Grantham Research Institute on Climate Change and the Environment

Centre for Climate Change Economics and Policy

To build or not to build? Capital stocks and climate policy

Elizabeth Baldwin, Yongyang Cai and Karlygash Kuralbayeva

January 2018

Centre for Climate Change Economics and Policy Working Paper No. 325
ISSN 2515-5709 (Online)

Grantham Research Institute on Climate Change and the Environment Working Paper No. 290
ISSN 2515-5717 (Online)

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

The Grantham Foundation for the Protection of the Environment

University of Leeds

LSE The London School of Economics and Political Science

The Centre for Climate Change Economics and Policy (CCCEP) was established by the University of Leeds and the London School of Economics and Political Science in 2008 to advance public and private action on climate change through innovative, rigorous research. The Centre is funded by the UK Economic and Social Research Council. Its second phase started in 2013 and there are five integrated research themes:

1. Understanding green growth and climate-compatible development
2. Advancing climate finance and investment
3. Evaluating the performance of climate policies
4. Managing climate risks and uncertainties and strengthening climate services
5. Enabling rapid transitions in mitigation and adaptation

More information about the Centre for Climate Change Economics and Policy can be found at: www.cccep.ac.uk.

The Grantham Research Institute on Climate Change and the Environment was established by the London School of Economics and Political Science in 2008 to bring together international expertise on economics, finance, geography, the environment, international development and political economy to create a world-leading centre for policy-relevant research and training. The Institute is funded by the Grantham Foundation for the Protection of the Environment and the Global Green Growth Institute. It has six research themes:

1. Sustainable development
2. Finance, investment and insurance
3. Changing behaviours
4. Growth and innovation
5. Policy design and evaluation
6. Governance and legislation

More information about the Grantham Research Institute on Climate Change and the Environment can be found at: www.lse.ac.uk/grantham.

This working paper is intended to stimulate discussion within the research community and among users of research, and its content may have been submitted for publication in academic journals. It has been reviewed by at least one internal referee before publication. The views expressed in this paper represent those of the authors and do not necessarily represent those of the host institutions or funders.

To Build or Not to Build? Capital Stocks and Climate Policy*

Elizabeth Baldwin† Yongyang Cai‡ Karlygash Kuralbayeva§

January 29, 2018

Abstract

We investigate how irreversibility in "dirty" and "clean" capital stocks affects optimal climate policy, from both theoretical and numerical perspectives. An increasing carbon tax will reduce investments in assets that pollute, and so reduce emissions in the short term: our "irreversibility effect". As such the "Green Paradox" has a converse if we focus on demand side capital stock effects. We also show that the optimal subsidy increases with the deployment rate: our "acceleration effect". Considering second-best settings, we show that, although carbon taxes achieve stringent targets more efficiently, in fact renewable subsidies deliver higher welfare when policy is more mild.

Keywords: Infrastructure; Clean and Dirty Energy Inputs; Renewable Energy; Stranded Assets; Carbon Budget; Climate Change Policies; Green Paradox

JEL codes: O44, Q54, Q58

*We thank Lassi Ahlvik, Alex Bowen, Maria Carvalho, Simon Dietz, Carolyn Fischer, Roger Fouquet, Reyer Gerlagh, Kenneth Gillingham, Niko Jaakkola, Per Krusell, Linus Mattauch, Armon Rezai, Daniel Spiro, Rick van der Ploeg, Till Requate, Frank Venmans, and participants at: the Sustainable Development workshop in Rimini (2017); GTAP 2017 (Purdue); the pre-EAERE workshop on Stranded Assets and Climate Policy; EAERE 2017 (Athens); OxCarre; LSE; BEEER 2017 (Bergen); INFORMS 2017 (Houston); CESifo annual Energy & Climate Economics conference; for helpful comments and suggestions. When this project started, Baldwin and Kuralbayeva were employed by the Grantham Research Institute of the London School of Economics and were supported by the UK's Economic and Social Research Council (ESRC), and the Grantham Foundation for the Protection of the Environment, and Cai was employed by the Becker Friedman Institute of the University of Chicago and a visiting fellow at Hoover Institution at Stanford University. Cai acknowledges support from the National Science Foundation (SES-0951576 and SES-1463644) under the auspices of the RDCEP project at the University of Chicago. Kuralbayeva also acknowledges support from Statoil via the Statoil Chair in Economics at NHH. The authors have no other relevant or material financial interests that relate to the research described in this paper.

†Department of Economics and Hertford College, Oxford University, UK; elizabeth.baldwin@economics.ox.ac.uk

‡Department of Agricultural, Environmental and Development Economics, The Ohio State University, USA; cai.619@osu.edu

§Department of Geography and Environment, London School of Economics and Political Science, UK; k.z.kuralbayeva@lse.ac.uk

1 Introduction

Irreversibility is an important feature of investment decisions. Many productive facilities are firm- or industry-specific and disinvestment is costly, if at all possible. Once capital is installed, it has little or no value unless used in production. It is obvious that with inability to disinvest, firms face more constrained conditions compared with firms undertaking reversible investments. This has important implications for investment decisions even without uncertainty[1]. The outstanding example of such a situation is provided by investment into fossil fuel-fired power plants. The world continues to make big investments into their construction, particularly coal plants: estimates suggest almost 1 trillion US dollars of such investments are planned (Shearer et al. 2016). Given the long lifetimes of fossil fuel based power plants, the emissions embodied in this infrastructure potentially undermine more stringent long-term climate objectives, such as the 2°C target (see Pfeiffer et al. 2016). As such, a fast coal phase-out strategy is considered as one of the necessary conditions to achieve a transformation in line with the Paris Agreement. Some countries (e.g. the UK, Finland, France) have significantly reduced their power production from coal in recent years and announced phasing out coal completely in the coming 10-15 years. In addition, production of electricity from renewable sources has become more competitive, expanding dramatically, primarily due to the decline in costs driven by economies of scale. These considerations prompt three natural questions. When is the optimal time to stop investment into fossil fuel based power plants when investments are irreversible? How much should we invest into the clean energy sector? And, which policy instrument (carbon tax or subsidy) is more efficient in terms of maximizing social welfare when only one instrument is available (second-best setting)?

In this paper, we study these questions both theoretically and numerically. Our analysis is in two complementary parts. First, we explore the properties of irreversible investment decisions (Arrow 1968, Arrow and Kurz 1970, Greenwood et al. 1997) in a simplified model. As well as serving intuition, the model presents messages that are of a general nature. They characterize optimal irreversible investment decisions when it is known that returns on this capital are due to fall. And similarly, we explore investments, returns and optimal subsidies when the price of investments undergoes learning-by-doing (Wright 1936, Arrow 1962). We then quantify the importance of irreversibility and learning-by-doing in a dynamic general equilibrium climate-economy model. This is based on DICE (Nordhaus 2014a) but deviates in two important ways. Firstly, the energy sector is modeled explicitly, incorporating both irreversibilities in a "dirty" sector and learning-by-doing in a "clean" sector. And secondly, as well as copying the damage function of Nordhaus (2014a), we also consider scenarios in which global temperature changes do not exceed 2°C. This stringent target makes both irreversibilities and learning-by-doing more important; it is better in line with current international aspirations. Given the two externalities present in our model (global warming and learning-by-doing), we consider cases in which both carbon tax and subsidy instruments (the first-best setting) or only one of the two instruments (the second-best) are available.

The four main findings of the paper are as follows. First, we establish a theoretical result on the relationship between investment in dirty capital stock and climate policies, which we call the "irreversibility effect": if dirty capital cannot be converted to other capital, then it is optimal to stop investing into dirty capital earlier (as compared with a case in which investment is reversible). Irreversibility in investment implies an earlier shift to investment into the clean sector, to avoid later stranding of assets in the dirty energy sector. It therefore reduces emissions in the short term. We thus demonstrate that irreversibility effects on the demand side *enhance* the effects of carbon tax in the short-term, and so reduce emissions in the short-term. This is in contrast with

[1] Irreversibility of investment features prominently in the modern theory of firm-level investment under uncertainty, e.g., Abel (1983), Pindyck (1991), Dixit (1992).

the standard Green Paradox (GP) effect,[2] which focuses on the *suppliers* of a fossil fuel resource and shows that the knowledge of an increasing carbon tax will increase extraction of fossil fuel and will thus counteract the effects of the carbon tax in the short-term. Moreover, at the time at which we stop investing into dirty fossil fuel infrastructure, returns on its existing stocks go above those of the general economy. From the perspective of an investor, this makes perfect sense. In the long-term, returns on this investment will fall, and thus the current investments are only attractive when short-term additional gains are sufficient to compensate for future losses.[3]

Second, we provide a simple expression for the optimal subsidy on technologies whose price evolves via "learning-by-doing". This subsidy depends on depreciation and the learning rate, and also on the rate of deployment of the technology: under normal parameterizations, the subsidy increases with the rate of deployment. We call this the "*acceleration effect*" for technology policy. Thus, if, for example, a carbon tax restricts investment in the dirty sector and enhances future deployment of clean technology, this implies an *additional* case for a greater subsidy in the short term. So the importance of learning-by-doing is accentuated by the early withdrawal from the dirty energy sector.

Third, quantitative results support our theoretical findings and illustrate that the net (of depreciation) rate of returns on dirty capital infrastructure with irreversible investments follows an unusual trajectory: initially matching the returns in the general economy, we see that it rises above the returns in the general economy when we stop investing in dirty capital and remains above it for some period of time; within this period and for some time thereafter, investment will be equal to zero, although the dirty capital is not underutilized. However, net returns on dirty capital will fall eventually, reaching zero once the capital is indeed underutilized. Quantitative results illustrate that the timing of these effects depends on the climate policy target: the irreversibility effect kicks in only if policy objectives are stringent enough.

Finally, we quantitatively explore which instrument - carbon tax or subsidies - under the second-best setting yields lowest welfare loss compared with the first-best situation. We show that under less ambitious climate policy, the economy is better off with the subsidy policy, while carbon pricing induces lower welfare loss compared with the subsidy policy if climate policies are more ambitious.

These results further relate to the literature in the following ways. On the theory side, the irreversibility and the acceleration effects are novel, to the best of our knowledge, and are related to two branches of the literature. First, our theoretical result linking investment irreversibility and an earlier end to investment in polluting infrastructure is closest in spirit to findings of Arrow (1968), who was the first one to study investment irreversibility in a deterministic setting. He showed that optimal irreversible investment is characterized by alternating periods of positive gross investment and zero gross investment.[4] In relation to these studies, we develop a stylized model to explicitly demonstrate this effect and related pattern in rates of return on irreversible investment and apply the results to the case of a polluting industry. Second, there is an extensive literature[5] that has explored if the Green Paradox effect remains robust by considering various extensions of the typical

[2]See e.g., Sinn (2008), Jensen et al. (2015), Sinn (2015).

[3]Such an extra premium on irreversible investment even without uncertainty is also called the commitment premium, see e.g., Bernstein and Mamuneas (2007).

[4]Similar result, but within the Ramsey model of optimal capital accumulation, was obtained by Arrow and Kurz (1970), who show that it is possible to have as an optimal solution practically any number of alternating intervals in which the nonnegativity constraint (on investment) is binding or not. As such solutions, as they conclude, become in essence a computational problem.

[5]For instance Gerlagh (2011) focuses on strong and weak GP and explores if increasing fossil extraction costs counteracts the (strong) GP, while imperfect energy substitutes may make the weak and the strong GP vanish. Michielsen (2014) investigates how the existence of a virtually non-exhaustible resource like coal can work against the GP mechanism. See also, e.g., van der Ploeg (2013), van der Ploeg and Withagen (2014).

resource model that underlies the GP. The irreversibility effect, even though complements other mechanisms against the GP discussed in the literature, is conceptually different as it focuses on the *demand* side of a fossil fuel resource.

On the quantitative side, the results relate to other two strands of research. On the one hand, there is an extensive literature that investigates relative merits of carbon tax and renewable subsidies to address climate change.[6] However, these studies generally abstract from consideration of different climate targets under second-best settings with irreversible investment decisions.[7] On the other hand, a rich and growing literature has developed integrated assessment models to study a number of different climate change issues. Papers assessing future emissions from the energy sector include Pfeiffer et al. (2016) and Davis et al. (2010). However, these are not dynamically optimizing frameworks, as in the economics literature. Other climate-economy models generally ignore the interplay between irreversible investment decisions, inertia in energy systems, and climate policies, on which this paper focuses.[8]

Finally, our paper belongs to the literature on path dependence and climate change.[9] We contribute to this literature by analyzing the implications of path dependence embodied in carbon-intensive infrastructure for the design of optimal climate change policies.

In terms of broader implications of the results, our paper speaks to the debate on characteristics of optimal policy to combat climate change. Some advocate a "gradual slope" in policy because economic growth implies that the current generation is poor relative to the future, and so should not bear the costs of emission reductions. Moreover, doing so reduces pressure for premature retirement of the existing dirty capital stock, and it provides valuable time to develop low-cost, low-carbon-emitting technology.[10] Others counter this line of reasoning by arguing that an effective way to reduce abatement costs is to accelerate learning-by-doing.[11] We find that early investment in the renewable sector is crucial, and not only to accelerate the decline in the costs of clean energy but also to prevent later stranding of assets using fossil fuel. Our quantitative results within the second-best setting emphasize the importance of adopting carbon pricing - an instrument that can facilitate a rapid decarbonization of the global power sector under ambitious climate policy target as set under the Paris Agreement. However, considering the past 10-20 years, relatively unambitious

[6]For instance the literature has argued that one of the advantages of using carbon pricing is that it can help to minimize the cost of pollution control. Fischer and Newell (2008) show that reliance on non-price policy instruments often leads to higher abatement costs. In a more recent study, Fischer et al. (2017) show that even with multiple market failures, pricing policy remains the most cost-effective option for reducing emissions. See also Gerlagh and van der Zwaan (2006), who use a top-down energy-economy model to compare five instruments, including carbon taxes and renewable subsidies, in terms of costs, efficiency and their impacts on the composition of the energy supply systems. See, e.g., also Baranzini et al. (2017) and references therein.

[7]A burgeoning theoretical literature investigates the forms of policy interventions in second-best settings. Examples include an analysis of optimal carbon taxation as part of distortionary fiscal policy (Barrage, 2014); policy intervention via carbon taxes and research subsidies as well as alternative policies to encourage the transition to a green economy (Acemoglu et al., 2016); analysis of how carbon taxes combined with green alternatives can increase fossil fuels abandonment (Rezai and van der Ploeg, 2016). None of these studies, however, has analyzed optimal policy interventions when investment decisions are irreversible.

[8]To the best of our knowledge, the only exception is Rozenberg et al. (2014), who however do not find the effects pertaining to the irreversibility and learning-by-doing as we do in this paper.

[9]e.g., Fouquet 2016; Aghion et al. 2014, and Aghion et al. 2016. The papers relevant to our analysis are Grubb et al. (1995), Wigley et al. (1996), Grubler and Messner (1998), Goulder and Mathai (2000), Vogt-Schilb et al. (2012) and Rozenberg et al. (2014).

[10]W. Nordhaus was one of those in the past who recommended delay, but he recently argued that a target with a limit of 2°C "appears to be unfeasible with reasonably accessible technologies" (Nordhaus 2016). Wigley et al. (1996), e.g., argue that the cost-effective emissions pathway is one that departs only gradually from the emissions baseline.

[11]Still, some authors find that leaning-by-doing has an ambiguous impact on the timing of emissions abatement (Tol 1999, Goulder and Mathai 2000).

policy has manifested in large part through subsidy on renewables; if that level of policy had been optimal, that choice of one instrument may well have been an excellent second best.

Finally, our paper speaks to the debate on stranded assets and climate policy. The literature so far has dominated by the studies that estimate the amount of existing fossil fuel reserves that would be required to remain in the ground to limit climate change to less than two degrees of warming. For instance, according to McGlade and Ekins (2015), an estimated third of oil reserves, half of gas reserves and more than 80% of known coal reserves are referred to as "stranded". As we show, the economics is different when one considers stranding of assets that use the fuel. Moreover, the question is not only about the shift away from fossil fuel energy and related physical capital, but also about how smoothly it is done. Our model is a perfect foresight model and investors form rational expectations. The model suggests that energy assets, such as coal fired power stations, become stranded or underutilized early as a rational response to stringent climate policy target. As such, a credible political signal of stringent climate policy is required, to avoid investments that will in fact be unprofitable. This in turn will lead to earlier reductions in emissions. The Paris Agreement could be such a signal, prompting investors to behave rationally and helping politicians to buy time, as they start implementing the policies which will be required to implement the necessary reductions in emissions.

The rest of the paper is organized as follows. In the next section we present a simple analytical model in which we characterize optimal irreversible investment decisions when we anticipate that returns on those investments will fall in the future. In Section 3 we consider a simple model of investment with learning by doing. Section 4 describes how we set up the full dynamic general equilibrium climate-economy model to quantify the theoretical results. Section 5 sets out the results from the simulations of the climate-economy model. The final section provides a discussion and some concluding comments. Details on the calibration, and proofs of technical results, are provided in Appendices.

2 A Simple Model of Irreversible Investments

First we model key features of the economy in isolation, in order to clearly present the theoretical results. The models analyzed here will be embedded in our full structure in Section 4. We consider the implications of irreversibility in investments in capital stocks whose economic productivity will decline (cf. Arrow 1968, Arrow and Kurz 1970). Our key example is investment in a fossil fuel using power stations, but many other illustrations can be found. For example, the design of cities may lock in high energy usage, in ways that are difficult and expensive to reverse; this motivates earlier sustainable design (cf. e.g. Hoornweg and Freire 2013).

2.1 The Household's problem

Consider a representative household, which holds k_t of a certain capital asset and can make an additional irreversible investment of i_t in each period t. The asset offers a period-t return of r_t and depreciates at rate δ. There are other opportunities for investment and other sources of income, written net as o_t, and the household's per-period consumption is c_t, so their budget constraint is $i_t + c_t = r_t k_t + o_t$ where $i_t = k_{t+1} - (1-\delta)k_t$ and $i_t \geq 0$.

Write the standard ratio from the Euler equation as $e_{t+1} := \frac{u'(c_t)}{\beta u'(c_{t+1})} - 1$ where u is a utility function and β is the utility discount factor. Make the minor assumptions that there exist $\epsilon > 0$ and $R \gg 0$ with $-\delta + \epsilon < e_t < R$ for all t, that is, e_t is bounded and bounded away from minus depreciation, $-\delta$.

In the following, we analyze this model (proofs are provided in Appendix A). First:

Proposition 2.1. *For any $s_0, s_1 \in \mathbb{Z}_+$, investment $i_t > 0$ holds for all $t \in \{s_0, \ldots, s_1\}$ only if $r_t - \delta = e_t$ for $t \in \{s_0 + 1, \ldots, s_1\}$.*

To understand this intuitively, suppose we have an asset, "general capital", in which there is non-zero investment in every period, and whose rate of return r_t^g may be treated as exogenous. Then $r_t^g - \delta = e_t$ for all $t \geq 1$ (the Euler equation). So for non-zero investment in two assets over a time period, their net returns must match.

We make the obvious point of Proposition 2.1 to contrast with the following, more interesting case. Suppose the net return from the asset drops *below* e_t at some time t: changing economic conditions mean that this capital asset is no longer as productive as it was. Then we stop investing at an *earlier* time, and reap *excess* returns for some of the intervening period. Write $\Delta_{t,s} = \prod_{s'=1}^{s} \frac{1}{1+e_{t+s'}}$ for the compound consumption discount factor. Then

Proposition 2.2. *Suppose that $i_0 > 0$, and that $r_t - \delta < e_t$ for $t \in \{s_1, \ldots, s_2\}$. Then there exists $s_0 \leq s_1 - 1$ such that $r_{s_0} - \delta > e_{s_0}$ and such that $i_t = 0$ for $t \in \{s_0, \ldots, s_2 - 1\}$. Moreover, then*

$$\sum_{s=1}^{s_1-1} (1-\delta)^{s-1} \Delta_{0,s}((r_s - \delta) - e_s) \geq \sum_{s=s_1}^{s_2} (1-\delta)^{s-1} \Delta_{0,s}(e_s - (r_s - \delta)) \tag{1}$$

Thus the net returns $r_t - \delta$ from this asset follow an unusual trajectory: initially matching the consumption discount rate path e_t, or the net returns from a "general capital" stock, we see that $r_t - \delta$ rises above e_t at some point before it falls beneath. Investment is zero while returns follow this pattern. (We illustrate this pattern in our simulations in Section 5: see Figure 2).

It is mathematically possible that the minimal s_0 found by Proposition 2.2 satisfies $s_0 = s_1 - 1$: investment merely ends one period before the net economic return drops below the general level in the economy. However, because equation (1) must hold, such a solution would be surprising when the time step (and so the depreciation rate) are moderately small. Recall from Proposition 2.1 that while investment is ongoing, the net return on the asset must match e_t. So the only non-zero term on the left hand size of equation (1) would be for $s = s_1 - 1$. The excess of $r_{s_1-1} - \delta$ above e_{s_1-1}, in that period alone, would have to be great enough to compensate for the entire (though discounted and depreciated) sequence of periods $s \in \{s_1, \ldots, s_2\}$ in which $r_s - \delta < e_s$. If there is only moderate change over time in both r_s and e_s, and if depreciation and discounting are not overwhelming, then returns will have to rise, and investment will have to cease, at some earlier point in time.

From the perspective of an investor, these short-term excess returns make perfect sense. If the investor knows that, in the long-term, returns on this infrastructure will fall, then it is not an attractive investment. However, the prospect short-term additional gains will compensate for long-term losses. These short-terms gains will indeed be realized if all other investors are similarly ending investment early.[12]

Both Propositions 2.1 and 2.2 follow straightforwardly from a technical lemma (Lemma A.1) presenting the shadow price on the irreversibility constraint, $i_t \geq 0$ as the net present value of investment in this asset, relative to the opportunity cost. See Appendix A.

Now we consider what this means for the quantity of total holdings of this capital stock. If there are L_0 households in the economy, each of size $\frac{L_t}{L_0}$, then this behavior simply scales up. We use capital letters to denote total capital K_t and total investment I_t for the asset under consideration.

[12] Our results can be illustrated with a historical example, for which we are grateful to Roger Fouquet. In the first half of the 19th century, the introduction of steam engines brought cheaper and more comfortable medium and longer distance travel than had previously been provided by stagecoaches (pulled by horses). Coach companies responded to this heightened competition from railways by ceasing investment into equipment and horses, driving their prices even higher. This inevitably accelerated the transition to railways (Fouquet, 2012).

We assume, as is implied by standard models, that in each period, returns r_t are monotone strictly decreasing in capital stock K_t. So the pattern of investment implied by Proposition 2.2 implies *a short-term decrease in the dirty energy capital stock, relative to a world in which investments are reversible (and so underutilization is never an issue).*

To explore this, consider an otherwise identical model in which we relax the constraint $i_t \geq 0$ – allowing holdings of this capital stock to be converted back into cash for consumption or other purposes. Refer to variables in this modified model as $\widetilde{K}_t$, $\widetilde{I}_t$, etc. We suppose that e_t is unchanged by relaxing the constraint $i_t \geq 0$, because the sector concerning the asset in question is very small in relation to the rest of the economy.

Corollary 2.3. *Suppose that $I_0 > 0$ and there exists $t_1 \geq 1$ such that $\tilde{I}_{t_1} < 0$. Then there exists $t_0 < t_1$ such that $I_{t_0} < \tilde{I}_{t_0}$ and such that $K_t < \tilde{K}_t$ for $t \in \{t_0 + 1, \ldots, t_1\}$.*

In the short term, less is invested in the irreversible capital stock, relative to a world in which investments are reversible.

2.2 The Irreversibility Effect in Climate Change Economics

In this paper we apply the observations of Section 2.1 to a model of climate change economics. We are particularly concerned with capital investments in installations, such as coal fired power stations, which will burn fossil fuels. The quantity of fuel demanded, and burnt, is associated with the quantity of appropriate capital infrastructure in the economy. If, in the extreme case, this relationship is Leontief, then Corollary 2.3 implies:

Corollary 2.4. ***[The Irreversibility effect]****Suppose emissions are directly proportional to dirty fossil infrastructure. Assume that investment in dirty fossil infrastructure is non-zero in the first period, but there exists a time $t_1 \geq 1$ such that this infrastructure would be globally divested if it could be. Then, for some period leading up to t_1, emissions are below the level they would reach if divestment were possible.*

That is, capital stock effects on the *demand* side for fossil fuels *enhance* the effect of the carbon tax in the short term.

Recall and contrast with the Green Paradox from Sinn (2008): if future climate policies are expected to be more stringent than those currently in place, then resource suppliers accelerate extraction of their fossil fuel stocks. That is, irreversibilities have opposing implications depending on whether we consider suppliers, or demanders, of fossil fuel. It is important to bear this distinction in mind when considering the question of stranded assets.

3 A Simple Model of Investing with Learning-By-Doing

Learning-by-doing is often cited as a rationale for subsidizing renewable electricity. The theory of learning-by-doing is motivated by simple observation: production performance (either in form of productivity or cost of technology) tends to improve with the accumulation of experience. We are particularly interested in the form that was specified both by Wright (1936) and Arrow (1962): each doubling of cumulative deployment reduces prices by the same factor, the "learning rate".[13]

[13]Wright (1936) was the first one to describe the concept of learning, after observing a uniform decrease in the number of direct labor hours required to produce an airframe for each doubling of the cumulative production of the plant under consideration.

Empirically, the existing literature has found evidence that the price of renewable energy evolves in this way, although causality may not yet be finally established.[14]

To model learning, we consider a renewable capital asset, H_t (writing h_t for household-level holdings as usual). The notation reminds us that this asset embodies human capital in the form of knowledge, as well as the infrastructure itself. The form of this knowledge is the price of installing this infrastructure: investments will have price p_t^H, which depends on the total installed capacity H_t, so $p_t^H = G(H_t)$. And we generally assume that learning follows Wright's Law: there exists a constant $\lambda > 0$ with

$$p_t^H = G(H_t) = p_0^H \left(\frac{H_t}{H_0}\right)^{-\lambda}. \tag{2}$$

Asset h_t is priced at p_t^H, so that $i_t^H = p_t^H(h_{t+1} - (1-\delta)h_t)$.

This learning is of course an externality. So we first explore the optimal program of investment found by a social planner. We contrast this with the behavior of households who act as price-takers, to better identify and understand the optimal subsidy.

3.1 The Social Planner's Case

The social planner optimizes total welfare $\sum_{t=0}^{\infty} \beta^t L_t u\left(\frac{C_t}{L_t}\right)$, where L_t is the population size and C_t is total consumption. This is subject to the budget constraints $I_t^H + C_t = f_t(H_t, O_t)$, where we have written $O_t = L_t o_t$ for the aggregate across the economy of "other" incomes, so that we can write $f_t(H_t, O_t)$ for the production function. In this simple model, the planner will treat O_t as exogenous, that is, the planner is constrained to provide the individually rational levels of "other" parts of the economy, and does not attempt to influence them. This assumption is benign if all externalities in the remainder of the economy have been internalized. Welfare maximization is also subject to the investment equations $I_t^H = p_t^H(H_{t+1} - (1-\delta)H_t)$; the investment bounds $I_t^H \geq 0$; and the price evolution given in (2) above.

We define the *shadow returns* on our capital stock as the discrete time version of the definition of Jorgenson (1967):

$$R_{t+1} := \frac{\mu_t^H - \beta(1-\delta)\mu_{t+1}^H}{\beta u'(C_{t+1}/L_{t+1})} \tag{3}$$

where μ_t^H is the shadow price on the investment equation $I_t^H = p_t^H(H_{t+1} - (1-\delta)H_t)$. This definition says that the shadow return tomorrow on investment made today is equal to the shadow value of additional capital tomorrow, less the discounted depreciated shadow value of this capital going further forward (as these gains will be realized further in the future). Naturally, everything is measured relative to the marginal value today of consumption tomorrow.

On the other hand, define the *direct return* (accounting for the price of capital) to be:

$$r_{t+1}^s := \frac{1}{p_{t+1}^H} \frac{\partial}{\partial H_{t+1}} f_{t+1}(H_{t+1}, O_{t+1}). \tag{4}$$

[14] Lindman and Soderholm (2012) use aggregate data and show that learning externalities are present in wind turbines and solar panel costs. Such studies based on aggregate data, however, unable to disentangle the effect of exogenous technological change from the effect of leaning-by-doing thus masking the diverse drivers of technology costs (see also Nordhaus 2014b). Nemet (2006) for instance finds that after accounting for measures of technological change and the cost of inputs, learning has only weak explanatory power for solar panel costs. Much more recently, Lafond et al. (2017) use hindcasting techniques to assess this model, finding it provides a very good fit. Bollinger and Gillingham (2014) provide evidence for cost reductions due to learning-by-doing across installation contractors of solar photovoltaics in California from 2002 to 2012.

We write r_t^s to distinguish from the notation for the market rate of return r_t, which we will use in the decentralized model below.

Proposition 3.1. *Suppose that investment into this sector will be non-zero next period, i.e.* $I_{t+1}^H > 0$. *Then:*

$$\frac{p_t^H}{p_{t+1}^H} R_{t+1} = \underbrace{r_{t+1}^s}_{\text{direct return}} - \underbrace{\frac{p_t^H - p_{t+1}^H}{p_{t+1}^H}(1-\delta)}_{\text{price effect}} - \underbrace{(H_{t+2} - (1-\delta)H_{t+1})\frac{G'(H_{t+1})}{p_{t+1}^H}}_{\text{learning effect}} \quad (5)$$

shadow return

First, we note that the shadow return R_{t+1} is multiplied by factor p_t^H/p_{t+1}^H because R_{t+1} values returns from the moment at which the investment is made, while we have defined the direct return relative to prices at the time at which we receive the return. Next, the shadow return relative to those prices is composed of three terms. There is an obvious "direct return" from renewable energy capital, taking into account the price of this capital.

Next, we observe a "price effect". Every unit of renewable capital we have in period $t+1$, is only worth p_{t+1}^H today, but was priced at p_t^H back when investment took place. Thus its value is reduced by this factor going forward, and hence the shadow return is also reduced. However, the reduction is mediated by the extent to which capital will in any case depreciate. So we have an incentive to delay the build-up of our capital holdings because prices will be lower tomorrow; but this is only important for persistent assets. This effect is important from a social planning perspective because we assume that prices are constant within each year, and so the gains from learning are only realized by investments in the following period. Thus, the size of the effect will depend on the time-step we take.

The final term, which we call the "learning effect", is the product of the change in the capital stock, and the marginal change in price achieved by making the final unit of investment, relative to the price p_{t+1}^H. One must not be confused by the negative sign: typically $G'(H) < 0$ (prices decrease with capacity), and $H_{t+1} > (1-\delta)H_t$ (investment is positive), so that the learning effect is typically positive.

The net effect of the price and learning effects may be positive or negative, and so the total return on renewables may be greater than, or less than, their direct net return. From Proposition 3.1, we have the following corollary:

Corollary 3.2. *Assume that* $G'(H) < 0$ *and* $H_{t+1} > (1-\delta)H_t$. *If* $\delta = 1$ *then* $\frac{p_t^H}{p_{t+1}^H} R_{t+1} > r_{t+1}^s$. *If* $\delta = 0$, *if* G *is convex, and if* $H_{t+2} - H_{t+1} = H_{t+1} - H_t$ *then* $\frac{p_t^H}{p_{t+1}^H} R_{t+1} < r_{t+1}^s$.

Regarding the case in which the price effect dominates the learning effect: the assumption of convexity for the function G giving the decline in prices, is very natural. The assumption that capital is increasing by a constant *amount*, rather than a constant *factor*, is less so; and increases in $H_{t+2} - H_{t+1}$ relative to $H_{t+1} - H_t$ will increase $\frac{p_t^H}{p_{t+1}^H} R_{t+1}$ relative to r_{t+1}^s. In general we expect the learning effect to dominate, but it is worth noting that when capital is very persistent, and when there is a considerably delay in realizing the benefits of learning, then the extent to which the learning effect pushes the shadow return above the direct return, is mitigated by the price effect.
An important difference between the price and learning effects is that the former will be taken into account by small rationally optimizing firms, whereas the latter will not: in our specification, learning-by-doing is a pure externality. As the learning effect is positive, it follows that investing in this capital stock is worthwhile from a social perspective before it is individually rational: if

investment will take place in the near future, it is socially optimal to start earlier than an individual would choose to.

So, as we will see next, the optimal subsidy in a decentralized model is equal to the learning effect.

3.2 Learning-By-Doing and the Acceleration Effect

We assume that households act as price-takers on the capital stock undergoing learning-by-doing. There will be under-investment without intervention. So we introduce a subsidy, τ_t; it is convenient to express this as a subsidy on the rate of return. Now we may write the household's budget constraint as $i_t^H + c_t = (r_t + \tau_t)p_t^H h_t + o_t$. These investments are characterized by $i_t^H = p_t^H(h_{t+1} - (1-\delta)h_t)$ and $i_t^H \geq 0$; recall o_t represents other sources of income. The subsidy is paid for out of lump sum taxation; as the households are price-takers, this taxation may be incorporated into o_t. Meanwhile, a final goods firm maximizes its profits $f_t(H_t, O_t) - r_t p_t^H H_t - p_t^o O_t$, where p_t^o is the price they must pay for access to other assets.

Again there are L_0 households in the economy, each of size $\frac{L_t}{L_0}$, so that the consumption of a representative individual is $\frac{L_0 c_t}{L_t}$.

Proposition 3.3. *Suppose that any externalities in o_t have been internalized. The subsidy τ_t which optimizes consumer welfare $\sum_{t=0}^{\infty} \beta^t L_t u\left(\frac{L_0}{L_t} c_t\right)$ is equal to the learning effect:*

$$\tau_t = -\left(H_{t+1} - (1-\delta)H_t\right) \frac{G'(H_t)}{p_t^H}.$$

Thus, the subsidy is a straightforward function of the growth rate of the renewables sector. Contrary to models which prescribe a short-term subsidy to this sector, the specification we use implies that this subsidy is positive as long as there is any investment in this sector, even only to replace depreciating stock.

The simple form (2) is what we generally assume for the relationship between price and accumulated capital. It implies:

Corollary 3.4. ***[The Acceleration Effect]*** *If $G(H_t) = p_0^H \left(\frac{H_t}{H_0}\right)^{-\lambda}$, then*

$$\tau_t = \lambda\left(\frac{H_{t+1}}{H_t} - (1-\delta)\right).$$

In particular, the subsidy τ_t increases with the deployment rate $\frac{H_{t+1}}{H_t}$.

Thus, if the capital asset in question becomes more attractive in the economy, and so starts to accumulate faster irrespective of the subsidy, we *also* increase the subsidy to this asset. We call this the *acceleration effect* for technology policy. Very natural contexts are discussed below, in Section 4.9.

4 The Full Model

This section outlines the full dynamic general equilibrium climate-economy model which is used for quantitative analysis. The derivations of the equations that define the solution of the model are given in Appendix D. To summarize, the model presents a climate-economy structure, where, unlike

other leading climate-economy models,[15] we differentiate between three capital stocks:[16] general capital, "clean" and "dirty", with irreversibility in investments characterizing the latter two capital stocks, as in Section 2 above. We allow underutilization of dirty capital stocks, once they become uncompetitive. In addition we assume that the "clean" sector is characterized by "learning-by-doing": costs of new technologies decline as a function of cumulative installed capacity in the sector, as in Section 3. The climate module uses the representation of the carbon cycle, temperature system, and climate-economy feedbacks based on the DICE framework (Nordhaus, 2014a), but calibrated to an annual time step (Cai et al., 2015, 2016).

There are five production sectors: final goods producing firms, aggregate electricity producing firms, the dirty electricity producing firms, the fossil-fuel extracting firm and the firms producing electricity from renewable sources. All firms operate under perfect competition. Notably, fossil fuel extracting firm maximizes the present value of its profits, subject to the depletion equation, internalizing the effect of depletion on future extraction costs and on present and future revenue. Producers of renewable energy maximize the present value of their profits, taking the market price of renewable energy and the stock of accumulated knowledge about using renewable energy as given.

Turning to the demand side of the economy, we are interested in the behavior of a representative household who does not internalize the learning-by-doing externality and treats all prices as given. Finally, there are three sources of carbon dioxide emissions: general output production, electricity production from dirty energy inputs, and land use. Climate change affects productivity in the final goods producing sector.

4.1 The households' problem

We are interested in the behavior of a representative household. There are L_0 households (defined as the population size of the economy at the initial period, which in our calibrated model is 2012), and the size of the family at time t is $\frac{L_t}{L_0}$, where L_t is the population size at period t.[17]

We consider all variables on a per-household basis, so we will write $k_t^g = \frac{K_t^g}{L_0}$, etc., where capital letters denote aggregate variables (over all households), for instance K_t^g is aggregate general capital stock and H_t is aggregate renewable energy knowledge and capital stock. The household seeks to maximize the sum of the welfare of family members as individuals, that is:

$$\sum_{t=0}^{\infty} \beta^t \frac{L_t}{L_0} u\left(\frac{C_t}{L_t}\right) = \sum_{t=0}^{\infty} \beta^t \frac{L_t}{L_0} u\left(\frac{L_0}{L_t} c_t\right) \tag{6}$$

where C_t is aggregate consumption and we write $c_t := \frac{C_t}{L_0}$ as the per-household consumption. The household owns a representative share of all three capital assets and the five sorts of companies. We denote r_t^D, r_t^H and r_t^g as the rate of return on capital assets in fossil (dirty) capital, renewable (clean) capital, and general capital (used in production of final-goods producing firms), respectively. Further, we write w_t for the wage, Π_t^g for the total profit from the sale of the final goods, Π_t^D for the total profit from the sale of dirty fuel based electricity, Π_t^H for the total profit from the sale of "clean" electricity, Π_t^{DE} for the total profit from the sale of fossil fuel, and Π_t^E for the total profit from the sale of aggregate electricity, so that the aggregate profit is $\Pi_t = \Pi_t^g + \Pi_t^D + \Pi_t^H + \Pi_t^{DE} + \Pi_t^E$, and the per-household profit is $\pi_t := \frac{\Pi_t}{L_0}$.

[15] See, e.g., Golosov et al. (2014), Barrage (2014), Acemoglu et al. (2016), Nordhaus (2008), Rezai and Van Der Ploeg (2017).

[16] In a similar way, but within a different context, Greenwood et al. (1997) developed the importance of investment into differentiated capital stocks for growth and technological change.

[17] Table A.2 in the Appendix B provides a summary of variables' notation and definition.

In each period, the household faces the following budget constraint:

$$i_t^g + i_t^D + i_t^H + c_t = \frac{L_t}{L_0} w_t + \pi_t + r_t^g k_t^g + r_t^D p_t^D k_t^D + r_t^H p_t^H h_t + \frac{1}{L_0} \left(\tau_t^D (D_t^E + D_t^g) - \tau_t^H p_t^H H_t \right) \tag{7}$$

where i_t^g is investment in general capital, i_t^D is investment into dirty capital used in production of dirty electricity, i_t^H is investment in capital used in the clean sector, k_t^g, k_t^D, h_t are capital stocks in the general, dirty and clean sectors respectively, τ_t^D is the carbon tax, τ_t^H is the subsidy, and D_t^E and D_t^g are carbon emissions in the dirty and general sectors. Since we measure fossil and renewable energy capital in gigawatt (GW), p_t^D and p_t^H are the prices of the fossil fuel and renewable capital in \$/GW. The price p_t^H of renewable energy capital falls with our embodied technological progress in renewable energy knowledge and capital stock, and evolves as (Arrow, 1962):

$$p_t^H = G(H_t) = p_0^H \left(\frac{H_t}{H_0} \right)^{-\lambda} \tag{8}$$

However, as we treat a household as very small, we assume that their investment in renewable energy does not influence its price, so that the learning-by-doing externality arises. That is, the household takes p_t^H as given. The price of fossil fuel capital will be fixed, so that $p_t^D = p^D$.

Finally, we assume that the household receives rebates on the taxes and pays for the subsidies (the last two terms in the right hand side of the budget constraint), but as we assume they are small they cannot affect these levels.

The capital stocks in the general, dirty and renewable sectors are accumulated according to the following equations respectively:

$$i_t^g = k_{t+1}^g - (1 - \delta^g) k_t^g \tag{9}$$

$$i_t^D = p_t^D (k_{t+1}^D - (1 - \delta^D) k_t^D) \tag{10}$$

$$i_t^H = p_t^H (k_{t+1}^H - (1 - \delta^H) k_t^H) \tag{11}$$

where δ^g, δ^D, δ^H are depreciation parameters, and

$$i_t^D \geq 0 \tag{12}$$

$$i_t^H \geq 0 \tag{13}$$

are the irreversibility assumptions - a non-negativity constraint on the rate of accumulation of dirty and clean capital.

4.2 The final-goods firms' problem

The final goods are produced by identical firms, but output is damaged by climate change. Because this sector exhibits constant returns to scale, we can work with aggregate variables, and so write output:

$$Y_t = \Omega(T_t) f(Y_t^g, E_t)$$

where T_t is the temperature change from pre-industrial levels and $\Omega(T_t)$ is the damage factor ($1 - \Omega(T_t)$ is the ratio of damage to output); and E_t is electricity and Y_t^g is "general" output (i.e. non-electricity).

The final-goods firms maximize

$$\sum_{t=0}^{\infty} q_t \Pi_t^g = \sum_{t=0}^{\infty} q_t \left(\Omega(T_t) f(Y_t^g, E_t) - r_t^g K_t^g - w_t L_t - p_t^e E_t - \Psi_t - p_t^{fuel} D_t^g \right) \tag{14}$$

where $q_t := \beta^t \frac{u'(c_t)}{u'(c_0)}$ is a compound discount rate for the relative price of consumption in period t, expressed in period 0 units.[18] To produce final goods, these firms rent (aggregate) capital K_t^g, hire labor L_t, purchase aggregate electricity E_t at price p_t^e, buy fossil fuel D_t^g from fossil fuel extracting firms at price p_t^{fuel}. The firms spend on abatement Ψ_t, which is assumed to abate the η_t fraction of emissions via the following relation:

$$\Psi_t = \frac{\phi_{1,t} \eta_t^{\phi_2}}{(1-\eta_t)^{\phi_3}} Y_t^g \tag{15}$$

so that they face the emissions constraint given by:

$$D_t^g = \sigma_t (1-\eta_t) Y_t^g \tag{16}$$

where ϕ_2 and ϕ_3 are parameters, σ_t represents the ratio of carbon-equivalent emissions to output, which along with the parameter $\phi_{1,t}$ evolve exogenously, as in Cai et al. (2016). Firms do not take into account their emissions' impact on the pollution stock and thus on productivity. In other words, firms take $\Omega(T_t)$ as a given. This, in a conjunction with the knowledge externality in the renewable sector, represents a "twin-market failure" (Jaffe et al. (2005)).

For the solution of the model, we assume that the function for production before damages takes constant elasticity of substitution (CES) form (Hassler et al., 2012):

$$f(Y_t^g, E_t) = \left[(1-\theta)(Y_t^g)^{1-1/\kappa} + \theta \, (E_t)^{1-1/\kappa} \right]^{\frac{1}{1-1/\kappa}}.$$

and

$$Y_t^g = f_t^g(K_t^g, L_t) = A_t^g (K_t^g)^{\alpha} (L_t)^{1-\alpha}.$$

Here, θ, κ, α are parameters, A_t^g is a technology process in the general sector, K_t^g is general capital and L_t is labor. Both A_t^g and L_t evolve exogenously in the same way as in Cai et al. (2016).

4.3 Aggregate electricity producing firms' problem

These firms again work at constant returns to scale, so we can work with aggregate variables. They produce aggregate electricity $E_t = f_t^E(H_t, \Gamma_t^{ED})$ which is a combination of fossil fuel production capacity Γ_t^{ED}, and clean production capacity H_t, with these inputs being priced at p_t^{EH} and p_t^{ED} respectively. They sell their output at price p_t^e, so that the firms maximize the present value of their profits :

$$\sum_{t=0}^{\infty} q_t \Pi_t^E = \sum_{t=0}^{\infty} q_t \left(p_t^e f_t^E(H_t, \Gamma_t^{ED}) - p_t^{EH} H_t - p_t^{ED} \Gamma_t^{ED} \right) \tag{17}$$

In modeling the electricity sector, we follow Papageorgiou et al. (2016)[19] and assume a CES production function of renewable production capacity H_t and dirty production capacity Γ_t^{ED}:

$$E_t = f_t^E(H_t, \Gamma_t^{ED}) = A_t^E \left(w H_t^{\xi} + (1-w)(\Gamma_t^{ED})^{\xi} \right)^{1/\xi}, \tag{18}$$

where A_t^E is a technology process in the electricity sector and w and ξ are CES parameters.

[18] See Appendix D.1 for more detailed discussion on derivation of compound interest for the firms' problems.

[19] We do not use their model for overall energy as a combination of electricity and "other dirty energy", as in their model the latter requires no capital input and so is disproportionately favored under optimization.

4.4 The dirty electricity producing firms' problem

The dirty electricity producing firms are fossil-fuel based power stations, which combines existing infrastructure (e.g., coal-based power plants) with fossil fuels via a Leontief production function. Again, due to constant returns, we work at the aggregate scale:

$$\Gamma_t^{ED} = \min[\zeta_t K_t^D, D_t^E/\nu] \tag{19}$$

where K_t^D is total capital in dirty electricity production, $\zeta_t \in [0,1]$ is the utilization rate, and ν is the conversion rate from fossil fuel to electricity. The Leontief function implies a fixed ratio between utilized fossil energy capital and dirty fuel use:

$$D_t^E = \nu \zeta_t K_t^D. \tag{20}$$

The firms buy fossil fuel D_t^E at price p_t^{fuel}, rent the dirty capital infrastructure at rate r_t^D, and sell their output Γ_t^{ED} to the aggregate electricity producing firms at price p_t^{ED}. So, the firms in the sector maximize the present value of their profits:

$$\sum_{t=0}^{\infty} q_t \Pi_t^D = \sum_{t=0}^{\infty} q_t \left(p_t^{ED}(\zeta_t K_t^D) - r_t^D p^D K_t^D - p_t^{fuel} D_t^E \right) \tag{21}$$

subject to emissions constraint (20), and a constraint on utilization rate: $\zeta_t \leq 1$.

4.5 The fossil-fuel extracting firm's problem

We treat fossil fuel extraction as being handled by a single large firm (to give rise to the Hotelling equation and the Green Paradox). This firm maximizes the present value of its profits, by taking the market price of fossil fuel, p_t^{fuel} as given and internalizing the effect of depletion on future extraction costs and resource availability:

$$\sum_{t=0}^{\infty} q_t \Pi_t^{DE} = \sum_{t=0}^{\infty} q_t [p_t^{fuel} - \tau_t^D - G^D(S_t)](D_t^E + D_t^g) \tag{22}$$

where τ_t^D is tax on production of fossil fuels. Fossil fuels are extracted from finite reserves; the stock remaining at time t is denoted S_t. The evolution of this stock follows from the standard depletion equation (e.g., Rezai and Van Der Ploeg (2017)):

$$S_{t+1} = S_t - (D_t^E + D_t^g). \tag{23}$$

The fossil fuel extraction cost per unit is given by:

$$G^D(S_t) = \gamma_1 \left(\frac{S_0}{S_t} \right)^{\gamma_2}, \tag{24}$$

so less stock will incur higher extraction cost, where γ_1 and γ_2 are parameters.

4.6 The renewable firms' problem

The renewable sector is composed of small firms, which do not internalize the learning-by-doing externality (8). That is, they take the stock of accumulated knowledge about using the renewable energy H_t as given, with a rent rate r_t^H. And they receive a subsidy of τ_t^H on their dollar-valued

holdings of renewable energy capital H_t. They sell their output to the aggregate electricity producing firm at price p_t^{EH}. The firms take all prices as given, so on aggregate they maximize:

$$\sum_{t=0}^{\infty} q_t \Pi_t^H = \sum_{t=0}^{\infty} q_t [p_t^{EH} - p_t^H (r_t^H - \tau_t^H)] H_t. \tag{25}$$

Note that in the "simple model" of Section 2 we did not model renewable firms explicitly, for simplicity there, and so in that model we wrote the subsidy as accruing to the householder, who also owns the capital.

4.7 Climate system, emissions and damages

The carbon dioxide emissions D_t have three sources: "general" output production D_t^g; electricity production from using fossil fuel D_t^E; and land use D_t^{land}.

$$D_t = D_t^E + D_t^g + D_t^{\text{land}} \tag{26}$$

Land-use emissions D_t^{land} are set exogenously as by Cai et al. (2016). We use the climate system of Cai et al. (2016), which adapts the climate system of DICE2013 (Nordhaus, 2014a) to an annual time step. As this component of our model has been described extensively elsewhere, we omit it here, instead we simply denote the mapping from emissions to temperature by:

$$T_t = \mathcal{W}_t(D_0, \ldots, D_{t-1}) \tag{27}$$

where T_t is global atmospheric temperature change over pre-industrial levels, D_s is fossil pollution at time $s < t$ and these are related via the warming function $\mathcal{W}_t$.

Finally, the damage factor for "DICE damages" is given by

$$\Omega_t(T_t) = \frac{1}{1 + \varsigma_1 T_t^{\varsigma_2}}, \tag{28}$$

where ς_1 and ς_2 are two parameters. The damage function in climate change economics is very controversial (see, e.g. Weitzman 2009, 2010; Cai et al. 2015). In fact there do not exist well-founded estimates of damages for even moderate temperature changes, and so their ability to dictate optimal climate policy is limited. However, a great deal of discussion in real-world policy-making focuses on limiting global temperature changes to 2°C. We simulate this constraint by letting

$$\Omega_t(T_t) = \frac{1}{(1 + \varsigma_1 T_t^{\varsigma_2})\,(1 + \varsigma_3\,(T_t/2)^{\varsigma_4})} \tag{29}$$

with a small positive parameter $\varsigma_3 = 0.001$ and a large exponent parameter $\varsigma_4 = 50$. Thus, when atmospheric temperature increase T_t is smaller than 2°C, the new damage factor (29) is almost the same as (28), but when T_t is larger than 2°C, the new damage factor will imply large damage. This new damage factor (29) will be referred as the stringent damage factor.

4.8 Decentralized equilibrium vs social planner's problem

To find an optimal solution of the decentralized model, we formulate it as that of a principal who must choose an allocation from among those that can be implemented as a decentralized equilibrium, bearing in mind how the other economic participants (the "agents") will respond. In the

optimal taxation literature such conditions imposed on the (Ramsey) principal are known as implementability conditions. We solve it using mathematical programming with equilibrium constraints. The details are in Appendix D.

The previous sections laid out the decentralized equilibrium model. To retrieve the values of optimal carbon tax and optimal subsidies that could replicate the first-best allocation in the decentralized equilibrium model, we also outline a social planner model where the social planner maximizes the social welfare given constraints describing the carbon cycle, temperature, damages and fossil fuel depletion, and the capital accumulation equations. See Appendix C for details.

4.9 Subsidy and carbon tax

In the decentralized equilibrium, there are two instruments: a subsidy on renewable capital, τ_t^H, and a carbon tax, τ_t^D. There are various scenarios related to the choice of policy instruments. We differentiate between four cases: (1) a no policy scenario in which we set $\tau_t^D = 0$ and $\tau_t^H = 0$; (2) the optimal policy version, in which both instruments are freely chosen to maximize the principal's objective; (3) $\tau_t^D = 0$ and subsidy is chosen freely to maximize the principal's objective; (4) $\tau_t^H = 0$ and carbon tax is chosen freely to maximize the principal's objective. Clearly, the second policy yields the same outcome as the social planner's problem, and it is the first-best, which we prove in the appendices. Cases (3) and (4) are situations with second-best policies.

Next, we define:

Definition 4.1. The *social cost of carbon*, χ_t is the shadow price on carbon emissions, relative to the shadow value of output. That is, if μ_t^D is the shadow price of Equation (26) constraining total emissions, then:

$$\chi_t := \frac{\mu_t^D}{u'\left(C_t/L_t\right)}.$$

Next, we prove (see Appendix D.8):

Proposition 4.2. *The decentralized equilibrium allocation coincides with the solution to the social planner's problem if carbon taxes are set as the social cost of carbon χ_t and if subsidies are set equal to the "learning effect":*

$$\tau_t^H = -(H_{t+1} - (1-\delta^H)H_t)\frac{G'(H_t)}{p_t^H} \tag{30}$$

This verifies that the theoretical insights on learning-by-doing from the "simple model" in Section 3 all carry across to the full model. That is, Corollary 3.4 holds and we have an "acceleration effect". In particular, an increase in the carbon tax which reduces investment in and utilization of dirty energy capital and so increases deployment of the substitute renewable energy capital, *also* implies an increase in the optimal renewable energy subsidy.

Naturally, Proposition 4.2 also shows that we can examine optimal policy by using a social planner's model, which is easier computationally. However, we do not restrict attention to this simpler case; we are also very interested in worlds without optimal (first-best) policy. If only the tax, or only the subsidy, are in use, then Proposition 4.2 does not apply. We explore such scenarios with our numerical results.

5 Quantitative results from the calibrated model

This section presents the quantitative results in three parts.[20] The first investigates the links between irreversible investment decisions and climate policies. We compare optimal policies with and without a stringent climate target (i.e., use the DICE damage factor (28) or the stringent damage factor (29)). In addition, we illustrate the importance of the irreversibility in investment decisions relative to the case when investments are reversible (the irreversibility effect). In the second part we study the acceleration effect pertaining to an early start of investment into the renewable sector. And finally, we study the implications of the second-best policies for the welfare and dynamics of the model.

The initial period in our model is 2012. The model could be run under various scenarios that can be differentiated along three different dimensions: (1) damage function: DICE damage factor (28) vs stringent damage factor (29); (2) irreversible vs reversible investments; and (3) the choice of policy instruments: optimal tax and subsidy vs second-best policies. The runs of the decentralized equilibrium under combined optimal tax and subsidy are equivalent to the runs of the social planner model (the first-best policy).

5.1 Irreversible investment and climate policies

First, we want to understand how the optimal paths of variables depend on the irreversibility assumption coupled with different climate policy targets. We notice that the effect of irreversibility (compared with when the investment is reversible) becomes quantitatively important only if the climate policy objective is ambitious enough. Figure 1 shows that the paths of the investment on dirty energy are almost the same with reversible and irreversible investments under a mild climate policy objective (i.e., the DICE damage factor) in this century, but they are distinctly different under more ambitious climate policy target (i.e., the stringent damage factor).

These results emphasize the importance of setting ambitious climate policies in inducing permanent fuel energy switching. The strong path dependence embodied in carbon-intensive infrastructure suggests that mild climate change policies (i.e., those induced by DICE damage factor) would not induce shifts away from dirty energy towards green energy, as would be required to meet the Paris Agreement objectives,[21] as we see in Figure 6 in Section 5.3.

Further, Figure 1 shows that with the irreversible investments and stringent policy, there is no investment in dirty energy after 2020. In contrast, when investment is reversible, the decumulation rate of dirty capital stock is unlimited, and we keep investing into this capital stock for another seven years until 2027, when we start shifting dirty capital stock into general capital stock, a process that continues until about 2075. However, we never entirely stop using the dirty capital stock because of the imperfect substitutability between dirty and clean energy in electricity production. So, since we decumulated the dirty capital stock sufficiently in the preceding decades, investment into dirty capital stock resumes after 2075 under reversible investment.

These dynamic patterns of investment into dirty energy with the (ir)reversible investments and the stringent damage factor correspond to the dynamics of return on those investments shown in Figure 2. The theoretical counterpart of this figure is Proposition 2.2 in Section 2. First, the figure shows that we end investment into the dirty capital stock when the investment is still attractive with the rate of return, $r_t^D - \delta^D$, exceeding the rate of return on the general economy $r_t^g - \delta^g$. This is

[20] The calibration of the model is described in Appendix B.

[21] This finding echoes the one in Meng (2016), who estimates the strength of path dependence in the electricity sector for the U.S. Midwest and shows that a permanent decline in U.S. electricity sector emissions would require shocks of larger magnitude and longer duration that that of recent natural gas prices.

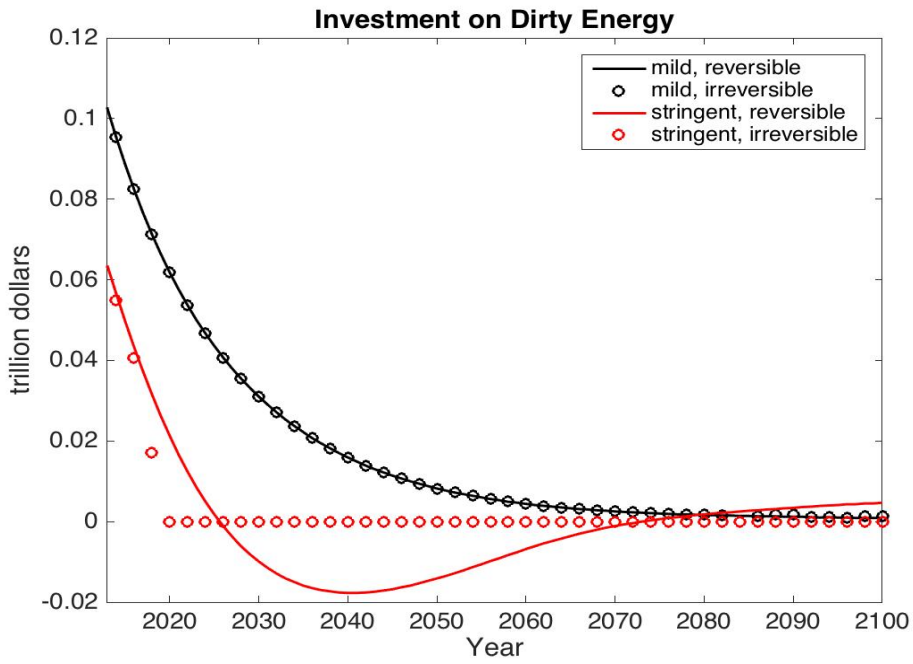

Figure 1: Investment on Dirty Energy

because we invest into infrastructure that will become obsolete in the future only if the short-term benefits from such investment compensate for future losses. Thus even without uncertainty, returns to irreversible investment require a premium.[22] Even we end investment around 2020, we continue to fully utilize the dirty capital stock for about another 25 years, till 2045, when the return on dirty capital (i.e., r_t^D) reaches zero and we start underutilizing the dirty capital stock.

Next, as we end investment "earlier" than in the counter-factual (i.e., when disinvestment is a viable option), the economy continues to hold less dirty capital stock under irreversibility than in the reversible case in the medium-term, till about 2037 (solid lines in Figure 3). After that year, however, the economy holds larger stocks of dirty capital in the long-run if investment is irreversible. This is due to the path dependence: capital cannot be converted into other forms of capital stock. But, if we take into consideration the underutilization of the dirty capital stock in the irreversible investment case (circles in Figure 3), then, in the long-run, the same total amount of the dirty capital stock will be *utilized* under both irreversible and reversible investment decisions (Figure 3).

5.2 Acceleration Effect

For convenience here we reproduce the theoretical result related to the optimal subsidy, when the optimal carbon tax also applies (Corollary 3.4):

$$\tau_t^H = \lambda \left(\frac{H_{t+1}}{H_t} - (1 - \delta^H) \right) \tag{31}$$

[22]Previous studies such as Bernstein and Mamuneas (2007) develop a simple model of production and investment with costly disinvestment to estimate the magnitude of the premium associated with irreversible investment in telecommunications industry, assuming future telecommunications capital acquisition prices are random variables. Their findings indicate that the premium increases the user cost of capital by 70%, with implies an average hurdle rate of 14% over the period 1986-2002. Using different methods and framework, Pindyck (2005) provides similar estimates of the telecommunications hurdle rate.

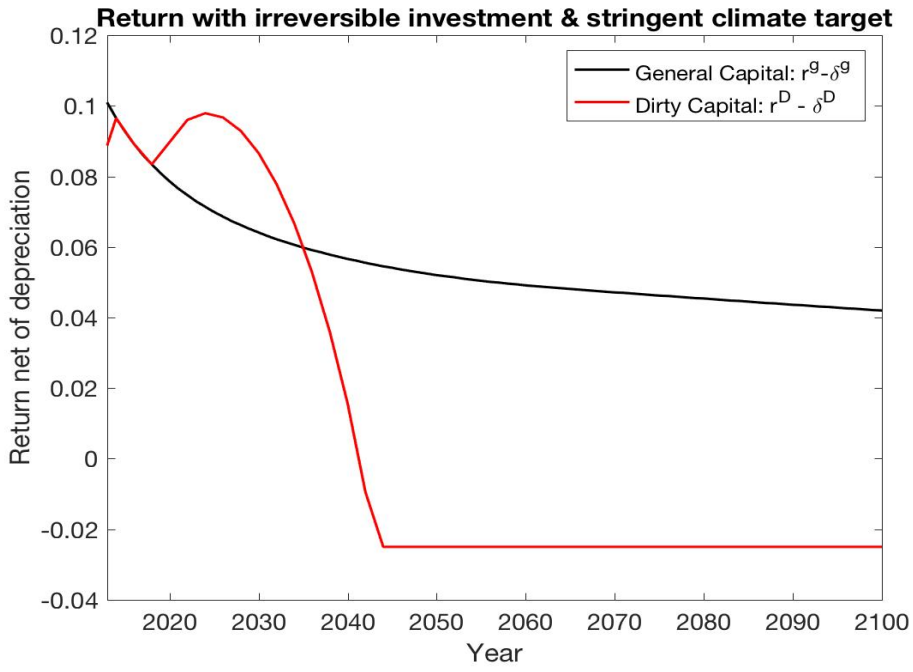

Figure 2: Return on general and dirty capital

This formula implies that (i) the subsidy continues as long as there is investment in the sector; (ii) the subsidy increases with learning coefficient λ; and (iii) the optimal subsidy is higher when renewable capital grows faster. There are three different mechanisms that can lead to higher capital accumulation in the renewable sector and consequently to a higher level of subsidies: (1) more stringent climate policies; (2) the dirty sector could be shrinking faster than otherwise due to the irreversibility effect; and (3) under second-best policies in the absence of a carbon tax, it could be optimal to grow the renewable sector faster to crowd out the dirty energy sector. We here investigate the first two of these channels. We consider second-best policies, which encompass many important effects, in Section 5.3.

Recall that (31) presents the subsidy to the rate of return on investment in H. It is useful also to consider the subsidy *level*, multiplying by the cost of investment.

5.2.1 Channel 1: stringent climate policy

Figure 4 plots the optimal subsidy to the rate of return under mild and stringent climate policy targets. Figure 5 plots the total level of this subsidy.

We observe in Figure 4 that the subsidy to the interest rate is always higher under the stringent climate policy. Because we use less fossil fuel in this scenario, we must generate more of our electricity from renewables, and so the latter sector is always growing faster than it is in the mild policy scenario. Thus, by the acceleration effect, the subsidy to the rate of return is always higher. Moreover, interestingly, the subsidy to the rate of return increases over time in all cases. The growth in H_t is increasing over time, because there are two reasons to switch to this clean sector: the fact that its own price is decreasing, and the withdrawal from the fossil sector.

Note that it does not follow that the *total* subsidy is always higher, however. Figure 5 makes this clear, giving the product $p_t^H \tau_t^H$ in each case. The decline in the price of H means that the

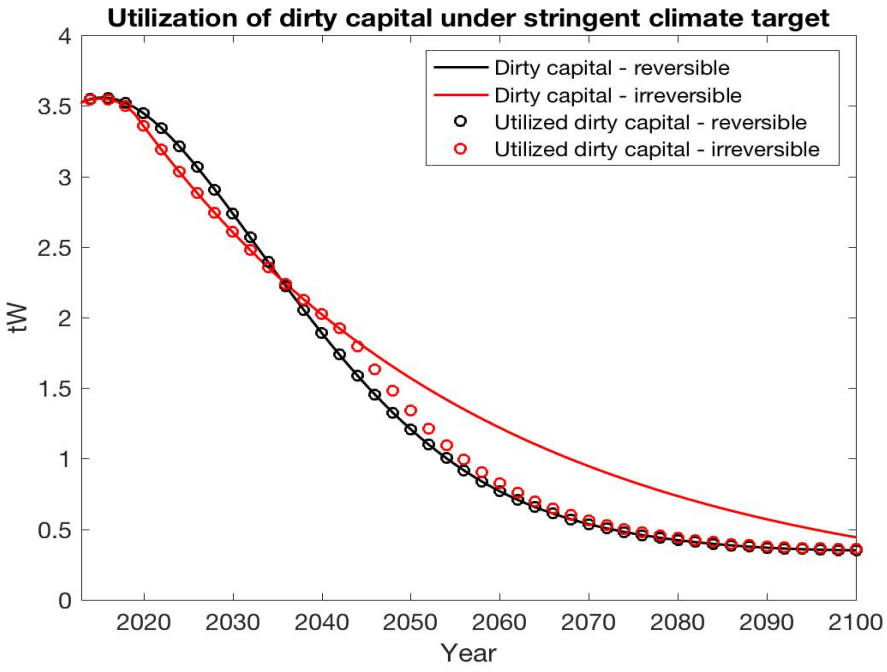

Figure 3: Utilization of dirty capital

total subsidy declines, rapidly in the early periods. Prices are eventually so much lower under the stringent policy that the total subsidy is lower – even though (recall from Figure 4) the subsidy to the interest rate is higher.

5.2.2 Channel 2: interaction with the irreversibility effect

Figures 4 and 5 plot the optimal subsidies under the stringent climate policy with both *irreversible* and *reversible* investment decisions. Due to the irreversibility effect discussed above, the dirty sector shrinks faster in the irreversible case. Once we start building fewer coal-based power plants, we need to develop other sources of energy generation. This increases deployment of the substitute renewable energy capital, which in turn *also* implies an increase in the optimal renewable energy subsidy relative to when the irreversibility effect is absent that is in the case of reversible capital. But once the renewable energy capital is built, due to this short-term aggressive policy, the opposite happens and the subsidy becomes lower than in the counter-factual.[23]

5.3 Second-best policies

The decentralized equilibrium with the optimal carbon tax on the externality created by fossil fuel use, and with the optimal subsidy on the leaning-by-doing externality in the renewable sector, implements the optimal allocation obtained in the social planner's problem (the first-best). In practice, however, one of those two policy instruments may be unavailable, and policy makers might have to rely on second-best policies. In this section we compare the relative performance of these two policy instruments under alternative climate policy objectives and (ir)reversible investment

[23]Related literature has investigated the optimal time path for innovation policy, see, e.g., Gerlagh et al. (2009) and Gerlagh et al. (2014). For instance, the latter show that if the patent lifetime is finite, the optimal subsidy starts at a high level, providing an incentive to accelerate R&D investments, and then falls over time.

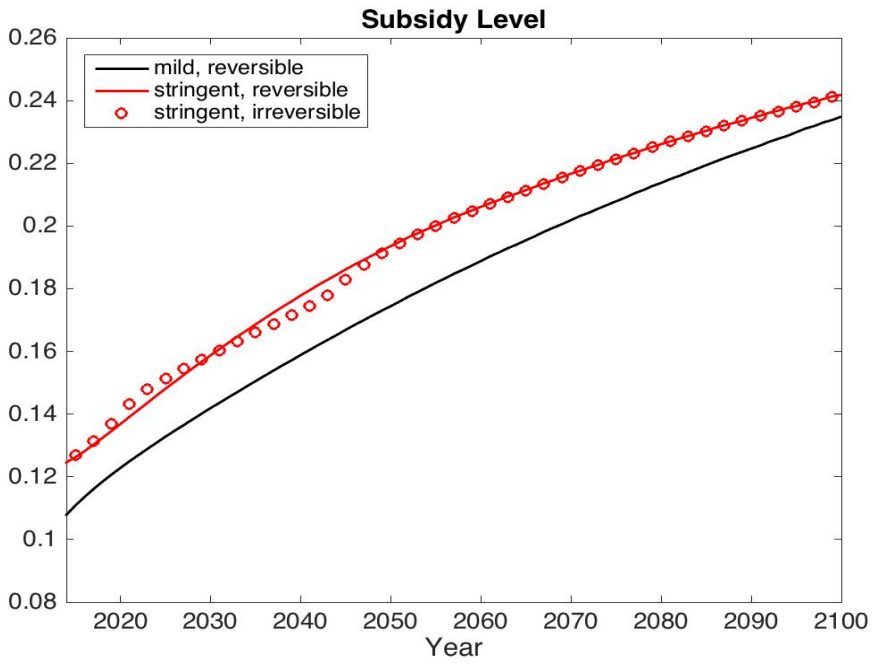

Figure 4: Optimal subsidy to the rate of return

decisions. This is an important exercise given debates on instruments to tackle climate change. There are two extreme views: on one hand, many argue that innovation policies with subsidies are sufficient for effective climate policy; on the other hand, some criticize adoption of subsidies as expensive and inefficient policies instead advocating carbon pricing (e.g., Helm 2012). In between, many advocate necessity of mixed policies, but stressing the critical importance of carbon pricing.[24]

We contribute to this debate, arguing that in a second-best world, which policy instrument should be used depends on how stringent climate policy objectives are. More specifically, under mild climate policy targets, as in case with 'DICE' damage factor (28), the economy is better off with optimal subsidy as an instrument for climate policy. In contrast, under more stringent climate policy targets, as in case with the stringent damage factor (29), the economy is better off if optimal carbon pricing policy is adopted (see Table 1).[25] As the results reported in the Table 1 further indicate, irreversibility in investment decisions does not affect the relative ranking of these policy instruments.

In what follows, we attempt to unpack the reasons why the economy is doing better with innovation policy (i.e., optimal subsidy and zero tax) in the low-damage case. And why it is less desirable to use the same policy under the stringent climate policy.

Figure 6 shows the temperature, emission, and tax levels under mild climate policy targets (the left panels) or stringent climate policy targets (the right panels), both assuming reversible investments. The top-left and middle-left panels show that with only carbon pricing, temperature

[24]Bowen (2011) argues that "other policies are needed, too, particularly to promote innovation and appropriate infrastructure investment, but cannot be relied upon by themselves to bring about the necessary reductions to emissions. Carbon pricing is crucial".

[25]These findings are in line with ones in Gerlagh and van der Zwaan (2006) who use a long-term top-down model with a decarbonization option through CCS to show that carbon taxes do better for stringent targets, and subsidies do better for modest targets. Instead, this paper analyzes the implications of the second-best instruments for climate policy in a transparent setting.

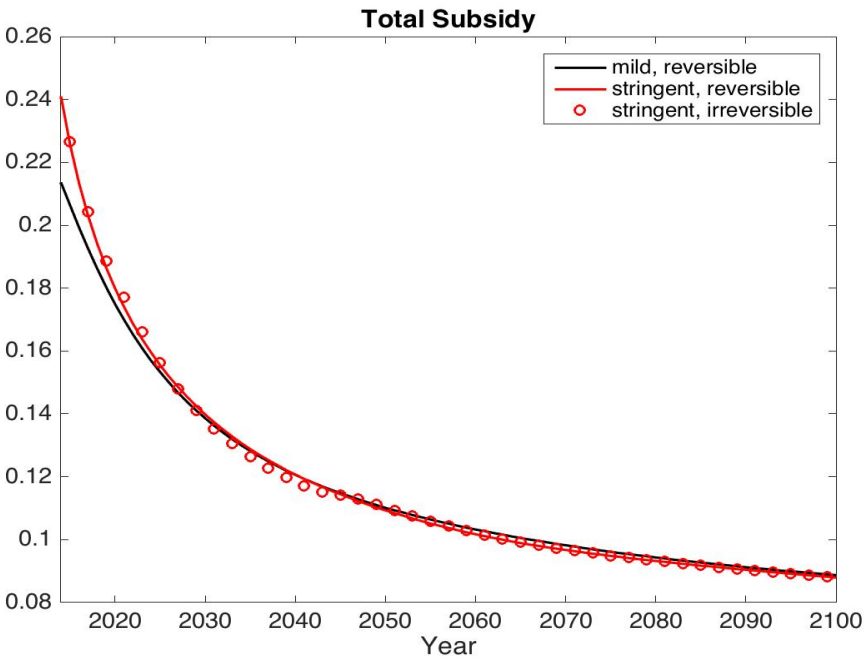

Figure 5: Optimal subsidy: total subsidy

	Optimal tax zero subsidy	Optimal subsidy zero tax
Reversible investment mild climate policy target	1.90%	1.59%
Reversible investment stringent climate policy target	2.52%	5.59%
Irreversible investment stringent climate policy target	2.49%	3.56%

Table 1: Second-best policies: welfare loss, % of initial period consumption

and emissions paths closely follow those under the first-best policy. This is accomplished with a (slightly) higher level of carbon tax than under the first-best scenario. If we consider the more stringent climate policy case (but still with reversible investments), we observe a similar pattern of paths for temperature and emissions: with carbon pricing only, they closely follow the paths of the first-best (the top-right and middle-right panels of Figure 6). The second-best tax level is again higher than the first-best counterpart. The intuition behind these results is as follows.

With only carbon pricing, there is a risk of lock-in into the ways of producing electricity which are currently cheap: coal-based power plants.[26] Meanwhile, the alternative of producing electricity from renewables, is currently more expensive and might not become competitive. As a result, the principal imposes a higher level of carbon taxes on the fossil fuel extracting firms compared with the first-best. But since the size of the dirty sector in the energy sector of the economy is large, this policy of making the sector "less competitive" through carbon taxes is relatively more costly, in terms of welfare, than the policy of making competitive the renewable sector through

[26] see, e.g., Unruh (2002) and Jaffe et al. (2005).

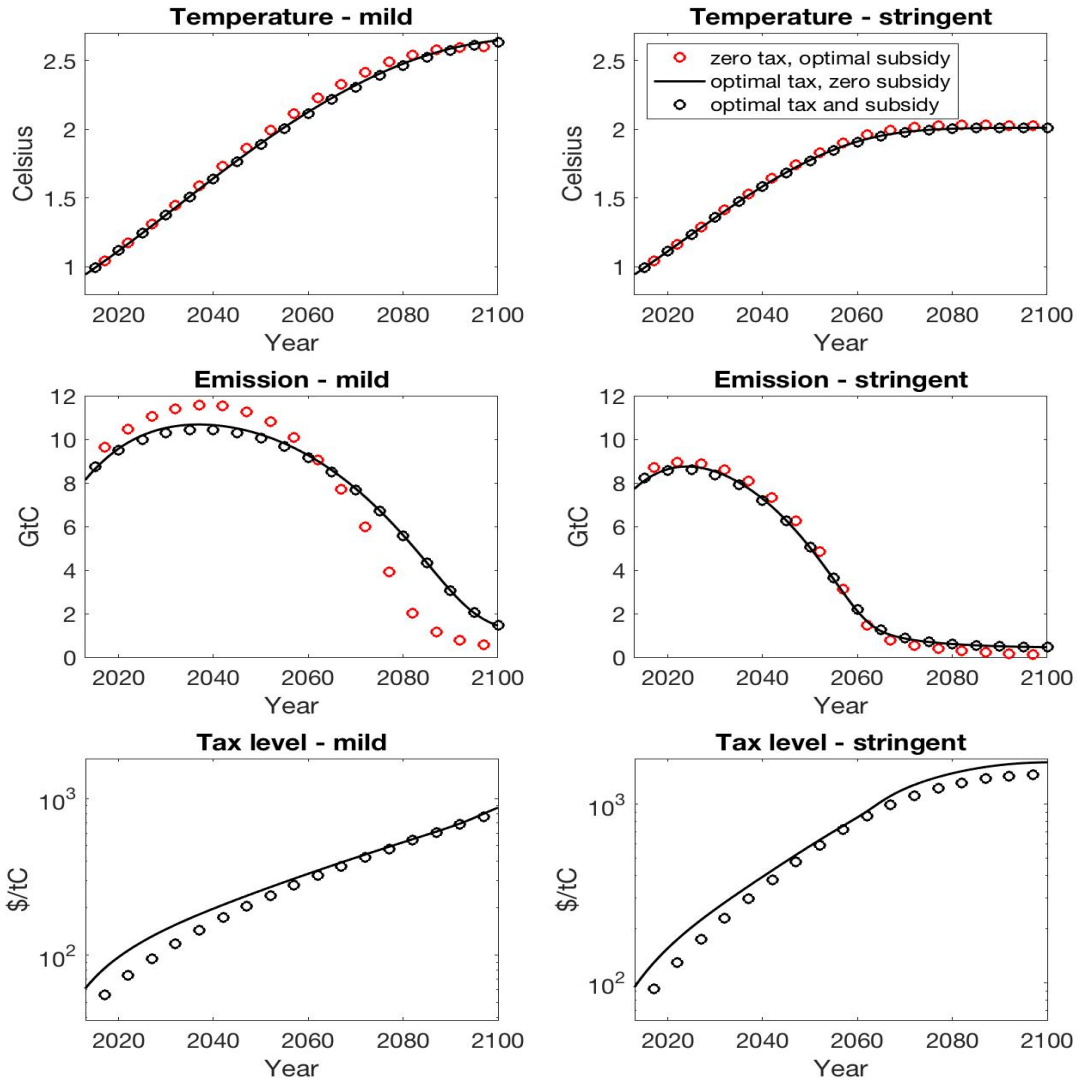

Figure 6: Temperature, Emission, and Tax level under the mild or stringent climate policy targets

direct subsidies. Contrary to carbon pricing, subsidies directly stimulate investment into renewable energy and, once clean technologies develop and become competitive, the renewable sector crowds out the dirty energy sector. Under less ambitious climate policy, this appears sufficient, as well as less costly than carbon pricing (given also the relatively smaller size of the clean sector).

On the other hand, achievement of the more stringent climate policy through innovation policy is extremely difficult as it requires decarbonization of the large dirty energy sector. Adoption of the instrument - carbon pricing - which directly targets that sector is an policy that is associated with relatively higher welfare.

Finally, the emissions and temperature paths with carbon pricing only, irrespective of the assumptions about stringency of climate policy, closely follow the ones of the first-best because carbon pricing internalizes the global warming externality, and thus is better suited to target climate policy

objectives.

6 Discussion

In this paper we have studied implications of two capital stock effects - path dependence in infrastructure and learning-by-doing effect in the renewable sector - for the design of optimal climate policies, using both simple analytical models and simulations of the full dynamic general equilibrium climate-economy model. We define path dependence as irreversibility of investment in both the clean and dirty energy sectors (as opposed to allowing aggregate divestment). We compare the simulation results from our model with irreversibility, with those coming from a model without inertia in the energy sectors.

The simulation results speak to the debate on the characteristics of optimal policy to combat climate change, which involve issues about the timing as well as choice of the instruments to address the problem. On the timing of the climate policy, the debate has centered on whether we should adopt a "gradual slope" approach to the policy, according which we should delay investment into low-carbon emitting technologies and focus on a carbon price that rises gradually. In an alternative approach, it has been urged to accelerate learning-by-doing and to reduce abatement costs of mitigation policies.

We demonstrate that it is optimal to stop investment into the dirty sector earlier - the irreversibility effect - and consequently start earlier investment into the renewable sector. Previous literature has justified the early investment in the renewable sector on the basis of the learning-by-doing effect (see, e.g., van der Zwaan et al. (2002)). We bring a new argument based on the existence of inertia in the dirty sector - the acceleration effect.

On the debate on instruments choice for effective climate policy, our results on the relative performance of carbon pricing versus subsidies in a second-best setting reflect the broad trends in the global climate political landscape. Nowadays we observe rapid expansion of the use of renewable energy technologies.[27] Renewable energy technologies are viewed today as tools to mitigate climate change, to improve local air quality, to advance economic development and to create jobs. Declining costs have played a pivotal role in the expansion of renewable energy technologies in the recent years. The stage for such expansion was set more than decades ago when a handful of countries, such as Germany, Denmark, Spain and the United States, created a critical market for renewables, which drove early economies of scale and led to the changes we witness today (REN21, 2014). During that period and effectively till 2016, when the Paris Agreement came into force, progress in the area of international climate policy had been modest at best. Although the European Union had started campaigning for the 2°C in the mid-1990s, this target was not formally adopted until 2010 at the UN Climate Change Conference in Cancun (Geden, 2013). As such, we could characterize the international climate policy up to 2015 as having unambitious climate policy objectives.

The Paris Agreement, however, rebooted climate political landscape, at least in theory. After Paris, there is a larger recognition of urgency of the measures to set up more ambitious emissions reductions. The agreement has also revived the discussion about importance of adopting carbon pricing to implement the emissions mitigation pledges submitted by 186 countries for the December 2015 Paris Agreement,[28] which is in line with the message from simulations of our model under the second-best setting that more ambitious climate policy should adopt carbon pricing.

[27] Renewables accounted for nearly half of all new power generation capacity in 2014, led by growth in China, the United States, Japan and Germany, with costs continuing to fall (EIA, 2015).

[28] Baranzini et al. (2017) provide a summary of the main arguments in favor of carbon pricing in a post-Paris world. See also Farid et al. (2016) who urge for carbon taxes (or equivalently carbon trading systems) for implementation of the Paris pledges.

Finally, our model has important implications for understanding the problem of stranded assets in the climate policy literature. The narrative on stranded assets that emerges through the results of our model has two principal components. First, to make low carbon alternatives widely available, there is a need to start investment in those technologies earlier enough, to take advantage of scale effects, as well as acceleration effect. Second, a stringent target is necessary as it helps to anchor expectations of both investors and politicians giving them sufficient flexibility and time to make rational decisions regarding investment into fossil fuel energy and related physical capital.

7 Conclusion

This paper has shown that capital stock effects of infrastructure such as coal based power plants are important for design of optimal climate policies. Specifically, we characterize and then quantify the optimal time of ending investments into fossil fuel power plants in a dynamic general equilibrium climate-economy model with irreversible "dirty" and "clean" investments. We find that for temperature changes to not exceed 2°C, investments in dirty infrastructure should end very soon.

We show that the "Green Paradox" – that future stringent climate policy raises short-term emissions – has a converse if we focus on demand side capital stock effects. If the dirty capital stock cannot be converted to other capital, then it is optimal to stop investing into the dirty capital stock earlier than when the capital investments are reversible.

Learning-by-doing significantly advances the timing of investment in renewables, not only to prevent later stranding fossil fuel assets but also to accelerate decline in the costs of clean energy.

The timing of these effects depends, of course, on the stringency of climate policy. Climate policy induces earlier shift to clean energy and away from dirty energy only if it is stringent. Otherwise, path dependence in energy systems and low substitutability between the dirty and clean sources imply a prolonged period of using the dirty capital stock.

References

A. B. Abel. Optimal investment under uncertainty. *American Economic Review*, 73(1):228–233, 1983.

D. Acemoglu, U. Akcigit, D. Hanley, and W. Kerr. Transition to clean technology. *Journal of Political Economy*, 124(1):52–104, 2016.

P. Aghion, C. Hepburn, A. Teytelboym, and D. Zenghelis. Path dependence, innovation and the economics of climate change. *New Climate Economy contributing paper*, 2014.

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

K. Arrow. The economic implications of learning by doing. *The Review of Economic Studies*, 29: 155–173, 1962.

K. Arrow. Optimal capital policy with irreversible investment. In J. N. Wolfe, editor, *Value, Capital and Growth, Essays in Honor of Sir John Hicks*. Edinburgh: Edinburgh University Press, 1968.

K. Arrow and M. Kurz. Optimal growth with irreversible investment in a Ramsey model. *Econometrica*, 38(2):331–344, 1970.

A. Baranzini, J. C. J. M. van den Bergh, S. Carattini, R. Howarth, E. Padilla, and J. Roca. Carbon pricing in climate policy: seven reasons, complementary instruments, and political economy considerations. *WIRES Climate Change*, 8:1–17, 2017.

L. Barrage. Optimal dynamic carbon taxes in a climate-economy model with distortionary fiscal policy. Mimeo., 2014.

J. I. Bernstein and T. P. Mamuneas. Irreversible investment, capital costs and productivity growth: Implications for telecommunications. *Review of Network Economics*, 6(3), 2007.

B. Bollinger and K. Gillingham. leraning-by-doing in solar photovoltaic installations. Available at SSRN: https://ssrn.com/abstract=2342406 or http://dx.doi.org/10.2139/ssrn.2342406, 2014.

A. Bowen. The case for carbon pricing. *Policy Brief, Grantham Research Institute on Climate and the Environment*, 2011.

Y. Cai, K. L. Judd, and T. S. Lontzek. The social cost of carbon with economic and climate risks. Technical Report arXiv:1504.06909v2, arXiv.org, April 2015.

Y. Cai, T. M. Lenton, and T. S. Lontzek. Risk of multiple interacting tipping points should encourage rapid CO2 emission reduction. *Nature Climate Change*, 6(5):520–525, 2016.

S. Davis, K. Caldeira, and H. D. Matthews. Future co2 emissions and climate change from existing energy infrastructure. *Science*, 329:1330–1333, 2010.

A. Dixit. Investment and hysteresis. *Journal of Economic Perspectives*, 6(1):107–132, 1992.

EIA. Energy and climate change: World energy special report. *Report*, 2015.

Energy and Environmental Economics, Inc. Generation cost model for China, December 2012.

M. Farid, M. Keen, M. Papaioannou, I. Parry, C. Pattillo, and A. Ter-Martirosyan. After Paris: Fiscal, macroeconomic, and financial implications of climate change. *IMF Staff Discussion Note, SDN/16/01*, 2016.

C. Fischer and R. Newell. Environmental and technology policies for climate mitigation. *Journal of Environmental Economics and Management*, 55(2):142–162, 2008.

C. Fischer, L. Preonas, and R. Newell. Environmental and technology policy options in the electricity sector: are we deploying too many? *Journal of the Association of Environmental and Resource Economics*, 4(4):959–984, 2017.

R. Fouquet. Trends in income and price elasticities of transport demand (1850-2010). *Energy Policy*, 50:62–71, 2012.

R. Fouquet. Path dependence in energy systems and economic development. *Nature Energy*, N. 16098, 2016.

O. Geden. Modifying the 2C target: Climate policy objectives in the contested terrain of scientific policy advice, political preferences, and rising emissions. *SWP Research Paper*, 2013.

R. Gerlagh. Too much oil. *CESifo Economic Studies*, 57(1):79–102, 2011.

R. Gerlagh and van der Zwaan. Options and instruments for a deep cut in CO_2 emissions: carbon dioxide capture or renewables, taxes or subsidies? *The Energy Journal*, 27(3):25–48, 2006.

R. Gerlagh, S. Kverndokk, and K. E. Rosendahl. Optimal timing of climate change policy: interaction between carbon taxes and innovation externalities. *Environmental and Resource Economics*, 43(3):369–390, 2009.

R. Gerlagh, S. Kverkdokk, and K. E. Rosendahl. The optimal time path of clean energy R&D policy when patents have finite lifetime. *Journal of Environmental Economics and Management*, 67(1): 71–88, 2014.

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

L. H. Goulder and K. Mathai. Optimal CO2 abatement in the presence of induced technological change. *Journal of Environmental Economics and Management*, 39:1–38, 2000.

J. Greenwood, Z. Hercowitz, and P. Krusell. Long-run implications of investment-specific technological change. *The American Economic Review*, 87(3):342–362, 1997.

M. Grubb, T. Chapuis, and M. Ha-Duong. The economics of changing course: Implications of adaptability and inertia for optimal climate policy. *Energy Policy*, 23 (4-5):417–431, 1995.

A. Grubler and S. Messner. Technological change and the timing of mitigation measures. *Energy Economics*, 20 (5-6):495–512, 1998.

J. Hassler, P. Krusell, and C. Olovsson. Energy-saving technical change. Working Paper 18456, National Bureau of Economic Research, October 2012.

D. Helm. *The Carbon Crunch: How We're Getting Climate Change Wrong–and how to Fix it.* Yale University Press, 2012.

D. Hoornweg and M. Freire. Building sustainability in an urbanizing world: A partnership report. Urban Development Series Knowledge Papers 17, World Bank, 2013.

A. B. Jaffe, R. G. Newell, and R. N. Stavins. A tale of two market failures: Technology and environmental policy. *Ecological Economics*, 54:164–174, 2005.

S. Jensen, K. Mohlin, K. Pittel, and T. Sterner. An introduction to the green paradox: the unintended consequences of climate policies. *Review of Environmental Economics and Policy*, 9: 246–265, 2015.

D. Jorgenson. The theory of investment behavior. In R. Ferber, editor, *Determinants of Investment behavior*, pages 129–175. NBER, 1967.

F. Lafond, A. G. Bailey, J. D. Bakker, D. Rebois, R. Zadourian, P. McSharry, and J. D. Farmer. How well do experience curves predict technological progress? A method for making distributional forecasts. *arXiv:1703.05979 [q-fin]*, mar 2017. arXiv: 1703.05979.

A. Lindman and P. Soderholm. Wind power learning rates: a conceptual review and meta-analysis. *Energy Economics*, 34(3):754–761, 2012.

C. McGlade and P. Ekins. The geographical distribution of fossil fuels unused when limiting global warming to 2°C. *Nature*, 517:187–190, 2015.

K. Meng. Estimating path dependence in energy transitions. *unpublished manuscript*, 2016.

T. Michielsen. Brown backstops versus the green paradox. *Journal of Environmental Economics and Management*, 68(1):87–110, 2014.

G. Nemet. Beyond the learning curve: factors influencing cost reductions in photovoltaics. *Energy Policy*, 34(17):3218–3232, 2006.

W. Nordhaus. *A Question of Balance*. Yale University Press, 2008.

W. Nordhaus. Estimates of the social cost of carbon: concepts and results from the DICE-2013R model and alternative approaches. *Journal of the Association of Environmental and Resource Economics*, 1(1/2):273–312, 2014a.

W. Nordhaus. The perils of the learning model for modeling endogenous technological change. *The Energy Journal*, Volume 35(1), Jan. 2014b.

W. Nordhaus. Projections and uncertaintities about climate change in an era of minimal climate policies. *Cowles Foundation discussion paper No. 2057*, 2016.

C. Papageorgiou, M. Saam, and P. Schulte. Substitution between clean and dirty energy inputs - a macroeconomic perspective. *Review of Economics and Statistics*, 2016.

A. Pfeiffer, R. Millar, C. Hepburn, and E. Beinhocker. The '2°C capital stock' for electricity generation: Committed cumulative carbon emissions from the electricity generation sector and the transition to a green economy. *Applied Energy*, 179:1395–1408, Oct. 2016.

R. S. Pindyck. Irreversibility, uncertainty, and investment. *Journal of Economic Literature*, 29(3): 1110–1148, 1991.

R. S. Pindyck. Pricing capital under mandatory unbundling and facilities sharing. *NBER working paper No. 11225*, 2005.

REN21. Renewable energy policy network for the 21st century. the first decade: 2004-2014. 10 years of renewable energy progress. *Report*, 2014.

A. Rezai and F. van der Ploeg. Second-best renewable subsidies to de-carbonize the economy: Commitment and the green paradox. Working Paper 5721, CESifo Group Munich, January 2016.

A. Rezai and F. Van Der Ploeg. Abandoning fossil fuel: How fast and how much. *The Manchester School*, 85:e16–e44, 2017.

J. Rozenberg, A. Vogt-Schilb, and S. Hallegatte. Transition to clean capital, irreversible investment and stranded assets. *Policy Research Working Paper, World Bank, No. 6859*, 2014.

C. Shearer, N. Ghio, L. Myllyvirta, A. Yu, and T. Nace. Boom and bust 2016: Tracking the global coal plant pipeline. *CoalSwarm, Greenpeace and Sierra Club*, 2016.

H.-W. Sinn. Public policies against global warming: a supply side approach. *International Tax and Public Finance*, 15 (4):360–394, 2008.

H.-W. Sinn. Introductory comment; the green paradox: A supply-side view of climate policy. *Review of Environmental Economics and Policy*, 9(2):239–245, 2015.

R. Tol. The optimal timing of greenhouse gas emission abatement, the individual rationality and intergenerational equity. In C. Carraro, editor, *International Environmental Agreements on Climate Change*. Kluwer Academic Publishers, 1999.

G. Unruh. Escaping carbon lock-in. *Energy Policy*, 30:317–325, 2002.

F. van der Ploeg. Cumulative carbon emissions and the green paradox. *Annual Review of Resource Economics*, 5(1):281–300, 2013.

F. van der Ploeg and C. Withagen. Growth, renewables, and the optimal carbon tax. *Internationl Economic Review*, 55(1):283–311, 2014.

B. van der Zwaan, R. Gerlagh, and L. Schrattenholzer. Endogenous technological change in climate change modelling. *Energy economics*, 24(1):1–19, 2002.

A. Vogt-Schilb, G. Meunier, and S. Hallegatte. How inertia and limited potentials affect the timing of sectoral abatements in optimal climate policy. *World Bank Policy Research*, 2012.

M. Weitzman. On modeling and interpreting the economics of catastrophic climate change. *The Review of Economics and Statistics*, 91(1):1–19, 2009.

M. Weitzman. What is the "damage function" for global warming - and what difference might it make? *Climate Change Economics*, 1:57–69, 2010.

T. M. L. Wigley, R. Richels, and J. A. Edmonds. Economic and environmental choices in the stabilization of atmospheric CO2 concentrations. *Nature*, 379 (6562):240–243, 1996.

T. P. Wright. Factors affecting the cost of airplanes. *Journal of the Aeronautical Science*, 3(2): 122–128, 1936.

Online Appendix for "To Build or Not to Build? Capital Stocks and Climate Policy" (For Online-Only Publication)

A Proofs of Theoretical Results: Simplified Model

To start with, we define:

$$P_t := \sum_{s=1}^{\infty} (1-\delta)^{s-1} \Delta_{t,s} \left(r_{t+s} - \delta - e_{t+s} \right). \tag{A.1}$$

This is the net present value of investment in the irreversible asset, infrastructure, relative to the opportunity cost. The following technical lemma is very illuminating:

Lemma A.1. *Given the framework above,*

1. $P_t \leq 0$ *for all* t.
2. $i_t > 0$ *only if* $P_t = 0$.
3. $i_t > 0$ *only if both* $r_t - \delta \leq e_t$ *and* $r_{t+1} - \delta \geq e_{t+1}$.
4. $i_t > 0$ *with* $r_{t+1} - \delta > e_{t+1}$ *only if* $i_{t+1} = 0$.

Proof of Lemma A.1. Write o_t for all other sources of income, net of any other investments (which may also be irreversible). We maximize

$$\sum_{t=1}^{\infty} \beta^t u(c_t) \tag{A.2}$$

subject to constraints

$$\mu_t^{bc} \qquad i_t + c_t = r_t k_t + o_t \tag{A.3}$$

$$\mu_t^{i} \qquad i_t \geq 0 \tag{A.4}$$

$$\mu_t^{k} \qquad i_t \geq k_{t+1} - (1-\delta) k_t \tag{A.5}$$

The Lagrangian is:

$$\mathcal{L}_t = \sum_{t=0}^{\infty} \beta^t \Big(u(c_t) - \mu_t^{bc}(i_t + c_t) + \mu_t^{bc}(r_t k_t + o_t) + \mu_t^{i} i_t \tag{A.6}$$

$$+ \mu_t^{k}(i_t - (k_{t+1} - (1-\delta)k_t)) \Big) \tag{A.7}$$

Leading to FOCs and complementary slack conditions

$$c_t \qquad u'(c_t) = \mu_t^{bc} \tag{A.8}$$

$$i_t \qquad \mu_t^{bc} = \mu_t^i + \mu_t^k \tag{A.9}$$

$$k_{t+1} \qquad \mu_t^k = \beta(\mu_{t+1}^{bc} r_{t+1} + \mu_{t+1}^k(1-\delta)) \tag{A.10}$$

$$\mu_t^i \geq 0 \tag{A.11}$$

$$\mu_t^i i_t = 0 \tag{A.12}$$

$$\mu_t^k \geq 0 \tag{A.13}$$

$$\mu_t^k(i_t - (k_{t+1} - (1-\delta)k_t)) = 0 \tag{A.14}$$

Substitute (A.8) and (A.9) into (A.10) and divide by $\beta u'(c_{t+1})$:

$$\frac{u'(c_t)}{\beta u'(c_{t+1})}\left(1 - \frac{\mu_t^i}{\mu_t^{bc}}\right) = r_{t+1} + \left(1 - \frac{\mu_{t+1}^i}{\mu_{t+1}^{bc}}\right)(1-\delta) \tag{A.15}$$

Write $e_{t+1} := \frac{u'(c_t)}{\beta u'(c_{t+1})} - 1$ and $\Delta_{t,s} = \prod_{s'=1}^{s} \frac{1}{1+e_{t+s'}}$. Re-arrange so that this will provide a forward-looking formula for $\frac{\mu_t^i}{\mu_t^{bc}}$:

$$\begin{aligned}
\frac{\mu_t^i}{\mu_t^{bc}} &= \frac{e_{t+1} - (r_{t+1} - \delta)}{e_{t+1}+1} + \frac{(1-\delta)}{(e_{t+1}+1)}\frac{\mu_{t+1}^i}{\mu_{t+1}^{bc}} \\
&= \frac{e_{t+1} - (r_{t+1} - \delta)}{e_{t+1}+1} + \frac{1-\delta}{e_{t+1}+1}\left(\frac{e_{t+2} - (r_{t+2} - \delta)}{e_{t+2}+1} + \frac{(1-\delta)}{(e_{t+2}+1)}\frac{\mu_{t+2}^i}{\mu_{t+2}^{bc}}\right) \\
&= \sum_{s=1}^{T}(1-\delta)^{s-1}\Delta_{t,s}\left(e_{t+s} - r_{t+s} + \delta\right) + (1-\delta)^T \Delta_{t,T}\frac{\mu_{t+T}^i}{\mu_{t+T}^{bc}}.
\end{aligned} \tag{A.16}$$

Next we will show that the final term in (A.16) tends to zero as $T \to \infty$. Since we assumed that there exist $\epsilon > 0$ and $R \gg 0$ with $-\delta + \epsilon < e_t < R$ for all t, it follows that $\frac{1-\delta}{1+e_t} < 1 - \frac{\epsilon}{1+e_t} < 1 - \frac{\epsilon}{R+1}$ for all t and so that $(1-\delta)^T \Delta_{t,T} \to 0$ as $T \to \infty$. Finally, $0 \leq \mu_T^i \leq \mu_T^{bc}$ for all T, by consideration of (A.9) and (A.13). It follows that $0 \leq \frac{\mu_T^i}{\mu_T^{bc}} \leq 1$, and hence the final term in (A.16) tends to 0 as $T \to \infty$, and we conclude:

$$\frac{\mu_t^i}{\mu_t^{bc}} = \sum_{s=1}^{\infty}(1-\delta)^{s-1}\Delta_{t,s}\left(e_{t+s} - r_{t+s} + \delta\right) =: -P_t \tag{A.17}$$

$$\text{with per-period equation:} \qquad \Delta_{t,1}^{-1}\frac{\mu_t^i}{\mu_t^{bc}} = (e_{t+1} - r_{t+1} + \delta) + (1-\delta)\frac{\mu_{t+1}^i}{\mu_{t+1}^{bc}} \tag{A.18}$$

Part 1 of Lemma A.1 follows from (A.17). Next, if $i_t > 0$, complementary slackness (A.12) tells us $\mu_t^i = 0$ and so Part 2 follows from (A.17).

If $i_t > 0$ then by (A.12) $\mu_t^i = 0$, and since $\mu_{t+1}^i \geq 0$ and $\mu_{t-1}^i \geq 0$, (A.18) implies $r_{t+1} - \delta \geq e_{t+1}$ and $r_t - \delta \leq e_t$. In addition, Part 4 follows, in the same way as the previous result: if $i_t > 0$ with $r_{t+1} - \delta > e_{t+1}$, then (A.18) implies $\mu_{t+1}^i > 0$ and then $i_{t+1} = 0$ from (A.12). □

Proof of Proposition 2.1. Immediate from Lemma A.1 Part 3. □

Proof of Proposition 2.2. If $r_{s_1} - \delta < e_{s_1}$ then $i_{s_1-1} = 0$ (by Lemma A.1 Part 3). However, by assumption, $i_0 > 0$. Let t_0 be maximal such that $t_0 < s_1$ and $i_{t_0} > 0$. Now, by Lemma A.1 Part 2, $P_{t_0} = 0$. So:

$$\begin{aligned} 0 = P_{t_0} &= \sum_{s=1}^{s_1-t_0} (1-\delta)^{s-1}\Delta_{t_0,s}(r_{t_0+s} - \delta - e_{t_0+s}) + \sum_{s=s_1-t_0+1}^{\infty} (1-\delta)^{s-1}\Delta_{t_0,s}(r_{t_0+s} - \delta - e_{t_0+s}) \\ &= \sum_{s=t_0+1}^{s_1} (1-\delta)^{s-t_0-1}\Delta_{t_0,s-t_0}(r_s - \delta - e_s) + \sum_{s=1}^{\infty}(1-\delta)^{s_1-t_0+s-1}\Delta_{t_0,s_1-t_0+s}(r_{s_1+s} - \delta - e_{s_1+s}) \end{aligned} \tag{A.19}$$

It is easy to show that, for any t_1, t_2, we have $\Delta_{0,t_1}\Delta_{t_1,t_2} = \Delta_{0,t_1+t_2}$. Thus $\Delta_{0,t_0}\Delta_{t_0,s_1-t_0+s} = \Delta_{0,s_1+s}$. It also follows that $\Delta_{0,s_1}\Delta_{s_1,s} = \Delta_{0,s_1+s}$, and that $\Delta_{0,t_0}\Delta_{t_0,s_1-t_0} = \Delta_{0,s_1}$. Putting these facts together we see that $\Delta_{t_0,s_1-t_0+s} = \Delta_{t_0,s_1-t_0}\Delta_{s_1,s}$ So, continuing from (A.19), we see

$$\begin{aligned} P_{t_0} = &\sum_{s=t_0+1}^{s_1} (1-\delta)^{s-t_0-1}\Delta_{t_0,s-t_0}(r_s - \delta - e_s) \\ &+ (1-\delta)^{s_1-t_0}\Delta_{t_0,s_1-t_0}\sum_{s=1}^{\infty}(1-\delta)^{s-1}\Delta_{s_1,s}(r_{s_1+s} - \delta - e_{s_1+s}) \\ = &\sum_{s=t_0+1}^{s_1} (1-\delta)^{s-t_0-1}\Delta_{t_0,s-t_0}(r_s - \delta - e_s) + (1-\delta)^{s_1-t_0}\Delta_{t_0,s_1-t_0}P_{s_1}. \end{aligned} \tag{A.20}$$

But $P_{s_1} \leq 0$ by Lemma A.1 Part 1. And $i_{t_0} > 0$ so $r_{t_0} - \delta \leq e_{t_0}$ by Lemma A.1 Part 3. Thus:

$$\sum_{s=t_0+1}^{s_1} (1-\delta)^{s-t_0-1}\Delta_{t_0,s-t_0}(r_s - \delta - e_s) \geq 0.$$

Since $r_{s_1} - \delta - e_{s_1} < 0$ it follows that there exists $s \in \{t_0+1, \ldots, s_1-1\}$ such that $r_s - \delta > e_s$. Letting s_0 be the minimal such s, it is clear that this meets our requirements.

Next, by exactly the same arguments as those used to prove (A.20), and by $P_{s_2} \leq 0$, it follows that

$$\begin{aligned} 0 = P_0 &= \sum_{s=1}^{s_2}(1-\delta)^{s-1}\Delta_{0,s}(r_s - \delta - e_s) + (1-\delta)^{s_2}\Delta_{0,s_2}P_{s_2} \\ &\leq \sum_{s=1}^{s_2}(1-\delta)^{s-1}\Delta_{0,s}(r_s - \delta - e_s) \end{aligned}$$

By splitting the sum into terms with $s \in \{1, \ldots, s_1-1\}$ and $s \in \{s_1, \ldots, s_2\}$, and rearranging, we obtain the expression given. □

Proof of Corollary 2.3. First, see that without the constraint $I_t \geq 0$ we have $\tilde{r}_t - \delta = e_t$ for all t.

Next, since $I_0 > 0$ we know $P_0 = 0$ by Lemma A.1 Part 2. If $r_t - \delta = e_t = \tilde{r}_t - \delta$ for all t then $K_t = \tilde{K}_t$ for all t, but this is not possible since $\tilde{I}_{t_1} < 0$ and $I_{t_1} \geq 0$. If we assume $r_t - \delta \geq e_t$ for all t we must conclude also $r_t - \delta > e_t$ for some t, whence $P_0 > 0$, which is a contradiction. So there exist some minimal s_1 such that $r_{s_1} - \delta < e_{s_1}$ and some maximal $s_2 \in \mathbb{R} \cup \{\infty\}$ such that $s_2 \geq s_1$ and $r_t - \delta < e_t$ for $t \in \{s_1, \ldots, s_2\}$. Applying Proposition 2.2 we conclude that there exists

$s_0 \leq s_1 - 1$ such that $r_{s_0} - \delta > e_{s_0}$ and such that $I_t = 0$ for $t \in \{s_0, \ldots, s_2 - 1\}$. Pick s_0 minimal with these properties.

We show that s_0 is minimal such that $r_t - \delta \neq e_t$. First, by definition of s_1, there is no $t < s_0$ with $r_t - \delta < e_t$. Next, if $r_t - \delta > e_t$ for $t < s_0$ then there exists $t' \in \{t, \ldots, s_0 - 1\}$ such that $I_{t'} > 0$ (for otherwise s_0 is not minimal as defined). But $P_0 = 0$ and $P_{t'} = 0$ imply that there must also exist $t'' \in \{1, \ldots, t'\}$ such that $r_{t''} - \delta < e_{t''}$, and we already know this is not so.

Since $r_t - \delta = e_t = \tilde{r}_t - \delta$ for $t \in \{0, \ldots, s_0 - 1\}$, it follows that $K_t = \tilde{K}_t$ for $t \in \{0, \ldots, s_0 - 1\}$ and so that $I_{t-1} = \tilde{I}_{t-1} \geq 0$ for $t \in \{0, \ldots, s_0 - 1\}$. So we know $t_1 \geq s_0$.

Next, $r_{s_0} - \delta > e_{s_0} = \tilde{r}_{s_0} - \delta$ so $K_{s_0} < \tilde{K}_{s_0}$; but $K_{s_0-1} = \tilde{K}_{s_0-1}$, so $I_{s_0-1} < \tilde{I}_{s_0-1}$. So set $t_0 := s_0 - 1$.

Finally, by definition $r_{s_1} - \delta < e_{s_1} = \tilde{r}_{s_1} - \delta$, which implies $K_{s_1} > \tilde{K}_{s_1}$. But $K_{t_0+1} < \tilde{K}_{t_0+1}$ and so, since $I_t = 0$ for $t \in \{t_0 + 1, \ldots, s_2 - 1\}$ we conclude that $K_t < \tilde{K}_t$ for $t \leq \{t_0 + 1, \ldots, \min(s_2 - 1, t_1)\}$. Since $s_1 \leq s_2 - 1$ and since $K_{s_1} > \tilde{K}_{s_1}$ we conclude that $\min(s_2 - 1, t_1) = t_1$, i.e. that $K_t < \tilde{K}_t$ for $t \in \{t_0 + 1, \ldots, t_1\}$ as required. □

The Social Planner's problem for Section 3.1 The planner optimizes

$$\sum_{t=0}^{\infty} \beta^t L_t u\left(\frac{C_t}{L_t}\right) \tag{A.21}$$

subject to the constraints:

$$\Lambda_t^s \qquad I_t + C_t = f_t(H_t, O_t) \tag{A.22}$$

$$\mu_t^I \qquad I_t \geq 0 \tag{A.23}$$

$$\mu_t^H \qquad I_t = p_t^H(H_{t+1} - (1-\delta)H_t) \tag{A.24}$$

$$\mu_t^p \qquad p_t^H = G(H_t) \tag{A.25}$$

where $O_t = L_t o_t$ represents all other factors of production in the economy. In our model the planner treats this as exogenous.

At time t, the Lagrangian is

$$\begin{aligned}\mathcal{L}_t = \sum_{t=0}^{\infty} \beta^t \Bigg(& L_t u\left(\frac{C_t}{L_t}\right) - \Lambda_t^s\Big(I_t + C_t - f_t(H_t, O_t)\Big) + \mu_t^I I_t \\ & + \mu_t^H(I_t - p_t^H(H_{t+1} - (1-\delta)H_t)) + \mu_t^p(p_t^H - G(H_t))\Bigg)\end{aligned}$$

the first order conditions are:

$$\partial C_t: \qquad \Lambda_t^s = u'\left(\frac{C_t}{L_t}\right) \tag{A.26}$$

$$\partial H_{t+1}: \qquad p_t^H \mu_t^H = \beta\left(\Lambda_{t+1}^s \frac{\partial f_{t+1}}{\partial H_{t+1}} + \mu_{t+1}^H p_{t+1}^H (1-\delta)\right) - \beta \mu_{t+1}^p G'(H_{t+1}) \tag{A.27}$$

$$\partial I_t: \qquad \Lambda_t^s = \mu_t^H + \mu_t^I \tag{A.28}$$

$$\partial p_t^H: \qquad \mu_t^p = \mu_t^H(H_{t+1} - (1-\delta)H_t) \tag{A.29}$$

together with the constraints above and the inequality $\mu_t^I \geq 0$, which is complementary slack with (A.23).

Proof of Proposition 3.1. Divide (A.27) through by $p_t^H \beta \Lambda_{t+1}^s$, substitute in (A.29) and re-arrange to obtain:

$$R_{t+1} = \frac{\mu_t^H - \beta(1-\delta)\mu_{t+1}^H}{\beta \Lambda_{t+1}^s} = \frac{1}{p_t^H}\frac{\partial f_{t+1}}{\partial H_{t+1}} + \frac{\mu_{t+1}^H}{\Lambda_{t+1}^s}\frac{p_{t+1}^H - p_t^H}{p_t^H}(1-\delta) - \frac{\mu_{t+1}^H}{\Lambda_{t+1}^s}\frac{H_{t+2} - (1-\delta)H_{t+1}}{p_t^H}G'(H_{t+1}) \tag{A.30}$$

if $I_{t+1} > 0$ then, by complementary slackness, $\mu_{t+1}^I = 0$ and so $\mu_{t+1}^H = \Lambda_{t+1}^s$. Thus, multiplying both sides by $\frac{p_t^H}{p_{t+1}^H}$, and substituting in the definition for direct returns we obtain the expression given. □

Proof of Proposition 3.3. Considering first the firm, there is no inter-temporal element to their objective function or constraints and so we can consider their optimization period-by-period; obviously the relevant first-order condition is that

$$\frac{\partial f_t}{\partial H_t} = r_t p_t^H. \tag{A.31}$$

Meanwhile, the household maximizes:

$$\sum_{t=0}^{\infty} \beta^t \frac{L_t}{L_0} u\left(\frac{L_0}{L_t} c_t\right) \tag{A.32}$$

subject to the constraints:

$$\Lambda_t \qquad i_t + c_t = (r_t + \tau_t) p_t^H h_t + o_t \tag{A.33}$$

$$\mu_t^i \qquad i_t \geq 0 \tag{A.34}$$

$$\mu_t^h \qquad i_t = p_t^H (h_{t+1} - (1-\delta)h_t) \tag{A.35}$$

Additionally, the price is constrained by $p_t^H = G(H_t)$, but the household does not take this into account. At time t, the Lagrangian is

$$\mathcal{L}_t = \sum_{t=0}^{\infty} \beta^t \left(\frac{L_t}{L_0} u\left(\frac{L_0}{L_t} c_t\right) - \Lambda_t \Big(i_t + c_t - (r_t + \tau_t) p_t^H h_t - o_t \Big) + \mu_t^i i_t + \mu_t^h (i_t - p_t^H (h_{t+1} - (1-\delta)h_t)) \right)$$

the first order conditions are:

$$\partial c_t : \qquad \Lambda_t = u'\left(\frac{L_0}{L_t} c_t\right) = u'\left(\frac{C_t}{L_t}\right) \tag{A.36}$$

$$\partial h_{t+1} : \qquad p_t^H \mu_t^h = \beta \Lambda_{t+1} (r_{t+1} + \tau_{t+1}) p_{t+1}^H + \beta \mu_{t+1}^h p_{t+1}^H (1-\delta) \tag{A.37}$$

$$\partial i_t : \qquad \Lambda_t = \mu_t^h + \mu_t^i \tag{A.38}$$

together with the constraints above and the inequality $\mu_t^i \geq 0$, which is complementary slack with (A.34).

Substitute (A.31) into (A.37) and rearrange: now this first order condition reads:

$$p_t^H \mu_t^h = \beta \left(\Lambda_{t+1} \frac{\partial f_{t+1}}{\partial H_{t+1}} + \mu_{t+1}^h p_{t+1}^H (1-\delta) \right) + \beta \Lambda_{t+1} \tau_{t+1} p_{t+1}^H \tag{A.39}$$

We seek the equation for τ_{t+1} that will lead to the same solution as in the social planner's problem; as derived above, this is defined by constraints (A.22)–(A.25), first order conditions (A.26)–(A.29) and the inequality $\mu_t^I \geq 0$, which is complementary slack with (A.23). Those equations are all counterparts to the equations of this model, with the exception of A.39: we wish this to imply (A.27). But this will be the case if we set (substituting in also (A.29))

$$\begin{aligned} \Lambda_{t+1} \tau_{t+1} p_{t+1}^H &= -\mu_{t+1}^h (H_{t+2} - (1-\delta) H_{t+1}) G'(H_{t+1}) \\ \Leftrightarrow \tau_{t+1} &= -\frac{\mu_{t+1}^h}{\Lambda_{t+1}} \frac{H_{t+2} - (1-\delta) H_{t+1}}{p_{t+1}^H} G'(H_{t+1}) \end{aligned} \tag{A.40}$$

So if $i_t > 0$, which implies $\mu_{t+1}^h = \Lambda_{t+1}$, then the two models are defined by the same first-order conditions in variables C_t, H_t and I_t. In each case p_t^H is defined by H_t, so if $O_t = L_0 o_t$ for all t then the solutions are equal – that is, this level of subsidy achieves the social optimum (subject to O_t).

We have treated o_t and O_t as exogenous for both the household and the social planner. More generally, a model will allow optimization in all factors of production and sources of income. However, if all externalities except for the learning-by-doing in p_t^H have been internalized, then by the Coase Theorem and the First Welfare Theorem, it follows that the optimal O_t^* for the planner satisfies $O_t^* = L_0 o_t^*$, where o_t^* is optimal for the household, so the solutions to the models coincide. □

Proof of Corollary 3.4. If $p_t^H = G(H_t) = p_0^H \left(\frac{H_t}{H_0}\right)^{-\lambda}$, then

$$G'(H_t) = -\lambda \frac{p_0^H}{H_0} \left(\frac{H_t}{H_0}\right)^{-\lambda-1} = -\lambda \frac{p_0^H}{H_t} \left(\frac{H_t}{H_0}\right)^{-\lambda} = -\frac{\lambda p_t^H}{H_t}$$

Hence, in this case,

$$\tau_t = \lambda \left(\frac{H_{t+1}}{H_t} - (1-\delta) \right)$$

B Calibration

This section describes calibration of the model. We build on the seminal "DICE 2013" climate-economy model of Nordhaus (2014a), which serves as benchmark in the literature and policy applications. Some of the parameter values are drawn from the existing studies, in particular, from Hassler et al. (2012), Papageorgiou et al. (2016) and Rezai and Van Der Ploeg (2017). All the parameter values are summarized in Table A.1. Details of the calibration are as follows:

B.1 Production

Labor L_0 is given for 2012 using United Nations data. We assume it continues to evolve as in DICE 2013. We set the value of elasticity of substitution between general output, Y^g, and electricity, E,

in the final-goods production function, $\kappa = 0.46$, following Rezai and Van Der Ploeg (2017), as a compromise between short-term insubstitutability (Hassler et al. (2012)) and longer-term substitutability. We take the value of θ from Papageorgiou et al. (2016) to be 0.003. The technology weightings A_0^g and A_t^E will be set to match other data. Subsequently, A_t^g evolves as in DICE, and A_t^E evolves in step with it. We set $\alpha = 0.4$ as an approximation of the values Papageorgiou et al. (2016) get in their various specifications, but this is also commonly-used value in the literature. We set the deprecation rate of the general capital stock at $\delta^g = 0.05$ following Rezai and Van Der Ploeg (2017).

In modeling the electricity sector we follow Papageorgiou et al. (2016): we set the value of w at 0.32 (across various specifications, they find $w = 0.19$ to 0.70, with a mean of 0.32). We set the value of the substitution parameter $\xi = 0.46$, in line with their estimates. We find the initial generating capital stocks for the dirty and renewable generating capacity from EIA data.[29] We set A_0^E so that electricity output in the first period matches the EIA data on electricity output in 2012.

In calibrating the prices of fossil and renewable energy capital p_t^D, p_t^H, we set p_t^D to be constant and to match the current price of new coal-fired power stations in China, as these may be the marginal new plants in consideration.[30] For p_t^H, see the section below. Exponential depreciation for fossil and renewable energy capital is calculated so that the net lifetime availability of capital is equal to the general expected lifetime of plants in this sector: 40 and 25 years respectively.

We know the initial value of K_t^D from EIA data for 2012, and D_t from Europe Union data. We assume that initially $\zeta_t = 1$.

The function form of fossil fuel extraction cost is taken from Rezai and Van Der Ploeg (2017), but we calibrate it differently because we are more concerned with the price of coal than oil. So we set γ_1 to represent the cost of coal in 2012 (IEA2014 data), which we have converted to give this cost as a price per GtC CO_2 pollution (so that fuel and pollution will be in a straightforward 1:1 ratio), to give a cost of 0.09 trillion 2010\$ / GtC. We take $S_0 = 2000$.[31] Using the IEA estimate of the cost of coal in 2040 along a given trajectory, and the additional fractional fossil stock use that this would represent, the second parameter of the resource cost equation is calculated to be $\gamma_2 = 1.64$.

We set the value of ϕ_2 in the mitigation expenditure function Ψ_t from DICE2013.

[29] All fossil generating capacity has been included on the 'dirty' side. For renewables, we exclude hydropower, because it is a relatively mature source of electricity (cost are not falling very fast) and its use is constrained by physical geography, with a large fraction of suitable sites already in use (its use cannot expand fast), so this technology does not well represent the features of interest in the model. Since extensive hydropower capacity already exists, the inclusion of existing capacity would severely bias the trajectory of the equation relating renewable capital to cost of renewable capital.

[30] Numbers taken from Energy and Environmental Economics, Inc. (2012).

[31] The proven resource of all fossil fuel resource may be estimated to be 1003 GtC using EIA data. However continued exploration will enlarge these stocks. We use stock figure of 2000 GtC.

Parameter	Value	Units	Definition
L_0	7.10	billion people	Population
A_0^g	2.53		productivity
K_0^g	150.00	trillion 2010$	initial 'general' capital stock
θ	0.003		energy share parameter, global output
α	0.4		share of capital, global output
κ	0.46		elas of substitution bwt energy and capital/labor
ξ	0.46		elas of subs between clean and dirty electricity capital
w	0.32		weight on renewable capital in electricity output
D_0	9.4	GtC	CO_2 emissions in year 2012
D_0^{land}	0.90	GtC	Land-use CO_2 emissions in year 2012
D_0^E	3.30	GtC	Electricity CO_2 emissions in year 2012
D_0^g	5.22	GtC	General economy CO_2 emissions in 2012
ν	0.91	GtC/(tW capacity)	Fuel use & emissions from dirty electricity production
S_0	2000	GtC	Existing stock of fossil fuel (as of 2012)
Y_0	60.11	trillion 2010$	initial gross world output
K_0^D	3.61	tW	initial capital stock of fossil technology
H_0	0.46	tW	Initial renewable-knowledge capital stock
p^D	0.57	trillion 2010$/tW	Price of dirty electricity capital
p_0^H	2.11	trillion 2010$/tW	Initial price of clean electricity capital
δ^g	0.05	year^{-1}	capital stock depreciation rate
δ^D	0.025	year^{-1}	Fossil energy capital depreciation
δ^H	0.04	year^{-1}	Renewable energy capital depreciation
γ_1	0.09	trillion 2010$/GtC	Parameter of fuel extraction costs
γ_2	1.64		Parameter of fuel extraction costs
A_0^E	6.93		productivity of energy production
λ	0.295		rate of learning.
ς_1	0.00267		damage function parameter.
ς_2	2		damage function parameter.
ς_3	0.001		damage function parameter.
ς_4	50		damage function parameter.
ϕ_2	2.8		mitigation expenditure parameter.
ϕ_3	0.01		mitigation expenditure parameter.
σ_0	0.0904	GtC/trillion 2010$	the carbon-equivalent emissions to output ratio.
$\phi_{1,0}$	0.041		backstop costs.

Table A.1: Parameter values

Variable	Definition
c_t	per-household consumption
L_t	population at period t
K_t^g	aggregate capital stock in general economy
K_t^D	aggregate dirty capital stock
H_t	aggregate clean (renewable) capital stock
I_t^g	aggregate investment in general economy
I_t^D	aggregate investment in dirty capital stock
I_t^H	aggregate investment in clean (renewable) capital stock
Ψ_t	abatement
S_t	fossil fuel stock at period t
$G^D(S_t)$	the fossil fuel extraction costs
r_t^D	rate of return on fossil (dirty) capital
r_t^H	rate of return on renewable (clean) capital
r_t^g	rate of return on general capital
Π_t^g	the total profits from sale of the final goods
Π_t^D	the total profits from sale of the dirty fuel based electricity
Π_t^H	the total profits from sale of the clean electricity
Π_t^{DE}	the total profits from sale of the fossil fuel
Π_t^E	the total profits from sale of the aggregate electricity
Π_t	the sum of all profits
π_t	the total profits per-household
p_t^D	the cost of fossil fuel capital
p_t^H	the cost of renewable energy capital
p_t^{EH}	the price of electricity generated by clean power stations
p_t^{ED}	the price of electricity generated by fossil fuel based power plants
p_t^e	the price of aggregate electricity
p_t^{fuel}	the price of dirty fossil fuel
Γ_t^{ED}	electricity generated by fossil-fuel based power plants
$Y_t = f(Y_t^g, E_t)$	total output before damages
Y_t^g	output of the general economy
$E_t = f_t^E(H_t, \Gamma_t^{ED})$	aggregate electricity
ψ_t	utilization rate of dirty capital stock
D_t^E	fossil fuel (e.g., coal) used in production of electricity
D_t^g	fossil fuel used in the general economy

Table A.2: Variables notation and definition

C The Setup of Social Planner's Problem

We will consider two alternative perspectives for returns on investment, which will be relevant in different contexts. First, as in the section 3.1, we define:

Definition C.1. The *shadow returns on investment in the general, dirty and renewable capital*

stocks are defined to be respectively R_t^g, R_t^D and R_t^H so that:

$$R_{t+1}^g := \frac{\mu_t^{Kg} - \beta(1-\delta^g)\mu_{t+1}^{Kg}}{\beta u'\left(C_{t+1}/L_{t+1}\right)} \tag{A.41}$$

$$R_{t+1}^D := \frac{\mu_t^{KD} - \beta(1-\delta^D)\mu_{t+1}^{KD}}{\beta u'\left(C_{t+1}/L_{t+1}\right)} \tag{A.42}$$

$$R_{t+1}^H := \frac{\mu_t^{KH} - \beta(1-\delta^H)\mu_{t+1}^{KH}}{\beta u'\left(C_{t+1}/L_{t+1}\right)} \tag{A.43}$$

where μ_t^{Kg}, μ_t^{KD} and μ_t^H are the shadow prices on the capital accumulation constraints as below.

On the other hand, one might consider the more immediate definitions for direct economic returns to investment:

Definition C.2. The *direct economic returns on investment in the general, dirty and renewable capital stocks* are defined respectively to be r_t^g, r_t^D and r_t^H so that:

$$r_{t+1}^g := \frac{\partial}{\partial K_{t+1}^g}\left(Y_{t+1} - \Psi_{t+1}\right) \tag{A.44}$$

$$r_{t+1}^D := \frac{1}{p_{t+1}^D}\frac{\partial}{\partial K_{t+1}^D}\left(Y_{t+1} - \Psi_{t+1}\right) \tag{A.45}$$

$$r_{t+1}^H := \frac{1}{p_{t+1}^H}\frac{\partial}{\partial H_{t+1}}\left(Y_{t+1} - \Psi_{t+1}\right) \tag{A.46}$$

Here we measure the direct effects of investment on output net of mitigation costs, and the output is

$$Y_t = \Omega(T_t) f(Y_t^g, E_t) \tag{A.47}$$

with $Y_t^g = f_t^g(K_t^g, L_t)$.

The social planner's problem is outlined below. Specifically, it maximizes the social welfare:

$$\sum_{t=0}^{\infty} \beta^t L_t u\left(\frac{C_t}{L_t}\right) \tag{A.48}$$

subject to constraints:

$$Y_t = I_t^g + I_t^D + I_t^H + C_t + G^D(S_t)(D_t^E + D_t^g) + \frac{\phi_{1,t}\eta_t^{\phi_2} Y_t^g}{(1-\eta_t)^{\phi_3}} \qquad \mu_t^{BC} \qquad \text{(A.49)}$$

$$S_{t+1} = S_t - D_t^E - D_t^g \qquad \mu_t^S \qquad \text{(A.50)}$$

$$D_t = D_t^E + D_t^{\text{land}} + D_t^g \qquad \mu_t^D \qquad \text{(A.51)}$$

$$T_t = \mathcal{W}_t(D_0, \ldots, D_{t-1}) \qquad \mu_t^W \qquad \text{(A.52)}$$

$$E_t = f_t^E(H_t, \zeta_t K_t^D) = A_t^E \left(w(H_t)^\xi + (1-w)(\zeta_t K_t^D)^\xi \right)^{1/\xi} \qquad \mu_t^E \qquad \text{(A.53)}$$

$$D_t^E = \nu \zeta_t K_t^D \qquad \mu_t^{DE} \qquad \text{(A.54)}$$

$$D_t^g = \sigma_t(1-\eta_t) Y_t^g \qquad \mu_t^{Dg} \qquad \text{(A.55)}$$

$$\zeta_t \leq 1 \qquad \mu_t^\zeta \qquad \text{(A.56)}$$

$$p_t^H = G(H_t) \qquad \mu_t^{pH} \qquad \text{(A.57)}$$

$$I_t^g = K_{t+1}^g - (1-\delta^g) K_t^g \qquad \mu_t^{Kg} \qquad \text{(A.58)}$$

$$I_t^D = p^D(K_{t+1}^D - (1-\delta^D) K_t^D) \qquad \mu_t^{KD} \qquad \text{(A.59)}$$

$$I_t^H = p_t^H(H_{t+1} - (1-\delta^H) H_t) \qquad \mu_t^{KH} \qquad \text{(A.60)}$$

$$I_t^D \geq 0 \qquad \mu_t^{ID} \qquad \text{(A.61)}$$

$$I_t^H \geq 0 \qquad \mu_t^{IH} \qquad \text{(A.62)}$$

(We do not need to specify $\zeta_t \geq 0$ as this will never be violated in the optimum.) So we calculate the Lagrangian $\mathcal{L}$ as

$$\begin{aligned}
\mathcal{L} \;=\; & \sum_{t=0}^{\infty} \beta^t \left[L_t u\left(\frac{C_t}{L_t}\right) - \mu_t^S(S_{t+1} - S_t + D_t^E + D_t^g) + \mu_t^D(D_t - D_t^E - D_t^{\text{land}} - D_t^g) \right] \\
& + \sum_{t=0}^{\infty} \beta^t \mu_t^W (T_t - \mathcal{W}_t(D_0, \ldots, D_{t-1})) \\
& + \sum_{t=0}^{\infty} \beta^t \mu_t^{BC} \left[\Omega(T_t) f(Y_t^g, E_t) - I_t^g - I_t^D - I_t^H - C_t - G^D(S_t)(D_t^E + D_t^g) - \frac{\phi_{1,t}\eta_t^{\phi_2} Y_t^g}{(1-\eta_t)^{\phi_3}} \right] \\
& - \sum_{t=0}^{\infty} \beta^t \left[\mu_t^E (E_t - f_t^E(H_t, \zeta_t K_t^D)) \right] \\
& + \sum_{t=0}^{\infty} \beta^t \left[\mu_t^{DE}(D_t^E - \nu \zeta_t K_t^D) + \mu_t^{Dg}(D_t^g - \sigma_t(1-\eta_t) Y_t^g) + \mu_t^{pH}(p_t^H - G(H_t)) + \mu_t^\zeta(1-\zeta_t) \right] \\
& + \sum_{t=0}^{\infty} \beta^t \left[\mu_t^{Kg} \left(I_t^g - K_{t+1}^g + (1-\delta^g) K_t^g \right) \right] \\
& + \sum_{t=0}^{\infty} \beta^t \left[\mu_t^{KD} \left(I_t^D - p^D K_{t+1}^D + p^D(1-\delta^D) K_t^D \right) + \mu_t^{ID} I_t^D \right] \\
& + \sum_{t=0}^{\infty} \beta^t \left[\mu_t^{KH} \left(I_t^H - p_t^H H_{t+1} + p_t^H(1-\delta^H) H_t \right) + \mu_t^{IH} I_t^H \right]
\end{aligned}$$

We obtain first order conditions (write as shorthand f_t for $f(Y_t^g, E_t)$, f_t^g for $f_t^g(K_t^g, L_t)$, etc.)

$$\partial C_t: \qquad u'\left(\frac{C_t}{L_t}\right) = \mu_t^{BC} \tag{A.63}$$

$$\partial S_{t+1}: \qquad \beta\mu_{t+1}^S = \mu_t^S + \beta\mu_{t+1}^{BC}\frac{dG^D}{dS}(S_{t+1})(D_{t+1}^E + D_{t+1}^g) \tag{A.64}$$

$$\partial D_t^E: \qquad \mu_t^{DE} = \mu_t^S + \mu_t^D + \mu_t^{BC}G^D(S_t) \tag{A.65}$$

$$\partial D_t^g: \qquad \mu_t^{Dg} = \mu_t^S + \mu_t^D + \mu_t^{BC}G^D(S_t) \tag{A.66}$$

$$\partial D_t: \qquad \mu_t^D = \sum_{m=0}^{\infty}\beta^m\mu_{t+m}^W\frac{\partial \mathcal{W}_{t+m}}{\partial D_t} \tag{A.67}$$

$$\partial T_t: \qquad \mu_t^W = -\mu_t^{BC}\Omega'(T_t)f_t \tag{A.68}$$

$$\partial E_t \qquad \mu_t^E = \mu_t^{BC}\Omega(T_t)\frac{\partial f_t}{\partial E_t} \tag{A.69}$$

$$\partial K_{t+1}^g: \qquad \mu_t^{Kg} = \beta\mu_{t+1}^{Kg}(1-\delta^g) + \beta\mu_{t+1}^{BC}\left(\Omega(T_{t+1})\frac{\partial f_{t+1}}{\partial Y_{t+1}^g} - \frac{\phi_{1,t+1}\eta_{t+1}^{\phi_2}}{(1-\eta_{t+1})^{\phi_3}}\right)\frac{\partial f_{t+1}^g}{\partial K_{t+1}^g} \tag{A.70}$$

$$\partial I_t^g \qquad \mu_t^{Kg} = \mu_t^{BC} \tag{A.71}$$

$$\partial K_{t+1}^D \qquad p^D\mu_t^{KD} = \beta p^D\mu_{t+1}^{KD}(1-\delta^D) + \beta\zeta_{t+1}\left(\mu_{t+1}^E\frac{\partial f_{t+1}^E}{\partial(\zeta_{t+1}K_{t+1}^D)} - \mu_{t+1}^{DE}\nu\right) \tag{A.72}$$

$$\partial I_t^D \qquad \mu_t^{KD} = \mu_t^{BC} - \mu_t^{ID} \tag{A.73}$$

$$\partial H_{t+1} \qquad p_t^H\mu_t^{KH} = \beta p_{t+1}^H\mu_{t+1}^{KH}(1-\delta^H) + \beta\mu_{t+1}^E\frac{\partial f_{t+1}^E}{\partial H_{t+1}} - \beta\mu_{t+1}^{pH}G'(H_{t+1}) \tag{A.74}$$

$$\partial I_t^H \qquad \mu_t^{KH} = \mu_t^{BC} - \mu_t^{IH} \tag{A.75}$$

$$\partial p_t^H: \qquad \mu_t^{pH} = \mu_t^{KH}(H_{t+1} - (1-\delta^H)H_t) \tag{A.76}$$

$$\partial\zeta: \qquad \mu_t^{\zeta} = K_t^D\left(\mu_t^E\frac{\partial f_t^E}{\partial(\zeta_t K_t^D)} - \mu_t^{DE}\nu\right) \tag{A.77}$$

$$\partial\eta_t: \qquad \sigma_t\mu_t^{Dg} = \mu_t^{BC}\frac{\phi_{1,t}\eta_t^{\phi_2-1}}{(1-\eta_t)^{1+\phi_3}}[\phi_2(1-\eta_t) + \eta_t\phi_3] \tag{A.78}$$

together with constraints (A.49)–(A.62) and inequalities $\mu_t^{\zeta} \geq 0$, $\mu_t^{ID} \geq 0$, $\mu_t^{IH} \geq 0$ which are complementary slack with corresponding equations (A.56) and (A.61)-(A.62).

Before we proceed, we substitute (A.63) and (A.71) into (A.70) and use Definitions C.1 and C.2 to prove that $R_{t+1}^g = r_{t+1}^g$. It has been presented in a more compact form from the observations that $Y_{t+1} = \Omega(T_{t+1})f_{t+1}$ and $\Psi_{t+1} = \frac{\phi_{1,t+1}\eta_{t+1}^{\phi_2}}{(1-\eta_{t+1})^{\phi_3}}Y_{t+1}^g$. The form that is most useful for further derivations is (from (A.71)):

$$R_{t+1}^g = r_{t+1}^g - \delta^g = \frac{\mu_t^{BC}}{\beta\mu_{t+1}^{BC}} - 1. \tag{A.79}$$

To prove the proposition 4.2 of the main text, we will use the following results.

Proposition C.3 (**The social cost of carbon**). *In an optimal solution:*

$$\chi_t = -u'\left(\frac{C_t}{L_t}\right)^{-1} \sum_{m=1}^{\infty} \beta^m u'\left(\frac{C_{t+m}}{L_{t+m}}\right) \frac{\partial Y_{t+m}}{\partial D_t}. \tag{A.80}$$

That is, the social cost of carbon is the marginal effect on future welfare of present emissions.

Proof of Proposition C.3 (Social Cost of Carbon). Substitute (A.68) into (A.67), and divide through by μ_t^{BC}, to obtain:

$$\frac{\mu_t^D}{\mu_t^{BC}} = -\sum_{m=0}^{\infty} \beta^m \left(\frac{\mu_{t+m}^{BC}}{\mu_t^{BC}} \Omega'(T_{t+m}) f_{t+m}\right) \frac{\partial \mathcal{W}_{t+m}}{\partial D_t} \tag{A.81}$$

$$= -\sum_{m=1}^{\infty} \beta^m \left(\frac{\mu_{t+m}^{BC}}{\mu_t^{BC}} \Omega'(T_{t+m}) f_{t+m}\right) \frac{\partial \mathcal{W}_{t+m}}{\partial D_t} \tag{A.82}$$

where the sum is from $m = 1$ because $\frac{\partial \mathcal{W}_t}{\partial D_t} = 0$. Next, note that

$$\frac{\partial Y_{t+m}}{\partial D_t} = \Omega'(T_{t+m}) \frac{\partial \mathcal{W}_{t+m}}{\partial D_t} f_{t+m} \tag{A.83}$$

Substituting (A.83), as well as (A.63), into (A.82), we obtain and write this as

$$\chi_t := \frac{\mu_t^D}{u'(C_t/L_t)} = \frac{\mu_t^D}{\mu_t^{BC}} = -u'\left(\frac{C_t}{L_t}\right)^{-1} \sum_{m=1}^{\infty} \beta^m u'\left(\frac{C_{t+m}}{L_{t+m}}\right) \frac{\partial Y_{t+m}}{\partial D_t}. \tag{A.84}$$

Since $\Omega'(T_{t+m}) < 0$, we have $\partial Y_{t+m}/\partial D_t < 0$, then $\chi_t > 0$. We call this term the social cost of carbon. It represents the marginal future welfare effect of emissions in terms of current welfare. □

Proposition C.4. ***[Hotelling with fossil stocks]*** *Write μ_t^S for the shadow price on Equation (A.50) constraining the stock of fossil fuel. Then:*

$$\frac{\mu_{t+1}^S}{u'\left(C_{t+1}/L_{t+1}\right)} = \frac{\mu_t^S}{u'\left(C_t/L_t\right)}(1 - \delta^g + r_{t+1}^g) + \frac{dG^D}{dS}(S_{t+1})(D_{t+1}^E + D_{t+1}^g) \tag{A.85}$$

$$\textit{and so} \quad \frac{\mu_t^S}{u'\left(C_t/L_t\right)} = -\sum_{s=1}^{\infty} \Delta_{t,s} (G^D)'(S_{t+s})(D_{t+s}^E + D_{t+s}^g) \tag{A.86}$$

where $\Delta_{t,s} = \prod_{s'=1}^{s} \frac{1}{1-\delta^g + r_{t+s'}^g}$ is the compound discount factor.

That is, the return on extracting a unit of fossil tomorrow should be equal to the return on extracting an extra unit today, selling it and getting return on it at the rate of interest, less the increase in future extraction cost.

Proof of Proposition C.4 (Hotelling with fossil stocks). Divide (A.64) through by μ_{t+1}^{BC}:

$$\beta \frac{\mu_{t+1}^S}{\mu_{t+1}^{BC}} = \frac{\mu_t^S}{\mu_t^{BC}} \frac{\mu_t^{BC}}{\mu_{t+1}^{BC}} + \beta \frac{dG^D}{dS}(S_{t+1}) \left(D_{t+1}^E + D_{t+1}^g\right)$$

Substitute in (A.79) and divide by β, to obtain the Hotelling rule:

$$\frac{\mu^S_{t+1}}{\mu^{BC}_{t+1}} = \frac{\mu^S_t}{\mu^{BC}_t}(1-\delta^g+r^g_{t+1}) + \frac{dG^D}{dS}(S_{t+1})\left(D^E_{t+1}+D^g_{t+1}\right) \tag{A.87}$$

That is, we proved Equation (A.85) as $\mu^{BC}_t = u'(C_t/L_t)$ from (A.63). To get the infinite sum, repeatedly substitute:

$$\begin{aligned}
\frac{\mu^S_t}{\mu^{BC}_t} &= \frac{1}{1-\delta^g+r^g_{t+1}}\left(\frac{\mu^S_{t+1}}{\mu^{BC}_{t+1}} - (G^D)'(S_{t+1})(D^E_{t+1}+D^g_{t+1})\right) && \text{(A.88)}\\
&= \frac{1}{1-\delta^g+r^g_{t+1}}\left(\frac{1}{1-\delta^g+r^g_{t+2}}\left(\frac{\mu^S_{t+2}}{\mu^{BC}_{t+2}} - (G^D)'(S_{t+2})(D^E_{t+2}+D^g_{t+2})\right)\right. && \text{(A.89)}\\
&\quad \left. - (G^D)'(S_{t+1})(D^E_{t+1}+D^g_{t+1})\right) && \text{(A.90)}\\
&= -\sum_{s=1}^{\infty}\Delta_{t,s}(G^D)'(S_{t+s})\left(D^E_{t+s}+D^g_{t+s}\right) && \text{(A.91)}
\end{aligned}$$

where

$$\Delta_{t,s} = \prod_{s'=1}^{s}\frac{1}{1-\delta^g+r^g_{t+s'}} \tag{A.92}$$

That is, we proved Equation (A.86). □

Proposition C.5. ***[Returns on dirty fuel]***

$$\frac{\partial Y_t}{\partial D^E_t} = \frac{\mu^S_t}{u'\left(C_t/L_t\right)} + \chi_t + G^D(S_t) + \frac{p^D R^D_t}{\zeta_t \nu}. \tag{A.93}$$

That is, in an optimal solution, the marginal productivity of fossil fuel in final output is equal to the shadow value of fossil stocks plus the social cost of carbon, the extraction cost and fraction of the rate of return on investment in K^D (gross of depreciation) which represents fuel use.

Proof of Proposition C.5 (Returns on dirty fuel). Now take (A.65), divide by μ^{BC}_t and substitute in (A.80):

$$\frac{\mu^{DE}_t}{\mu^{BC}_t} = \frac{\mu^S_t}{\mu^{BC}_t} + \chi_t + G^D(S_t)$$

For R^D_{t+1}, divide (A.72) by $\beta p^D \mu^{BC}_{t+1}$ and substitute (A.69), and then (A.73) and (A.63) to obtain

$$\begin{aligned}
\frac{\mu^{KD}_t}{\beta\mu^{BC}_{t+1}} &= \frac{\mu^{KD}_{t+1}}{\mu^{BC}_{t+1}}(1-\delta^D) + \frac{\zeta_{t+1}}{p^D}\left(\Omega(T_{t+1})\frac{\partial f_{t+1}}{\partial E_{t+1}}\frac{\partial f^E_{t+1}}{\partial(\zeta_{t+1}K^D_{t+1})} - \frac{\mu^{DE}_{t+1}}{\mu^{BC}_{t+1}}\nu\right)\\
\Rightarrow \frac{\zeta_{t+1}}{p^D}\left(\frac{\partial Y_{t+1}}{\partial(\zeta_{t+1}K^D_{t+1})} - \frac{\mu^{DE}_{t+1}}{\mu^{BC}_{t+1}}\nu\right) &= \frac{\mu^{KD}_t}{\beta\mu^{BC}_{t+1}} - \frac{\mu^{KD}_{t+1}}{\mu^{BC}_{t+1}}(1-\delta^D) = R^D_{t+1} \qquad \text{(A.94)}
\end{aligned}$$

which could be written as:

$$R^D_{t+1} = \frac{\zeta_{t+1}}{p^D}\left(\Omega(T_{t+1})\frac{\partial f_{t+1}}{\partial E_{t+1}}\frac{\partial f^E_{t+1}}{\partial(\zeta_{t+1}K^D_{t+1})} - \nu\left(\frac{\mu^S_{t+1}}{\mu^{BC}_{t+1}} + \chi_{t+1} + G^D(S_{t+1})\right)\right). \tag{A.95}$$

Now, differentiating (A.47) by D^E_t and multiplying by ν:

$$\nu\frac{\partial Y_{t+1}}{\partial D^E_{t+1}} = \Omega(T_{t+1})\frac{\partial f_{t+1}}{\partial E_{t+1}}\frac{\partial f^E_{t+1}}{\partial(\zeta_{t+1}K^D_{t+1})} \tag{A.96}$$

So:

$$R^D_{t+1} = \frac{\nu\zeta_{t+1}}{p^D}\left(\frac{\partial Y_{t+1}}{\partial D^E_{t+1}} - \frac{\mu^S_{t+1}}{\mu^{BC}_{t+1}} - \chi_{t+1} - G^D(S_{t+1})\right) \tag{A.97}$$

$$\Rightarrow \frac{\partial Y_t}{\partial D^E_t} = \frac{\mu^S_t}{\mu^{BC}_t} + \chi_t + G^D(S_t) + \frac{p^D R^D_t}{\zeta_t\nu} \tag{A.98}$$

□

Lemma C.6. *In the optimal social planner's solution, If $I^H_t > 0$ and $I^H_{t+1} > 0$ then:*

$$\frac{p^H_{t+1}}{p^H_t}r^H_{t+1} = 1 + r^g_{t+1} - \delta^g - \frac{p^H_{t+1}}{p^H_t}(1-\delta^H) + \frac{(H_{t+2} - (1-\delta^H)H_{t+1})}{p^H_t}G'(H_{t+1}) \tag{A.99}$$

Proof of Lemma C.6. Consider the equation for renewable capital (A.74). Dividing by $\beta p^H_t \mu^{BC}_{t+1}$, and substituting in equations (A.69) and (A.76) as well as (A.75), we see

$$\begin{aligned}\frac{\mu^{KH}_t}{\beta\mu^{BC}_{t+1}} &= \frac{p^H_{t+1}}{p^H_t}\frac{(\mu^{BC}_{t+1} - \mu^{IH}_{t+1})}{\mu^{BC}_{t+1}}(1-\delta^H) + \frac{\Omega(T_{t+1})}{p^H_t}\frac{\partial f_{t+1}}{\partial E_{t+1}}\frac{\partial f^E_{t+1}}{\partial H_{t+1}} \\ &\quad - \frac{(\mu^{BC}_{t+1} - \mu^{IH}_{t+1})}{\mu^{BC}_{t+1}}\frac{(H_{t+2} - (1-\delta^H)H_{t+1})}{p^H_t}G'(H_{t+1}). \\ &= \left(1 + \frac{p^H_{t+1} - p^H_t}{p^H_t}\right)\left(1 - \frac{\mu^{IH}_{t+1}}{\mu^{BC}_{t+1}}\right)(1-\delta^H) + \frac{1}{p^H_t}\frac{\partial Y_{t+1}}{\partial H_{t+1}} \\ &\quad - \left(1 - \frac{\mu^{IH}_{t+1}}{\mu^{BC}_{t+1}}\right)\frac{(H_{t+2} - (1-\delta^H)H_{t+1})}{p^H_t}G'(H_{t+1}).\end{aligned}$$

From (A.75) and (A.79), we have

$$\frac{\mu^{KH}_t}{\beta\mu^{BC}_{t+1}} = \frac{\mu^{BC}_t - \mu^{IH}_t}{\beta\mu^{BC}_{t+1}} = (1 + r^g_{t+1} - \delta^g)\left(1 - \frac{\mu^{IH}_t}{\mu^{BC}_t}\right) \tag{A.100}$$

Combining the above two equations and Definition C.2, we have

$$(1+r^g_{t+1}-\delta^g)\left(1-\frac{\mu^{IH}_t}{\mu^{BC}_t}\right)=\left(1+\frac{p^H_{t+1}-p^H_t}{p^H_t}\right)\left(1-\frac{\mu^{IH}_{t+1}}{\mu^{BC}_{t+1}}\right)(1-\delta^H)+r^H_{t+1}\frac{p^H_{t+1}}{p^H_t}$$
$$-\left(1-\frac{\mu^{IH}_{t+1}}{\mu^{BC}_{t+1}}\right)\frac{(H_{t+2}-(1-\delta^H)H_{t+1})}{p^H_t}G'(H_{t+1}).$$

This gives the more general form; when $I^H_t>0$ and $I^H_{t+1}>0$, implying $\mu^{IH}_t=\mu^{IH}_{t+1}=0$, then the version given in the lemma follows. □

D Decentralized Equilibrium

A representative household maximizes:

$$\sum_{t=0}^{\infty}\beta^t\frac{L_t}{L_0}u\left(\frac{L_0}{L_t}c_t\right) \tag{A.101}$$

subject to the constraints:

$$\Lambda_t \qquad i^g_t+i^D_t+i^H_t+c_t=\frac{L_t}{L_0}w_t+\pi_t+r^g_tk^g_t+r^D_tp^D_tk^D_t+r^H_tp^H_th_t$$
$$+\frac{1}{L_0}\left(\tau^D_t(D^E_t+D^g_t)-\tau^H_tp^H_tH_t\right) \tag{A.102}$$
$$\mu^{iD}_t \qquad i^D_t\geq 0 \tag{A.103}$$
$$\mu^{iH}_t \qquad i^H_t\geq 0 \tag{A.104}$$
$$\mu^{kg}_t \qquad i^g_t=k^g_{t+1}-(1-\delta^g)k^g_t \tag{A.105}$$
$$\mu^{kD}_t \qquad i^D_t=p^D_t(k^D_{t+1}-(1-\delta^D)k^D_t) \tag{A.106}$$
$$\mu^{kH}_t \qquad i^H_t=p^H_t(k^H_{t+1}-(1-\delta^H)k^H_t) \tag{A.107}$$

At time t, the Lagrangian is

$$\mathcal{L}_t=\sum_{t=0}^{\infty}\beta^t\Bigg(\frac{L_t}{L_0}u\left(\frac{L_0}{L_t}c_t\right)-\Lambda_t\left(i^g_t+i^D_t+i^H_t+c_t\right)+\Lambda_t\left(\frac{L_t}{L_0}w_t+\pi_t+r^g_tk^g_t+r^D_tp^D_tk^D_t+r^H_tp^H_th_t\right)$$
$$+\frac{\Lambda_t}{L_0}\left(\tau^D_t(D^E_t+D^g_t)-\tau^H_tp^H_tH_t\right)+\mu^{iD}_ti^D_t+\mu^{iH}_ti^H_t+\mu^{kg}_t\left(i^g_t-(k^g_{t+1}-(1-\delta^g)k^g_t)\right)$$
$$+\mu^{kD}_t(i^D_t-p^D_t(k^D_{t+1}-(1-\delta^D)k^D_t))+\mu^{kH}_t(i^H_t-p^H_t(h_{t+1}-(1-\delta^H)h_t))\Bigg) \tag{A.108}$$

the first order conditions are:

$$\partial c_t : \qquad \Lambda_t = u'\left(\frac{L_0}{L_t}c_t\right) = u'\left(\frac{C_t}{L_t}\right) = \left(\frac{C_t}{L_t}\right)^{-\psi} \tag{A.109}$$

$$\partial k^g_{t+1} : \qquad \mu^{kg}_t = \beta(\Lambda_{t+1} r^g_{t+1} + \mu^{kg}_{t+1}(1-\delta^g)) \tag{A.110}$$

$$\partial k^D_{t+1} : \qquad p^D_t \mu^{kD}_t = \beta\left(\Lambda_{t+1} p^D_{t+1} r^D_{t+1} + \mu^{kD}_{t+1} p^D_{t+1}(1-\delta^D)\right) \tag{A.111}$$

$$\partial h_{t+1} : \qquad p^H_t \mu^{kH}_t = \beta\left(\Lambda_{t+1} p^H_{t+1} r^H_{t+1} + \mu^{kH}_{t+1} p^H_{t+1}(1-\delta^H)\right) \tag{A.112}$$

$$\partial i^g_t : \qquad \Lambda_t = \mu^{kg}_t \tag{A.113}$$

$$\partial i^D_t : \qquad \Lambda_t = \mu^{kD}_t + \mu^{iD}_t \tag{A.114}$$

$$\partial i^H_t : \qquad \Lambda_t = \mu^{kH}_t + \mu^{iH}_t \tag{A.115}$$

together with the constraints above and the inequalities $\mu^{iD}_t \geq 0$, $\mu^{iH}_t \geq 0$, which are complementary slack with (A.103) and (A.104).

As usual we combine (A.110) with (A.113) to write:

$$\frac{\Lambda_t}{\beta\Lambda_{t+1}} = 1 - \delta^g + r^g_{t+1} \tag{A.116}$$

Substitute (A.115) into (A.112), divide by Λ_{t+1}, and then substitute in (A.116) and divide by β:

$$p^H_t\left(\frac{\Lambda_t}{\Lambda_{t+1}} - \frac{\mu^{iH}_t}{\Lambda_{t+1}}\right) = \beta\left(p^H_{t+1} r^H_{t+1} + \left(1 - \frac{\mu^{iH}_{t+1}}{\Lambda_{t+1}}\right) p^H_{t+1}(1-\delta^H)\right) \tag{A.117}$$

$$\Leftrightarrow p^H_t(1-\delta^g + r^g_{t+1})\left(1 - \frac{\mu^{iH}_t}{\Lambda_t}\right) = p^H_{t+1} r^H_{t+1} + \left(1 - \frac{\mu^{IH}_{t+1}}{\Lambda_{t+1}}\right) p^H_{t+1}(1-\delta^H) \tag{A.118}$$

$$\Leftrightarrow p^H_{t+1} r^H_{t+1} = p^H_t(1-\delta^g + r^g_{t+1})\left(1 - \frac{\mu^{iH}_t}{\Lambda_t}\right) - p^H_{t+1}(1-\delta^H)\left(1 - \frac{\mu^{iH}_{t+1}}{\Lambda_{t+1}}\right) \tag{A.119}$$

Recall that also $\mu^{iH}_t i^H_t = 0$. So we will be able to combine this result with others below to obtain equations determining i^H_t, and thus we will be able to scale up the household's problem.

Similarly, considering dirty capital, we can substitute (A.114) into (A.111), then substitute in (A.116) to obtain:

$$p^D_{t+1} r^D_{t+1} = p^D_t(1-\delta^g + r^g_{t+1})\left(1 - \frac{\mu^{iD}_t}{\Lambda_t}\right) - p^D_{t+1}(1-\delta^D)\left(1 - \frac{\mu^{iD}_{t+1}}{\Lambda_{t+1}}\right) \tag{A.120}$$

And, again, $\mu^{iD}_t i^D_t = 0$.

Of course, if investment is ongoing ($\mu^{iH}_t = \mu^{iH}_{t+1} = \mu^{iD}_t = \mu^{iD}_{t+1} = 0$) then these two equations are identities between variables we are claiming are "exogenous". In that case, these provide necessary conditions on investment being non-zero (and non-infinite).

Moreover, because the economy is made up of identical agents behaving in this same way, we

may sum complementary slack equations over all these agents to obtain

$$\mu_t^{iH} I_t^H = 0 \tag{A.121}$$
$$\mu_t^{iD} I_t^D = 0 \tag{A.122}$$

Moreover, now we have equations for the solution to the maximization problem, we can scale up from the household level. We have determined that, given prices and rates of return (equations for which follow) aggregate consumption C_t and investments I_t^g, I_t^D, I_t^H are determined by (also using that $p_t^D = p^D$):

$$\begin{aligned} I_t^g + I_t^D + I_t^H + C_t &= L_t w_t + \pi_t + r_t^g K_t^g + r_t^D p_t^D K_t^D + r_t^H p_t^H H_t \\ &\quad + \left(\tau_t^D (D_t^E + D_t^g) - \tau_t^H p_t^H H_t\right) \end{aligned} \tag{A.123}$$
$$I_t^D \geq 0 \tag{A.124}$$
$$I_t^H \geq 0 \tag{A.125}$$
$$I_t^g = K_{t+1}^g - (1-\delta^g) K_t^g \tag{A.126}$$
$$I_t^D = p^D \left(K_{t+1}^D - (1-\delta^D) K_t^D\right) \tag{A.127}$$
$$I_t^H = p_t^H \left(K_{t+1}^H - (1-\delta^H) K_t^H\right) \tag{A.128}$$
$$\frac{u'\left(C_t/L_t\right)}{\beta u'\left(C_{t+1}/L_{t+1}\right)} = 1 - \delta^g + r_{t+1}^g \tag{A.129}$$
$$r_{t+1}^D = (1 - \delta^g + r_{t+1}^g)\left(1 - \frac{\mu_t^{iD}}{u'\left(C_t/L_t\right)}\right) - (1-\delta^D)\left(1 - \frac{\mu_{t+1}^{iD}}{u'\left(C_{t+1}/L_{t+1}\right)}\right) \tag{A.130}$$
$$\mu_t^{iD} \geq 0 \tag{A.131}$$
$$I_t^D \mu_t^{iD} = 0 \tag{A.132}$$
$$r_{t+1}^H = \frac{p_t^H}{p_{t+1}^H}(1 - \delta^g + r_{t+1}^g)\left(1 - \frac{\mu_t^{iH}}{u'\left(C_t/L_t\right)}\right) - (1-\delta^H)\left(1 - \frac{\mu_{t+1}^{iH}}{u'\left(C_{t+1}/L_{t+1}\right)}\right) \tag{A.133}$$
$$\mu_t^{iH} \geq 0 \tag{A.134}$$
$$I_t^H \mu_t^{iH} = 0 \tag{A.135}$$

D.1 Compound interest for the firms' problems

Recall our term $\Pi_t = \Pi_t^g + \Pi_t^D + \Pi_t^H + \Pi_t^{DE} + \Pi_t^E$. We treated that as a lump-sum above. However, in fact the firms are owned by the households, so they choose their activity to maximize the utility pay-off to the households. Thus, for example, the final-goods firms seek to maximize

$$\sum_{t=0}^{\infty} \beta^t \Lambda_t \Pi_t^g \tag{A.136}$$

subject to its production constraints, where Λ_t is exactly the shadow price on the household's budget constraint above. It is equivalent to divide by Λ_0 and so to use a compound discount rate of $q_t := \beta^t \frac{\Lambda_t}{\Lambda_0} = \beta^t \frac{u'(c_t)}{u'(c_0)}$ for the relative price of consumption in period t, expressed in period 0 units.

Moreover, recall from (A.116) that $\frac{\Lambda_t}{\Lambda_{t+1}} = \beta(1 - \delta^g + r^g_{t+1})$. Thus

$$q_t = \beta^t \frac{\Lambda_t}{\Lambda_0} = \frac{\beta\Lambda_t}{\Lambda_{t-1}} \cdot \frac{\beta\Lambda_{t-1}}{\Lambda_{t-2}} \cdots \frac{\beta\Lambda_1}{\Lambda_0} = \prod_{j=1}^{t} \frac{1}{1 - \delta^g + r^g_j} \tag{A.137}$$

$$\frac{q_{t+1}}{q_t} = \frac{1}{1 - \delta^g + r^g_{t+1}} \tag{A.138}$$

D.2 The final-goods firms' problem

The final-good firms maximize

$$\sum_{t=0}^{\infty} q_t \Pi^g_t = \sum_{t=0}^{\infty} q_t \left(\Omega(T_t) f(Y^g_t, E_t) - r^g_t K^g_t - w_t L_t - p^e_t E_t - \frac{\phi_{1,t}\eta_t^{\phi_2}}{(1-\eta_t)^{\phi_3}} Y^g_t - p^{fuel}_t D^g_t \right) \tag{A.139}$$

(remember that $Y^g_t \equiv f^g_t(K^g_t, L_t)$) where D^g_t are fossil fuels used by these firms, p^e_t is the price of electricity and $\frac{\phi_{1,t}\eta_t^{\phi_2}}{(1-\eta_t)^{\phi_3}} Y^g_t$ is spending on abatement by these firms, so that firms face emissions constraint given in every period by:

$$D^g_t = \sigma_t(1 - \eta_t) Y^g_t \tag{A.140}$$

The first order conditions are then:

$$\partial K^g_t : \qquad \Omega(T_t) \frac{\partial f}{\partial Y^g_t} \frac{\partial f^g_t}{\partial K^g_t} = r^g_t + \frac{\phi_{1,t}\eta_t^{\phi_2}}{(1-\eta_t)^{\phi_3}} \frac{\partial f^g_t}{\partial K^g_t} + p^{fuel}_t \sigma_t (1 - \eta_t) \frac{\partial f^g_t}{\partial K^g_t} \tag{A.141}$$

$$\partial L_t : \qquad \Omega(T_t) \frac{\partial f}{\partial Y^g_t} \frac{\partial f^g_t}{\partial L_t} = w_t + \frac{\phi_{1,t}\eta_t^{\phi_2}}{(1-\eta_t)^{\phi_3}} \frac{\partial f^g_t}{\partial L_t} + p^{fuel}_t \sigma_t (1 - \eta_t) \frac{\partial f^g_t}{\partial L_t} \tag{A.142}$$

$$\partial E_t : \qquad \Omega(T_t) \frac{\partial f_t}{\partial E_t} = p^e_t \tag{A.143}$$

$$\partial \eta_t : \qquad p^{fuel}_t \sigma_t = \frac{\phi_{1,t}\eta^{\phi_2 - 1}}{(1-\eta_t)^{1+\phi_3}} [\phi_2(1 - \eta_t) + \eta_t \phi_3] \tag{A.144}$$

Equation (A.141) is an optimal condition for demand of aggregate capital and states that the return on capital is the marginal product of capital minus additional spending on abatement to clean a given fraction of extra emissions and costs of fuel. Equation (A.141) is the counterpart of equation (A.142) for labor demand. Equation (A.143) is an optimal condition for demand of electricity. Finally, equation (A.144) says that the firm reacts to the price of fuel (implicitly to carbon tax) by choosing the level of abatement (equivalently the level of emissions) such that the price of fuel would be equal to the marginal cost of emissions reduction.

D.3 Aggregate electricity producing firms' problem

The firms produce aggregate electricity by combining both electricity generated by fossil-fuel based power plants and electricity generated by renewable energy based power stations. Note that we are taking the output from these two plants, in GW, as inputs priced by p^{EH}_t and p^{ED}_t respectively and

so we do not need to convert by p_t^H and p^D here.

$$\sum_{t=0}^{\infty} q_t \Pi_t^E = \sum_{t=0}^{\infty} q_t \left(p_t^e f_t^E(H_t, \zeta_t K_t^D) - p_t^{EH} H_t - p_t^{ED}(\zeta_t K_t^D) \right) \tag{A.145}$$

FOCs are:

$$p_t^e \frac{\partial f_t^E}{\partial H_t} = p_t^{EH} \tag{A.146}$$

$$p_t^e \frac{\partial f_t^E}{\partial (\zeta_t K_t^D)} = p_t^{ED} \tag{A.147}$$

D.4 The dirty electricity producing firms' problem

The dirty electricity producing firms are fossil-fuel based power stations, which combine existing infrastructure (e.g., coal-based power plants) with fossil fuel, and so maximizes:

$$\sum_{t=0}^{\infty} q_t \Pi_t^D = \sum_{t=0}^{\infty} q_t \left(p_t^{ED}(\zeta_t K_t^D) - r_t^D p^D K_t^D - p_t^{fuel} D_t^E \right) \tag{A.148}$$

where firms face emissions constraint: $D_t^E = \nu \zeta_t K_t^D$, and constraint $\zeta_t \leq 1$. So the Lagrangian is (making the obvious substitution)

$$\sum_{t=0}^{\infty} q_t \left(p_t^{ED}(\zeta_t K_t^D) - r_t^D p^D K_t^D - p_t^{fuel} \nu \zeta_t K_t^D + \mu_t^\zeta (1 - \zeta_t) \right) \tag{A.149}$$

And the first order conditions and constraints are

$$\partial K_t^D : \qquad r_t^D p^D = \left(p_t^{ED} - p_t^{fuel} \nu \right) \zeta_t \tag{A.150}$$

$$\partial \zeta_t : \qquad \mu_t^\zeta = K_t^D \left(p_t^{ED} - p_t^{fuel} \nu \right) \tag{A.151}$$

$$\mu_t^\zeta (1 - \zeta_t) = 0 \tag{A.152}$$

$$\mu_t^\zeta \geq 0 \tag{A.153}$$

where μ_t^ζ is Lagrangian multiplier attached to the above constraint. Thus, if $\zeta < 1$ then $p_t^{ED} = p_t^{fuel} \nu$, and $r_t^D p^D = 0$ or $r_t^D = 0$. Intuitively, when there is underutilization, the market pushes the return on dirty energy capital to zero.

D.5 The fossil-fuel extracting firm's problem

The firm maximizes

$$\sum_{t=0}^{\infty} q_t \Pi_t^{DE} = \sum_{t=0}^{\infty} q_t [p_t^{fuel} - \tau_t^D - G^D(S_t)](D_t^E + D_t^g) \tag{A.154}$$

where τ^D is tax on production of fossil fuels. The firm faces the constraint:

$$S_{t+1} = S_t - (D_t^E + D_t^g) \tag{A.155}$$

to which we assign the shadow price $\tilde{\mu}_t$. So the Lagrangian is

$$\mathcal{L}_t = \sum_{t=0}^{\infty} q_t \Big([p_t^{fuel} - \tau_t^D - G^D(S_t)](D_t^E + D_t^g) \tag{A.156}$$

$$- \tilde{\mu}_t \left(S_{t+1} - S_t + (D_t^E + D_t^g) \right) \Big) \tag{A.157}$$

FOCs are:

$$\partial(D_t^E + D_t^g): \qquad \tilde{\mu}_t = p_t^{fuel} - \tau_t^D - G^D(S_t) \tag{A.158}$$

$$\partial S_{t+1}: \qquad q_t \tilde{\mu}_t = q_{t+1} \left(\tilde{\mu}_{t+1} - (D_{t+1}^E + D_{t+1}^g) \left(G^D\right)'(S_{t+1}) \right) \tag{A.159}$$

Combining the firm's first order conditions yields the standard Hotelling condition, into which we then substitute from (A.138)

$$p_t^{fuel} - \tau_t^D - G^D(S_t) = \frac{q_{t+1}}{q_t} \left(p_{t+1}^{fuel} - \tau_{t+1}^D - G^D(S_{t+1}) - (D_{t+1}^E + D_{t+1}^g) \left(G^D\right)'(S_{t+1}) \right) \tag{A.160}$$

$$= \frac{1}{1 - \delta^g + r_{t+1}^g} \left(p_{t+1}^{fuel} - \tau_{t+1}^D - G^D(S_{t+1}) - (D_{t+1}^E + D_{t+1}^g) \left(G^D\right)'(S_{t+1}) \right) \tag{A.161}$$

which states that the return on extracting an extra unit of fossil fuels, selling and getting a return on it must be equal to the expected capital gain from keeping an extra unit of fossil fuels in the earth, but extracting it tomorrow minus the increase in future extraction costs. As before, we may repeatedly substitute forward to obtain

$$p_t^{fuel} - \tau_t^D - G^D(S_t) = -\sum_{s=1}^{\infty} \Delta_{t,s} (D_{t+s}^E + D_{t+s}^g) \left(G^D\right)'(S_{t+s}) \tag{A.162}$$

$$\text{where } \Delta_{t,s} := \prod_{s'=1}^{s} \frac{1}{1 - \delta^g + r_{t+s'}^g} \tag{A.163}$$

D.6 The renewable firms' problem

In contrast to other sectors, we assume that the firms in the renewable sector are small in the sense that they take the the stock of accumulated knowledge about using the renewable energy H_t as given. The renewable firms receive subsidy of τ_t^H on its dollar-valued holdings of renewable energy capital H_t. The firms take all prices as given, so they maximize:

$$\sum_{t=0}^{\infty} q_t \Pi_t^H = \sum_{t=0}^{\infty} q_t [p_t^{EH} - p_t^H (r_t^H - \tau_t^H)] H_t. \tag{A.164}$$

The first order condition is just:

$$p_t^{EH} = p_t^H (r_t^H - \tau_t^H) \tag{A.165}$$

D.7 The Principal's Problem

In this section we collect all equations we need to solve the decentralized equilibrium model and formulate it as the principal-agent problem:

$$\max_{\tau^D,\tau^H} \sum_{t=0}^{\infty} \beta^t L_t u\left(\frac{C_t}{L_t}\right) \tag{A.166}$$

subject to:

$$\begin{aligned}
I_t^g + I_t^D + I_t^H + C_t &= L_t w_t + \Pi_t + r_t^g K_t^g + r_t^D p^D K_t^D + r_t^H p_t^H H_t \\
&\quad + \left(\tau_t^D (D_t^E + D_t^g) - \tau_t^H p_t^H H_t\right) && \text{(A.167)}\\
I_t^D &\geq 0 && \text{(A.168)}\\
I_t^H &\geq 0 && \text{(A.169)}\\
I_t^g &= K_{t+1}^g - (1-\delta^g) K_t^g && \text{(A.170)}\\
I_t^D &= p^D \left(K_{t+1}^D - (1-\delta^D) K_t^D\right) && \text{(A.171)}\\
I_t^H &= p_t^H \left(K_{t+1}^H - (1-\delta^H) K_t^H\right) && \text{(A.172)}\\
p_t^H &= G(H_t) && \text{(A.173)}\\
D_t^E &= \nu \zeta_t K_t^D && \text{(A.174)}\\
D_t^g &= \sigma(1-\eta_t) Y_t^g && \text{(A.175)}\\
\frac{u'\left(C_t/L_t\right)}{\beta u'\left(C_{t+1}/L_{t+1}\right)} &= 1 - \delta^g + r_{t+1}^g && \text{(A.176)}\\
r_{t+1}^D &= (1-\delta^g + r_{t+1}^g)\left(1 - \frac{\mu_t^{iD}}{u'\left(C_t/L_t\right)}\right) - (1-\delta^D)\left(1 - \frac{\mu_{t+1}^{iD}}{u'\left(C_{t+1}/L_{t+1}\right)}\right) && \text{(A.177)}\\
\mu_t^{iD} &\geq 0 && \text{(A.178)}\\
I_t^D \mu_t^{iD} &= 0 && \text{(A.179)}\\
r_{t+1}^H &= \frac{p_t^H}{p_{t+1}^H}(1-\delta^g + r_{t+1}^g)\left(1 - \frac{\mu_t^{iH}}{u'\left(C_t/L_t\right)}\right) - (1-\delta^H)\left(1 - \frac{\mu_{t+1}^{iH}}{u'\left(C_{t+1}/L_{t+1}\right)}\right) && \text{(A.180)}\\
\mu_t^{iH} &\geq 0 && \text{(A.181)}\\
I_t^H \mu_t^{iH} &= 0 && \text{(A.182)}\\
r_t^g &= \left(\Omega(T_t)\frac{\partial f}{\partial Y_t^g} - \frac{\phi_{1,t}\eta_t^{\phi_2}}{(1-\eta_t)^{\phi_3}} - p_t^{fuel}\sigma_t(1-\eta_t)\right)\frac{\partial f_t^g}{\partial K_t^g} \\
&= \Bigg(\Omega(T_t)(1-\theta)\left[(1-\theta)(Y_t^g)^{1-1/\kappa} + \theta(E_t)^{1-1/\kappa}\right]^{\frac{1/\kappa}{1-1/\kappa}}(Y_t^g)^{-\frac{1}{\kappa}} - \frac{\phi_{1,t}\eta_t^{\phi_2}}{(1-\eta_t)^{\phi_3}} \\
&\quad - p_t^{fuel}\sigma_t(1-\eta_t)\Bigg) A_t^g \alpha (K_t^g)^{\alpha-1}(L_t)^{1-\alpha} && \text{(A.183)}
\end{aligned}$$

$$
\begin{aligned}
w_t &= \left(\Omega(T_t)\frac{\partial f}{\partial Y_t^g} - \frac{\phi_{1,t}\eta_t^{\phi_2}}{(1-\eta_t)^{\phi_3}} - p_t^{fuel}\sigma_t(1-\eta_t)\right)\frac{\partial f_t^g}{\partial L_t} \\
&= \Bigg(\Omega(T_t)(1-\theta)\left[(1-\theta)(Y_t^g)^{1-1/\kappa} + \theta(E_t)^{1-1/\kappa}\right]^{\frac{1/\kappa}{1-1/\kappa}}(Y_t^g)^{-\frac{1}{\kappa}} - \frac{\phi_{1,t}\eta_t^{\phi_2}}{(1-\eta_t)^{\phi_3}} \\
&\qquad - p_t^{fuel}\sigma_t(1-\eta_t)\Bigg)A_t^g(1-\alpha)(K_t^g)^{\alpha}(L_t)^{-\alpha} \qquad \text{(A.184)}
\end{aligned}
$$

$$
p_t^e = \Omega(T_t)\frac{\partial f_t}{\partial E_t} = \Omega(T_t)\theta\left[(1-\theta)(Y_t^g)^{1-1/\kappa} + \theta(E_t)^{1-1/\kappa}\right]^{\frac{1/\kappa}{1-1/\kappa}}E_t^{-1/\kappa} \qquad \text{(A.185)}
$$

$$
p_t^{fuel}\sigma_t = \frac{\phi_{1,t}\eta_t^{\phi_2-1}}{(1-\eta_t)^{1+\phi_3}}[\phi_2(1-\eta_t) + \eta_t\phi_3] \qquad \text{(A.186)}
$$

$$
p_t^{EH} = p_t^e\frac{\partial f_t^E}{\partial H_t} = p_t^e A_t^E w H_t^{\xi-1}\left(wH_t^{\xi} + (1-w)(\Gamma_t^{ED})^{\xi}\right)^{\frac{1-\xi}{\xi}} \qquad \text{(A.187)}
$$

$$
p_t^{ED} = p_t^e\frac{\partial f_t^E}{\partial(\zeta_t K_t^D)} = p_t^e A_t^E(1-w)(\zeta_t K_t^D)^{\xi-1}\left(wH_t^{\xi} + (1-w)(\Gamma_t^{ED})^{\xi}\right)^{\frac{1-\xi}{\xi}} \qquad \text{(A.188)}
$$

$$
p^D r_t^D = \left(p_t^{ED} - p_t^{fuel}\nu\right)\zeta_t \qquad \text{(A.189)}
$$

$$
\mu_t^{\zeta} = K_t^D\left(p_t^{ED} - p_t^{fuel}\nu\right) \qquad \text{(A.190)}
$$

$$
\mu_t^{\zeta}(1-\zeta_t) = 0 \qquad \text{(A.191)}
$$

$$
\mu_t^{\zeta} \geq 0 \qquad \text{(A.192)}
$$

$$
p_t^{EH} = p_t^H(r_t^H - \tau_t^H) \qquad \text{(A.193)}
$$

$$
p_t^{fuel} - \tau_t^D - G^D(S_t) = -\sum_{s=1}^{\infty}\Delta_{t,s}(D_{t+s}^E + D_{t+s}^g)\left(G^D\right)'(S_{t+s}) \qquad \text{(A.194)}
$$

$$
\Delta_{t,s} = \prod_{s'=1}^{s}\frac{1}{1-\delta^g + r_{t+s'}^g} \qquad \text{(A.195)}
$$

$$
D_t = D_t^E + D_t^{\text{land}} + D_t^g \qquad \text{(A.196)}
$$

$$
T_t = \mathcal{W}_t(D_0, \ldots, D_{t-1}) \qquad \text{(A.197)}
$$

$$
S_{t+1} = S_t - (D_t^E + D_t^g) \qquad \text{(A.198)}
$$

D.8 Social planner problem versus decentralized equilibrium

Proof of proposition 4.2 First, from (A.185) and (A.188), we note that:

$$
p_t^{ED} = \Omega(T_t)\frac{\partial f_t}{\partial E_t}\frac{\partial f_t^E}{\partial(\zeta_t K_t^D)} = \nu\frac{\partial Y_t}{\partial D_t^E} \qquad \text{(A.199)}
$$

From (A.189) it follows that:

$$
\frac{p^D r_t^D}{\zeta_t\nu} = \frac{p_t^{ED}}{\nu} - p_t^{fuel} \qquad \text{(A.200)}
$$

And substituting here from the above implies:

$$\frac{\partial Y_t}{\partial D_t^E} = \frac{p^D r_t^D}{\zeta_t \nu} + p_t^{fuel} \tag{A.201}$$

And substituting the expression for p_t^{fuel} from (A.194), we obtain:

$$\frac{\partial Y_t}{\partial D_t^E} = \frac{p^D r_t^D}{\zeta_t \nu} + \tau_t^D + G^D(S_t) - \sum_{s=1}^{\infty} \Delta_{t,s} (D_{t+s}^E + D_{t+s}^g) \left(G^D\right)' (S_{t+s}) \tag{A.202}$$

Recall that in the social planner case solution, the returns on dirty fuel are equal to (see Proposition C.5):

$$\frac{\partial Y_t}{\partial D_t^E} = \frac{\mu_t^S}{u'\left(C_t/L_t\right)} + \chi_t + G^D(S_t) + \frac{p^D R_t^D}{\zeta_t \nu} \tag{A.203}$$

where (see Proposition C.4)

$$\frac{\mu_t^S}{u'\left(C_t/L_t\right)} = -\sum_{s=1}^{\infty} \Delta_{t,s} (G^D)'(S_{t+s}) (D_{t+s}^E + D_{t+s}^g) \tag{A.204}$$

Expression (A.202) is identical to (A.203) when taxes are equal to the social cost of carbon, and when $r_t^D = R_t^D$.

Next, we find the value of subsidies under which the solutions of the social planner's problem and decentralized equilibrium coincide. First, if the investment into the renewable sector continues then $\mu_t^{iH} = \mu_{t+1}^{iH} = 0$, from (A.180) it follows that:

$$r_{t+1}^H = \frac{p_t^H}{p_{t+1}^H} (1 - \delta^g + r_{t+1}^g) - (1 - \delta^H) \tag{A.205}$$

or

$$\frac{p_{t+1}^H}{p_t^H} r_{t+1}^H = (1 - \delta^g + r_{t+1}^g) - \frac{p_{t+1}^H}{p_t^H} (1 - \delta^H) \tag{A.206}$$

Using (A.185), (A.187) and (A.193), we can also write that:

$$r_{t+1}^H = \frac{1}{p_{t+1}^H} \frac{\partial Y_{t+1}}{\partial H_{t+1}} + \tau_{t+1}^H \tag{A.207}$$

Next, we denote the return on clean investment in the social planner's case as $\tilde{r}_{t+1}^H$. Recall that in the social planner solution (Lemma C.6):

$$\frac{p_{t+1}^H}{p_t^H} \tilde{r}_{t+1}^H = (1 + r_{t+1}^g - \delta^g) - \frac{p_{t+1}^H}{p_t^H} (1 - \delta^H) + \frac{H_{t+2} - (1 - \delta^H) H_{t+1}}{p_t^H} G'(H_{t+1}) \tag{A.208}$$

and

$$\tilde{r}_{t+1}^H = \frac{1}{p_{t+1}^H} \frac{\partial Y_{t+1}}{\partial H_{t+1}} \tag{A.209}$$

Comparison of (A.207) with (A.209) yields the value of subsides:

$$\tau_{t+1}^H = r_{t+1}^H - \tilde{r}_{t+1}^H \tag{A.210}$$

But comparison of (A.206) with (A.208), further yields that:

$$\frac{p_{t+1}^H}{p_t^H}\left(r_{t+1}^H - \tilde{r}_{t+1}^H\right) = -\frac{H_{t+2} - (1-\delta^H)H_{t+1}}{p_t^H} G'(H_{t+1}) \tag{A.211}$$

and the level of subsidies:

$$\tau_t^H = -(H_{t+1} - (1-\delta^H)H_t)\frac{G'(H_t)}{p_t^H} \tag{A.212}$$

Finally note that it is straightforward to show that the budget constraint (A.167) is identical to the economy's aggregate constraint as in the social planner's problem after substituting expressions for profits and returns on capital and labor. □