 - \eta/2) - \lambda \left( \frac{\pi_c - \pi_d}{\Lambda\tau} \right) < 0$$

characterizes the region for which  $dW/dv < 0$ . The resulting polynomial has only imaginary roots when

$$\lambda \left( \frac{\pi_c - \pi_d}{\Lambda\tau} \right) > \frac{1}{2}$$

which is precisely the condition that  $\lambda \geq \lambda^*$ .

## Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.jpubeco.2022.104802>.

## References

Abrego, L., Perroni, C., 2002. Investment subsidies and time-consistent environmental policy. *Oxford Econ. Papers* 54, 617–635.

Acemoglu, D., 2007. Equilibrium bias of technology. *Econometrica* 75 (5), 1371–1409.

Acemoglu, D., Aghion, P., Bursztyn, L., Hémous, D., 2012. The environment and directed technical change. *Am. Econ. Rev.* 102 (1), 131–166.

Acemoglu, D., Akcigit, U., Hanley, D., Kerr, W., 2016. Transition to clean technology. *J. Polit. Econ.* 124 (1), 52–104.

Aghion, P., Dechezleprêtre, A., Hémous, D., Martin, R., Reenen, J.V., 2016. Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry. *J. Polit. Econ.*, 124 (1), 1–51.

Aghion, P., Griffith, R., 2005. *Competition and Growth: Reconciling Theory and Evidence*. MIT Press, Zeuthen Lectures.

Aghion, P., Howitt, P., 1992. A model of growth through creative destruction. *Econometrica* 60 (2), 323–351.

Akcigit, U., Kerr, W., 2018. Growth through heterogeneous innovation. *J. Polit. Econ.* 126 (4), 1374–1443.

Barrett, S., 2008. The incredible economics of geoengineering. *Environ. Resour. Econ.* 39 (1), 45–54.

Battaglini, M., Harstad, B., 2016. Participation and duration of environmental agreements. *J. Polit. Econ.* 124 (1), 160–204.

Bellman, R., 1957. *Dynamic Programming*. Princeton University Press.

BNEF (2018). *State of Clean Energy Investment*. Tech. rep., Bloomberg New Energy Finance.

Bollinger, B., Gillingham, K., 2014. Learning-by-doing in solar photovoltaic installations. Working Paper.

Borenstein, S., Bushnell, J., Wolak, F.A., Zaragoza-Watkins, M., 2019. Expecting the unexpected: Emissions uncertainty and environmental market design. *Am. Econ. Rev.* 109 (11), 3953–3977.

Bovenberg, A.L., Smulders, S., 1995. Environmental quality and pollution-augmenting technological change in a two-sector endogenous growth model. *J. Public Econ.* 57 (3), 369–391.

Brunner, S., Flachsland, C., Marschinski, R., 2012. Credible commitment in carbon policy. *Climate Policy* 12 (2), 255–271.

Burtraw, D., 1996. Cost savings sans allowance trades? Evaluating the SO<sub>2</sub> emission trading program to date. Resources for the Future Discussion Paper, 95–30-REV.

Calvo, G., 1978. On the time consistency of optimal policy in a monetary economy. *Econometrica* 46, 1411–1428.

- Cook, J., Oreskes, N., Doran, P., Anderegg, W., Verheggen, B., Maibach, E., Carlton, J.S., Lewandowsky, S., Skuce, A., Green, S., Nuccitelli, D., Jacobs, P., Richardson, M., Winkler, B., Painting, R., Rice, K., 2016. Consensus on consensus: A synthesis of consensus estimates on human-caused global warming. *Environ. Res. Lett.* 11 (4), 1–7.
- Dell, M., Jones, B., Olken, B., 2012. Temperature shocks and economic growth: Evidence from the last half century. *Am. Econ. J.: Macroecon.* 4 (3), 66–95.
- Desmet, K., Rossi-Hansberg, E., 2015. On the spatial economic impact of global warming. *J. Urban Econ.* 88, 16–37.
- Emmerling, J., Tavoni, M., 2018. Climate engineering and abatement: A 'flat' relationship under uncertainty. *Environ. Resour. Econ.* 69 (2), 395–415.
- EPA (2018). Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990–2016. Annex 2 (Methodology for estimating CO<sub>2</sub> emissions from fossil fuel combustion), Table A-41. Tech. Rep. 430-R-18-003, U.S. Environmental Protection Agency, Washington, D.C.
- Flannery, T., 2015. *Atmosphere of Hope: Solutions to the climate crisis*. Penguin.
- Gertler, C., O'Gorman, P., 2019. Changing available energy for extratropical cyclones and associated convection in northern hemisphere summer. *Proc. Natl. Acad. Sci.*, 116 (10), 4105–4110.
- Golosov, M., Hassler, J., Krusell, P., Tsyvinski, A., 2014. Optimal taxes on fossil fuel in general equilibrium. *Econometrica* 82 (1), 41–88.
- Goulder, L., Mathai, K., 2000. Optimal CO<sub>2</sub> abatement in the presence of induced technological change. *J. Environ. Econ. Manage.* 39 (1), 1–38.
- Greenstone, M., Kopits, E., Wolverton, A., 2013. Developing a social cost of carbon for US regulatory analysis: A methodology and interpretation. *Rev. Environ. Econ. Policy* 7 (1), 23–46.
- Grimaud, A., Lafforgue, G., Magné, B., 2011. Climate change mitigation options and directed technical change: A decentralized equilibrium analysis. *Resour. Energy Econ.* 33 (4), 938–962.
- Grossman, G., Helpman, E., 1991. Quality ladders in the theory of growth. *Rev. Econ. Stud.* 58 (1), 43–61.
- Hartley, P., Medlock, K., Temzelides, T., Zhang, X., 2016. Energy sector innovation and growth: An optimal energy crisis. *Energy J.* 37 (1), 233–258.
- Hassler, J., Krusell, P., Olovsson, C., 2012. Energy-saving technical change. NBER Working Paper 18456.
- Hassler, J., Krusell, P., Smith, T., 2016. Environmental macroeconomics. In: Taylor, J., Uhlig, H. (Eds.), *Handbook of Macroeconomics*, vol. 2. Elsevier, pp. 1893–2008.
- Helm, D., Hepburn, C., Mash, R., 2003. Credible carbon policy. *Oxford Rev. Econ. Policy* 19 (3), 438–450.
- Heutel, G., Moreno-Cruz, J., Shayegh, S., 2018. Solar geoengineering, uncertainty, and the price of carbon. *J. Environ. Econ. Manage.* 87, 24–41.
- Hsiang, S., Burke, M., Miguel, E., 2013. Quantifying the influence of climate on human conflict. *Science* 341 (6151).
- IEA (2018). *World Energy Balances 2018*. Tech. rep., International Energy Agency, Paris, France.
- IPCC (2014). *Climate Change 2014: Mitigation of Climate Change*. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, United Kingdom: Cambridge University Press.
- Jensen, S., Mohlin, K., Pittel, K., Sterner, T., 2015. An introduction to the Green Paradox: The unintended consequences of climate policies. *Rev. Environ. Econ. Policy* 9 (2), 246–265.
- Jevons, W., 1866. *The Coal Question: An Inquiry Concerning the Progress of the Nation, and the Probable Exhaustion of our Coal-mines*. Macmillan & Co., London.
- Jones, C., 1995. R&D-based models of economic growth. *J. Polit. Econ.* 103 (4), 759–784.
- Karp, L., Newbery, D., 1993. Intertemporal consistency issues in depletable resources. In: Kneese, A., Sweeney, J.L. (Eds.), *Handbook of Natural Resource and Energy Economics*, vol. 3. Elsevier, pp. 881–931.
- Keith, D., 2013. *A Case for Climate Engineering*. MIT Press.
- Klette, T.J., Kortum, S., 2004. Innovating firms and aggregate innovation. *J. Polit. Econ.* 112 (5), 986–1018.
- Kopp, R., Kemp, A., Bittermann, K., Horton, B., Donnelly, J., Gehrels, W.R., Hay, C., Mitrovica, J., Morrow, E., Rahmstorf, S., 2016. Temperature-driven global sea-level variability in the Common Era. *Proc. Natl. Acad. Sci.*, 113 (11), 1434–1441.
- Kydland, F., Prescott, E., 1977. Rules rather than discretion: The inconsistency of optimal plans. *J. Polit. Econ.* 85 (3), 473–492.
- Laffont, J.-J., Tirole, J., 1996. Pollution permits and compliance strategies. *J. Public Econ.* 62, 85–125.
- Lawrence, M., Crutzen, P., 2017. Was breaking the taboo on research on climate engineering via albedo modification a moral hazard, or a moral imperative? *Earth's Future* 5 (2), 136–143.
- Lemoine, D., 2017. Green expectations: Current effects of anticipated carbon pricing. *Rev. Econ. Stat.* 99 (3), 499–513.
- Lemoine, D., Traeger, C., 2014. Watch your step: Optimal policy in a tipping climate. *Am. Econ. J.: Econ. Policy* 6 (1), 137–166.
- Lenton, T., Vaughan, N., 2009. The radiative forcing potential of different climate geoengineering options. *Atmos. Chem. Phys.* 9, 5539–5561.
- Mann, M., Rahmstorf, S., Kornhuber, K., Steinman, B., Miller, S., Coumou, D., 2017. Influence of anthropogenic climate change on planetary wave resonance and extreme weather events. *Scientific Reports*.
- Marsiliani, L., Renström, T., 2000. Time inconsistency in environmental policy: Earmarking as a commitment solution. *Econ. J.* 110, C123–C138.
- Meier, F., Traeger, C., 2020. Solar geoengineering in a regional analytic climate economy. In: *Economic Association Annual Meeting: Association of Environmental and Resource Economists Paper Session*. American, San Diego, CA.
- Moreno-Cruz, J., 2015. Mitigation and the geoengineering threat. *Resource and Energy Economics* 41 (248–263).
- Moreno-Cruz, J., Keith, D., 2013. Climate policy under uncertainty: A case for solar geoengineering. *Clim. Change* 121 (3), 431–444.
- Morrow, D., 2014. Ethical aspects of the mitigation obstruction argument against climate engineering research. *Philos. Trans. Royal Soc. A* 372 (2031).
- Morton, O., 2015. *The Planet Remade*. Princeton University Press.
- National Research Council (2015). *Climate Intervention: Reflecting Sunlight to Cool Earth*. Washington, DC: The National Academies Press.
- Newell, R., Jaffe, A., Stavins, R., 1999. The induced innovation hypothesis and energy-saving technological change. *Quart. J. Econ.* 114 (3), 941–975.
- NOAA (2019). Trends in atmospheric carbon dioxide, National Oceanic & Atmospheric Administration, Earth System Research Laboratory, [https://www.esrl.noaa.gov/gmd/ccgg/trends/gl\\_data.html](https://www.esrl.noaa.gov/gmd/ccgg/trends/gl_data.html).
- Nordhaus, W., 1991. To slow or not to slow: The economics of the greenhouse effect. *Econ. J.* 101 (407), 920–937.
- Nordhaus, W., 2008. *A Question of Balance: Weighing the Options on Global Warming Policies*. Yale University Press.
- Nordhaus, W., 2017. Revisiting the social cost of carbon. *Proceedings of the National Academy of Sciences*, 114 (7), 1518–1523.
- Nordhaus, W., Moffat, A., 2017. A survey of global impacts of climate change: Replication, survey methods, and a statistical analysis. NBER Working Paper, (23646).
- Peltzman, S., 1975. The effects of automobile safety regulation. *J. Polit. Econ.* 83 (4), 677–726.
- Phaneuf, D., Requate, T., 2017. *A Course in Environmental Economics*. Cambridge University Press.
- Phelan, C., Stacchetti, E., 2001. Sequential equilibria in a Ramsey tax model. *Econometrica* 69 (6).
- Popp, D., 2002. Induced innovation and energy prices. *Am. Econ. Rev.* 92 (1), 160–180.
- Popp, D., 2019. Environmental policy and innovation: A decade of research. NBER Working Paper, (25631).
- Quaas, M., Quaas, J., Rickels, W., Boucher, O., 2017. Are there reasons against open-ended research into solar radiation management? A model of intergenerational decision-making under uncertainty. *J. Environ. Econ. Manage.* 84, 1–17.
- Requate, T., Unold, W., 2001. On the incentives created by policy instruments to adopt advanced abatement technology if firms are asymmetric. *J. Inst. Theor. Econ.* 157 (4), 536–554.
- Requate, T., Unold, W., 2003. Environmental policy incentives to adopt advanced abatement technology: Will the true ranking please stand up? *Eur. Econ. Rev.* 47, 125–146.
- Schmalensee, R., Stavins, R., 2013. The SO<sub>2</sub> allowance trading system: The ironic history of a grand policy experiment. *J. Econ. Perspect.* 27 (1), 103–122.
- Sinn, H.-W., 2012. *The Green Paradox*. M.I.T. Press.
- Spence, A.M., Starrett, D., 1975. Most rapid approach paths in accumulation problems. *Int. Econ. Rev.* 16 (2), 388–403.
- Weitzman, M., 2015. A voting architecture for the governance of free-driver externalities, with an application to geoengineering. *Scand. J. Econ.* 117 (4), 1049–1068.
- World Bank (2019). *World Bank Development Indicators: GDP (current USD)*. <https://data.worldbank.org/indicator/NY.GDP.MKTP.KD>.
- Yan, Q., Wei, T., Korty, R., Kossin, J., Zhang, Z., Wang, H., 2016. Enhanced intensity of global tropical cyclones during the mid-Pliocene warm period. *Proc. Natl. Acad. Sci.* 113 (46), 12963–12967.