Mirage on the Horizon: Geoengineering and Carbon Taxation Without Commitment

Daron Acemoglu  
MIT  
Will Rafey  
MIT  
August 10, 2017

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 are separate from energy producers’ investments and reduce the negative environmental effects of the existing stock of carbon or newly-emitted carbon, decrease future carbon taxes, thus discouraging investments in conventional clean technology. We characterize the conditions under which these advances diminish—rather than improve—environmental quality and welfare. Crucially, the negative consequences of geoengineering in our model arise not because the promise of these new technologies may not be realized, but precisely because it is expected to be realized.

JEL Classification: Q01, Q4, Q54, Q55, Q58, O30, O31, O33, C65  
Keywords: carbon tax, geoengineering, environment, innovation, time inconsistency

Work in Progress. Comments Welcome.
1 Introduction

There is increasing recognition that a transition to cleaner technology has to be the bedrock of future reductions in carbon emissions. 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 subsidies can contribute to the advances in these technologies, some experts and policy-makers instead pin their hopes on geoengineering breakthroughs, such as large-scale carbon sequestration, ocean fertilization, and solar radiation management. 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 against climate change. As the Royal Society (2009, p. 45) concludes in its report on geoengineering:

There is often an assumption that geoengineering represents a moral hazard, and could undermine popular and political support for mitigation or adaptation. Although this prospect should be taken seriously, there is yet little empirical evidence on whether the prospect of climate intervention galvanises or undermines efforts to reduce emissions.

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 changes 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 by 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.

---


2We 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 such as governments). Existing research in economics focuses almost entirely focuses on these conventional technologies (Weitzman (2015), is a notable exception we discuss below). Newell et al. (1999), Popp (2002) and Hassler et al. (2012) provide evidence that energy-saving technologies are responsive to the price of energy, while Aghion et al. (2016) find a large impact of carbon taxes on the direction of both current and future innovation in the automobile industry. Acemoglu et al. (2012) and Acemoglu et al. (2016) provide theoretical frameworks for analyzing optimal policies (under commitment) to encourage the development of cleaner technologies.

3For example, Keith (2013), Flannery (2015), and Morton (2015).

4This type of externality arises naturally in almost all models of endogenous technology, including the quality
Second, we assume that policy is chosen by a social planner without the ability to commit to future policy promises. 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. In environmental economics, though it has not been studied extensively, its importance is well recognized. For example, in reviewing several economic frameworks for climate policy evaluation, William Nordhaus (2007, p. 693) observes that

[N]one of these approaches touch on the structure of actual intertemporal decision-making, in which this generation cannot decide for or tie the hands of future generations. Instead, each generation is in the position of one member of a relay team, handing off the baton of capital to the next generation, and hoping that future generations behave sensibly and avoid catastrophic choices by dropping or destroying the baton…. but this is largely uncharted territory in economic growth theory.

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 Pigovian benchmark (where the carbon taxes set equal to the marginal damage from one more unit of carbon). However, when the (social) planner would also like to encourage 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 with-

---

5Throughout, 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). For example, in the context of fiscal policy, time-inconsistency naturally arises because of interactions between the government and the agents as noted by, among others, Chari and Kehoe (1990), Benhabib and Rustichini (1997), Phelan and Stacchetti (2001), and Klein et al. (2008).

6Several noteworthy about-turns in environmental policy in the OECD illustrate the relevance of time-inconsistency concerns in this area. Major political reversals include Canada’s withdrawal from the Kyoto Accord in 2011, Australia’s repeal of its carbon tax in 2014, and the rollback of the U.S. EPA’s Clean Power Plan in 2017. Examples of regulators revising energy pricing schemes in response to technological change include the Spanish solar-feed-in-tariff, where the government reneged on solar subsidies after the production costs fell by more than regulators anticipated, and the U.K.’s moves in late 2011 to cut solar subsidies under the 2008 Energy Act by 55%.

7As emphasized in Acemoglu et al. (2012), the presence of clean technology innovations or investments necessitates carbon taxes to be augmented by direct subsidies to the development or adoption of these technologies. In our model, in contrast, we are assuming that the planner uses a carbon tax to encourage these investments. This
out commitment to future policies, firms will anticipate that any promised taxes above the Pigovian level will be revised, and underinvest in clean technology. The extent of this underinvestment is regulated (and limited) by the responsiveness of future Pigovian taxes to higher stock of carbons.

It is in this setting that we introduce the prospect of geoengineering. For clarity, we distinguish between three different types of geoengineering technologies (fully recognizing that real-world technologies are often a mixture of these three types): type I technologies, carbon removal, which correspond to a downward shift of the environmental damage function (or, equivalently, reduce the effective stock of carbon that enters the damage function by a constant amount); type II, climate adaptation, which reduce marginal damages from carbon in the atmosphere; and type III, carbon capture, which reduce the amount of carbon that goes into the atmosphere from a given quantity of emissions. Examples of type I technologies include all forms of large-scale carbon dioxide removal, like mass afforestation, biochar, ambient air capture, and ocean fertilization. Examples of type II technologies include solar radiation management, such as albedo enhancement, space reflectors, or stratospheric aerosols. Type III technologies correspond to large-scale on-site carbon sequestration or regulations that change the carbon-intensity of electricity generation. Though each type of geoengineering technology has somewhat different implications, they all work in a similar manner. Without an equilibrium response, all three types of geoengineering technology reduce future damages, and thus future Pigovian carbon taxes. But since future Pigovian carbon taxes determine today’s investment in conventional clean technologies, we show that the advent of geoengineering technologies increases underinvestment in these socially valuable technologies.

More specifically, we demonstrate that type I geoengineering technologies reduce investment in conventional clean technology exactly so much as to leave the total amount of damages the same as in the world without the geoengineering technology. The intuition for this result is simple. To restore incentives for the adoption of conventional clean technologies—to satisfy the “technology IC” constraint—the Pigovian tax (marginal damage of carbon) needs to be at a certain level. With the marginal value of damages remaining the same, the equilibrium value of overall damages cannot change. Even though overall damages remain con-

---

8See, for example, Lenton and Vaughan (2009) for a scientific assessment of these technologies for global climate dynamics and Socolow et al. (2011) and Lackner et al. (2012) for cost estimates for direct carbon removal.

9See, for example, the analysis in Rasch et al. (2008) and the survey in National Research Council (2015).

10See, for example, reviews of existing carbon capture and sequestration technologies in the United States (U.S. Department of Energy, 2014) and Europe (McCulloch et al., 2016).

11The reasoning developed here applies beyond geoengineering, for example, in the context of the “pollution haven hypothesis” in the international trade literature (e.g., Copeland and Taylor, 1994), or the problem of “carbon
stant, 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 in fact, type III geoengineering technologies always lead to greater damages. The logic in these cases is similar, except that now with the marginal value of damages changing, the overall amount of damages must shift to restore the Pigovian carbon tax to the value consistent with the adoption of conventional clean technologies. With these types of geoengineering technologies, negative welfare effects are even more likely than type I technologies.

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. Instead, our framework identifies potential inefficiencies from geoengineering 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 geoengineering 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 setting is that there exists a unique level of the carbon tax that satisfies the technology IC—making 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

---

12 The results would be equivalent if the clean technology investment of the incumbent reduces the cost of clean technology investment for the entrant.
satisfy the technology IC. In this light, the implications of various different types of geoengineering technologies become straightforward to see. A type I geoengineering 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 exactly to offset the benefits from geoengineering. The implications of the other types of geoengineering technologies can be analyzed and understood similarly.

After expositing our main ideas in a transparent manner in this static model, 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 the results discussed in the previous two paragraphs do not depend on assuming a static setting.

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 equilibrium with a time-inconsistent planner in this setting. 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).

We also discuss how relaxing the assumption that firms are ex-ante identical does not change our qualitative results. The technology IC still makes the fraction of firms that transition to clean technology a function of the carbon tax, and provided that this relationship is sufficiently responsive to the level of the carbon tax, the same reasoning applies.

Our work relates to the small but 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, 1996), Goulder and Mathai (2000), Goulder and Schneider (1999), Grimaud et al. (2011), Hartley et al. (2016), Hassler et al. (2012), Popp (2002, 2004), and van der Zwaan et al. (2002) also discuss endogenous technology in the context of environmental policy and climate change.

Our work is also related to a very small literature on environmental policy without commitment (surveyed in Karp and Newbery (1993)). Laffont and Tirole (1996a,b) study pollution permit markets and innovation in a two-period setting with asymmetric information,
where the regulator cannot commit and the usual hold-up problem arises. Harstad (2012) and Harstad and Battaglini (2016) study incentives to invest in clean technologies in the presence of multiple regulators without commitment, and show there will be underinvestment because additional investments reduce each regulator’s bargaining position against the others. Also notable is Harstad (2016) who analyzes environmental policy under hyperbolic discounting. We are not aware of any papers that model or note how with time-inconsistency, improvements in technologies can lead to declines in welfare.

Finally, there are only a few papers in the economics literature on geoengineering. Most relevant to our work are Barrett (2008) and Weitzman (2015), who focus on the international political economy dimensions of geoengineering technologies, and emphasize the risks that unilateral adoption might pose if local climate modifications create externalities for other states.

The remainder of the paper is organized as follows. Section 2 introduces our model and characterizes the equilibrium. Section 3 extends our results to an infinite-horizon setting in continuous time. Section 4 concludes. Appendix A contains the proofs omitted from the text, while online Appendix B demonstrates the robustness of our main results to various variations.

2 Baseline Model

In this section, we introduce our baseline static model. In the next section, we consider a dynamic model which provides a clearer microfoundation for some of the assumptions utilized in this section 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 $[0, 1]$. For simplicity, we take the output of 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 final good as input, it can produce

$$f_d(k_i)$$

units of output. 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., $\lim_{k \to 0} f'_d(k) = \infty$ and $\lim_{k \to \infty} f'_d(k) = 0$). Since all activities are perfectly substitutable,
firms will act competitively and we choose the price of energy output as numéraire (normalizing it to 1).\textsuperscript{13}

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 with production function

\[
f_c(k_i)
\]

units of output (where \(f_c\) is also twice continuously differentiable, increasing and concave, and satisfies \(\lim_{k \to 0} f_c'(k) = \infty\) and \(\lim_{k \to \infty} f_c'(k) = 0\)), but only \(\gamma k_i\) units of carbon (where \(\gamma < 1\)). We think of clean firms as switching to a technology that produces energy output from cleaner sources such as wind or solar energy, or upgrading their existing plant’s efficiency to reduce emissions per unit of energy production.

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 the cleaner technology, the entrant inherits it.\textsuperscript{14} We further simplify the notation by assuming that the entrant has access to exactly 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).\textsuperscript{15}

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

\[
E = q \gamma k_c + (1 - q) k_d,
\]

where \(k_c\) is the equilibrium production level of clean technology and \(k_d\) is the equilibrium production 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

\[
S = (1 - \delta) S_0 + (1 - \zeta) E,
\]

where \(S_0 \geq 0\) is the initial level of carbon, \(\delta \in [0, 1]\) is the “depreciation” of this stock of carbon (for example, by being absorbed by oceans and forest cover), and \(\zeta \in [0, 1]\) captures one of

\textsuperscript{13}Though different activities are perfect substitutes, the Inada conditions on the production functions ensure that all of them will be produced in equilibrium.

\textsuperscript{14}Nothing in our qualitative results below change if we instead assumed 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.

\textsuperscript{15}This structure of entrants replacing incumbents will be further micro-founded in the context of the dynamic model in the next section.
the dimensions of geoengineering advances discussed below; to start with, we set $\zeta = 0$. This formulation is chosen to create continuity with the dynamic model in the next section. The damages from carbon in the atmosphere are denoted by

$$D(S; \zeta, \nu),$$

where $D$ is an increasing, twice continuously differentiable and strictly convex function, and for now, damages are taken to be additive, and the parameters $\zeta$ and $\nu$ are also introduced to model the effects of other types of geoengineering advances on environmental damages. For now, we suppress these parameters, writing environmental damages simply as $D(S)$.\(^{16}\)

### 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

$$\pi_d(\tau) = \max_{k \geq 0} f_d(k) - (1 + \tau)k$$

and

$$\pi_c(\tau) = \max_{k \geq 0} f_c(k) - (1 + \gamma\tau)k$$

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 choose higher levels of input usage.\(^{17}\) 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

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

\(^{16}\)In Appendix B, we show that our qualitative results are unaffected if damages affect productivity as in Nordhaus (1991, 2008); Golosov et al. (2014), or affect utility in a non-additively separable manner.

\(^{17}\)This will always be the case if $f_c = f_d$, but not necessarily otherwise.
2.3 Clean Technology Decisions

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

\[
\Psi(\tau) = \pi_c(\tau) - \pi_d(\tau)
\]

\[
= [f_c(k_c(\tau)) - (1 + \gamma \tau) k_c(\tau)] - [f_d(k_d(\tau)) - (1 + \tau) k_d(\tau)]
\]

\[
= [f_c(k_c(\tau)) - k_c(\tau)] - [f_d(k_d(\tau)) - k_d(\tau)] + \tau \Lambda(\tau),
\]

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 (which has 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 geo-engineering technologies and the equilibrium tax rate on this Technology IC 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)**

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

(Technology IC)

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

\[
\Psi(\tau) = \frac{\Gamma}{1 - \lambda},
\]

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

The following lemma shows that higher taxes make (Technology IC Constraint) more likely to hold (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 the fact that both clean and dirty firms are choosing profit-maximizing levels of input usage. That this effect is positive is a consequence of Assumption 1. This result greatly simplifies our analysis by ensuring that the \( \Psi \) function 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 input, welfare can be written as

\[
W = (1-q)[f_d(k_d) - k_d] + q[f_c(k_c) - k_c] - q\Gamma - D(S)
\]

(6)

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

There are three important observations. 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 of the current producer being replaced by new entrants, \( \lambda \), 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.

In what follows, we assume that the planner has access to a single instrument—a carbon tax, \( \tau \).

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 equation (1) as \(E(\tau^*, q^*)\) and the level of stock of carbon in the atmosphere from equation (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 to the case in which the planner can commit to carbon taxes as we will see below.

More formally, we have:

Proposition 1 In equilibrium, the carbon tax is given as

\[
\tau^* = D'(S(\tau^*, q^*)). \tag{7}
\]

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 some others do not.\(^{18}\)

\(^{18}\)This equilibrium can also be labeled “asymmetric” because some ex ante identical firms switch to clean technology, while others do not. We show in Appendix B3, however, that this asymmetry is not the important feature,
Assumption 2 (Conditions for Interior Equilibrium) We have
\[
\frac{\Gamma}{1-\lambda} \in (\Psi(\tau), \Psi(\bar{\tau}))
\]
where \(\bar{\tau} = D'(1-\delta)S_0 + \gamma k_c(\tau)\) and \(\tau = D'(1-\delta)S_0 + k_d(\tau)\).

When this assumption does not hold, there exists a unique equilibrium in which all firms switch to the clean technology or no firms switch to the clean technology, and in neither case do we have interesting comparative statics of investment in clean technology. Thus Assumption 2 restricts the analysis to the interesting part of the parameter space, where the equilibrium is interior. This is also empirically reasonable—since in practice only a limited fraction of energy producers have made the transition to clean technology.

Intuitively, as explained above, for an interior equilibrium, we need the condition (5) to hold. This will only be the case if, 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 \(\tau\) 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).

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 is the carbon tax, the more likely they are to make this transition. However, the carbon tax is determined by the planner after the technology decisions are made—and herein lies the time-inconsistency problem in our model. 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, which is exactly \(\hat{\tau}\) given in Proposition 2. Given the convexity of damages in (3), the greater is the stock of carbon, the and similar results obtain when firms are ex ante heterogeneous in terms of their cost of switching to clean technology. The important feature, thus, is that in our equilibrium, the transition to clean technology is not complete.
higher is the carbon tax rate that the planner will set. Thus for \( \hat{\tau} \) to emerge as the planner’s choice, the stock of carbon in the atmosphere needs to take a specific value, \( S(\hat{\tau}, \hat{\eta}) \). But then from the emission decisions of clean and dirty firms, this translates into a specific fraction of firms that switch to clean technology, \( \hat{\eta} \) (recall that \( S(\tau, q) \) is increasing in \( q \)). If more firms than this switch to clean technology, there will be less carbon in the atmosphere than \( S(\hat{\tau}, \hat{\eta}) \), and consequently, the planner will choose a lower carbon tax than \( \hat{\tau} \), violating (5). Likewise, if fewer firms than \( \hat{\eta} \) made the switch, the carbon tax rate would be higher than \( \hat{\tau} \), once again violating (5). This reasoning also provides the intuition for why the equilibrium is unique. The more firms switch to clean technology, the lower is the carbon tax and the weaker are incentives for such a switch, ruling out complementarities and multiplicity.

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 liked to do if she could commit to the carbon taxes ex ante, as we show next.

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. There are then two cases to consider. First, following this tax rate, the equilibrium is still interior—a fraction \( q^{SB} \in (0, 1) \) of firms switch to the clean technology. Second, following this tax rate, we have \( q^{SB} = 1 \) (the third case in which \( q^{SB} = 0 \) is ruled out by Assumption 2). The next proposition shows that, regardless of which case we are in, 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 \( q^{SB} > \hat{\eta} \), 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 $\tau^{SB} > D'(S^{SB})$. The reason for this is that when $\lambda > 0$, there is underinvestment in clean technology—because firms do not take into account the benefit they create on others who will build on their clean technology investments. As a result, in the second-best where she cannot directly control technology investments, the planner would like to encourage greater investment in clean technology by setting higher carbon taxes than the Pigovian benchmark, and this will 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., $q^{SB} > \hat{q}$.

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, setting the right price of carbon—i.e., the Pigovian tax—is sufficient to induce the right level of technology investment, and thus the planner has no reason to resort to a non-Pigovian carbon 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. Thus any further increase in the stock of carbon has a first-order negative impact on welfare.

### 2.10 The Effects of Geoengineering

We next study the implications of various geoengineering technologies on equilibrium carbon taxes, investment in clean technologies, environmental damages and welfare. By geoengineering technologies, we refer to technological advances that either reduce the damages from a given stock of carbon or reduce the impact of emissions on the stock of carbon. Crucially, these are different from those of firms’ investments in clean technology, and are operated or introduced by other entities (which could be the “government” or other firms in the economy).

We will distinguish between three different types of geoengineering technologies, which

---

19Note, 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).
we first enumerate and motivate. We then analyze their implications separately. Actual geoengineering breakthroughs are likely to combine features from these three types, but it is useful for our purposes to exposit their implications separately.

To incorporate all three types of geoengineering technologies, let us write the law of motion of the stock of carbon in the atmosphere given in (2) with $\zeta$ not necessarily equal to 0, substitute this into (3), and make the role of the different parameters explicit as follows

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

where $\tilde{D}$ is a base damage function, and changes in the parameters $\xi \geq 0$, $\upsilon \in [0, 1)$ and $\zeta \in [0, 1)$ lead to shifts of 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 downward shift of the environmental damage function as shown in Figure 2.10. 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 or geoengineering technology of type II, corresponds to an increase in $\upsilon$, a proportional downward shift of the environmental damage function as also illustrated in Figure 2.10. We interpret this class as representing a range of technologies related to solar radiation management, which aim to slow down temperature rise at a given emissions level. The most well-known example is the injection of sulfur dioxide into the stratosphere, as 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.

Finally, our geoengineering technology of type III (or carbon capture) corresponds to an increase in $\zeta$, reducing the rate of emissions from the use of a given amount of inputs. Technologies that involve various carbon capture and sequestration strategies that operate externally to firms fall within this category, and so do any improvements in the carbon emission rate of various energy sources such as coal, shale gas or diesel fuel.

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

**Proposition 4 (Implications of 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
• $d\hat{\tau} = 0$ (there is no effect on the equilibrium carbon tax).
• $d\hat{q} = -d\xi / \Lambda(\hat{\tau})$ (investment in clean technology declines).
• $dE = d\xi > 0$ (emissions increase, through lower $\hat{q}$).
• $dD = 0$ (environmental damages remain constant).
• $dW < 0$ 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 we introduce the geoengineering technology, the technology IC, (5), still pins down the carbon tax rate at $\hat{\tau}$—because with a small change in $\xi$, Assumption 2 will continue to hold and the equilibrium has to be interior. But $\hat{\tau}$ is an equilibrium outcome only if the planner still prefers to set it as the carbon tax. This implies that the marginal environmental damage needs to remain constant. Since geoengineering shifts the damage function downward by $d\xi$, and $D$ is strictly convex, the total stock of carbon must increase by $d\xi$. In our setup, this can only happen if fewer firms make the switch to clean technology (since the carbon tax is constant). At the end, emissions and the total stock of carbon increase just enough to keep the total environmental damages constant, which then also ensures that marginal damages remain constant. These results thus show that the geoengineering technology backfires and some of its beneficial effects are undone.

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 in general $\lambda > 0$, 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. More specifically, 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 more than 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.

We should also note that, if instead of a small increase in $\xi$, we consider a large increase, we may violate Assumption 2, and if so, the planner may wish to deviate from (5), forgoing any investment in clean technology. In this case, damages may again decline and welfare may increase, though this is again not guaranteed. A similar caveat applies to geoengineering technologies of type II and III discussed next.

**Remark 1 (Carbon leakage)** Though our focus is on geoengineering technologies, Proposition 4 applies identically to 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 $d\xi > 0$. This reduction in emissions 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 somewhat more involved, but broadly similar, as they can also increase emissions and reduce welfare.

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

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

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

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

17
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 total environmental damages constant. However, with the proportional change in the environmental damage function, there needs to be a further increase in emissions to keep the marginal damage constant. The implications of these changes for environmental damages is ambiguous and depends on the elasticity of the marginal damage function, \( \eta \). If this elasticity is high, it means that marginal damages 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 total environmental damages.

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 \( \eta^{II} \)) 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 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: with sufficiently high \( \eta \), total environmental damages decline by a large amount, while with very small \( \lambda \) private investment in clean technology is nearly optimal, and a further reduction in these investments only has second-order welfare costs.

Remark 2 (Quadratic damage 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. Consequently, in this case, the condition

\[
\lambda \geq \lambda^* = \frac{1}{2} \left( \frac{\Lambda \hat{\tau}}{\pi_c - \pi_d} \right)
\]

is necessary and sufficient for welfare to decline (note that \( \lambda^* < 1/2 \)).

Finally, we turn to geoengineering technologies of type III, which always lead to an increase in environmental damages and are even more likely to reduce welfare than the other
Figure 1: Environmental damage before and after Type I ("carbon removal") and Type II ("climate adaptation") geoengineering, respectively.
two types of geoengineering advances.

**Proposition 6 (Implications of Geoengineering Technologies of Type III)** Suppose that Assumptions 1 and 2 hold. Consider a geoengineering technology improvement of type III that increases \( \zeta \) by a small amount \( d\zeta > 0 \). Then we have

- \( d\hat{\tau} / d\zeta = 0 \) (there is no effect on the equilibrium carbon tax).
- \( dS / d\zeta > 0 \) (the total stock of carbon increases).
- \( d\hat{q} / d\zeta < 0 \) (investment in clean technology declines).
- \( dE / d\zeta > 0 \) (emissions increase, through lower \( \hat{q} \)).
- \( dD / d\zeta > 0 \) (damages increase).
- \( dW / d\zeta < 0 \) if and only if

\[
\eta \leq \eta^{III}(\lambda) \equiv \begin{cases} 
\left( \frac{\Lambda \tau}{\lambda (\pi_c - \pi_d)} - 1 \right)^{-1} \frac{b}{1 - \zeta} & \text{if } \Lambda \tau > \lambda (\pi_c - \pi_d) \\
+\infty & \text{otherwise}.
\end{cases}
\]

(welfare may decrease). Here, \( b \equiv S/E \geq 1 \).

**Proof.** See Appendix A. \( \blacksquare \)

The intuition for this result is similar to those of Propositions 4 and 5, but the result is even sharper. The reason for this is that an increase in \( \zeta \) reduces the Pigovian tax through two channels—first, because a given amount of emissions creates less damage, inducing a lower Pigovian tax, and second, the reduced impact of emissions on the total stock of carbon decreases marginal damages because \( D \) is convex. This type of geoengineering thus necessitates an even larger increase in emissions to restore the Pigovian tax to \( \hat{\tau} \), and consequently, environmental damages always increase. This also implies that the condition for welfare to decline is even less restrictive than before—in particular, it is strictly weaker than the conditions under which welfare declines with geoengineering technologies of type I or II.

Together, Propositions 4–6 form the main result 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 total environmental damages (Proposition 4), geoengineering technologies that flatten the damage function sometimes increase environmental damage (Proposition 5), and geoengineering technologies that reduce the rate at which emissions enter the atmosphere always increase environmental damage (Proposition 6). It is also
noteworthy that when $\lambda$ (the probability of new entrants replacing existing producers) is sufficiently large, all three geoengineering improvements reduce welfare.\textsuperscript{20}

3 Dynamic Model

We now extend our static model to a dynamic economy where production decisions are made continuously, firms enter and exit, 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 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 geoengineering technologies on the BGP are essentially identical to those derived in the static model.

3.1 Production, Entry and Environmental Damages

As in the static model, we consider an economy with a unique final energy good, produced by a continuum of perfectly substitutable activities indexed by $[0, 1]$. Time $t$ is infinite and discrete, of length $\Delta > 0$. In what follows, we simplify the exposition by taking $\Delta \to 0$ and thus 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, and 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 $t$ has access to the production technology

$$A^{n_{it}} f_d(k_{it}), \quad (8)$$

where $k_{it} \equiv K_{it} / A^{n_{it}}$ is “normalized investment,” $K_{it}$ is investment (again in terms of the final 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).\textsuperscript{21} We also assume that the dirty production technology emits $K_{it}$ units of carbon given investment $K_{it}$.

\textsuperscript{20}In Appendix B, we show that the main insights from the static model generalize if we make damages non-separable with consumption or allow for firms to be ex ante heterogeneous in terms of their cost of switching to clean technology.

\textsuperscript{21}As in the static model, the Inada conditions imply that, despite productivity differences across activities, all activities will produce positive output.
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

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

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 emissions at time $t$ are

$$E_t = \int_0^1 K_{it} di.$$

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

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

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

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

$$\dot{S}_t = (1 - \zeta) E_t A_t - \delta S_t,$$

where $S_0 \geq 0$, $\zeta$ is the same geoengineering parameter introduced in the static model, and $\delta > 0$ is the environmental regeneration rate, while environmental damages are

$$A_t D(S_t; \xi, \upsilon),$$

where

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

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

$$D(S; \xi, \upsilon) \equiv (1 - \upsilon) \bar{D}(S - \xi)$$

as in the static model, and $\bar{D}(\cdot)$ is increasing, strictly concave, and twice continuously differentiable in the stock of carbon $S$. We set the geoengineering parameters as $(\xi, \upsilon, \zeta) = 0$ and omit them from our notation until the final subsection of this section. Note that damages are multiplied by average productivity, while emissions are divided by average productivity.
This formulation captures the fact that when the productivity or consumption 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, who discounts the future at the exponential rate $\rho > 0$. In the text we simplify our 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

$$\sum_{s=t}^{\infty} \left[ C_s + A_s \right] D(S_s; \xi, \upsilon) e^{-\rho(s-t)},$$

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

$$\int_t^{\infty} \left[ C_s - A_s D(S_s; \xi, \upsilon) \right] e^{-\rho(s-t)} ds. \quad (14)$$

The switch from dirty to clean technology has a fixed cost of $A_n \Gamma > 0$ in terms of the final good for activity $i$ with productivity $A_n$, and is incurred only once for each activity (because once an activity switches to clean technology, all future productivity improvements build on the existing clean technology). 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 R&D in order to improve over existing products. R&D uses a scarce input, say scientists, which has an inelastic supply of $Z > 0$.\(^{22}\) 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 matched to one of the activities in $[0, 1]$. A successful innovation for activity $i$ currently with productivity $A_n$ enables the entrant to replace the incumbent producer of this activity with a new technology with productivity $A_n + 1$. Let us denote R&D effort (scientists hired) at time $t$ by $z_t$. Then the (Poisson) arrival rate of a successful innovation is

$$\lambda_t = \phi z_t. \quad (15)$$

\(^{22}\)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.
where $\varphi > 0$. The cost of R&D effort of $z_t$ is $z_tw_t$, where $w_t$ denotes the equilibrium wage for scientists. This wage is determined from the market-clearing condition for scientists given by

$$z_t = Z \quad \text{for all } t \geq 0. \quad (16)$$

This naturally ensures that in equilibrium

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

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

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

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

### 3.2 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

$$\Pi_{idt} = \max_{k \geq 0} A^{n_t} [f_d(k) - (1 + \tau_t)k] = A^{n_t} [f_d(k_d(\tau_t)) - (1 + \tau_t)k_d(\tau_t)] \quad (17)$$

and

$$\Pi_{ict} = \max_{k \geq 0} A^{n_t} [f_c(k) - (1 + \gamma \tau_t)k] = A^{n_t} [f_c(k_c(\tau_t)) - (1 + \gamma \tau_t)k_c(\tau_t)] \quad (18)$$

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 $\pi_j(\tau_t) \equiv A^{n_t} / A^{n_t} \Gamma$ 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 the 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 the “dividend” on its asset of $V_{ct}(n)$, given by the profit flow, 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, this interest rate is always equal to the discount rate, i.e.,

$$r_t = \rho,$$

and thus this expected net present discounted value can be expressed as

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

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 the cost $A^n \Gamma > 0$, 

$$V_{dt}(n) = \max \left\{ V_{ct}(n) - A^n \Gamma, \frac{A^n \pi_d(\tau_1) + \dot{V}_{dt}(n)}{\rho + \lambda} \right\}.$$

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).

Equations (19) and (20) show that $V_{jt}(n)/A^n$ is independent of $n$ (for $j \in \{c,d\}$), and we define $v_{jt} \equiv V_{jt}(n)/A^n$ as the normalized value function.

### 3.3 Clean Technology and R&D Decisions

Equation (20) immediately gives us the equivalent of (Technology IC) in the static model. Firms are happy to switch to clean technology only when the max operator in this expression picks the first term, or, put in terms of normalized value functions, when

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

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

$$\begin{cases} v_{dt} = v_{ct} - \Gamma & \implies q_t \in [0, 1] \\ v_{dt} > v_{ct} - \Gamma & \implies q_t = 0 \end{cases}$$

(Dynamic Technology IC)

which resembles the static model.$^{23}$ Unlike the static condition (Technology IC), however, there is no case in which $v_{dt} < v_{ct} - \Gamma$, since $v_{dt} = \max\{v_{ct} - \Gamma, (\rho + \lambda)^{-1}(\pi_d - \dot{V}_{dt})\}$ implies that $v_{dt} \geq v_{ct} - \Gamma$ for all $t > 0$. Naturally, the equilibrium involves $q_t = 1$ when the max operator always strictly picks the first term in (20). We provide conditions for this not to be the case in equilibrium in Assumption 2’ below.
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 (15), 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 (19) and (20), $w_t$ 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

$$q_t v_{ct} + (1 - q_t) v_{dt} = \frac{w_t}{\varphi A_t}.$$  (22)

At each $t$, the wage for scientists, $w_t$, adjusts to satisfy (22) (so $z_t = Z$).

### 3.4 Planner’s Problem

As in the static model, the (social) planner is benevolent, and therefore maximizes the same objective as the representative household, (14). She will seek to achieve this objective by choosing 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 via carbon taxes 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, $\lambda$, with the main difference now that this flow rate is endogenous.

### 3.5 Definition of Equilibrium

We focus on Markovian equilibria where no agent can condition its strategy at $t$ on the history of play except through the state variables $(S_t, q_t, \{n_{it}\}_{i \in [0,1]})$. 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.24

---

24In our setup, this would take the form of firms expecting prohibitively high taxes following a deviation to lower than promised taxes by the planner at any point. Though such schemes are not always feasible, they nevertheless complicate the analysis.
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 \( (\tau_t^*)_{t \geq 0} \) maximize household utility (14) at each \( t \geq 0 \),
- Given \( (\tau^*_t)_{t \geq 0} \), clean technology decisions \( (q^*_t)_{t \geq 0} \) satisfy (Dynamic Technology IC),
- Given \( \tau_t^* \), input choices \( k^*_dt \) and \( k^*_ct \) maximize, respectively, \( \pi^*_dt \) and \( \pi^*_ct \) in (17) and (18), 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 (16) and free entry (22) for each \( t \geq 0 \).

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

\[
g \equiv \alpha \lambda = \alpha \varphi Z,
\]

where the presence of the term \( \alpha = A - 1 \) follows from the properties of the Poisson process.\(^{25}\) We will also see that in a BGP, \( S_t = \hat{S} \) for all \( t \). When this causes no confusion, we will also include \( \hat{S} \) in the definition of a BGP (or \( S^*_t \) in the definition of an equilibrium).

### 3.6 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 (14), subject to the evolution of the stock of carbon given in (11),

\[
H_t(K_t, S_t) = C_t - A_t D(S_t) - \mu_t \left[ E_t / A_t - \delta S_t \right],
\]

\(^{25}\)Each \( n_t \) is a sample path of a Poisson process with intensity \( \lambda t \), so \( \int_0^1 A^{n_t} dt \) corresponds to the expectation of \( A^{N_t} = \exp(N_t \log A) \), where \( N_t \sim \text{Pois}(\lambda t) \) (and \( \mathbb{E}[e^{\phi N_t}] = \exp(\lambda t(e^\phi - 1)) \) for any \( \phi \in \mathbb{R} \)).
where \( \mu_t \) is the costate variable associated with the stock of carbon in the atmosphere.\(^{26}\) Since emissions are divided by average productivity, \( A_t \), the shadow value of carbon emissions is given by

\[
p_t = \frac{\mu_t}{A_t}. \tag{24}
\]

Under this reparameterization, the necessary and (with the usual transversality condition) sufficient first-order condition for optimality is

\[
\frac{\partial H_t}{\partial S} = \dot{\mu}_t - \rho \mu_t,
\]

which yields a simple form for the shadow price of carbon emissions provided that the planner’s maximization problem in (23) 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,

\[
\dot{p}_t = -D'(S_t) + (\delta + \rho - \alpha \lambda) p_t \tag{25}
\]

and thus

\[
p_t = \int_t^\infty D'(S_s) e^{-(\delta + \rho - \alpha \lambda)(s-t)} ds, \tag{26}
\]

for all \( t \geq 0 \).

**Proof.** See Appendix A. \( \blacksquare \)

We refer to the tax trajectory implied by (26) 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 taxes characterized in (26), at least once clean technology converges.\(^{27}\)

\(^{26}\)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.

\(^{27}\)The main technical detail, showing that equilibrium clean technology indeed always converges in finite time, is stated and proven as Lemma A1 in the Appendix.
Proposition 7 (Pigovian best-response) There exists $T < \infty$ such that equilibrium taxes are given by

$$\tau_t = p_t$$

for all $t \geq T$.

Proof. See Appendix A.

Proposition 7 shows that, despite the complicated dependence of clean technology and R&D decisions on taxes, equilibrium carbon taxes take a simple form. In fact, (26), 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 any possibility of the planner choosing a tax sequence that is ex post distortionary (different from this Pigovian benchmark), 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 proposition.\[28\]

Remark 3 (Counterexample to pure Pigovian taxes) Proposition 7 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 may prefer to set a tax rate strictly less than the Pigovian one in the interval $[0, T]$ because, by doing so, she increases future Pigovian taxes and encourages a faster switch to clean technology. We analyze the conditions under which this possibility could arise in Appendix A, 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; whereas we can choose $\lambda$ to be lower than the threshold given in (A15) and still have geoengineering technologies reduce welfare).

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

3.7 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

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

\[28\]This result is reminiscent of the generic time-inconsistency result of Calvo (1978). It is also simplified due to continuous time, which removes the possibility of choosing a distortionary tax today in order to affect behavior until the taxes are adjusted tomorrow.
This assumption enables us to develop another parallel with the static model. Analogously with (4), let us define
\[
\Psi(\tau_t) \equiv \pi_c(\tau_t) - \pi_d(\tau_t)
\]
\[
= f_c(k_c(\tau_t)) - k_c(\tau_t) - (f_d(k_d(\tau_t)) - k_d(\tau_t)) + \Lambda(\tau_t)\tau_t
\]
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),
\[
v_{dt} - v_{ct} = \frac{\Psi(\hat{\tau})}{\rho + \lambda} = \Gamma,
\]
where 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 (21).

**Assumption 2' (Conditions for Dynamic Interior Equilibrium)** Let the initial carbon stock be \(S_0\). Then for all \(t \geq 0\),
\[
\Gamma \in \left(\int_{t}^{\infty} \Psi(\tau_s)e^{-(\rho+\lambda)(s-t)}ds, \int_{t}^{\infty} \Psi(\tau_s)e^{-(\rho+\lambda)(s-t)}ds\right)
\]
where
\[
\tau_t = \int_{t}^{\infty} D'(S_0e^{-\delta s} + \int_{0}^{s} k_d(\tau_v)e^{-\delta(s-v)}dv) e^{-(\delta+\rho-\alpha\lambda)(s-t)}ds
\]
and
\[
\tau_t = \int_{t}^{\infty} D'(S_0e^{-\delta s} + \int_{0}^{t} k_d(\tau_v)e^{-\delta(t-v)}dv) e^{-(\delta+\rho-\alpha\lambda)(s-t)}ds + \gamma \int_{t}^{s} k_c(\tau_v)e^{-\delta(s-v)}dv e^{-(\delta+\rho-\alpha\lambda)(s-t)}ds.
\]

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 far-sighted Pigovian taxes and R&D intensities that will prevail when no firm ever switches to clean technology, \((\tau_t)_{t \geq 0}\), or all firms switch to clean technology, \((\tau_s)_{s \geq t}\). As in its static analogue, Assumption 2, the conditions 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 (11), this limiting value of the stock of carbon must satisfy
\[
\hat{q}\gamma k_c(\hat{\tau}) + (1 - \hat{q})k_d(\hat{\tau}) = \delta \hat{S}.
\]
Using (26) and (27), the stationary Pigovian tax \( \hat{\tau} \) is given by

\[
\hat{\tau} = \frac{D'(\hat{S})}{\delta + \rho - \alpha \lambda}.
\] (31)

These two equations together with (29) determine \((\hat{S}, \hat{\tau}, \hat{q})\). The next proposition establishes that such a BGP exists and is unique.

**Proposition 8 (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 (29), (30), and (31).

**Proof.** See Appendix A. ■

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 equation (29), there exists a unique \( \hat{q} \) that solves (31). 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, \( \hat{q} \in (0,1) \), or in other words the equilibrium is “interior.” This, in particular, ensures that in the BGP, (29) 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 8, and does so by some \( T < \infty \).

**Proposition 9 (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:

1. for all \( t \in [0, T) \), \( \tau_t \) and \( S_t \) grow continuously and \( q_t = 0 \).

2. 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 8.

**Proof.** See Appendix A. ■

Figure 3.7 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 (29) holds, at which point clean technology leaps from zero to \( \hat{q} \).
The proof of Proposition 9 is provided in the Appendix. Here we give some intuition. Proposition 8 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 9 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, (29), so that we have $\tau_t = \hat{\tau}$ for all $t \geq T$.

3.8 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 $\lambda_t > 0$, the second-best deviates from Pigovian taxation. The main differences are that the condition that
\( \lambda_t > 0 \) 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 10 (Dynamic second-best)**  
1. 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 \( q^t_{SB} > \hat{q} \).

2. 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 \( q^t_{SB} = \hat{q} \).

**Proof.** See Appendix A. ■

### 3.9 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 8, and show that the results are essentially identical to the effects of geoengineering in the static model, derived in Section 2.10. We again distinguish between the three types of geoengineering advances, captured by the parameters \( \xi, \upsilon \) and \( \zeta \) in the general damage function \( (1 - \upsilon)D(S_t - \xi) \), with the associated law of motion of the stock of carbon in the atmosphere \( \dot{S}_t = \delta S_t + (1 - \zeta)E_t/A_t \).

**Proposition 11 (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).
- \( dW/d\xi < 0 \iff \frac{1}{\rho + \lambda}(\hat{\pi}_c - \hat{\pi}_d) > \Lambda \hat{\tau} \) (welfare may decline).

**Proof.** See Appendix A. ■

This proposition shows that any geoengineering advance of type I results in conclusions similar to Proposition 4—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 two propositions give the dynamic analogues of Propositions 5 and 6. In each case, there is a negative effect from geoengineering on the BGP carbon tax, and welfare may decline.

**Proposition 12 (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 \(\nu\) by a small amount \(d\nu > 0\), and let \(\eta = \hat{SD}'(\hat{S}) / D'(\hat{S})\) be the elasticity of the marginal damage function. Then

- \(d\hat{\tau} / d\nu = 0\) (taxes do not change).
- \(d\hat{S} / d\nu = \frac{D'(\hat{S})}{(1-v)D''(\hat{S})} > 0\) (the stock of carbon increases).
- \(d\hat{q} / d\nu < -\frac{d\hat{S}}{(1-v)\Lambda \hat{q}} < 0\) (clean technology declines).
- \(dW / d\nu < 0 \iff \eta < \eta^{II}(\lambda)\), where
  \[
  \eta^{II}(\lambda) \equiv a \left( \delta(\rho - g + \delta) \frac{\hat{\pi}_c - \hat{\pi}_d}{\rho + \lambda \Lambda\hat{\tau}} + \frac{\delta}{\rho - g + \delta} \right)
  \]
  and \(a \equiv \hat{SD}'(\hat{S}) / D(\hat{S}) > 1\) (welfare may decline).

**Proof.** See Appendix A.

**Proposition 13 (Dynamic Implications of Type III 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 III that increases \(\zeta\) by a small amount \(d\zeta > 0\), and let \(\eta = \hat{SD}''(\hat{S}) / D'(\hat{S})\) be the elasticity of the marginal damage function. Then

- \(d\hat{\tau} / d\zeta = 0\) (taxes do not change).
- \(d\hat{S} / d\zeta = \frac{D'(\hat{S})}{(1-\zeta)D''(\hat{S})}\) (the stock of carbon increases).
- \(d\hat{q} / d\zeta < -\delta \frac{d\hat{S}}{(1-\zeta)\Lambda} \left(1 + \frac{1}{1-\zeta} \frac{1}{\eta}\right)\) (clean technology falls).
- \(dW / d\zeta < 0 \iff \eta < \eta^{III}(\lambda)\), where we define
  \[
  \eta^{III}(\lambda) \equiv \left[ \frac{\delta}{\rho - g + \delta} - \delta(\rho - g + \delta) \frac{\hat{\pi}_c - \hat{\pi}_d}{\rho + \lambda \Lambda\hat{\tau}} \right]^{-1} \left( \delta(\rho - g + \delta) \frac{\hat{\pi}_c - \hat{\pi}_d}{\rho + \lambda \Lambda\hat{\tau}} + \frac{\delta}{\rho - g + \delta} \right)
  \]
  when \(\lambda < \frac{\rho + \lambda}{(\rho - g + \delta)^2} (\hat{\pi}_c - \hat{\pi}_d)\), and \(\eta^{III}(\lambda) \equiv +\infty\) otherwise (welfare can fall).
Proof. See Appendix A. ■

We note in addition that the conditions for welfare to decline as a result of a geoengineering advance of type II or type III are again very similar to those we have obtained in the static model (particularly in Propositions 5 and 6). In particular, as in the static model, if $\hat{\lambda}$ is sufficiently large, welfare declines following any one of the three types of geoengineering advances.

4 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. In this paper, we have proposed an alternative perspective on the possible dark side of geoengineering. We have argued, theoretically, that geoengineering may damage the environment and welfare precisely because it is expected to, and it will, materialize. 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), and this will discourage current investments in conventional cleaner technology (where our emphasis on “conventional” is to distinguish it from geoengineering technologies).

To advance this argument, we have developed a model of endogenous transition to clean technology with policy-making without commitment. Both of these elements are relatively new in the environmental literature, and important for our argument. Though transition to various types of clean technology (including wind, solar, and geothermal) is generally seen as a bedrock of any reduction in the pace of buildup of carbon in the atmosphere, there are relatively few analyses of this process in the economics literature (see the references in the Introduction). The modeling of the transition to conventional clean technology is critical for understanding the potential adverse effects of geoengineering, because it is these types of investments that may be discouraged if future carbon taxes are expected to be low. Lack of commitment to future policies in general and carbon taxes in particular is also an evident reality, but most economic analyses of environmental policy have stayed away from the time-inconsistency issues that arise in the absence of such commitment. It also plays a pivotal role in our setting because it is this lack of commitment that makes it impossible for future carbon
taxes to remain high when geoengineering advances materialize.

We start with a static model in which existing energy producers can undertake costly investments to switch to clean technology and once these technology investments are made, a benevolent planner sets the carbon tax. Lack of commitment to policies means that the planner cannot deviate from the Pigovian carbon tax once technology investments are sunk. But because such investments create a positive externality—for other firms can also build on them—the planner would have preferred to commit to a carbon tax greater than the Pigovian level, had this been possible. Furthermore, we restrict attention to parameters such that the equilibrium is an “interior” one where some firms switch to clean technology, while others do not. In the static model, an interior equilibrium is feasible only if the rate of carbon taxation takes a specific value.

We then introduce geoengineering breakthroughs into this framework. For simplicity, we distinguish between three different types of geoengineering advances. Type I, which corresponds to various technologies aiming at carbon removal from the atmosphere, shifts the damage function from the stock of carbon in the atmosphere downwards in a parallel fashion—and is thus equivalent to a decline in the effective stock of carbon. If no economic decisions changed following this type of geoengineering breakthrough, the marginal and overall damages would decline, leading to lower carbon taxes in the future and higher welfare. But anticipating a lower rate of carbon taxation, all firms would then abandon their investments in clean technology. This would increase emissions and the stock of carbon. Provided that the geoengineering breakthrough is not so large as to destroy the interior equilibrium, we must then have sufficiently higher emissions so that marginal damages are restored to their pre-geoengineering level and energy producers are incentivized to invest in conventional clean technology again. In this case, therefore, total environmental damages remain constant despite the geoengineering breakthrough, and overall welfare may decrease. With geoengineering advances of type II, which correspond to climate adaptation technologies including solar radiation management, environmental damages decline proportionately. In this case, we show that a similar reasoning leads to an increase or decrease in the overall environmental damages depending on the elasticity of the marginal damage function. Intuitively, the stock of carbon in the atmosphere has to increase again so that the marginal environmental damage and future carbon taxes do not decline, and depending on the aforementioned elasticity, this may necessitate a large or a small increase in the stock of carbon in the atmosphere (the greater is the elasticity, the larger is the requisite change in the stock of carbon). As a result, welfare may again decrease. Finally, geoengineering breakthroughs of type III, which
include technologies that are aimed at external carbon capture and sequestration, neutralize some portion of emissions directly. In this case, because this type of technology directly reduces the impact of emissions, it reduces future carbon taxes even more sharply, and as a result, environmental damages necessarily increase and welfare is even more likely to decrease. Overall, these three different types of geoengineering breakthroughs all generate countervailing negative effects, and may make the problem of reducing and controlling carbon emissions much more difficult.

We show that the general insights are not dependent on specific assumptions made for tractability and clarity in our model. Adding heterogeneity or changing the way in which damages are modeled does not change our qualitative conclusions. More importantly, similar results apply in the context of a dynamic model in which the stock of carbon in the atmosphere and technology evolve gradually. In this dynamic model, the positive externalities from switching to clean technology have a more compelling microfoundation: technological progress takes the form of firms going up a quality ladder, and investments for switching to a clean technology enable further improvements on that ladder to build on the foundations laid by this clean technology. We show that the BGP equilibrium in this dynamic model has a very similar structure to our static equilibrium, and the effects of the three types of geoengineering breakthroughs are essentially identical to what we described in the previous paragraph.

We see this paper as a first step both in the investigation of the implications of policy-making without commitment in the context of environmental policies and in the study of the consequences of 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 missing element from our analysis has been interactions between different countries and jurisdictions, which would bring political economy considerations, in addition to the issues of policy-making without commitment. Perhaps even 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.

**Appendix A: Omitted Proofs**

**Proof of Proposition 2 (interior equilibrium).** *(Interiority).* From Proposition 1, \( \hat{\tau} = D'(\cdot; \hat{\theta}) \).

Assumption 2, which imposes that \((1 - \lambda)\Psi(\tau) < \Gamma\) and \((1 - \lambda)\Psi(\tau) > \Gamma\), 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 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 $\frac{\partial S}{\partial q} = \Lambda$ and the fact that in the interior, $f_c(k_c) - k_c - (f_d(k_d) - k_d) + \Lambda \tau = \frac{\Gamma}{1-\lambda} = \Gamma - \frac{\lambda}{1-\lambda} \Lambda \tau$, becomes

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

(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 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 $-\Lambda dq = d\xi$.

(Welfare). We can calculate the total derivative of welfare, $W = q(f_c(k_c) - k_c - \Gamma) + (1 - q)(f_d(k_d) - k_d) - D((1-\delta)S_0 - \xi + E)$, with respect to $\xi$ as

$$\frac{dW}{d\xi} = \left[ q(f_c'(k_c) - 1) \frac{dk_c}{d\tau} + (1 - q)(f_d'(k_d) - 1) \frac{dk_d}{d\tau} \right] \frac{d\hat{\tau}}{d\xi}$$

$$+ [f_c(k_c) - k_c - \Gamma - (f_d(k_d) - k_d)] \frac{dq}{d\xi} + D' - \frac{dE}{d\xi} D'$$

$$= [f_c(k_c) - k_c - \Gamma - (f_d(k_d) - k_d)] \frac{dq}{d\xi} + D' - \frac{dE}{d\xi} D'$$

(A2)
where the second line uses $d\tau/d\zeta = 0$ and the third uses $dE/d\zeta = 1$. (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

$$\frac{dW}{d\zeta} < 0 \iff \lambda(\pi_c - \pi_d) < \Lambda\tau.$$  

\[
\text{Proof of Proposition 5 (type II geoengineering). (I. Taxes). As in the proof of Proposition 4, only } d\tau/d\upsilon = 0 \text{ sustains IC.}
\]

\[
(II. \text{Environmental damage). Differentiating total environmental damage, } (1 - \upsilon)D(S), \text{ with respect to } \upsilon, \text{ we obtain}
\]

$$\frac{dD(\cdot)}{d\upsilon} = D(S) + (1 - \upsilon)D'(S) \frac{dS}{d\upsilon}.$$

(A3)

To calculate $dS/d\upsilon$, note that because $d\tau/d\upsilon = 0$, we can differentiate $(1 - \upsilon)D'(S) = \hat{\tau}$ with respect to $\upsilon$ to obtain

$$-D'(S) + \frac{dS}{d\upsilon}(1 - \upsilon)D''(\cdot) = 0 \implies \frac{dS}{d\upsilon} = \frac{1}{1 - \upsilon} \frac{D'(S)}{D''(S)}.$$  

The total effect in (A3) then becomes

$$\frac{dD(\cdot)}{d\upsilon} = -D(S) + \frac{1}{1 - \upsilon} \frac{D'(S)}{D''(S)} (1 - \upsilon)D'(S)$$

$= -D(S) + \frac{1}{1 - \upsilon} \eta \cdot SD'(S)$

(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

$$\eta < \eta^* \iff dD/d\upsilon > 0.$$  

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

$$\frac{dW}{d\upsilon} = \frac{\partial}{\partial q} \left[ q[f_c(k_c) - k_c] + (1 - q)[f_d(k_d) - k_d] - q\Gamma \right] \frac{d\hat{q}}{d\upsilon} - \frac{dD(\cdot)}{d\upsilon}$$

$$= [f_c(k_c) - k_c - [f_d(k_d) - k_d] - \Gamma] \frac{d\hat{q}}{d\upsilon} - \frac{dD(\cdot)}{d\upsilon}$$

$$= \left[ \frac{\lambda}{1 - \lambda} \Gamma - \Lambda\tau \right] \frac{d\hat{q}}{d\upsilon} - \frac{dD(\cdot)}{d\upsilon}.$$
where the last substitution follows from (5).

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

\[
\frac{dq}{d\nu} = \left[ \frac{\partial E}{\partial q} \right]^{-1} \frac{dS}{d\nu} = \frac{1}{\Lambda(1 - \nu)} \frac{D'(S)}{D''},
\]

or equivalently,

\[
\frac{dq}{d\nu} = -\frac{1}{\Lambda(1 - \nu)} \frac{S}{\eta}.
\]

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

\[
\left[ \frac{\lambda}{1 - \lambda} \frac{\Gamma - \Lambda \tau}{1 - \lambda} \right] \frac{dq}{d\nu} = -\frac{\lambda}{1 - \lambda} \frac{\Gamma}{1 - \lambda} \cdot \frac{1}{\Lambda(1 - \nu)} \frac{S}{\eta} + \frac{1}{\eta} S D'(S).
\]

From above, the total effect on environmental damage is

\[
-\frac{dD(\cdot)}{d\nu} = D(S) - \frac{1}{\eta} S D'(S).
\]

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

\[
\frac{dW}{d\nu} = D(S) - \frac{\lambda}{1 - \lambda} \frac{\Gamma}{1 - \lambda} \cdot \frac{1}{\Lambda(1 - \nu)} \frac{S}{\eta}.
\]

From (5), we have

\[
\Gamma = (1 - \lambda)(\pi_c - \pi_d),
\]

and multiplying everything by \( \Lambda(1 - \nu) D'(S) / D(S) \), we obtain

\[
\frac{dW}{d\nu} < 0 \iff \Lambda \tau - \lambda(\pi_c - \pi_d) \frac{1}{\eta} \frac{S D'(S)}{D(S)} < 0
\]

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

\[
\eta < \eta^I(\lambda) \equiv a\lambda \left( \frac{\pi_c - \pi_d}{\Lambda \tau} \right)
\]

characterizes the family of damage functions for which \( dW/d\nu < 0 \).

\textbf{Proof of Remark 2 (Quadratic damages).} \textit{(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), \( dD / d\nu > 0 \iff -D(S) + \eta^{-1} SD'(S) \), so

\[
dD / d\nu > 0 \iff -1 + \eta / 2 + 1 / \eta > 0,
\]
or \( dD / dv > 0 \iff \eta^2 / 2 - \eta + 1 > 0 \). 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 \( \lambda \geq \lambda^* \).

Proof of Proposition 6 (type III geoengineering). As in the proofs of propositions 4 and 5, \( d\tau / d\zeta = 0 \) to sustain IC. To obtain \( dq / d\zeta \), totally differentiate

\[
(1 - \zeta)D'(1 - \delta)S_0 + (1 - \zeta)E(q) = \hat{\tau}
\]

with respect to \( \zeta \), using \( d\hat{\tau} / d\zeta = 0 \), to obtain

\[
-D'(S) + \left[ -E + (1 - \zeta) \frac{\partial E}{\partial q} dq \right] (1 - \zeta)D''(S) = 0,
\]

and rearranging, with \( \partial E / \partial q = -\Lambda \), gives

\[
\frac{dq}{d\zeta} = - \frac{(1 - \zeta)^{-1} D'(S) / D''(S) + E}{(1 - \zeta)^{-1}} < 0,
\]

(A6)

so equilibrium technology always declines. Environmental damage increases according to

\[
\frac{d}{d\zeta} D(\cdot) = \frac{dS}{d\zeta} D'(S) = \left[ \frac{1}{1 - \zeta} \frac{D'(S)}{D''(S)} \right] D'(S) > 0.
\]

(A7)

Using (A7) and the derivative of output with respect to \( q \) obtained in the proof of Proposition 4,

\[
\frac{dW}{d\zeta} = \left( \frac{\lambda}{1 - \lambda} - \lambda \tau \right) \frac{dq}{d\zeta} - \left( \frac{1}{1 - \zeta} \frac{D'(S)}{D''(S)} \right) D'(S)
\]

and as \( \tau = (1 - \zeta)D'(S) \), using (A6), we obtain

\[
\frac{dW}{d\zeta} = \left( \frac{\lambda}{1 - \lambda} \right) \frac{dq}{d\zeta} + ED'(S)
\]

or

\[
\frac{dW}{d\zeta} = - \left( \frac{\lambda}{1 - \lambda} \right) (1 - \zeta)^{-1} \frac{D'(S) / D''(S) + E}{(1 - \zeta)^{-1}} + ED'(S)
\]
which, using \((1 - \lambda)^{-1} \Gamma = \pi_c - \pi_d\), and \(\tau = (1 - \zeta)D'(S)\), can be rewritten as
\[
\frac{dW}{d\zeta} < 0 \iff \lambda(\pi_c - \pi_d) \left( \frac{b}{1 - \zeta} \frac{1}{\eta} + 1 \right) > \Lambda \tau
\]
where \(b \equiv S/E \geq 1\). This yields the threshold
\[
\eta < \eta^{III}(\lambda) \equiv \begin{cases} 
\left( \frac{\Lambda \tau}{\lambda(\pi_c - \pi_d)} - 1 \right)^{-1} \frac{b}{1 - \zeta} & \text{if } \Lambda \tau > \lambda(\pi_c - \pi_d) \\
\infty & \text{otherwise.}
\end{cases}
\] (A8)

**Proof of Lemma 3 (the shadow cost of carbon).** In equilibrium, the costate variable \(\mu_t\) will satisfy the Euler-Lagrange condition \(\partial H_t/\partial S = \dot{\mu}_t - \rho \mu_t\), which we can write as
\[
\mu_t = -A_tD'(S_t) + (\rho + \delta)\mu_t.
\]
Dividing this equation by \(A_t\), we obtain
\[
\frac{\dot{\mu}_t}{A_t} = -D'(S_t) + (\rho + \delta)\mu_t.
\] (A9)
Since \(p_t = \mu_t/A_t\), we have
\[
\dot{p}_t = \frac{\dot{\mu}_t}{A_t} - \frac{\mu_t}{A_t^2} \frac{dA_t}{dt} = \frac{\dot{\mu}_t}{A_t} - p_t a \lambda,
\]
where the second equality uses the fact that \(dA_t/dt = a \lambda A_t\). Hence (A9) becomes
\[
\dot{p}_t = -D'(S_t) + (\delta + \rho - a \lambda) p_t,
\]
which is exactly (25). Furthermore, with the transversality condition,
\[
\lim_{t \to \infty} p_t S_t e^{-\rho t} = 0
\] (A10)
and the initial condition \(S_0\), we obtain
\[
p_0 = \int_0^\infty D'(S_t) e^{-(\delta + \rho - a \lambda)t} dt,
\]
and the differential equation (25) admits the unique solution (26). ■

**Proof of Proposition 7 (Pigovian best-response).** We start with a crucial lemma.

**Lemma A1 (Convergence of clean technology)** Suppose that Assumptions 1’, 2’, and 3 hold.
Then, \(q_t \to \hat{q}\) by some finite time \(T < \infty\).
Proof of Lemma A1. The sequence \((q_t)_{t \geq 0}\) lives in the compact set \([0,1]\), and \((q_t)_{t \geq 0}\) is monotone since clean technology decisions are irreversible. Hence \(q_t \to \tilde{q}\).

Suppose, to obtain a contradiction, that \(q_t < \tilde{q}\) for all \(t\). Any tax trajectory that sustains an equilibrium in which \(\dot{q}_t > 0\) for every \(t' < \infty\) must satisfy (21) infinitely often, which requires that for this for some \(t \geq t'\), we have

\[
\int_t^{\infty} \Psi(\tau_s) e^{-(p+\lambda)s} \, dt = \Gamma.
\]

Let \((\tau_t)_{t \geq 0}\) be a sequence of taxes with this property.

Furthermore, the planner can attain \(\tilde{q}\) by always setting Pigovian taxes, which she prefers to any \(\tilde{q} < q\) (this is immediate, and also follows directly from Proposition 10, which shows that the planner prefers to induce a transition to \(q_t \to \tilde{q}\)). Hence we can focus on the case where \(\tilde{q} > \tilde{q}\). Assumption 2' and \(\hat{q} > \hat{q}\) together imply that for every \(t' < \infty\), there is a subset of \([t', \infty)\) of positive measure for which \(\tau_t > p_t\).

We next use this fact and construct a deviation from \((\tau_t)_{t \geq t'}\) to \((p_t)_{t \geq t'}\) that induces \(q_t' \neq \tilde{q}\) for all \(t' < \infty\) (with \(q_t \to \tilde{q}\)), completing the contradiction argument. To verify this, observe that deviating at \(t'\) to \(p_t\) forever (calculated with reference to the deviation path \((S'_t)_{t \geq t'}\)) will fix \(q_t = q_t\) for all \(t \geq t'\). Therefore, the time-\(t'\) deviation will yield welfare

\[
U(t') = \int_{t'}^{\infty} [q_t' [f_c(k_c(p_t)) - k_c(p_t)] + (1 - q_t') [f_d(k_d(p_t)) - k_d(p_t)] - D(S_t')] e^{-(p-\gamma)t} \, dt,
\]

while the original path with \(q_t \to \tilde{q}\) will yield

\[
Z(t') = \int_{t'}^{\infty} [q_t [f_c(k_c(\tau_t)) - k_c(\tau_t)] + (1 - q_t) [f_d(k_d(\tau_t)) - k_d(\tau_t)] - D(S_t)] e^{-(p-\gamma)t} \, dt,
\]

where the stock of carbon after the deviation, \(S'_t\), satisfies \(S'_t = q_t' y k_c(p_t) + (1 - q_t') k_d(p_t) - \delta S_S\), the original stock of carbon \(S_t\) satisfies \(S_t = q_t y k_c(\tau_t) + (1 - q_t) k_d(\tau_t) - S_t\), and \(S'_t = S_t\).

Hence the gain from the time-\(t'\) Pigovian deviation is

\[
U(t') - Z(t') \geq \int_{t'}^{\infty} \left\{ q_t [f_c(k_c(p_t)) - k_c(p_t)] - \left[ f_c(k_c((\tau_t)) - k_c((\tau_t))] \right. \\
\left. + (1 - q_t') [f_d(k_d(p_t)) - k_d(p_t)] - \left[ f_d(k_d((\tau_t)) - k_d((\tau_t))] \right. \\
- [D(S(q_t', \tau_t)) - D(S(q_t', p_t))] \right\} e^{-(p-\gamma)t} \, dt \\
- \|q_t - q_t'\| \int_{t'}^{\infty} (f_c(\tau_t) - f_d(\tau_t) + \Lambda(\tau_t) D'((S(q_t', \tau_t))) e^{-(p-\gamma)t} \, dt
\]

(A11)

using \(\Gamma > 0\) and the bound \(D(S(q_t', \tau_t)) - D(S(q_t', p_t)) \leq \|q_t - q_t'\| \Lambda(\tau_t) D'((q_t', \tau_t))\) that follows from convexity of \(D(\cdot)\). As \(p_t < \tau_t\) for a set of positive measure, and \(p_t \leq \tau_t\) always,\(^{29}\) the

\(^{29}\)The planner will always set \(\tau_t \geq p_t\) for all \(t \geq t'\), as \(\tau_t < p_t\) reduces clean technology incentives and lowers welfare from net consumption.
first integral in (A11) is strictly positive by definition of \( p_t \), while the final term can be made arbitrarily small in finite time, since \( \| q_t - q_{t'} \| \leq \| \tilde{q} - q_{t'} \| \) for all \( t \geq t' \); thus, for any \( \varepsilon > 0 \), there exists a \( t' < \infty \) such that \( \| \tilde{q} - q_{t'} \| < \varepsilon \). But because \( \Psi(\cdot) \) is strictly increasing and \( \tau_t > p_t \) infinitely often and always on a subset of positive measure, we deduce that

\[
\int_{t}^{\infty} \Psi(p_s)e^{-(p+\lambda)(t-s)}ds < \int_{t}^{\infty} \Psi(\tau_s)e^{-(p+\lambda)(t-s)}ds = \Gamma
\]

for all \( t \geq t' \), which contradicts (21) unless \( \tilde{q}_t = 0 \) for all \( t \geq t' \). Therefore \( q_t \to \tilde{q} \) in finite time.

Finally, any limit \( \tilde{q} \neq \tilde{q} \) cannot be part of an equilibrium, because after \( \tilde{q} \) is reached, from Proposition 7, \( \tau_t = p_t \), and thus (29), (30), and (31) need to hold, and thus \( \tilde{q} = \tilde{q} \in (0,1) \) (where 0 and 1 or ruled out by Assumption 2').

Let \( T \) be given as in the above lemma. First, consider \( t_0 \geq T \). The planner’s objective is

\[
\int_{t_0}^{\infty} \left[ \tilde{q}A_t \left( f_c(k_{ct}) - k_{ct} \right) + (1 - \tilde{q})A_t \left( f_d(k_{dt}) - k_{dt} \right) - A_tD(S_t) \right] e^{-\rho(t-t_0)}dt
\]

which admits the (normalized) Hamiltonian

\[
\hat{H}_t = \tilde{q} \left[ f_c(k_{ct}) - k_{ct} \right] + (1 - \tilde{q}) \left[ f_d(k_{dt}) - k_{dt} \right] - p_t \left[ \gamma\tilde{q}k_{ct} + (1 - \tilde{q})k_{dt} \right]
\]

where \( p_t \) is given by (26). Socially optimal input levels must satisfy the necessary first-order conditions

\[
\frac{\partial\hat{H}_t}{\partial k_{ct}} = \tilde{q} \left[ f'_c(k_{ct}) - 1 \right] - \gamma\tilde{q}p_t = 0
\]

and

\[
\frac{\partial\hat{H}_t}{\partial k_{dt}} = (1 - \tilde{q}) \left[ f'_d(k_{dt}) - 1 \right] - (1 - \tilde{q})p_t = 0
\]

for all \( t \geq t_0 \), which are also sufficient because \( \hat{H}_t \) is strictly concave. Comparing these to the first-order conditions of firms, which are

\[
f'_c(k_{ct}) - 1 = \gamma\tau_t \quad \text{and} \quad f'_d(k_{dt}) - 1 = \tau_t \quad \text{for all} \ t \geq t_0,
\]

we conclude that \( \tau_t = p_t \) for all \( t \geq T \).

Second, consider \( t < T \). We prove \( \tau_t \leq p_t \) by backwards induction, and do this before taking the limit \( \Delta \to 0 \) for convenience. The planner’s utility at \( T - \Delta \), given by the discrete version of (14) and normalized by \( 1/A_t \), is

\[
\sum_{s=T}^{\infty} \left[ c_{T-\Delta+\Delta(s-T)} - D(\hat{S}_{T-\Delta+\Delta(s-T)}) \right] e^{-\Delta(p-g)(s-T)},
\]

which, since \( q_T = \hat{q} \), we can represent recursively as

\[
c_{T-\Delta} - D(\hat{S}_{T-\Delta}) + e^{-(p-g)\Delta\hat{V}_T(\hat{S}_T, \hat{q})} \quad \text{(A12)}
\]
where
\[ V_t(S_t, q_t) = \max_{\{T_i\}_{i=1}^{N}} \sum_{s=1}^{\infty} \left[ c_{t+\Delta(s-t)} - D(S_{t+\Delta(s-t)}) \right] e^{-(\rho - g)\Delta(s-t)} \]

is the planner’s continuation value conditional on stock of carbon \( S \) and aggregate clean technology \( q \). The first-order condition of (A12) is
\[
\frac{\partial c_{t-\Delta}}{\partial T_{t-\Delta}} + e^{-(\rho - g)\Delta} \frac{\partial V_t(S_t, \hat{q}_t)}{\partial T_{t-\Delta}} = 0,
\]
and, since \( q_t = \hat{q} \) for all \( t \geq T \),
\[
e^{-(\rho - g)\Delta} \frac{\partial V_t(S_t, \hat{q}_t)}{\partial T_{t-\Delta}} = -\sum_{s=0}^{\infty} \frac{\partial S_{t+s\Delta}}{\partial T_{t-\Delta}} D'(S_{t+s\Delta}) e^{-(\rho - g)\Delta s}
\]
\[
= \left( \gamma q_{t-\Delta} \frac{\partial k_{c,t-\Delta}}{\partial T_{t-\Delta}} + (1 - q_{t-\Delta}) \frac{\partial k_{d,t-\Delta}}{\partial T_{t-\Delta}} \right) p_{t-\Delta}.
\]
Consequently, as
\[
\frac{\partial c_{t-\Delta}}{\partial T_{t-\Delta}} = q_{t-\Delta} [f'_c(k_{c,t-\Delta}) - 1] \frac{\partial k_{c,t-\Delta}}{\partial T_{t-\Delta}} + (1 - q_{t-\Delta}) \left[ f'_d(k_{d,t-\Delta}) - 1 \right] \frac{\partial k_{d,t-\Delta}}{\partial T_{t-\Delta}},
\]
and, by firm-level optimization,
\[
f'_c(k_{c,t-\Delta}) - 1 = \gamma t_{t-\Delta} \quad \text{and} \quad f'_d(k_{d,t-\Delta}) - 1 = t_{t-\Delta}
\]
so only \( t_{t-\Delta} = p_{t-\Delta} \) is a best response for the planner. But then (Dynamic Technology IC) can only hold at \( T - \Delta \) if \( q_{t-\Delta} = 0 \). To see this, observe that the net present discounted value of switching to clean technology becomes
\[
v_{c,t-\Delta} - v_{d,t-\Delta} = \Psi(t_{t-\Delta}) + \sum_{s=0}^{\infty} \Psi(t_{t+\Delta s}) e^{-(\rho + \lambda)(s+1)\Delta}
\]
\[
= \Psi(p_{t-\Delta}) + e^{-(\rho + \lambda)\Delta} \frac{\Psi(\hat{t})}{\rho + \lambda}
\]
using our result on \( t \geq T \) to deduce that \( t_{t+\Delta s} = \hat{t} \) for all \( s \geq 0 \). Therefore by (21), equation (A13) gives
\[
v_{c,t-\Delta} - v_{d,t-\Delta} < \frac{\Psi(\hat{t})}{\rho + \lambda} = \Gamma
\]
using that \( \Psi'(\cdot) > 0 \) and \( p_{t-\Delta} < p_{t} = \hat{t} \) (the latter follows from (26) since \( S_T - S_{T-\Delta} > 0 \) and \( S_{T'} = S_T = \hat{S} \) for all \( T' \geq T \)). Hence \( q_{t-\Delta} = 0 \) in equilibrium. An identical argument implies that \( q_{t-2\Delta} = 0 \). Inductively, then, \( q_t = 0 \) for all \( t < T \) and \( t_t \leq p_t \) on \( t < T \).

We also note that we cannot show \( t_t \leq p_t \) for all \( t < T \), but the discussion of Remark 3 below gives precise conditions for which indeed \( t_t \geq p_t \), and therefore \( t_t = p_t \) for all \( t \geq 0 \).
Discussion of Remark 3 (counterexample to taxes being always Pigovian). Consider the Pigovian equilibrium candidate in which \( t_t = p_t \) for all \( t \), and

\[
q_t = \begin{cases} 
0 & \text{for } t < T \\
\hat{q} & \text{for } t \geq T 
\end{cases}
\]

where \( T \) is defined by \( p_t = \hat{\tau} \), for \( \hat{\tau} \) from the unique BGP. Let \( (S_t)_{t \geq 0} \) be the pollution stock associated with this Pigovian trajectory.

The net benefit to a deviation \( (t'_t)_{t \leq T} \) inducing a faster carbon trajectory \( S'_t \geq S_t \) on \([0, T]\) and an earlier transition at \( t' < T \), equals

\[
G(t', \lambda) = \int_0^T \left[ f(k_d(t'_t)) - k_d(t'_t) - f(k_d(p_t)) - k_d(p_t) \right] - D(S'_t) + D(S_t) \right] e^{-\gamma t} dt
\]

where the earlier transition creates additional flow profits on \([T', T]\) of

\[
\hat{q} \left( f_c(\hat{k}_c) - \hat{k}_c - f_d(\hat{k}_d) - \hat{k}_d \right) + f_d(\hat{k}_d) - f_d(k_d(p_t)) + k_d(p_t)
\]

incurs the fixed cost \( \Gamma \) at \( T' \), and by (21),

\[
f_c(\hat{k}_c) - \hat{k}_c - f_d(\hat{k}_d) - \hat{k}_d - \Gamma = \frac{\lambda(\hat{\pi}_c - \hat{\pi}_d)}{\rho + \lambda} - \Lambda \hat{\tau}.
\]

Equation (A14) implies that this strategy is preferred if and only if

\[
\lambda \geq \bar{\lambda} = \inf\{\lambda > 0 : G(t'_0, \lambda) > 0\}
\]

where \( t'_0 = \arg\max_{t \leq T} G(t, \lambda) \). Note that if \( G(t'_0, \lambda) \leq 0 \) for all \( \lambda \), then \( \bar{\lambda} = +\infty \). In particular, since the integral in (A14) is negative by the definition of \( p_t \), it must be that

\[
\frac{\bar{\lambda}(\hat{\pi}_c - \hat{\pi}_d)}{\rho + \bar{\lambda}} > \Lambda \hat{\tau}
\]

verifying the remark. ■

Proof of Proposition 8 (uniqueness of BGP). Using (30), the (farsighted) Pigovian tax given by (31) becomes

\[
\hat{\tau} = \frac{1}{\delta + \rho - \alpha \bar{\lambda}} D' \left( \frac{\hat{q} \gamma k_c(\hat{\tau}) + (1 - \hat{q}) k_d(\hat{\tau})}{\delta} \right),
\]

which is a decreasing function of \( \hat{q} \). Hence there exists at most one \( \hat{q} \in (0, 1) \) that solves (31) when \( \hat{\tau} \) is given by (29). ■

Proof of Proposition 9 (dynamic equilibrium). We know that \( q_t \rightarrow \hat{q} \) in finite time by Lemma A1. In addition,
Lemma A2 (Monotone pollution) In equilibrium, \((S_t)_{t \geq 0}\) is everywhere nondecreasing.

Proof of Lemma A2. Assumption 2' implies that \(S_0 \leq \hat{S}\). Suppose also that \(S_t \leq \hat{S}\) for all \(t\) (we prove this below). Now suppose, in order to obtain a contradiction, that \(\dot{S}_t < 0\) for some \(t\). But

\[
\dot{S}_t < 0 \iff \tau_t > \hat{\tau} \geq p_t
\]

because

\[
0 = \dot{q}k_c(\hat{\tau}) + (1 - \dot{q})k_d(\hat{\tau}), \dot{q} - \delta \hat{S} \\
\leq q_t k_c(\tau_t) + (1 - q_t)k_d(\tau_t) - \delta S_t \quad \text{for all } t \leq T,
\]

where the first line follows from the definition of \(\hat{\tau}\) and \(\hat{q}\), and the second follows from \(q_t \leq \hat{q}\) (by monotone convergence of \(q_t \uparrow \hat{q}\), proven in Lemma A1), \(\tau_t \leq \hat{\tau}\), and \(S_t \leq \hat{S}\) by assumption. But (A17) can never be optimal: deviating downwards to \(\hat{\tau}\) will strictly improve welfare since \(\hat{\tau} \geq p_t\) and \(\hat{\tau}\) cannot affect \((q_t)_{t \geq 0}\).

Finally, \(S_t \uparrow \hat{S}\), and until \(T\), at which point \(\dot{S}_T = 0\) and the economy is on the BGP. Hence \(S_0 \leq \hat{S}\) implies \(S_t \leq \hat{S}\) for all \(t\), completing the proof of the lemma.

To conclude the proof of Proposition 9, note also that \(S_t\) is in fact increasing when either \(q_t < \hat{\tau}\) and \(\tau_t < \hat{\tau}\) from (A18), and thus we can conclude that there exists \(T < \infty\) such that \(S_T = \hat{S}\), \(p_T = \hat{\tau}\) and \(q_T = \hat{q}\), which completes the proof of the proposition.

Proof of Proposition 10 (second-best dynamic policy with commitment). Let \(T^*\) denote the first-best switching time (when the planner controls both input decisions and technology choices) and

\[
T \geq T^*
\]

denote the equilibrium switching time without commitment specified in Proposition 9.

Observe that \(T = T^*\) (and therefore \(\hat{\tau}_t = \tau_t^{SB} = \tau_t^* = p_t\) for all \(t \geq 0\) and \(\hat{q} = q_t^{SB} = q^*\)) if and only if \(\lambda = 0\). This follows by comparing the first-order condition for the planner to those of firms. In particular, consider the full Hamiltonian that incorporates the constraint on the evolution of the stock of clean technology, which is

\[
\bar{H}_t = q_t [f_c(k_{ct}) - k_{ct}] + (1 - q_t) [f_d(k_{dt}) - k_{dt}] - q_t \Gamma + \rho_k (\gamma k_{ct} + (1 - q_t)k_{dt}) + Q_t q_t
\]

and note that the first-order conditions \(\bar{H}_q_t = Q_t - \Gamma = 0\) and \(Q_t = -\bar{H}_q_t + \rho Q_t\) imply

\[
f_c(k_c(\tau_t)) - k_c(\tau_t) - [f_d(k_d(\tau_t)) - k_d(\tau_t)] + p_t \Lambda(\tau_t) - \rho \Gamma = 0,
\]

(A19)
with complementary slackness. Now we can see that this coincides with firms’ first-order conditions,

\[ \int_t^\infty \left[ \pi_c(\tau_s) - \pi_d(\tau_s) + \tau_s \Lambda(\tau_s) \right] e^{-(\rho + \lambda)(s-t)} ds \leq \Gamma \]

(with equality if \( q_t > 0 \)) when \( \tau_t = p_t \) if and only if \( \lambda = 0 \). This proves the second part of the proposition.

Otherwise, when \( \lambda > 0 \), \( (A19) \) illustrates that the planner prefers a tax policy that induces convergence to \( q^{SB} > \hat{q} \), and in particular, she will obtain strictly higher welfare by committing to the strategy \( \tau^{SB}_t \) that maximizes welfare subject to \( (21) \) at \( T^{SB} \), i.e., subject to

\[ \int_{T^{SB}}^\infty \Psi(\tau^{SB}_t) e^{-(\rho + \lambda)(t - T^{SB})} dt = \Gamma, \]  \( (A20) \)

in order to induce a transition \( q^{SB} > \hat{q} \) on \([T^{SB}, \infty)\), where \( T^{SB} \leq T \). Note that \( \tau^{SB}_t \geq p^{SB}_t \) for all \( t \geq 0 \), since any \( \tau'_t < p_t \) will reduce clean technology incentives by \( \Psi' > 0 \) (Assumption 1') and reduce welfare by definition of \( p^{SB}_t \). Finally, Assumption 2' implies that \( \tau^{SB}_t > p^{SB}_t \) for some \( t \in [T^{SB}, \infty) \) if \( (A20) \) holds, completing the proof of the proposition.

\[ \text{Proof of Proposition 11 (dynamic geoengineering type I).} \]

\( \text{(Taxes, clean technology).} \) Recall that Type I geoengineering corresponds to an increase in \( \xi \), so that the damages function now equals \( D(S - \xi) \). Differentiating the BGP equilibrium tax

\[ \hat{\tau} = \frac{D'(S(\hat{\tau}, \hat{q}) - \xi)}{\rho + \delta - \alpha \lambda}, \]

with respect to \( \xi \), we obtain

\[ \frac{d\hat{\tau}}{d\xi} = \frac{D''(\cdot)}{\rho - g + \delta} \left[ \frac{\partial S \frac{d\hat{\tau}}{d\xi} - \Lambda \frac{dq}{d\xi} - 1}{\frac{\partial \tau}{d\xi} \frac{d\hat{\tau}}{d\xi}} + \frac{\alpha D'(\cdot)}{(\rho - g + \delta)^2} \frac{d\lambda}{d\xi} \right], \]  \( (A21) \)

Noting that \( d\lambda/d\xi = 0 \) because the supply of scientists is fixed, and \( d\hat{\tau} / d\xi = 0 \), equation \( (A21) \) implies that

\[ \frac{d\hat{q}}{d\xi} = -\frac{\Lambda}{\delta} < 0 \quad \text{and} \quad \frac{d\hat{\tau}}{d\xi} = 0. \]

\( \text{(Welfare).} \) The derivative of flow utility from production with respect to \( \hat{q} \) is

\[ f_c(\hat{k}_c) - \hat{k}_c - (f_d(\hat{k}_d) - \hat{k}_d) - \rho \Gamma \]

which we can write as \( \lambda \Gamma - \Lambda \hat{\tau} \), or equivalently

\[ \frac{\lambda}{\lambda + \rho} (\pi_c - \pi_d) - \Lambda \hat{\tau} \]
since \( \frac{\lambda}{\pi - \rho} (\pi_c - \pi_d) = \lambda \Gamma \) via rearranging the condition that \( \pi_c - \pi_d = (\rho + \lambda) \Gamma \). Total flow environmental damages respond as

\[
\frac{d}{d\xi} D(\hat{S} - \xi) = -D'(\cdot) + \frac{d\hat{S}}{d\xi} D'(\cdot) = 0
\]

using \( d\frac{\hat{S}}{d\xi} = 1 \).

**Proof of Proposition 12 (dynamic geoengineering type II).** Now the total derivative of the stationary farsighted Pigovian tax, \( \hat{\tau} = (1 - v)(\rho - g + \delta)^{-1} D'(\hat{S}) \), equals

\[
\frac{d\hat{\tau}}{dv} = -\frac{D'(\hat{S})}{\rho - g + \delta} + (1 - v) \frac{D''(\hat{S})}{\rho - g + \delta} \left[ \frac{\partial S}{\partial \tau} \cdot \frac{d\hat{\tau}}{dv} - \Lambda \frac{d\hat{q}}{dv} \right] + \frac{\alpha D'(\hat{S})}{(\rho - g + \delta)^2} \frac{d\lambda}{dv}
\]

which, with \( d\hat{\tau}/dv = 0 \) and \( d\lambda/dv = 0 \), implies that

\[
\frac{d\hat{q}}{dv} = -\frac{1}{1 - v} \frac{\delta \hat{S}}{\Lambda \eta}
\]

(A22)

The total effect on the stock of carbon equals

\[
(1 - v) \frac{d\hat{S}}{dv} = \frac{D'(\hat{S})}{D'(\hat{S})} = \frac{\hat{S}}{\eta}
\]

(A23)

so that the welfare-relevant flow term equals

\[
-(1 - v) \frac{d\hat{S}}{dv} D'(\hat{S}) + D(\hat{S}) = D(\hat{S}) - \frac{\hat{S} D'(\hat{S})}{\eta}
\]

Using \( d\hat{\tau}/dv = 0 \) and \( d\hat{q}/dv \) from (A22), and using the fact that flow output shifts as

\[
\left( \frac{\lambda}{\rho + \lambda} (\pi_c - \pi_d) - \Lambda \hat{\tau} \right) \frac{d\hat{q}}{dv},
\]

then

\[
-\Lambda \tau \frac{d\hat{q}}{dv} = \frac{\delta}{\rho - g + \delta} \frac{SD'(S)}{\eta}
\]

which we can combine with the previous expression, dividing by \( \hat{S} D'(\hat{S}) \),

\[
\frac{dW}{dv} < 0 \iff -\frac{\delta}{\eta} \left( (\rho - g + \delta) \frac{\lambda}{\rho + \lambda} (\hat{\pi}_c - \hat{\pi}_d) \frac{1}{\Lambda \hat{\tau}} \right) - \frac{\delta}{\rho - g + \delta} \frac{1}{\eta} + \frac{1}{a} < 0
\]

with \( a \equiv \hat{S} D'(\hat{S}) / D(\hat{S}) > 1 \) as in the static case. This reduces to \( dW/dv < 0 \) if and only if

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

which yields the proposition. ■

49
Proof of Proposition 13 (dynamic geoengineering type III). To obtain $d\hat{S}/d\zeta$, differentiate $(\rho - g + \delta)\hat{\tau} = (1 - \zeta)D'(\hat{S})$, so that

$$
\frac{d\hat{S}}{d\zeta} = \frac{1}{1 - \zeta} \frac{D'(\hat{S})}{D''(\hat{S})}
$$

(A24)

Noting that $\delta \hat{S} = (1 - \zeta)E_t/A_t$ in the stationary state, totally differentiating

$$
\hat{\tau} = (1 - \zeta) \frac{D'(1 - \zeta)E_t/A_t}{\rho - g + \delta}
$$

yields

$$
\frac{d\hat{\tau}}{d\zeta} = - \frac{D'(\hat{S})}{1 - \zeta} \left( \frac{D''(\hat{S})}{\rho - g + \delta} \left[ \frac{(1 - \zeta)\Lambda d\hat{q}}{\delta \hat{S}} + \hat{S} \right] + \frac{\alpha(1 - \zeta)D'(\hat{S}) d\lambda}{(\rho - g + \delta)^2 d\zeta} \right)
$$

which with $d\hat{\tau}/d\zeta = 0$ and $d\lambda/d\zeta = 0$ implies that

$$
\frac{d\hat{q}}{d\zeta} = - \left( 1 + \frac{1}{\eta} \right) \frac{\delta \hat{S}}{(1 - \zeta)^2 \Lambda}
$$

Flow welfare effects will be

$$
\frac{dW}{d\zeta} = \left( \frac{\lambda(\hat{\pi}_c - \hat{\pi}_d)}{\rho + \lambda} - \Lambda \hat{\tau} \right) \frac{d\hat{q}}{d\zeta} - \frac{d}{d\zeta} D(\hat{S}).
$$

Using that $dD/d\zeta = [d\hat{S}/d\zeta]D'$, equation (A24), and the expression for $d\hat{q}/d\zeta$ with $d\hat{\tau}/d\zeta$ fixed to zero, this yields

$$
\frac{dW}{d\zeta} = - \left( \frac{\lambda(\hat{\pi}_c - \hat{\pi}_d)}{\rho + \lambda} \right) \left( 1 + \frac{1}{\eta} \right) \frac{\delta \hat{S}}{(1 - \zeta)^2 \Lambda} + \frac{D'(\hat{S})}{\rho - g + \delta} \left( 1 + \frac{1}{\eta} \right) \frac{\delta \hat{S}}{1 - \zeta} - \frac{\delta D'(\hat{S})}{1 - \zeta} \frac{1}{\eta}.
$$

Dividing everything by $(1 - \zeta)^{-1}\hat{S}D'(\hat{S})$, we obtain

$$
\frac{dW}{d\zeta} < 0 \iff - \left( \frac{\lambda(\hat{\pi}_c - \hat{\pi}_d)}{\rho + \lambda} \right) \left( 1 + \frac{1}{\eta} \right) \frac{\delta(\rho - g + \delta)}{\Lambda \hat{\tau}} + \frac{\delta}{\rho - g + \delta} \left( 1 + \frac{1}{\eta} \right) - \frac{1}{\eta} < 0
$$

and rearranging, we obtain

$$
\frac{dW}{d\zeta} < 0 \iff \eta < \eta^{\text{III}}(\lambda)
$$

for

$$
\eta^{\text{III}}(\lambda) \equiv \left[ \frac{\delta}{\rho - g + \delta} - \frac{\delta(\rho - g + \delta)}{\rho + \lambda} \frac{\hat{\pi}_c - \hat{\pi}_d}{\Lambda \hat{\tau}} \right]^{-1} \left( \frac{\delta}{\rho - g + \delta} \frac{\hat{\pi}_c - \hat{\pi}_d}{\Lambda \hat{\tau}} + \frac{\delta}{\rho - g + \delta} \right)
$$

if its denominator is nonnegative; otherwise, $dW/d\zeta < 0$ always, i.e., $\eta^{\text{III}}(\lambda) \equiv +\infty$. 

50
References


Online Appendix B (Not for Publication)

In this appendix, we extend our economy in four directions. First, we allow the stock of carbon to directly affect the production technology in the static model, as in Nordhaus (2008) and Golosov et al. (2014). Second, we relax the assumption in the static model that environmental damages and producer surplus are additively separable, and instead assume that society obtains welfare from the ratio of consumption to environmental damage, which turns out to be isomorphic to the economy studied in (B1) up to a monotone transformation of the damages function. Third, we allow arbitrary heterogeneity across firms in the static model, modeled as fixed cost of switching. Fourth, we consider nonlinear flow utility over consumption (net of environmental damage) in our dynamic model.

B1 Alternative Specification of Environmental Damages

Abbreviating output (net of the fixed costs of clean technology) by

\[ Y(k, q) = q(f_c(k_c) - k_c - \Gamma) - (1 - q)(f_d(k_d) - k_d), \]

we let welfare equal

\[ [1 - D(S)]Y(k, q), \]  

(B1)

where \( D(\cdot) \) is the same increasing, convex, and twice continuously differentiable damage function as before. Firms of type \( j \in \{c, d\} \) access production technologies

\[ [1 - D(S)](f_j(k) - k) \]

and the fixed cost of switching is \([1 - D(S)]\Gamma\).

Lemma B1 (Pigovian taxation) In economy B.1, optimal taxation satisfies

\[ \hat{\tau} = D'(S)Y(k, q). \]  

(B2)

Proof. Differentiate (B1) with respect to \( k \) to obtain

\[ [1 - D(S)][q(f'_c(k_c) - 1) + (1 - q)(f'_d(k_d) - 1)] - Y(k, q)D'(S)(1 - q + q\gamma) = 0. \]

At a per-unit carbon tax \( \tau \), each firm maximizes its profits \([1 - D(S)][f_j(k) - k] - \tau \gamma j k\) by setting

\[ [1 - D(S)][f'_c(k_c) - 1] = \gamma \tau \quad \text{and} \quad [1 - D(S)][f'_d(k_d) - 1] = \tau \]

so (B2) will implement the Pigovian allocation.  \( \blacksquare \)
Lemma B2 (Equilibrium technology) \(\) If
\[
\frac{\lambda}{1-\lambda} \Gamma < \Lambda \left( \frac{D''(S)}{D'(S)} + D'(S) \right),
\] (B3)
then the interior equilibrium \((\hat{\tau}, \hat{q})\) in economy B.1 is unique.

**Proof.** The proof of Proposition 2 (the interior equilibrium) carries through if \(\hat{\tau}\) and \(\hat{q}\) are still strict strategic substitutes. Indeed, \(\Psi'(<\hat{\tau}\rangle > 0\) as before, so the equilibrium IC is unique. However, there may exist multiple \(q\)'s that satisfy this, since the second term in
\[
\frac{\partial \hat{\tau}}{\partial q} = -\Lambda D''(S) + \frac{\partial Y}{\partial q} D'(S)
\]
can make \(\tau(q)\) not everywhere decreasing in \(q\). However, as
\[
\frac{\partial Y}{\partial q} = \frac{\lambda}{1-\lambda} \Gamma - \Lambda \tau
\]
in the interior, assuming (B3) will guarantee that \(\tau'(q) < 0\).

What happens after geoengineering? When environmental damages affect production directly, geoengineering’s effect on the stock of carbon is still weak, but is no longer exactly zero:

**Proposition B1 (Geoengineering Type I)** \(\) Suppose that Assumptions 1 and 2 hold. In economy B.1, consider a geoengineering technology improvement of type I that increases \(\xi\) by a small amount \(d\xi\). Then we have
- \(d\hat{\tau} = 0\) (there is no effect on the equilibrium carbon tax).
- \(d\hat{q} = -d\xi / \Lambda(\hat{\tau})\) (investment in clean technology declines).
- \(dE = d\xi > 0\) (emissions increase, through lower \(\hat{q}\)).
- \(dD = 0\) (environmental damages remain constant).
- \(dW < 0\) if and only if \(\lambda(\pi_c - \pi_d) > \Lambda \tau\) (welfare may decline).

**Proof.** The incentive-compatible tax satisfies
\[
\hat{\tau} = YD'(S - \xi),
\]
so totally differentiating with respect to \(\xi\) and using that \(d\hat{\tau} / d\xi = 0\) in the interior yields
\[
0 = \frac{dY}{d\xi} D'(\cdot) + \frac{dS}{d\xi} YD''(\cdot) - YD''(\cdot)
\]
or an increase in the stock of carbon (through \( \hat{q} \)) by

\[
\frac{dS}{d\xi} = 1 - \frac{D'(\cdot)}{YD''(\cdot)} \frac{dY}{d\xi}, \tag{B4}
\]

which differs from our benchmark case by the presence of the second term that depends on \( dY/d\xi \).

Welfare, which equals \( (1 - D(\cdot))Y \), will shift as

\[
\frac{dW}{d\xi} = (1 - D(\cdot)) \frac{dY}{d\xi} - Y \frac{dD(\cdot)}{d\xi} + YD'(\cdot).
\]

Noting that \( dD/d\xi = [dS/d\xi] D'(\cdot) \), and using (B4), the direct geoengineering gain \( YD'(\cdot) \) cancels with the first term of \( dS/d\xi \) and the total effect on welfare reduces to

\[
\frac{dW}{d\xi} = \left(1 - D(\cdot) + \frac{(D'(\cdot))^2}{D''(\cdot)}\right) \frac{dY}{d\xi},
\]

and it is apparent that

\[
\frac{dW}{d\xi} > 0 \iff \frac{dY}{d\xi} = (\lambda(\pi_c - \pi_d) - \Lambda \tau) > 0,
\]

which is the same condition in our benchmark economy. ■

The intuition is that exactly as before, the (IC) constraint still pins down \( \hat{\tau} \), so \( d\hat{\tau} = 0 \). To sustain the incentive-compatible \( \hat{\tau} \) after an exogenous removal of \( dv > 0 \) carbon from the initial stock, emissions \( E \) need to increase to sustain the previous level of \( S \), through smaller \( q \). However, with fewer clean firms, the economy incurs fewer fixed costs, increasing output. However, the optimal tax formula (B2) increases in output, this helps to increase the post-geoengineering Pigovian price back to \( \hat{\tau} \). Hence emissions need not respond by “as much,” i.e., \( dE \in (0, d\xi) \).

A type II modification will require a lower \( q \) to increase emissions and sustain the incentive-compatible \( \hat{\tau} \), as exactly as in proposition 4. However, lowering \( q \) will affect total producer surplus, which as in the type I comparative static will affect the tax level through the income effect. Our analogue to Proposition 5 is therefore:

**Proposition B2 (Geoengineering Type II in Economy B1)** Suppose that Assumptions 1 and 2 hold. In economy B1, consider a geoengineering technology improvement of type II with \( v \in [0, 1) \), and let \( \eta = SD''(S)/D'(S) \) be the elasticity of the marginal damage function. 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).
• \( \frac{dq}{d\nu} < 0 \) (investment in clean technology declines).

• \( \frac{dE}{d\nu} > 0 \) (emissions increase, through lower \( \hat{q} \)).

• \( \frac{dD}{d\nu} > 0 \) if and only if \( \eta \leq \eta^* \) (environmental damage increases if the damage function is not too convex), where \( \eta^* \geq 1 \).

• \( \frac{dW}{d\nu} < 0 \) if and only if \( \eta \leq \eta_B^H(\lambda) \) defined below (welfare may decline).

Proof. As before, \( \hat{\tau} / d\nu = 0 \). Differentiating

\[ \hat{\tau} = (1 - \nu)YD'(\cdot), \]

we obtain

\[ 0 = \frac{dY}{d\nu} (1 - \nu)D'(S) - YD'(S) + (1 - \nu)Y \frac{dS}{d\nu} D''(S), \]

from which we conclude that

\[ \frac{dS}{d\nu} = \frac{D'(S)/D''(S)}{1 - \nu} - \frac{D'(S)}{YD''(S)} \frac{dY}{d\nu}. \] 

(B5)

Welfare satisfies

\[ W = [1 - (1 - \nu)D(S)]Y, \]

so

\[ \frac{dW}{d\nu} = \frac{dY}{d\nu} (1 - \nu)(1 - D(S)) + YD(S) - (1 - \nu)Y \frac{dD(S)}{d\nu}, \]

and using that

\[ \frac{dD}{d\nu} = \frac{dS}{d\nu} D'(S) \]

and our formula for \( dS/d\nu \) in (B5), we obtain

\[ \frac{dW}{d\nu} = \left( 1 - (1 - \nu)D + (1 - \nu) \frac{(D'(S))^2}{D''(S)} \right) \frac{dY}{d\nu} + YD(S) - Y \frac{1}{\eta} SD'(S) \]

so that the condition for environmental damages is the same as in the benchmark economy, and the threshold for welfare to decline becomes

\[ \eta < \eta_B^H(\lambda) \equiv \left( a + \frac{1 - (1 - \nu)D(S) dY}{SD'(S)} \frac{dY}{d\nu} \right)^{-1} \left( 1 + (1 - \nu) \frac{dY}{d\nu} \frac{1}{Y} \right) \]

where \( dY/d\nu = [\lambda(\pi_c - \pi_d) - \Lambda\tau][d\hat{q}/d\nu] \), and \( a \equiv SD'(S)/D(S) > 1 \) as before. ■

Proposition B3 (Geoengineering Type III in Economy B1) Suppose that Assumptions 1 and 2 hold. In economy B1, consider a geoengineering technology improvement of type III with \( \zeta \in [0, 1) \). Then we have

• \( \frac{d\hat{\tau}}{d\zeta} = 0 \) (there is no effect on the equilibrium carbon tax).
\[ dS / d\zeta > 0 \] (the total stock of carbon increases).

\[ d\hat{q} / d\zeta < 0 \] (investment in clean technology declines).

\[ dE / d\zeta > 0 \] (emissions increase, through lower \( \hat{q} \)).

\[ dD / d\zeta > 0 \] (damages increase).

\[ dW / d\zeta < 0 \] if and only if \( \eta \leq \eta_B^{III}(\lambda) \) defined below (welfare may decline).

**Proof.** The optimal tax satisfies

\[ \hat{\tau} = (1 - \zeta)YD' ((1 - \delta)S_0 + (1 - \zeta)E) \]

and as \( d\hat{\tau} / d\zeta = 0 \), we obtain

\[ 0 = (1 - \zeta) \frac{dY}{d\zeta} D' - (1 - \zeta)E \cdot YD'' - YD' + (1 - \zeta)Y \frac{dS}{d\zeta} D'' \]

after total differentiation. Thus

\[ \frac{dS}{d\zeta} = E + \frac{1}{1 - \zeta} \frac{D'}{D''} - \frac{dY}{d\zeta} \frac{D'}{YD''} \tag{B6} \]

Welfare shifts as

\[ \frac{dW}{d\zeta} = (1 - D) \frac{dY}{d\zeta} - \frac{dD}{d\zeta} Y \]

and noting that

\[ \frac{dD}{d\zeta} = \frac{dS}{d\zeta} D' = \left( E + \frac{1}{1 - \zeta} \frac{D'}{D''} - \frac{dY}{d\zeta} \frac{D'}{YD''} \right) D' \]

from equation (B6), we obtain

\[ \frac{dW}{d\zeta} = \left( 1 - D + \frac{(D')^2}{D''} \right) \frac{dY}{d\zeta} - \left( E + \frac{1}{1 - \zeta} S \right) YD' \]

translating into the threshold

\[ \eta < \eta_B^{III}(\lambda) = \left( \frac{1 - D}{YSD'} \frac{dY}{d\zeta} - \frac{1}{b} \right)^{-1} \left( \frac{1}{1 - \zeta} - \frac{1}{Y} \frac{dY}{d\zeta} \right) \]

provided that \( b^{-1} \equiv E / S < \frac{1 - D}{YSD'} \frac{dY}{d\zeta} \). If instead \( b^{-1} > \frac{1 - D}{YSD'} \frac{dY}{d\zeta} \), and also that \( \frac{1}{1 - \zeta} - \frac{1}{Y} \frac{dY}{d\zeta} > 0 \),

then \( \eta_B^{III} = +\infty \); that is, welfare always decreases. \( \blacksquare \)
B2 Another Alternative Environmental Damage Specification

We next briefly discuss the robustness of our results to a specification in which environmental damages directly affect consumption preferences. In particular, suppose now there exists a representative household who consumes all of the output in each period,

\[ C = q(f_c(k_c) - k_c - \Gamma) + (1 - q)(f_d(k_d) - k_d) \]  

(B7)

and garners increasing, concave, and differentiable utility

\[ U \left( \frac{C}{D(S)} \right) . \]  

(B8)

Differentiating (B8) with respect to \( k_j \), we obtain the first-order conditions

\[ U'(\cdot)\frac{f'_j(k_j)}{D(S)} - U'(\cdot)\gamma_jD'(S)\frac{C}{[D(S)]^2} = 0 \]

for \( j \in \{c,d\} \), which can be implemented as a decentralized equilibrium with a per-unit-carbon tax of

\[ \hat{\tau} = \frac{D'(S)}{D(S)}C. \]  

(B9)

Observe that if we transform damages into \( D(S) = \exp \hat{D}(S) \), so that \( D'(S)/D(S) = \hat{D}'(S) \), then expression (B9) is exactly the tax as in economy B1. Moreover, since \( U' \) will not alter the sign of \( dW \), all of our results from economy B1 go through, except with the modified damage function. Noting that in particular, the elasticity of \( D(S) = \exp \hat{D}(S) \) satisfies

\[ \eta = SD'(S) + \frac{D''(S)}{D'(S)}S, \]

the exponential transformation increases the curvature of our damages function.

B3 Robustness to Ex Ante Heterogeneity

We now extend our static economy in a third direction. Suppose that firms are differentiated in terms of fixed costs of transitioning to clean technology. In particular, suppose that

\[ \Gamma_i = \Gamma + \omega_i \]

with \( E[\omega_i] = 0 \) and \( H(x) \equiv \mathbb{P}(\omega_i \leq x) \). We call this “economy H.” While before, equilibrium technology adoption was the jump-discontinuous function

\[ \hat{q}(\tau) = \hat{q}1_{\{(1-\lambda)\Psi(\tau) = \Gamma\}} + 1_{\{(1-\lambda)\Psi(\tau) > \Gamma\}}, \]
the effect of heterogeneity is to smooth equilibrium technology,

\[ \hat{q}(\tau) = H \left( (1 - \lambda) \Psi(\tau) - \Gamma \right). \]

The equilibrium \( \hat{\tau} \) is now the fixed point of

\[ \tau = D'( (1 - \delta) S_0 + k_d - \Lambda \cdot H \left( (1 - \lambda) \Psi(\tau) - \Gamma \right) ), \tag{B10} \]

which will be unique if \( H(\cdot) \) is increasing in the neighborhood of the solution to (B10), since \( H(\cdot) \) is always nondecreasing. Define the derivative of \( H \) from the left by \( \hat{h}(\cdot) \).

We can build some intuition for our geoengineering comparative statics by totally differentiating \( \hat{\tau} \) with respect to \( \xi \) (Type I), to obtain

\[ \frac{\partial \hat{\tau}}{\partial \xi} = -\frac{1}{1/D''(S)} \frac{1}{[\partial k_d / \partial \tau] + \Lambda^2 (1 - \lambda) \hat{h}(\cdot)} < 0 \tag{B11} \]

and

\[ \frac{dS}{d\xi} = 1 - \frac{1}{D''(S)} \frac{\partial \hat{\tau}}{\partial \xi}. \tag{B12} \]

When either \( \hat{h}(\cdot) \to +\infty \) (the case of ex ante identical firms) or \( D'' \to 0 \), the RHS of (B11) vanishes: the Pigovian tax is totally invariant to geoengineering. Otherwise, the more concentrated the distribution of firms is at \( (1 - \lambda) \Psi(\tau) - \Gamma \), the closer \( [\partial \hat{\tau} / \partial \xi] \) gets to zero.

Types I–III geoengineering effects go through just as before, subject to a restriction on heterogeneity that we calculate exactly.

**Proposition B4 (Geoengineering Type I with Heterogenous Firms)** Consider economy \( H \). For every \( \epsilon > 0 \), there exists a \( \ell^I \geq 0 \) such that if \( h(\hat{q}) \geq \ell^I \), then

- \( |d\hat{\tau}/d\xi| < \epsilon \).
- \( d\hat{S}/d\xi > 1 - \left( \frac{1}{D'\tau} \right) \epsilon \).
- \( d\hat{q}/d\xi < -1/\Lambda + \left( \frac{1}{\Lambda D'\tau} - \frac{1}{\Lambda} \frac{\partial \hat{E}}{\partial \tau} \right) \epsilon \).
- \( dW/d\xi < 0 \iff \lambda(\hat{\tau}_c - \hat{\tau}_d) > \Lambda \hat{\tau} + O(\epsilon) \).

In particular,

\[ \ell^I(\epsilon) = \frac{1}{(1 - \lambda)^2 \Lambda} \left[ \frac{1}{\epsilon} - \frac{1}{D''} + \frac{\partial k_d}{\partial \tau} \right]. \]

**Proposition B5 (Geoengineering Type II with Heterogenous Firms)** Consider economy \( H \). For every \( \epsilon > 0 \), there exists an \( \ell^{II} \geq 0 \) such that if \( h(\hat{q}) \geq \ell^{II} \) then

- \( |d\hat{\tau}/d\nu| < \epsilon \).
\[ \frac{dS}{dv} > \frac{D'}{(1-v)D''} + \left( \frac{1}{(1-v)D''} \right) \epsilon \]

\[ \frac{dq}{dv} < -\frac{s}{(1-v)\Lambda} \left( \frac{1}{\Lambda} \frac{1}{D''} - \frac{1}{\Lambda} \frac{\partial E}{\partial \tau} \right) \epsilon \]

\[ \frac{dW}{dv} < 0 \iff \eta < \eta^I(\lambda) + O(\epsilon), \text{ where } \eta^I(\lambda) \text{ is defined in Proposition 5.} \]

In particular,

\[ \ell^I(\epsilon) = \frac{1}{(1-v)(1-\Lambda)^2} \left[ \frac{1}{\epsilon} \frac{S}{\eta} - \frac{1}{D''} + \frac{1}{1-v} \frac{\partial k_d}{\partial \tau} \right] . \]

**Proposition B6 (Geoengineering Type III with Heterogenous Firms)** Consider economy \( H \).

For every \( \epsilon > 0 \), there exists an \( \ell^I \geq 0 \) such that if \( h(\hat{q}) \geq \ell^I \), then

\[ \frac{|d\tau}{d\zeta}| < \epsilon \]

\[ \frac{dS}{d\zeta} > \frac{D'}{(1-\zeta)D''} + \left( \frac{1}{(1-\zeta)D''} \right) \epsilon \]

\[ \frac{dq}{d\zeta} < -\frac{s}{(1-\zeta)\Lambda} \left( 1 + \frac{1}{1-\zeta} \frac{1}{\Lambda} \right) + \left( \frac{1}{(1-\zeta)D''} - \frac{1}{\Lambda} \frac{\partial E}{\partial \tau} \right) \epsilon \]

\[ \frac{dW}{d\zeta} < 0 \iff \eta < \eta^I(\lambda) + O(\epsilon), \text{ where } \eta^I(\lambda) \text{ is defined in Proposition 6.} \]

In particular,

\[ \ell^I(\epsilon) = \frac{1}{(1-\zeta)^2(1-\Lambda)^2} \left[ \frac{1}{\epsilon} \frac{S}{\eta} + \frac{1}{1-\zeta} \frac{E}{\epsilon} - \frac{1}{D''} + \frac{1}{1-\zeta} \frac{\partial k_d}{\partial \tau} \right] . \]

### B4 Concave Preferences

In the text, we focused on linear preferences, even in the dynamic model, which greatly simplified the analysis. We now show that our results generalize when consumers have a concave utility function, so that their dynamic preferences are given at each \( t \) by

\[ \int_t^\infty U(\frac{C_s - A_s D(S_s; \zeta, v)}{\epsilon}) e^{-\rho(s-t)} ds, \]

where the utility function \( U(\cdot) \) is increasing, twice continuously differentiable, and concave. The specification where damages and consumption are additive arguments of \( U(\cdot) \) is similar to Greenwood et al. (1988). Our results can also be extended to different formulations, but those introduce additional income effects, further complicating the relevant conditions. In addition, to sustain a BGP, we also assume a constant elasticity of marginal consumption (relative risk aversion), \( \theta = -xU''(x)/U'(x) \), so that \( U \) can be represented by

\[ U(x) = \frac{\chi^{1-\theta}}{1-\theta} \]

B-8
for $\theta \neq 1$ and the limit $U(x) = \log(x)$ if $\theta = 1$ (and $\theta = 0$ recovers the linear specification of the main text.)

The household thus maximizes (B4)

$$
\frac{1}{1 - \theta} \int_t^\infty (\bar{C}_s)^{1-\theta} e^{-\rho(s-t)} ds,
$$

at each time $t \geq 0$, where we define net consumption as $\bar{C}_s = C_s - A_s D(S_s; \xi, \upsilon)$.

**B4.1 Household Optimization**

The Ramsey equation (derived from household intertemporal optimization with endogenous savings) yields a market interest rate of

$$
r_t = \rho + \theta g_t \tag{B13}
$$

where $g_t \equiv \dot{\bar{C}}_t / \bar{C}_t$ denotes the growth rate of average net consumption, defined as $\bar{C}_t = C_t - A_t D(S_t; \xi, \upsilon)$. We derive (B13) in Appendix B4.7.

**B4.2 Firms**

Now that the interest rate differs in general from $\rho$, we must define firms’ value functions slightly differently. Equation (19) for the value of a clean incumbent with quality $n$ at time $t$ becomes

$$
V_{ct}(n) = \mathbb{E}_t \left[ A^n \int_t^\infty \pi_c(\tau_s) e^{-r_s(s-t)} \mathbf{1}_{\{\text{extant at } s\}} ds \right]
= A^n \int_t^\infty \pi_c(\tau_s) e^{-(r_s + \lambda)(s-t)} ds, \tag{B14}
$$

Likewise, the expected net present discounted value at $t$ for a $d$-type firm with quality $n$ becomes

$$
V_{dt}(n) = \sup_{\nu \geq 0} \left[ [V_{ct+v}(n) - A^n \Gamma] e^{-\int_t^{t+\nu} (r_s + \lambda) ds} + A^n \int_t^{t+\nu} \pi_d(\tau_s) e^{-(r_s + \lambda)(s-t)} ds \right] \tag{B15}
$$

since they can switch at any time $t + \nu \geq t$. With these new value functions, and their normalized counterparts $v_{ct}$ and $v_{dt}$, R&D, input and clean technology switching decisions all go through just as in the benchmark case.

**B4.3 Optimal Taxation**

The planner maximizes household utility, (B4). The Hamiltonian which generalizes (23) of the main text is

$$
H_t(k_t, S_t) = U(\bar{C}_t) - \mu_t [E_t - \delta S_t].
$$
Optimal input decisions $k_{ct}, k_{dt}$ satisfy $\frac{\partial H_t}{\partial k_{ct}} = \frac{\partial H_t}{\partial k_{dt}} = 0$, or

$$A_tU'(\tilde{C}_t) \left[ f'_d(k_{dt}) - 1 \right] = \mu_t \quad \text{and} \quad A_tU'(\tilde{C}_t) \left[ f'_c(k_{ct}) - 1 \right] = \gamma \mu_t,$$

which coincide with private firm input decisions when the planner levies a per-unit emissions tax of $\mu_t / (A_tU'(\tilde{C}_t))$. Consequently, we use the normalization

$$p_t = \frac{\mu_t}{A_tU'(\tilde{C}_t)},$$

which differs from that in the main text only by the presence of the non-constant marginal utilities. Along the equilibrium path, the shadow price of the stock of carbon $\mu_t$ will satisfy the Euler-Lagrange condition $\frac{\partial H_t}{\partial S} = \dot{\mu}_t - \rho \mu_t$. The Pigovian shadow cost of carbon emissions is identical to that of the main text, except with an endogenous interest rate $(r_t)_{t \geq 0}$. However, we can weaken Assumption 3 of the main text to the following:

**Assumption 3' (Growth with concave preferences)** $(1 - \theta)\alpha \varphi Z \leq \rho + \delta$.

Note that Assumption 3' always holds if $\theta \geq 1$, regardless of the other parameter values. Consequently,

**Lemma B3 (Shadow cost of carbon with concave utility)** Suppose Assumption 3' holds. Then, along any optimal path,

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

for all $t \geq 0$, where $(r_s)_{s \geq 0}$ denotes the Ramsey interest rate given by (B13).

**Proof.** See Appendix B4.7. ■

**B4.4 Dynamic Equilibrium**

We define an equilibrium in the concave economy exactly as in the text in section 3.5, except appended with the natural additional condition that

- Given each of the other equilibrium objects, the interest rate $r^*_t$ satisfies (B13).

Moreover, to be precise, we need to rewrite our assumption for interiority such that it takes into account the dependence of farsighted Pigovian taxes on the interest rate.

**Assumption 2” (Conditions for Dynamic Interior Equilibrium)** For all $t \geq 0$,

$$\Gamma \in \left( \int_t^\infty \Psi(\tau_s) e^{-(r_s+\lambda)(s-t)} ds, \int_t^\infty \Psi(\tau_s) e^{-(r_s+\lambda)(s-t)} ds \right)$$
where
\[
\tau_t = \int_t^\infty D' \left( S_0 e^{-\delta s} + \int_0^s k_d(\tau_v) e^{-\delta(s-v)} dv \right) e^{-(\delta + r_s - \alpha \lambda)(s-t)} ds
\]
and
\[
\bar{\tau}_t = \int_t^\infty D' \left( S_0 e^{-\delta s} + \int_0^t k_d(\tau_v) e^{-\delta(t-v)} dv + \gamma \int_t^s k_c(\tau_v) e^{-\delta(s-v)} dv \right) e^{-(\delta + r_s - \alpha \lambda)(s-t)} ds,
\]
and \((r_s)_{s \geq 0}\) satisfies (B13).

**B4.5 BGP**

Just as before, if the stock of carbon converges to some \(\hat{S}\), (30) will hold just as before for all \(t\) hence. Hence growth reduces to as before
\[
g = (A - 1) \lambda
\]
and the interest rate (B13) simplifies to \(\hat{r} = \rho + \theta \alpha \lambda\). The BGP farsighted Pigovian condition (31) becomes
\[
\hat{\tau} = \frac{D'(\hat{S})}{\delta + \rho - (1 - \theta) \alpha \lambda}, \quad (B17)
\]
equation (21) becomes
\[
\frac{\Psi(\hat{\tau})}{\rho + (1 + \theta \alpha) \lambda} = \Gamma, \quad (B18)
\]
and we obtain an analogue to Proposition 8.

**Proposition B7 (Existence, uniqueness of the BGP)** Suppose Assumptions 1', 2'', and 3' hold. Then there exists a unique interior solution \((\hat{S}, \hat{q}, \hat{\tau})\) to (30), (B17), and (B18).

**Proof.** Omitted. ■

**B4.6 Geoengineering with Concave Preferences**

We conclude by noting that the versions of the dynamic geoengineering propositions in the main text, augmented to account for the endogenous interest rate dynamics, are substantively identical.

**B4.7 Proofs omitted above**

**Proof of the Ramsey equation (B13).** Suppose that our household can save at a rate \(r_t\), with savings \(a_t\), and costate \(s(t)\). Then
\[
\dot{a}_t = r_t a_t - A_t c_t + \int_0^1 A^n_i (f(k_{ii}) - k_{ii}) di,
\]
where recall $c_t = C_t/A_t$. Denoting normalized net consumption by $\tilde{c}_t = c_t - D(S_t)$, the household’s Hamiltonian may be written as

$$H = \frac{1}{1-\theta}A_t^{-\theta}(\tilde{c}_t)^{1-\theta} + s(t) \left[ r_t\alpha_t - A_t\tilde{c}_t + \int_0^1 A^n(f(k_{it}) - k_{it})di \right]$$

with discount rate $\rho$, or rather, since

$$A_t^{1-\theta} = \exp\{(1-\theta)(A-1)t\lambda\},$$

we can write the transformed

$$\tilde{H} = \frac{\tilde{c}_t^{1-\theta}}{1-\theta} + s(t) \left[ r_t\alpha_t - A_t\tilde{c}_t + \int_0^1 A^n(f(k_{it}) - k_{it})di \right]$$

with a transformed discount rate $\tilde{\rho} \equiv \rho + (1 - A)\lambda$. The first-order savings conditions of $\tilde{H}$ are

$$H_{\tilde{c}_t} = \tilde{c}_t^{-\theta} - A_t\tilde{s}(t) = 0 \quad \text{(B19)}$$

$$\tilde{s}(t) = -H_{\alpha_t} + \tilde{\rho}\tilde{s}(t). \quad \text{(B20)}$$

Differentiating (B19) with respect to time, we obtain

$$-\theta(\tilde{c}_t)^{-1-\theta}\dot{\tilde{c}}_t = A_t\dot{s}(t) + (A - 1)\lambda A_t\tilde{s}(t)$$

which we divide by $A_t\tilde{s}(t)$ to obtain via (B19)

$$\frac{\dot{s}(t)}{s(t)} = -\theta \frac{\dot{\tilde{c}}_t}{\tilde{c}_t} - (A - 1)\lambda$$

so that, using $H_{\alpha_t} = r_t$, the definition of $\tilde{\rho}$, and the differential equation (B20), we obtain

$$-\theta \frac{\dot{\tilde{c}}_t}{\tilde{c}_t} - (A - 1)\lambda = -r_t + \rho + (1 - \theta)(1 - A)\lambda$$

hence

$$r_t = \rho + \theta \frac{\dot{\tilde{c}}_t}{\tilde{c}_t} + \theta\lambda \quad \text{(B21)}$$

is the equilibrium interest rate. Noticing that

$$\frac{\dot{c}_t}{c_t} = \frac{\dot{C}_t}{C_t} - (A - 1)\lambda$$

yields equation (B13).}

**Proof of Lemma B3 (the shadow cost of carbon in GHH).** In equilibrium, we can write the necessary Euler-Lagrange condition $\partial H_t/\partial S = \dot{\mu}_t - \rho\mu_t$ as

$$\dot{\mu}_t = -A_t U'(\tilde{C}_t)D'(S_t) + (\rho + \delta)\mu_t.$$
Dividing this equation as before by $A_tU'(\tilde{C}_t)$, we obtain

$$\frac{\dot{\mu}_t}{A_tU'(\tilde{C}_t)} = -D'(S_t) + (\rho + \delta)p_t.$$  \hfill (B22)

Recall our change-of-variables

$$p_t = \frac{\mu_t}{A_tU'(\tilde{C}_t)}$$

and note in particular that it satisfies

$$\dot{p}_t = \frac{\dot{\mu}_t}{A_tU'(\tilde{C}_t)} - \frac{\mu_t}{(A_tU'(\tilde{C}_t))^2} \frac{d}{dt} [A_tU'(\tilde{C}_t)]$$

$$= \frac{\dot{\mu}_t}{A_tU'(\tilde{C}_t)} - p_t \left( \theta \frac{\dot{\tilde{c}}_t}{\tilde{c}_t} + (1 - A)(1 - \theta)\theta \right)$$

using

$$\frac{d}{dt} [A_tU'(\tilde{C}_t)] = \frac{d}{dt} A_t^{1 - \theta} \frac{d}{dt} \tilde{c}_t^{-\theta} = \frac{d}{dt} A_t^{1 - \theta} + \frac{d}{dt} \tilde{c}_t^{-\theta}.$$  

and $A_t = \exp\{t\lambda(A - 1)\}$. With (B21), equation (B22) becomes

$$\dot{p}_t = -D'(S_t) + (\delta + r_t - a\lambda) p_t,$$

which with $p_0$ determined by transversality gives (B16).  \hfill ■