Transition to Clean Technology

Daron Acemoglu
Ufuk Akcigit
Douglas Hanley
William Kerr

Working Paper

Transition to Clean Technology[*]

Daron Acemoglu   Ufuk Akcigit   Douglas Hanley   William Kerr

December 3, 2014

Abstract

We develop a microeconomic model of endogenous growth where clean and dirty technologies compete in production and innovation—in the sense that research can be directed to either clean or dirty technologies. If dirty technologies are more advanced to start with, the potential transition to clean technology can be difficult both because clean research must climb several rungs to catch up with dirty technology and because this gap discourages research effort directed towards clean technologies. Carbon taxes and research subsidies may nonetheless encourage production and innovation in clean technologies, though the transition will typically be slow. We characterize certain general properties of the transition path from dirty to clean technology. We then estimate the model using a combination of regression analysis on the relationship between R&D and patents, and simulated method of moments using microdata on employment, production, R&D, firm growth, entry and exit from the US energy sector. The model's quantitative implications match a range of moments not targeted in the estimation quite well. We then characterize the optimal policy path implied by the model and our estimates. Optimal policy makes heavy use of research subsidies as well as carbon taxes. We use the model to evaluate the welfare consequences of a range of alternative policies.

JEL Classification: O30, O31, O33, C65.
Keywords: carbon cycle, directed technological change, environment, innovation, optimal policy.

1 Introduction

Recent economic research recognizes the importance of the transition to clean technology in controlling and reducing fossil fuel emissions and potentially limiting climate change,[1] while empirical evidence suggests that innovation may switch away from dirty to clean technologies in response to changes in prices and policies. For example, Newell, Jaffe and Stavins (1999) show that following the oil price hikes, innovation in air-conditioners turned towards producing more energy-efficient units compared to the previous focus on price reduction; Popp (2002) finds that higher energy prices are associated with a significant increase in energy-saving innovations; Hassler, Krusell and Olovsson (2011) estimate a trend break in factor productivities in the energy-saving direction following the era of higher oil prices; and Aghion et al. (2014) find a sizable impact of carbon taxes on the direction of innovation in the automobile industry and further provide evidence that clean innovation has a self-perpetuating nature feeding on its past success. Based on this type of evidence, Acemoglu et al. (2012) suggest that a combination of (temporary) research subsidies and carbon taxes can successfully redirect technological change towards cleaner technologies. Several conceptual and quantitative questions remain, however. The first is whether, in the context of a micro-founded quantitative model, reasonable policies can secure a transition to clean technology. The second is whether there is an important role for significant research subsidies conditional on optimally-chosen carbon taxes. The third concerns how rapidly the transition to clean technology should take place under optimal policy.

A systematic investigation of these questions necessitates a microeconomic model of innovation and production where clean and dirty technologies compete given the prevailing policies, and the direction of technological change is determined as a function of these policies.[2] It also necessitates a combination of micro data for the modeling of competition in production and innovation with a quantitative model flexible enough to represent realistic dynamics of carbon emissions and potential climate change. This paper is an attempt in this direction.

Our first contribution is to develop a tractable microeconomic model for this purpose. In our model, which we view as an abstract representation of the energy production and delivery sectors, a continuum of intermediate goods can be produced either using a dirty or clean technology, each of which has a knowledge stock represented by a separate quality ladder. Given production taxes—which differ by type of technology, and thus can act as a "carbon

---

[1]On climate change, see, e.g., Stott et al. (2004) on the contribution of human activity to the European heatwave of 2003; Emanuel (2005) and Landsea (2005) on the increased impact and destructiveness of tropical cyclones and Atlantic hurricanes over the last decades; and Nicholls and Lowe (2006) on sea-level rise. On economic costs of climate change, see Mendelsohn et al. (1994), Pizer (1999), and Weitzman (2009). On economic analyses of climate change, see Golosov et al. (2014), Hassler and Krusell (2012), Krusell and Smith (2009), MacCracken et al. (1999), Nordhaus (1994), Nordhaus and Boyer (2000), Nordhaus (2008), and Stern (2007). On endogenous technology and climate change, see, Acemoglu et al. (2012), Bovenberg and Smulders (1995, 1996), Goulder and Mathai (2000), Goulder and Schneider (1999), Grimaud et al. (2011), Hartley et al. (2011), Hassler, Krusell and Olovsson (2011), Popp (2002, 2004), and Van der Zwaan et al. (2002).

[2]Acemoglu et al. (2012) assume that clean and dirty inputs are combined with a constant elasticity of substitution, which allows for a limited form of competition between clean and dirty technologies.

tax"—final good producers choose which technology to utilize. Profit-maximizing firms also decide whether to conduct research to improve clean or dirty technologies. Clean research, for example, leads to an improvement over the leading-edge clean technology in one of the product lines, though there is also a small probability of a breakthrough which will build on and surpass the dirty technology when the dirty technology is the frontier in the relevant product line. Research and innovation decisions are impacted by both policies and the current state of technology. For example, when clean technology is far behind, most research directed to that sector will generate incremental innovations that cannot be profitably produced (unless there are very high levels of carbon taxes). However, if clean research can be successfully maintained for a while, it gradually becomes self-sustaining as the range of clean technologies that can compete with dirty ones expands as a result of a series of incremental innovations.

Our second contribution is to estimate parameters of this model using microdata on R&D expenditures, patents, sales, employment and firm entry and exit from a sample of US firms in the energy sector. The data we use for this exercise are from the Census Bureau's Longitudinal Business Database and Economic Censuses, the National Science Foundation's Survey of Industrial Research and Development, and the NBER Patent Database. We design our sample around innovative firms in the energy sector that are in operation during the 1975-2004 period.[3] We use our sample to estimate two of the key parameters of the model with regression analysis using R&D and patents. We also estimate the initial distribution of productivity gaps between clean and dirty technologies in the economy by allocating the patent stocks of firms innovating in these technology areas across the three-digit industries in which the firms are operating. The remaining key parameters are estimated using simulated method of moments. We show that, despite its parsimony, the fit of the model to a rich and diverse set of moments not targeted in the estimation is fairly good.

We then combine this structure with a flexible model of the carbon cycle (also used in Golosov et al., 2014). Our final and main contribution is to use this estimated quantitative model for the analysis of optimal policy and for a range of counterfactual policy experiments. Though it is intuitive to expect that carbon taxes should do most of the work in the optimal allocation—because they both reduce current emissions and encourage R&D directed to clean technologies—we find a major role for both carbon taxes and research subsidies. The research subsidy is initially more aggressive and then declines over time, while we find that optimal carbon taxes are backloaded (but also start declining after about 130 years). Despite the differences between the models, the reason for the major role for research subsidies is related to the one emphasized in Acemoglu et al. (2012).[4] Research subsidies are powerful in redirecting

---

[3]See Popp (2006) and Jaffe et al. (2010) for background on technology, R&D and innovation in the energy sector.

[4]Major differences between the models include: [1] here the damage from atmospheric carbon is modeled as impacting production, along the lines of previous literature, rather than directly utility; [2] there is no "environmental disaster" threshold, making it possible for us to calibrate the parameters more closely to data and without taking a position on carbon emissions in the rest of the world; [3] in contrast to the constant

technological change, and given this, it is not worth distorting the initial production too much by introducing very high carbon taxes. It is important to emphasize that research subsidies are not used simply to correct a market failure (or an uninternalized externality) in research. In fact, in our model, in the absence of externalities from carbon, or in the special case in which there is only a dirty or a clean sector, the social planner does not have a reason to use research subsidies. This is because a scarce factor, skilled labor, is being used only for research, and thus the social planner cannot increase the growth rate by subsidizing research.[5] The reason why the social planner relies heavily on research subsidies is because when carbon emissions create negative externalities, inducing a transition to clean technology is an effective way of reducing future carbon emissions.

In terms of counterfactual policies, we investigate the welfare costs of just relying on carbon taxes and delaying intervention. Delaying optimal policy by 50 years has a welfare cost equivalent to a permanent 1.7% drop in consumption. The costs of relying only on the carbon tax, without any research subsidies, is similar—1.9%. Another useful comparison is to current US policies. We estimate the effective research subsidy from the differential between clean and dirty firms in our sample in the use of federally funded R&D expenditure. Utilizing this estimate and different values of effective carbon tax at the moment and its likely values in the future, our estimated optimal policies are quite different from their US counterparts, and we show that under current US policies, climate change dynamics will be significantly worse.

We also consider several variations and robustness checks to show which aspects of the model are important for our main theoretical and quantitative results. In particular, we investigate the implications of using different discount rates and estimates of the damage of carbon concentration on economic activity, allowing for alternative modelling of emissions, different degrees of distortions from research subsidies, different estimates of the microeconomic elasticities in the R&D technology, and different distributions of initial productivity gaps between clean and dirty technologies. Overall, most of the main qualitative and quantitative features of optimal policy are fairly robust across a range of plausible variations, though a few variations lead to optimal policies that are more aggressive and involve carbon taxes reaching higher levels and remaining high for 200 years or more.

Our model combines elements from four different lines of research, and is thus related to each of these four lines. First, we build on the growing literature on quantitative general equilibrium models of climate change, such as Golosov et al. (2014), Hassler and Krusell (2012),

---

elasticity of substitution formulation of Acemoglu et al. (2012), dirty and clean sectors are not (q-)complements in our model, but explicitly compete in each product line; [4] because of breakthrough innovations, even with fairly aggressive research subsidies and carbon taxes, the dirty sector is eliminated very slowly (in more than 1000 years); and [5] we also allow for research subsidies to create distortions, which plays some role in our quantitative policy conclusions. The third one is one of the most important microeconomic difference, making the production structure more realistic and enabling us to use microdata on innovation and production.

[5] The social planner could affect growth by reallocating research activity between incumbents and entrants, but this is neither quantitatively important nor allowed in our analysis (because we restrict subsidies/taxes on the incumbents and entrants to be the same).

Adao et al. (2012), Krusell and Smith (2009), Nordhaus (1994), Nordhaus and Boyer (2000), Nordhaus (2008), and Stern (2007). We follow these papers in introducing a simple model of the carbon cycle and the economic costs of carbon emissions in a general equilibrium model, and then characterizing optimal policy. Second, we introduce endogenous and directed technological change along the lines of Acemoglu (1998, 2002) in a model where producers have a choice between clean and dirty production methods. In combining these two first lines of research, we are following Acemoglu et al. (2012) as well as several other papers listed in footnote 1 above. Third, we develop a tractable but rich model of competition between dirty and clean technologies building on the literature on step-by-step competition as in Harris and Vickers (1995), Aghion et al. (2001), and Acemoglu and Akcigit (2012). Fourth, we model the microeconomics of innovation, employment and output dynamics building on Klette and Kortum (2004), where each firm consists of a number of products and technologies (different from other applications, technologies here are different from products because of the competition between clean and dirty sectors).

In estimating a general equilibrium model of firm-level innovation and employment dynamics, we follow Lentz and Mortensen (2008), Akcigit and Kerr (2010), and Acemoglu et al. (2013). We differ from existing work in this area in three important respects, however. First, we combine this type of estimation strategy with a model of clean and dirty technologies and estimate some of the parameters of the R&D technology directly from microdata. Second, rather than focusing on steady-state comparisons, we study non-steady-state dynamics, which is crucial for the question of transitioning to clean technology. Third, we characterize optimal policies in such a framework.

The remainder of the paper is organized as follows. Section 2 introduces our model and characterizes the equilibrium. Section 3 describes the dataset we use for estimation and quantitative evaluation, outlines the different components of our estimation strategy, and presents the estimates of some of the parameters we obtain from our micro data. Section 4 presents the simulated method of moments estimates of our parameters and discusses the fit of the model. Section 5 quantitatively characterizes the structure of optimal environmental policy. In this section, we also conduct a range of counterfactual exercises. Section 6 discusses a range of robustness exercises intended to convey which sorts of assumptions and parameters are important for the qualitative and quantitative results of the paper. Section 7 concludes, while the Online Appendix contains additional details and proofs.

2 Model

This section presents our baseline model, which is a simple dynamic general equilibrium setup where final output combines intermediates produced either using a clean or dirty technology, and the dirty technology also uses an exhaustible resource such as oil. The productivity of the dirty and clean technology for each intermediate is represented by a quality ladder. Production

is also subject to taxes, so profit-maximizing final good producers choose whether to use clean or dirty intermediates as a function of taxes and the productivity gap between the two. Research is conducted both by entrants and incumbent firms, which already hold a portfolio of products and technologies, and is directed towards clean or dirty technology. Finally, dirty technology contributes to carbon emissions, which create potential economic damage. We next describe each aspect of the model in turn and characterize the equilibrium.

2.1 Preferences and Endowments

We model an infinite-horizon closed economy in continuous time. Since the consumer side is not our focus, we simplify the discussion by modeling it with a representative household with a logarithmic instantaneous utility function and lifetime utility given by

$$U_0 = \int_0^\infty e^{-\rho t} \ln C_t dt,$$

where C_t is the representative household's consumption at time t and $\rho > 0$ is the discount rate. The representative household is comprised of unskilled workers, with measure normalized to 1, and "scientists," with measure L^s, who will be employed in R&D activities. All workers supply one unit of labor inelastically. The household owns all the firms in the economy, so it maximizes lifetime utility subject to the following budget constraint

$$w_t^u + w_t^s L^s + \Pi_t - T_t \geq C_t,$$

and the usual no Ponzi condition. Here Π_t is the total sum of corporate profits net of R&D expenses, w_t^u and w_t^s are the wage rates (and thus wage incomes) of the unskilled and R&D workers, and T_t is the net lump-sum tax (or transfer) used for balancing the government budget. Because research subsidies may create additional distortions (denoted by D_t), there may be a wedge between consumption and output, so that $C_t + D_t = Y_t$.

2.2 Final Good Technology, Intermediate Production and Pricing

The final good is produced by combining a measure one of intermediates with an elasticity of substitution equal to one. In addition, its production is negatively affected by the amount of atmospheric carbon concentration, which we denote by S_t. We follow the formulation suggested by Golosov et al. (2014), which builds on earlier work by Mendelsohn et al. (1994), Nordhaus (1994, 2008) and Stern (2007), and assume

$$\ln Y_t = -\gamma \left(S_t - \bar{S} \right) + \int_0^1 \ln y_{i,t} di, \tag{1}$$

where $\bar{S} > 0$ is the pre-industrial level of the atmospheric carbon concentration, $\gamma \geq 0$ is a scale parameter, and $y_{i,t}$ is the quantity of intermediate good i. When $\gamma = 0$, (1) gives the standard aggregate production function for combining intermediates to produce a final good

5

with unit elasticity of substitution. When $\gamma > 0$, levels of atmospheric carbon concentration above the pre-industrial level reduce productivity with an elasticity of γ.

A feature of (1), which will play a central role in our quantitative exercise, is worth noting: the proportional cost of a unit increase in atmospheric carbon concentration is independent of its current level. Though non-linearities, or even major threshold effects, are likely to be present in the impact of atmospheric concentration on economic activity, this functional form is consistent with assumptions made by other economic approaches to climate change (e.g., Nordhaus, 1994, 2002, 2007; Nordhaus and Boyer, 2000; Stern, 2007; Golosov et al., 2014).

Each intermediate $i \in [0,1]$ can be produced with either a dirty or a clean technology, and when it is produced with the clean (dirty) technology, we denote it by $y_{i,t}^c$ ($y_{i,t}^d$). We will sometimes refer to clean and dirty technologies as clean and dirty "sectors," and also use the terms "intermediates" and "product lines" interchangeably.

Firm f can produce intermediate i with either a clean or dirty technology ($j \in \{c,d\}$). The production function for clean technology is

$$y_{i,t}^c(f) = q_{i,t}^c(f) \, l_{i,t}^c(f),$$

where $l_{i,t}^c(f)$ is employment of production workers and $q_{i,t}^c(f)$ is the labor productivity of the technology that this firm has access to for producing with clean technology in product line i.[6] Focusing on the firm with the most productive clean technology for this intermediate and suppressing firm indices, we often write this as $y_{i,t}^c = q_{i,t}^c l_{i,t}^c$.

The production function for dirty technology is similar, except that it also uses an exhaustible resource $e_{i,t}$:

$$y_{i,t}^d(f) = q_{i,t}^d(f) \, l_{i,t}^d(f)^{1-\nu} \, e_{i,t}(f)^\nu, \tag{2}$$

where $\nu \in (0,1)$. We assume that the exhaustible resource is owned by a set of competitive firms, which can extract it with the following technology relying on unskilled labor:

$$e_{i,t} = \zeta l_{i,t}^e,$$

where $\zeta > 0$. This implies that the marginal cost of extraction is $c_{e,t} = w_t^u/\zeta$. The stock of exhaustible resource, R_t, evolves according to the law of motion

$$\dot{R}_t = -\int_0^1 e_{i,t} di. \tag{3}$$

Profit maximization of resource owners combined with this law of motion implies that the price of the exhaustible resource, $p_{e,t}$, follows the Hotelling rule. In particular, this price minus the cost of extraction must grow at the rate of interest (i.e., $p_{e,t} - \zeta w_t^u$ grows at the rate r_t).

---

[6]Capital can be introduced as an additional factor of production following Golosov et al. (2014), but would further complicate our setup. Given our focus on technology choice and endogenous innovation, we have not pursued this generalization.

The cost minimization problem of producers using a dirty technology implies

$$e_{i,t} = \left(\frac{\nu}{1-\nu} \frac{w_t^u}{p_{e,t}} \right) l_{i,t}^d. \tag{4}$$

Combining this with (2) gives a linear relationship between the output of dirty producers and labor

$$y_{i,t}^d = \left(\frac{\nu}{1-\nu} \frac{w_t^u}{p_{e,t}} \right)^\nu \times q_{i,t}^d l_{i,t}^d.$$

Only firms with the most advanced technology for intermediate i within the clean or dirty sector will find it profitable to produce it. However, because of taxes τ_t^j on sector j at time t, it is not necessarily the most advanced technology between these two sectors that will be active. Given these taxes, the marginal cost of production in line i is

$$MC_{i,t} = \begin{cases} \frac{(1+\tau_t^c)w_t^u}{q_{i,t}^c} & \text{if produced with clean technology} \\ \frac{\left(1+\tau_t^d\right)w_t^u \tilde{P}_{e,t}}{q_{i,t}^d} & \text{if produced with dirty technology} \end{cases},$$

where w_t^u is the wage rate of unskilled workers and

$$\tilde{P}_{e,t} \equiv (1-\nu)^{\nu-1} \left(\frac{p_{e,t}}{w_t^u \nu} \right)^\nu \tag{5}$$

is the normalized price of the exhaustible resource at time t. In equilibrium, only the technology with the lower marginal cost inclusive of taxes—or equivalently the one with the higher tax-adjusted labor productivity—will produce. Equilibrium production decisions are:

$$\text{produce intermediate } i \text{ with } \begin{cases} \text{clean technology if } \frac{q_{i,t}^c}{1+\tau_t^c} > \frac{q_{i,t}^d}{\left(1+\tau_t^d\right)\tilde{P}_{e,t}} \\ \text{dirty technology if } \frac{q_{i,t}^c}{1+\tau_t^c} < \frac{q_{i,t}^d}{\left(1+\tau_t^d\right)\tilde{P}_{e,t}} \end{cases}. \tag{6}$$

We assume that if condition (6) holds with equality, each technology produces with probability 50% at any point in time.[7]

Finally, we also assume that at the initial date $t = 0$, for each leading technology of quality $q_{i,0}^j$, there also exists an intermediate good of quality $q_{i,0}^j/\lambda$. This ensures that markups in the initial date will be equal to λ, and this result will hold endogenously in all subsequent dates.

2.3 Innovation, the Quality Ladder and Dynamics

Labor productivity for each intermediate (for each technology) evolves as a result of innovation. Research is directed towards clean or dirty technologies. Successful research leads to one of two types of innovation. The first is an *incremental* innovation, which takes place with probability $1 - \alpha$; and the second is a *breakthrough* innovation, which takes place with probability α (independently of all other events).

---

[7]This condition will hold as an inequality "generically," i.e., for almost all values of the normalized price of the exhaustible resource, $\tilde{P}_{e,t}$.

If research directed to sector $j \in \{c, d\}$ leads to an incremental innovation, then the innovator advances by one rung in the quality ladder over the current leading-edge technology of type j in a randomly chosen intermediate. Breakthrough innovations, on the other hand, enable the successful innovator to improve by one rung over the frontier technology, even if this frontier is set by the alternative technology—i.e., a breakthrough clean innovation improves over the dirty technology even if the latter is far ahead of the clean sector, thus allowing the clean sector to leapfrog the dirty one.[8]

We assume that each rung in the quality ladder corresponds to a proportional improvement of $\lambda > 1$. Consequently, labor productivity of technology j in intermediate i at time t can be expressed as $q_{i,t}^{j} = \lambda^{n_{i,t}^{j}}$, where $n_{i,t}^{j} \in \mathbb{Z}_{+}$ is the effective number of steps that this technology has taken since time $t = 0$ (where the initial levels, $q_{i,0}^{j}$'s, are set equal to 1 for all i and j without loss of any generality). This also implies that an incremental innovation during an interval of time Δt leads to a new technology with $q_{i,t+\Delta t}^{j} = \lambda q_{i,t}^{j}$ (i.e., an improvement of proportional amount λ over the leading technology in the same sector), while a breakthrough innovation in sector j that is behind sector $-j$ (i.e., with $q_{i,t}^{j} < q_{i,t}^{-j}$) leads to a new technology with $q_{i,t+\Delta t}^{j} = \lambda q_{i,t}^{-j}$, meaning that it builds on the more-advanced technology level of sector $-j$.

Given this specification, the relative productivity of dirty to clean technology in intermediate i at time t can be written as

$$\frac{q_{i,t}^{d}}{q_{i,t}^{c}} = \lambda^{n_{i,t}},$$

where $n_{i,t} \equiv n_{i,t}^{d} - n_{i,t}^{c} \in \mathbb{Z}$ is defined as the technology gap between dirty and clean sectors in product line i at time t.

Denoting by z_{t}^{j} the aggregate innovation rate (the sum of incumbents' and entrants' innovation rates) in technology j, the law of motion for the technology gap $n_{i,t}$ can be expressed as

$$n_{i,t+\Delta t} = \begin{cases} n_{i,t} - 1 & \text{with probability} & (1-\alpha)\, z_{t}^{c}\Delta t + o(\Delta t) & \forall n_{i,t} \\ n_{i,t} + 1 & \text{with probability} & (1-\alpha)\, z_{t}^{d}\Delta t + o(\Delta t) & \forall n_{i,t} \\ -1 & \text{with probability} & \alpha z_{t}^{c}\Delta t + o(\Delta t) & \text{if } n_{i,t} > 0 \\ n_{i,t} - 1 & \text{with probability} & \alpha z_{t}^{c}\Delta t + o(\Delta t) & \text{if } n_{i,t} \leq 0 \\ 1 & \text{with probability} & \alpha z_{t}^{d}\Delta t + o(\Delta t) & \text{if } n_{i,t} \leq 0 \\ n_{i,t} + 1 & \text{with probability} & \alpha z_{t}^{d}\Delta t + o(\Delta t) & \text{if } n_{i,t} > 0 \\ n_{i,t} & \text{otherwise} & & \end{cases},$$

where $o(\Delta t)$ represent second-order terms that disappear faster than Δt as Δt goes to zero.

Let us next denote the price-adjusted policy gap by m_{t}, such that

$$m_{t} \equiv \frac{1}{\ln \lambda}\left[\ln\left(\frac{(1+\tau_{t}^{d})}{(1+\tau_{t}^{c})}\tilde{P}_{e,t}\right)\right],$$

---

[8]Note that innovations here have a creative destruction element (e.g., Aghion and Howitt, 1992, Grossman and Helpman, 1991) because, by improving over an existing product typically operated by another firm, they transfer the leading-edge technology to the current innovator.

where $\tilde{P}_{e,t}$ is given by (5). The tax-adjusted technology gap can therefore be written as

$$\frac{q_{i,t}^d}{q_{i,t}^c} \frac{\left(1 + \tau_t^d\right)}{\left(1 + \tau_t^c\right)} \frac{1}{\tilde{P}_{e,t}} = \lambda^{n_{i,t} - m_t}.$$

The leading-edge tax-adjusted technology in intermediate i is dirty if $n_{i,t} > m_t$; the two technologies are neck and neck if $n_{i,t} = m_t$; and clean is the leading technology otherwise.

2.4 Firms, R&D and Free Entry

Following Klette and Kortum (2004), we define a firm as a collection of leading-edge technologies. Let u_f^j denote the number of intermediates where firm f has the leading-edge technology within sector $j \in \{c, d\}$, though these may not be the most advanced technology for the intermediate in question because there is also a competing technology $-j$. We assume that u_f^j captures the stock of knowledge on which the firm can build for further innovations with technology $j \in \{c, d\}$. In view of this and without any loss of generality, we simplify the exposition by assuming that each firm specializes in either clean or dirty technologies. Each firm can then combine its knowledge stock u_f^j (in sector/technology j) with scientists (R&D workers) H_f^j in order to generate a Poisson flow rate of X_f^j new innovations according to the production function

$$X_f^j = \theta \left(H_f^j\right)^\eta \left(u_f^j\right)^{1-\eta}, \tag{7}$$

where $\eta \in (0, 1)$ is the R&D elasticity with respect to scientists and $\theta > 0$. The *variable* cost of generating a flow rate of X^j is therefore $w_t^s u h^j$, where

$$h^j(x^j) = \left(\frac{x^j}{\theta}\right)^{\frac{1}{\eta}} \tag{8}$$

is the demand for skilled workers (as scientists) given innovation intensity per product line of $x^j \equiv X^j / u^j$ and w_t^s is the wage rate of scientists (we have dropped the firm index to simplify notation). There is in addition a fixed cost associated with R&D, which is again in terms of skilled workers. In particular, to perform any R&D firm f will need to hire $F_{I,i}$ scientists per product line, where $F_{I,i,t} \in [(1 - \xi) F_I, (1 + \xi) F_I]$ is an iid draw (across firms and over time) with mean F_I and $\xi \in (0, 1)$.[9] Hence, the total cost of R&D for firm i performing R&D directed at technology $j \in \{c, d\}$ at time t is $C_t\left(u, x^j\right) = w_t^s u \left(h^j + F_{I,i,t}\right)$.

Entrants can also undertake R&D directed to either sector by paying a variable cost as in (8) and a fixed cost in terms of $F_E \geq F_I$ scientists. We denote the endogenously determined mass of entrants performing R&D directed to technology j at time t by E_t^j.

---

[9] This heterogeneity in fixed costs is useful for smoothing out the dynamics. In particular, because of "creative destruction," the equilibrium path in this class of models involves some firms stopping R&D (as clean firms will do without policy change and dirty firms do under our optimal policy). This will lead to a discontinuous behavior shortly before this point because of the expectation that creative destruction will cease. Heterogeneity in fixed costs smooths this transition.

On the policy side, R&D for sector j receives a proportional government subsidy at the rate $s_t^j \in [0,1]$. Though the government uses lump-sum taxes to finance its budget shortfall, subsidies to specific types of research create various distortionary effects (e.g., because of the inefficiency of picking winners or misdirection of resources). We allow for this by introducing the parameter χ measuring the extent of such wastage and specifying the aforementioned wedge between output and consumption as $D_t = \chi S_t$. This also implies that the government budget at time t is

$$(1 + \chi) S_t = T_t.$$

2.5 The Carbon Cycle

While clean intermediate production $y_{i,t}^c$ creates no carbon emission, dirty production $y_{i,t}^d$ emits κ units of carbon per intermediate output.[10] This implies that total amount of carbon emission at time t is

$$K_t = \kappa Y_t^d, \tag{9}$$

where $Y_t^d = \int_0^1 y_{i,t}^d di$ is the total output of the dirty sectors such technology at time t. We follow Archer (2005) and Golosov et al. (2014) in assuming that the atmospheric carbon concentration S_t is determined as follows

$$S_t = \int_0^{t-T} (1 - d_l) K_{t-l} dl, \tag{10}$$

where $t = T$ is the first date when emission started and

$$d_l = (1 - \varphi_P) \left[1 - \varphi_0 e^{-\varphi l} \right]$$

is the amount of carbon emitted l years ago still left in the atmosphere; $\varphi_P \in (0,1)$ is the share of emission that remains permanently in the atmosphere; $(1 - \varphi_P) \varphi_0 \in (0,1)$ is the fraction of the transitory component that remains in the first period; and $\varphi \in (0,1)$ is the rate of decay of carbon concentration over time. As explained in Archer (2005) and Golosov et al. (2014), this specification approximates the complex dynamics of carbon concentration in the atmosphere and provides a good match to the observed dynamics of atmospheric carbon concentration as we show below.

---

[10] The assumption that carbon emissions are per unit of intermediate good produced with dirty technology is a plausible baseline. Given our formulation with two types of technological changes, assuming that emissions are directly from the the use of the exhaustible resource would amount to imposing that even dirty technological improvements reduce effective emissions; this is because the same amount of dirty output, which in our empirical work will correspond to energy output, can be obtained with less and less pollution over time—even without any clean innovations. The main motivation for our baseline choice is to avoid this blurring of the distinction between clean and dirty technologies. Alternative emission specifications, where it is the use of the exhaustible resource that creates emissions, and where the dirty sector does not use the exhaustible resource, are investigated in Section 6.

2.6 Prices and Profits

The aggregate production function (1) implies a unit elastic demand for intermediates which, taking the final good as numeraire, can be written as

$$y_{i,t} = \frac{Y_t}{p_{i,t}}. \tag{11}$$

To characterize equilibrium prices, we define the tax- and energy-price-adjusted (for exhaustible resource) qualities by

$$\tilde{q}_{i,t}^d \equiv \frac{q_{i,t}^d}{\left(1 + \tau_t^d\right)\tilde{P}_{e,t}} \text{ and } \tilde{q}_{i,t}^c \equiv \frac{q_{i,t}^c}{1 + \tau_t^c}.$$

Therefore, intermediate i will be produced using technology $j \in \{c, d\}$ only if $\tilde{q}_{i,t}^{-j} \leq \tilde{q}_{i,t}^j$. Moreover, as explained in the previous section, if the leading technology for intermediate i at time t is $q_{i,t}^j$, another firm will have access to technology $q_{i,t}^j/\lambda$ for free, and therefore, equilibrium markups can never exceed λ. This implies that the price and quantity are given as

$$p_{i,t}^j = \min\left\{\frac{\lambda w_t^u}{\tilde{q}_{i,t}^j}, \frac{w_t^u}{\tilde{q}_{i,t}^{-j}}\right\}, \text{ and } y_{i,t}^j = \max\left\{\frac{\tilde{q}_{i,t}^j}{\lambda w_t^u}, \frac{\tilde{q}_{i,t}^{-j}}{w_t^u}\right\} \times Y_t. \tag{12}$$

Equilibrium profits can then be computed as a function of m_t and n as

$$\pi_{n,t}^d = \tilde{\pi}\left(n - m_t\right)Y_t \text{ and } \pi_{n,t}^c = \tilde{\pi}\left(m_t - n\right)Y_t, \tag{13}$$

where

$$\tilde{\pi}(k) \equiv \begin{cases} 0 \text{ if } k \leq 0, \\ \frac{\lambda(k)-1}{\lambda(k)} \text{ otherwise,} \end{cases}$$

and $\lambda(k) \equiv \lambda^k$ if $k \in [0, 1]$ and $\lambda(k) = \lambda$ if $k > 1$ summarizes equilibrium markups.

2.7 Value Functions and Innovation Incentives

We now characterize innovation incentives. Let us focus on a single firm, drop the firm and time indices to simplify notation, and let $\vec{n}^j \equiv \left[n_1^j, ..., n_{u^j}^j\right]$ denote the vector of product lines where the firm in question holds the leading-edge technology of type $j \in \{c, d\}$ and n_i^j is the technology gap compared to technology $-j$ within the same product line. Let $\vec{n}_{-i}^j$ denote the same vector $\vec{n}^j$ without its ith element n_i^j. Then the value of the firm satisfies the Hamilton-Jacobi-Bellman (HJB) equation:

$$
\begin{aligned}
& rV_{\vec{n}^j}^j - \dot{V}_{\vec{n}^j}^j \\
&= \sum_{i=1}^u \left[\begin{array}{c} \pi_{n_i}^j + z^j\left(V_{\vec{n}_{-i}^j}^j - V_{\vec{n}^j}^j\right) + \\ z^{-j}\left(1 - \alpha + \mathbb{I}_{(n_i^j \leq 0)}\alpha\right)\left(V_{\vec{n}_{-i}^j \cup \{n_i^j-1\}}^j - V_{\vec{n}^j}^j\right) + \mathbb{I}_{(n_i^j>0)}z^{-j}\alpha\left(V_{\vec{n}_{-i}^j \cup \{-1\}}^j - V_{\vec{n}^j}^j\right) \end{array} \right] \\
& + \int \max_{x^j \geq 0}\left[u^j x^j\left(\mathbb{E}_n V_{\vec{n}^j \cup \{n_{u+1}^j\}}^j - V_{\vec{n}^j}^j\right) - (1-s^j)u^j w^s\left((x^j)^{\frac{1}{\eta}}\theta^{-\frac{1}{\eta}} + \mathbb{I}_{(x^j>0)}F_I\right)\right]dF_I.
\end{aligned} \tag{14}
$$

11

The first term on the right-hand side is the profits generated from u^j product lines. In addition, at the flow rate rate z^j, each product line i will be lost to an innovation by another firm from the same technology j, in which case i is taken out of firm's portfolio (so that the firm's portfolio becomes $\vec{n}^j_{-i}$). If instead production line i experiences an innovation from the alternative technology $-j$, which happens at the rate z^{-j}, then there are two possibilities. First, if the innovation is incremental (probability $1 - \alpha$) or technology j is already behind $\left(n^j_i \leq 0\right)$, the technology gap declines by one step to $n^j_i = n^j_i - 1$. Second, if the innovation is a breakthrough (probability α) *and* technology j was leading $\left(n^j_i > 0\right)$, then the firm falls behind by one step to $n^j_i = -1$. Finally, the firm invests in R&D itself and innovates at the flow rate $X^j = u^j x^j$, and the expected return from this R&D (inclusive of costs net of the R&D subsidy for this sector, s^j) is the final line of the right-hand side, with the integral accounting for the fixed costs being stochastic. When successful, the firm adds a new product line so that its portfolio becomes $\vec{n}^j \cup \{n^j_{u+1}\}$.

Let us next denote by $\mu_{n,t}$ the fraction of product lines where the dirty lead is exactly equal to n steps at time t. The next lemma provides a convenient re-expression of this HJB equation:

Lemma 1 *Equation* (14) *can be re-expressed as* $V^j_{\vec{n}^j,t} = \left(\sum_{i=1}^u v^j_{n_i,t}\right) \times Y_t$, *where*

$$\tilde{r}_t v^j_{n_i,t} - \dot{v}^j_{n_i,t} = \tilde{\pi}^j_{n_i} - z^j_t v^j_{n_i,t} + z^{-j}_t \left(1 - \alpha + \mathbb{I}_{(n^j_i \leq 0)}\alpha\right)\left(v^j_{n_i-1,t} - v^j_{n_i,t}\right) + \mathbb{I}_{(n^j_i>0)} z^{-j}_t \alpha \left(v^j_{-1,t} - v^j_{n_i,t}\right)$$
$$+ \int \max_{x^j_t \geq 0}\left[x^j_t \bar{v}^j_t - \left(1 - s^j_t\right)\tilde{w}^s_t\left(\left(x^j_t\right)^{\frac{1}{\eta}}\theta^{-\frac{1}{\eta}} + \mathbb{I}_{(x^j_t>0)} F_{I,t}\right)\right]dF_{I,t}, \tag{15}$$

where $\tilde{r}_t \equiv r_t - g_{Y,t}$, *is the effective interest rate (net of output growth),* $\tilde{w}^s_t \equiv w^s_t/Y_t$ *is the normalized skilled wage and* $\bar{v}^j_t$ *is the expected per product value of innovation defined as*

$$\bar{v}^d_t \equiv \sum_{n \geq 0}\mu_{n,t} v^d_{n+1,t} + \sum_{n < 0}\mu_{n,t}\left[(1 - \alpha)v^d_{n+1,t} + \alpha v^d_{1,t}\right], \quad and \tag{16}$$

$$\bar{v}^c_t \equiv \sum_{n \leq 0}\mu_{n,t} v^c_{n-1,t} + \sum_{n > 0}\mu_{n,t}\left[(1 - \alpha)v^c_{n-1,t} + \alpha v^c_{-1,t}\right].$$

Proof of Lemma 1. See the Online Appendix. ∎

Intuitively, because there are no technological or product market linkages between the different product lines in which the firm has a lead, its value can be written as the sum of the values of each one of its product lines. An important implication of this result is that incumbent innovation rates per product line are independent of their portfolios:

$$x^j_t = \mathbb{I}_{(x^j_{I,t}>0)}\left(\frac{\eta\theta^{\frac{1}{\eta}}\bar{v}^j_t}{\left(1 - s^j_t\right)\tilde{w}^s_t}\right)^{\frac{\eta}{1-\eta}} \quad \text{for } j \in \{c,d\}, \tag{17}$$

where $v^j_{n,t}$ is the solution to (15), $\bar{v}^j_t$ is given as in (16), and $\tilde{w}^s_t$ is the normalized skilled wage defined in Lemma 1. A number of important qualitative conclusions follow from (17):

12

I. Higher average values of innovation and lower scientist wages increase research effort.

II. Subsidies to research increase research effort. This will be important in encouraging clean innovation by means of research subsidies.

III. Carbon taxes increase clean research effort (and reduce dirty research effort). This can be seen by considering higher values of m in (13) which, given the distribution of technology gaps $\mu_{n,t}$, increases $\pi_{n,t}^c$ and $\bar{v}_t^c$, because production shifts from dirty to clean technologies (and neck-and-neck sectors shift to positive markups for clean technologies). This shows that carbon taxes alone may be sufficient to encourage clean innovation and thus a transition to clean technology. Whether they are optimal is an empirical and quantitative question we consider below.

IV. Perhaps most importantly, the nature of innovation is path-dependent in this economy. Consider the fraction of sectors in which clean producers are already making positive profits. The average profits for clean producers is $\Gamma_t^c \equiv \sum_{n<m} \mu_{n,t} \pi_t (m_t - n)$. Since the value function is determined as the discounted sum of these average profits, the sequence $\{\Gamma_t^c\}_{t=0}^{\infty}$ has a direct impact on innovation incentives through $\bar{v}_t^c$. When there are large technology gaps between dirty and clean, Γ_t^c is small, discouraging clean innovation and encouraging dirty innovation. But if clean innovation is maintained for a while, then the fraction of sectors with $n \leq m$ increases, raising the probability that clean innovation will improve over a profitable intermediate. Thus clean innovation can naturally self-reinforce over time, but this is a slow process because the fraction of sectors with $n \leq m$ changes slowly.

2.8   Free Entry and Labor Market Clearing

The previous subsection characterized the R&D decisions of incumbents as a function of the state of the economy and policies. The other component of R&D comes from entrants. Using similar reasoning to that for incumbent R&D incentives, the free entry condition for entrants for technology $j \in \{c, d\}$ is

$$\max_{x_{E,t}^j \geq 0} \left\{ x_{E,t}^j \bar{v}_t^j Y_t - \left(1 - s_t^j\right) w_t^s \left[h\left(x_{E,t}^j\right) + F_E \right] \right\} \leq 0, \tag{18}$$

holding as equality if $E_t^j > 0$. Inspection of (18) establishes that at time t, there can be positive entry into sector j only if the "policy-adjusted" value of innovation is higher in sector j than in sector $-j$. In other words, entrants will direct their R&D to the clean technology if $\bar{v}_t^c / \left(1 - s_t^c\right) > \bar{v}_t^d / \left(1 - s_t^d\right)$ and to the dirty technology if the reverse inequality holds. Conditional on entry, the optimal value of $x_{E,t}^j$ is given by equation (17), except that the indicator function now conditions on R&D by entrants in sector j being positive, i.e., $\mathbb{I}_{(x_{E,t}^j > 0)}$.

The labor market clearing condition for skilled workers, combining demand from incum-

bents and entrants, is

$$L^s = \sum_{j \in \{c,d\}} \left[\mathbb{I}_{(x_{E,t}^j > 0)} \left(h(x_{E,t}^j) + F_E \right) E_t^j + \int_0^1 \mathbb{I}_{(x_{i,t}^j > 0)} \left(h(x_{I,t}^j) + F_{I,i,t} \right) di \right], \quad (19)$$

where $h(\cdot)$ is defined in (8) above, while $x_{E,t}^j$ and $x_{I,t}^j$ depend on the normalized skilled wage $\tilde{w}_t^s$ via (17). This labor market clearing condition thus pins down this normalized skilled wage and indicates that the equilibrium normalized skilled wage will be higher when R&D is more profitable and is subsidized more heavily.

We next characterize labor market clearing for unskilled workers. From the equilibrium production decisions in (12), demand for production workers is

$$l_{i,t} = \begin{cases} \frac{Y_t}{(1+\tau_i^c)\lambda(m_t-n_{i,t})w_t^u} & \text{if } n_{i,t} \leq m_t \\ \frac{Y_t}{(1+\tau_i^d)\lambda(n_{i,t}-m_t)w_t^u} & \text{otherwise} \end{cases}.$$

Adding the labor used in the extraction of the exhaustible resource, unskilled labor market clearing can be written as

$$1 = \frac{Y_t}{w_t^u} \times \left\{ \sum_{n \leq m_t} \frac{\mu_{n,t}}{(1+\tau_t^c)\lambda(m_t-n)} + \left[\nu + (1-\nu) \left(\frac{1/\zeta}{\tilde{P}_{e,t}} \right) \right] \sum_{n > m_t} \frac{\mu_{n,t}}{(1+\tau_t^d)\lambda(n-m_t)} \right\}. \quad (20)$$

This equation shows both the impact of taxes on labor demand (which reduce labor demand and thus wages) and the distribution of technology gaps (because these affect markups). It also shows that if there were only one type of technology (and no extraction of the exhaustible resource), an increase in the tax rate would have no impact on production, just reducing the unskilled wage rate. This is no longer true with two types of technologies, because a tax on dirty technology, for example, would also change the prices of intermediates produced with dirty technology relative to those produced by clean technology, thus impacting production.

Next substituting the optimal quantities (12) into the final good production function (1), we obtain

$$w_t^u = \bar{Q}_t \Lambda_t, \quad (21)$$

where $\bar{Q}_t \equiv \exp\left(\int \ln \tilde{q}_{it} di \right)$ is the quality index of active tax-adjusted labor productivities, and $\Lambda_t \equiv \prod_n \lambda(n-m_t)^{-\mu_{n,t}}$ is an inverse function of equilibrium markups.

We can next use (20) to express aggregate output as a function of the quality index of active tax-adjusted labor productivities:

$$Y_t = \bar{Q}_t \Lambda_t \Omega_t^{-1} \exp\left(-\gamma \left(S_t - \bar{S} \right) \right), \quad (22)$$

where

$$\Omega_t \equiv \sum_{n \leq m_t} \frac{\mu_{n,t}}{(1+\tau_t^c)\lambda(m_t-n)} + \left[\nu + (1-\nu) \left(\frac{1/\zeta}{\tilde{P}_{e,t}} \right) \right] \sum_{n > m_t} \frac{\mu_{n,t}}{(1+\tau_t^d)\lambda(n-m_t)}$$

14

is an adjustment for labor demand coming both from taxes and markups.

As noted above, pollution is caused by total dirty intermediate production at time t, which can be expressed as

$$Y_t^d = \frac{Y_t}{\left(1 + \tau_t^d\right) w_t^u \tilde{P}_{e,t}} \left[\frac{1}{2} Q_{m,t}^d + \frac{1}{\lambda \left(n - m\right)} \sum_{n > m_t} Q_{n,t}^d \right], \tag{23}$$

where we broke up the quality index by step-size differential n and define $Q_{n,t}^d \equiv \int_{i \in \mu_n} q_{i,t}^d di$ (with a slight abuse of notation where $i \in \mu_n$ denotes intermediates where the technology gap is n steps). The evolution of the quality indices $Q_{n,t}^d$ and $Q_{n,t}^c$ are described in the Online Appendix.

Intertemporal maximization yields the standard Euler equation,

$$g_{C,t} = r_t - \rho, \tag{24}$$

where $g_{C,t}$ is the growth rate of consumption and r_t is the interest rate at time t (and in addition we impose the usual transversality condition).

Hotelling rule implies that the exhaustible resource price net of the cost of extraction must grow at the rate r_t (i.e., at the interest rate), and thus, the exhaustible resource price net of the cost of extraction relative to the unskilled wage, $\frac{p_{e,t} - \zeta w_t^u}{w_t^u}$, must grow at the rate $r_t - g_{w,t}$. The implied path of the normalized exhaustible resource price is therefore

$$\frac{p_{e,t}}{w_t^u} = \left[\frac{p_{e,0}}{w_0^u} - \zeta\right] \int_0^t e^{r_s - g_{w,s}} ds + \zeta. \tag{25}$$

An additional boundary condition is given by the requirement that the price of the exhaustible price must satisfy

$$\lim_{t \to \infty} \left[p_{e,t} - \zeta w_t^u\right] R_t = 0,$$

so that either the entire stock of the exhaustible resource is utilized as $t \to \infty$ or $p_{e,t} = \zeta w_t^u$ for all $t \geq 0$.

Finally, the evolution of the distribution of technology gaps represented by $\mu_{n,t}$ can be derived from the following differential equations (with some initial condition $\{\mu_{n,0}\}_{n=-\infty}^{\infty}$). For $n > 1$, we have

$$\dot{\mu}_{n>1,t} = z_t^d \mu_{n-1,t} + (1 - \alpha) z_t^c \mu_{n+1,t} - z_t \mu_{n,t}, \tag{26}$$

where $z_t^j \equiv \left(1 + E_t^j\right) x_t^j$ and $z_t \equiv z_t^d + z_t^c$. Intuitively, the fraction of intermediates with the technology gap n changes because of the difference between inflows and outflows. There will be inflows into state n from $n - 1$ when a dirty innovation occurs and from $n + 1$ when a clean innovation occurs without leapfrogging. Outflows from the technology gap of n steps will take place due to both clean or dirty innovation (bringing the state into $n + 1$, $n - 1$, or -1

depending on the innovation type). For $n \leq 1$, a similar reasoning identifies

$$
\begin{aligned}
\dot{\mu}_{1,t} &= z_t^d \mu_{0,t} + (1-\alpha) z_t^c \mu_{2,t} + \alpha z_t^d \sum_{n<0} \mu_{n,t} - z_t \mu_{1,t} \\
\dot{\mu}_{0,t} &= (1-\alpha) z_t^d \mu_{-1,t} + (1-\alpha) z_t^c \mu_{1,t} - z_t \mu_{0,t} \\
\dot{\mu}_{-1,t} &= z_t^c \mu_{0,t} + (1-\alpha) z_t^d \mu_{-2,t} + \alpha z_t^c \sum_{n>0} \mu_{n,t} - z_t \mu_{-1,t} \\
\dot{\mu}_{n<-1,t} &= z_t^c \mu_{n+1,t} + (1-\alpha) z_t^d \mu_{n-1,t} - z_t \mu_{n,t}.
\end{aligned}
\tag{27}
$$

We now summarize the dynamic equilibrium path using the equations we have derived in this section. For any given time path of policies $\left[\tau_t^j, s_{I,t}^j, s_{E,t}^j\right]_{t=0}^{\infty}$, a dynamic equilibrium path is characterized by time path of $[y_{i,t}^j, p_{i,t}^j, x_{I,t}^j, x_{E,t}^j, Y_t, w_t^s, w_t^u, e_{i,t}, p_{e,t}, R_t, E_t^j, \{\mu_{n,t}\}_{n=-\infty}^{\infty}$, $\{Q_{n,t}^d\}_{n=-\infty}^{\infty}, r_t, S_t]_{t=0}^{\infty}$ such that $[1]$ $y_{i,t}^j$ and $p_{i,t}^j$ maximize profits as in (12); $[2]$ $x_{I,t}^j$ and $x_{E,t}^j$ satisfy (17); $[3]$ w_t^u is determined by (21); $[4]$ aggregate output Y_t is given by (22); $[5]$ w^s is determined from the free-entry condition (18) when there is positive entry and from skilled labor market clearing (19) when there is no positive entry; $[6]$ E_t^j is determined from the skilled labor market clearing (19) when there is positive entry; $[7]$ technology gap shares $\{\mu_{n,t}\}_{n=-\infty}^{\infty}$ satisfy $(26)-(27)$; $[8]$ total productivity of the sectors with n-step gap $Q_{n,t}^d$ evolves according to the innovation rates in (17) and (18);[11] $[9]$ the interest rate satisfies the Euler equation (24); $[10]$ exhaustible resource quantity and price satisfy (4) and (25); $[11]$ R_t evolves according to the law of motion (3); and $[12]$ S_t is given by (9) and (10) with Y_t^d given by (23).

3   Empirical Strategy and Data

Our model has 18 parameters/variables, $\{\rho, \bar{S}, \gamma, \varphi, \varphi_0, \varphi_P, \kappa, \nu, \zeta, R_0, L^s, \alpha, \eta, \theta, \lambda, F_I, F_E, \chi\}$, and the initial distribution of technology gaps between clean and dirty technologies, $\{\mu_{0,t}\}_{n=-\infty}^{\infty}$, to be determined. As will become clearer below, given $\{\mu_{0,t}\}_{n=-\infty}^{\infty}$, the remaining parameters can be estimated without knowledge of taxes and subsidies, and also without any information on $\bar{S}, \gamma, \varphi, \varphi_0$, and φ_P. These parameters become relevant only for our policy analysis. Nevertheless, for completeness we also specify our choices for all these parameters.

We proceed in four steps. First, we choose from external sources the parameters of the carbon cycle and the discount rate: $\rho, \bar{S}, \gamma, \varphi, \varphi_0, \varphi_P, \kappa, \nu$ and χ. Second, we directly estimate L^s, α, and η from microdata. Third, we choose the initial distribution of technology gaps to match the distribution of patents between firms innovating primarily with clean and primarily with dirty technologies as we explain below. Finally, we estimate the remaining parameters $\theta, \lambda, R_0, \zeta, F_I$ and F_E using simulated method of moments, with moments being selected to model the firm-level R&D behavior, growth rates, entry/exit rates for the energy sector as we describe below, and from the time path of carbon emissions. The model performs well and is able to replicate these moments reasonably closely.

---

[11] See the Online Appendix for the explicit equations for the evolution of quality indices.

Throughout our focus is on the energy sector, the behavior of which has motivated our theoretical model. The energy sector is defined as firms involved in the sourcing, refinement and delivery of energy inputs for residential and industrial applications (e.g., oil and gas, electricity), firms that provide complementary inputs and equipment into this energy-production process (e.g., drilling equipment, power plant technologies), and firms that interface with the energy inputs for residential and industrial use (e.g., motor manufacturers). As such, the 1576 firms that make up our sample include oil and gas producers, mining and exploration firms, engine manufacturers, power companies building upon multiple techniques, energy equipment manufacturers, and similar. The data we use for estimation come from the Census Bureau, and we design our sample around innovative firms in the energy sector that are in operation during the 1975-2004 period.

3.1 External Calibration

We choose the parameters $\bar{S}, \gamma, \varphi, \varphi_0, \varphi_P, \kappa$ and ν to link our model to the carbon cycle and its impact on aggregate output following Golosov et al.'s (2014) approach. This approach takes into account the current level of carbon stock and its increase since pre-industrial times; the rate at which new emissions enter the atmosphere, the terrestrial biosphere or shallow oceans, and the deep oceans; how that movement and the various reservoirs of carbon influence the earth's temperature; and how higher temperatures and environmental damage hurt the economy. This approach is similar to prior work in environmental economics (e.g., Nordhaus, 2008; Nordhaus and Boyer, 2000), though with some important differences: in particular, it is more flexible in allowing non-linear absorption of atmospheric carbon, but it does not allow any delay on the impact of this carbon content on economic outcomes and temperature changes (e.g., because ocean temperature changes at a different rate than land temperature) and does not separately keep track of the dynamics of the atmospheric concentration of carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O).

The value of the pre-industrial stock of carbon dioxide in the atmosphere $\bar{S}$ is 581 GtC (gigatons of carbon). To model how emission increases the atmospheric stock of carbon, we define the three parameters φ, φ_0, and φ_P as follows. First, φ_P is the portion of new emissions that will remain in the atmosphere for a very long time, likely for thousands of years, and estimates of this parameter from Archer (2005) and IPCC (2007) are about 20%. The other two parameters, φ and φ_0, govern the short- and medium-term movement of the emitted carbon, which influence the earth's temperature over short horizons but are ultimately absorbed into the deep oceans. To identify these parameters, we utilize the 30 year half-life of carbon and match the carbon stock evolution under emissions during the 1900-2008 period. We use the law of motion of the atmospheric carbon concentration S_t, given by (10), over the period 1900-2008.[12] The emission data, $\{K_t\}_{t=1900}^{2008}$, are depicted in Figure 1.

---

[12] Namely, $S_t = \int_0^{t-1900} (1 - d_l) K_{t-l} dl + S_{1900}$.
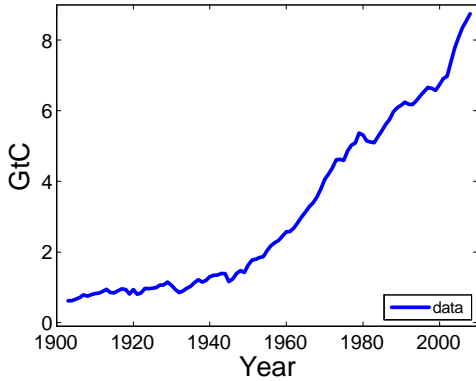
Figure 1. World Carbon Emissions
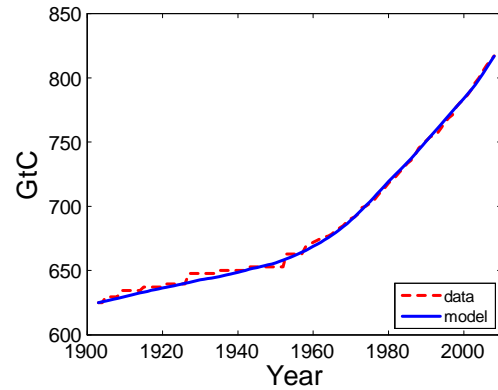(source: Boden et al., 2011)

Figure 2. Atmospheric Carbon
Concentration: Data and Model
Implications

Figure 2 then shows the close fit of (10) at the parameter values $\varphi = 0.0313$ and $\varphi_0 = 0.7661$ (the blue line) to the actual data on atmospheric carbon concentration (the red dotted line).

As already noted, our damage function also follows Golosov et al. (2014) and we choose the same γ parameter as they do, 5.3×10^{-5} GtC^{-1}. Because this number may be too low, we investigate the robustness of our results to higher values of γ in Section 6.

In addition, the κ parameter is chosen to link current world emissions levels to the baseline level of dirty output in the model, Y_d as given by (23), so that implied future increases in Y_d will translate to correspondingly higher levels of emissions. This formulation assumes that both emissions and technological change in our model can proxy those in the global economy (e.g., because new technologies can be utilized globally).[13] For the elasticity parameter ν in the dirty production function, we follow Golosov et al. (2014) and set $\nu = 0.04$.

As noted above, the parameter χ allows for distortionary effects of research subsidies, for example, because it is difficult for the government to know *ex ante* which projects are most worthy of public support. For instance, the US Department of Energy (DOE) provided $6.3 billion in funding for the support of clean energy research and implementation projects, and there have been some high-profile business and technology failures associated with this program (DOE, 2012). A well-known example is that of Solyndra Corporation, a solar panel manufacturer, which received approximately $500 million in support from the DOE, before declaring bankruptcy without achieving large-scale production. Motivated by this, we set our baseline value for χ to 10% (implying that out of the $6.3 billion, $630 million would be entirely wasted). We will show that our results are robust when χ is equal to 0% or 20%.

---

[13]This modeling strategy abstracts from several important issues: [1] carbon taxes and research subsidies vary across countries, especially in view of the prevailing lack of international cooperation (see Hassler and Krusell, 2012, for possible implications of this), and [2] new technologies, often first developed in advanced economies (in particular, the United States whose microdata we are using), will spread only slowly to other countries.

18

The final piece of our external calibration is the (private and social) discount rate, which we take as $\rho = 1\%$, close to the 1.5% chosen by Nordhaus. We also show that our results are quite similar, but of course involve more aggressive policies, when the (social) discount rate is 0.1% as assumed by Stern (2007).

3.2 Sample Construction and Data Sources

We use the individual records of patents granted by the USPTO from 1975-2009. We collect patents that are filed by inventors living in the United States at the time of the patent application, and our assigned to industrial firms. We identify patents related to the energy sector through patent technology codes. The technology codes are the most disaggregated level of the USPTO's classification scheme and number over 150,000. We adopt the prior classifications developed by Popp (2002) and Popp and Newell (2012). This authoritative prior work is particularly helpful in that they provide classifications into various types of energy technologies. We are interested, however, in several technologies (e.g., nuclear power) not considered by Popp and Newell (2012) and thus extend their list through three additional steps: [1] using environment-related technologies identified by the OECD under the IPC system, [2] examining patents made by prominent firms with energy-related R&D data from the NSF R&D Survey, and [3] manual searches on key phrases.

Our firm-level operating data are from the US Census Bureau's Longitudinal Business Database, which contains annual employment levels for every private-sector establishment with payroll from 1976 onward. We also employ Economic Censuses that are conducted every five years to collect additional plant and firm operation data (e.g., sales). We match patent data to these operating data using firm-name and geographic-location matching algorithms. We focus our sample on the years in which Economic Censuses are conducted. This structure matches our data's features and accounts for the lumpiness of innovation outcomes. We thus measure variables using the average of observed values for firms in five-year windows surrounding Economic Census years and have six time periods covering 1975-2004.

Our third data source is the NSF's Survey of Industrial Research and Development (R&D Survey) that collects information from all firms conducting over one million dollars in R&D and subsamples firms beneath this level. The R&D Survey provides us with information on many firms' R&D expenditures and employments of science and engineering workers. We use the data, along with the patenting of the firm, to calculate the innovation production function for the sector (e.g., the η and α parameters). These calculations only utilize firm observations for which we always observe reported data. For our broader moments on firm dynamics, we need a complete distribution that encompasses firms only subsampled beneath the threshold. Our patent data are universally observed, and we thus use patents to impute R&D values for firms that are less than the threshold of one million dollars R&D and not subsampled. Overall, our moments combine the R&D and patent data into a single measure of innovation (in R&D

terms) that accords with the model.

Our sample requires that a firm either patent or have measured R&D in the first period of its life, and does not condition on innovative activity before 1975-1979 for firms that are present at the sample's start. Because our model allows for firms to transition out of R&D, we include firms that stop innovation, though we do not include non-innovative incumbent firms that switch to innovation.[14] Our final sample requires the firm have positive employment and obtain three or more patents in the energy sector during 1975-2004. We also require that 10% of the firm's total patenting be in energy-related fields. The 10% bar is more substantial than it may initially appear as we have been fairly conservative in terms of defining energy-sector patents.

Thus, our compiled dataset includes 6228 observations from 1576 innovative firms in the energy sector from 1975-2004. While focused on a single sector, our firm-level panel contains 19% of all US R&D industrial expenditures and accounts for about 70% of industrial patents for the U.S. energy sector. Across all activity in the economy, our sample contains 1% of establishments, 5% of employment, and 10% of sales. Our sample accounts for a substantial amount of activity in several of the main sectors responsible for emissions (e.g., Mueller et al., 2011).[15] While our sample does not include many firms directly from two high-emission sectors, agriculture and transportation, it includes many of the manufacturers of products that are key inputs to these sectors.

We define patents as related to dirty technologies if they are connected to the extraction, refinement or use of fossil-fuel based energy, including oil, coal, natural gas, and shale technologies. We group into clean-energy patents fields that are related to geothermal, hydroelectric, nuclear, solar, and wind energy. We also include patents for conservation and utilization of energy. Results are robust to reclassifications of border group types.

Finally, to determine the model's initial conditions, we identify whether firms are primarily operating in dirty- or clean-energy applications. We first classify an observation (a firm in a given period) as focused on clean energy if 25% or more of its energy-sector patents are devoted to clean-energy fields; otherwise the firm is classified as a dirty-energy firm in the period. We use the 25% threshold as our assignments of clean-energy fields are conservative compared to dirty-energy fields. We then describe the firm overall as a clean-energy firm if half or more of its time periods achieve this clean-energy focus. The distribution between clean and dirty uses at the firm level is fairly bimodal—96% of observations have 75% or more of their patents in one technology—making the exact details of these procedures less important. In total, 11% of our

---

[14] As the probability that an existing, non-innovative firm commences R&D or patenting over the ensuing five years (conditional on survival) is only about 1%, this exclusion is reasonable. These procedures define the base pool of innovative firms in the energy sector.

[15] In the 1992-1997 period, for example, we account for 59% of sales in industries related to coal and oil extraction, refinement, and shipment; 33% of sales in industries related to electricity production; and 21% of manufacturing sales. Among manufacturing industries, our sample contains higher shares in industries more closely linked with emissions (e.g., 64% in petroleum refinement, 31% in primary metals).

firms are classified in the clean-energy sub-sector; 14% of energy-sector patents are classified as clean energy.

3.3 Estimation and Choice of Parameters from Microdata

We first estimate the η parameter from our innovation production function, $X_f = \theta(H_f)^\eta(u_f)^{1-\eta}$, which can be rewritten as $\ln(X/u_f) = \ln(\theta) + \eta \cdot \ln(H/u_f)$. We measure X by the firm's count of patents, H by the firm's R&D expenditures or scientist counts, and u through four proxies of firm size given below. Our patent count measure is weighted by citations (normalized by the average citations achieved by other patents in the same patent class and application year).

To estimate the η elasticity as accurately as possible, we use the panel nature of our data and later return to estimating the θ parameter. As noted earlier, we only use for this exercise firms that have a full panel of reported R&D data. To focus on higher-quality data for our differenced estimations, we also require that the firm be present in at least three periods. We first estimate a linear regression with year fixed effects δ_t as follows:

$$\ln(Patents/product_{f,t}) = \beta \cdot \ln(R\&D/product_{f,t}) + \delta_t + \epsilon_{f,t}, \tag{28}$$

with standard errors clustered by firm. We then extend the estimation to allow for firm fixed effects by first-differencing to obtain the following specification:

$$\Delta\ln(Patents/product_{f,t}) = \beta \cdot \Delta\ln(R\&D/product_{f,t}) + \delta_t + \epsilon_{f,t}. \tag{29}$$

Panel A of Table 1 summarizes eight variants of the OLS levels regression (28). The rows indicate four measures of firm size u_f: SIC3 industry counts, SIC4 industry counts, sales, and establishment counts. Column headers indicate whether R&D inputs are being measured through R&D expenditures or counts of scientists. The eight coefficients are from eight separate estimations of regression (28). The average is 0.69 with a range of 0.63-0.76.

TABLE 1. OLS ESTIMATES FOR η PARAMETER

Firm Size u_f:	A. Levels		B. First-differenced	
	R&D	Scientists	R&D	Scientists
SIC3 Counts	0.632 (0.042)	0.653 (0.048)	0.342 (0.056)	0.286 (0.052)
SIC4 Counts	0.625 (0.043)	0.644 (0.048)	0.353 (0.057)	0.296 (0.053)
Sales	0.761 (0.053)	0.751 (0.048)	0.405 (0.075)	0.348 (0.065)
Establishments	0.714 (0.039)	0.732 (0.041)	0.505 (0.058)	0.455 (0.054)

Note: This table presents variants of regressions (28) and (29). Standard errors in parentheses.

Panel B similarly summarizes eight estimation variants of the first-differenced regression (29). The average across these variants is lower at 0.37, with a range of 0.29 to 0.51. We set the baseline value of η to 0.5, the mid-point within the range of estimates in Table 1.[16]

---

[16] We also find comparable η parameters in robustness checks off of this sample platform. For example,

We next turn to the α parameter that describes technology leapfrogging. This process is challenging to model empirically, and we are unfortunately unable to identify exact races between clean and dirty technologies directly within the patent data. This limitation is due to the narrowness of the technology codes that are entirely clean or dirty in application, while patent class divisions are too broad and few in number. We thus identify the rate at which patents with exceptional quality emerge using patent citations and quantify the rate at which patents enter and quickly establish high levels of citations compared to incumbent peers.

Specifically, we start with our dataset of all energy-sector patents granted to US inventors during the post-1975 period. We calculate among these energy-sector patents the citation count distribution of incumbent patents by year, excluding within-firm citations. Incumbent patents are defined to be those that were applied for 5-10 years before the focal year; we cap at 10 years prior so that we can have a stable window across a time period from 1985 onwards for analysis. Citations are coming from patents being applied for in the focal year. By focusing on patents receiving a citation in a given year, we are effectively looking at technologies that are being actively used, with many incumbent patents dropping out as no one is building on them. We then calculate for new patents the citations they receive by year. We designate a major entrant as any patent that has a citation count that exceeds the 90th percentile of the incumbent distribution in any of its first three years. This evaluation approach is designed to keep the incumbent groups (5-10 years earlier) separate from the entrant groups (max of three years earlier). 4.2% of entrants achieve this level of major entrant. We find a slightly lower estimate at 4.0% using Popp and Newell's (2012) definitions, and a rate of 4.1% when making the citation distributions specific to each patent class. Based upon these findings, we set $\alpha = 4\%$.

Finally, for L^s, which is the supply of scientists and engineers involved in R&D-type activities in the model (relative to unskilled workers), we use 5.5%. We calculate this share from Census IPUMS using the 2000 5% sample, focusing on non-group quartered workers aged 20-65, employed in industries closely related to the energy sector. We also require 20 weeks worked within the year and usual hours worked of 20 or more during each week. 5.5% is the share of these workers with college education or more, employed in occupations related to science and engineering.

restricting the sample to firms with energy patents as more than 30% of their patent portfolio yields levels and first-differences estimates of 0.744 (0.065) and 0.384 (0.100), respectively. Restricting our sample to firms identified from Popp and Newell's (2012) codes yields levels and first-differences estimates of 0.704 (0.049) and 0.301 (0.071), respectively. Relaxing our requirement that the firm be present in three periods yields levels and first-differences estimates of 0.614 (0.043) and 0.338 (0.056), respectively. We likewise find similar outcomes when incorporating industry-year fixed effects, using unweighted patent counts, or similar adjustments. Finally, the Online Appendix provides similar estimates from Poisson models that allow for zero patenting outcomes.

3.4 Initial Technology Gaps

To provide the initial distributions of the model, we develop estimates of the cumulative stock of technologies invented by clean- and dirty-energy firms using SIC3 industries as approximations of product lines. We develop these distributions in three steps. The first step is to calculate the sum of patents by each firm during the 1975-2004 period and the firm's distribution of employment across SIC3 industries in these sectors over the same period. We then apportion the firm's cumulative patent stock across SIC3 industries using the firm's employment distributions. For each SIC3 industry, we finally sum the apportioned patents made by clean- and dirty-energy firms. This sum of patents across all firms, active or inactive, reflects the quality ladders structure of our model.

These calculations provide us with over 400 estimates of comparative clean- and dirty-energy stocks. Across these SIC3 industries, clean-energy firms are estimated to have a higher cumulative patent stock in 13.1% of industries. For data quality and Census Bureau disclosure restrictions, we focus on the upper half of the industry distribution in terms of cumulative clean and dirty patent counts, which has 13.0% of industries being led by the clean-energy stock (within manufacturing and energy production specifically, this share is 12.5%).

The average gap relative to the frontier in the 13% of cases where clean patents have the lead is 39%, while it is 76% in the 87% of cases where dirty patents have the lead. We convert the patent gaps into the technology gaps in the model using a conversion factor between the patents rates in the data and the innovation rates in the model. In particular, in our model the annual innovation flow of incumbents is 24.1% per product line (the sum of $x^c = 4.8\%$ and $x^d = 19.3\%$), while in the data the mean of patent flow is approximately 43 per line. Hence we divide the empirical patent distribution of clean and dirty (which consists of patents registered between 1975-2004) by the conversion factor of 43/0.241, and then round the resulting technology to the nearest integer. This gives us the initial number of improvements $n_{j,0}^d$ and $n_{j,0}^c$. We then compute the initial productivities as $q_{j,0}^d = \lambda^{n_{j,0}^d}$ and $q_{j,0}^c = \lambda^{n_{j,0}^c}$ to provide the initialization values.

Figure 3 plots the density of the resulting distribution of initial technology gaps between dirty and clean technologies.[17] It shows that in most product lines the dirty technology is only a few steps ahead of clean technology, but there is a long tail of product lines with a large gap between dirty and clean, and a small set in which clean is ahead of dirty. The fraction of product lines with a non-zero gap in terms of step sizes is 90%. Clean energy leads by one or more step sizes in 9% of cases. Dirty energy has a lead of 20 and 50 step sizes or more in 11%

---

[17]See the Online Appendix for additional summary statistics.

and 2% of technologies, respectively. We later consider alternative initial distributions.
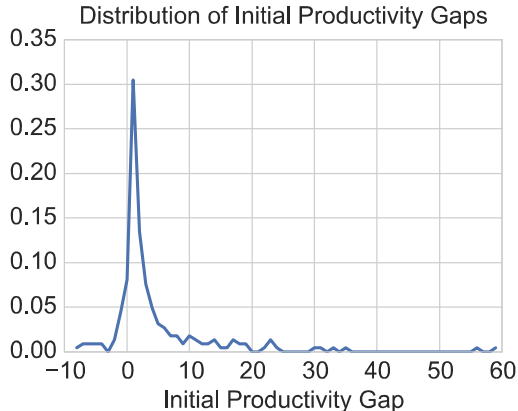

Figure 3. The distribution of initial productivity gaps between clean and dirty technologies across product lines (a positive number indicates the dirty technology having the lead).

3.5 Simulated Method of Moments

The remaining parameters $\theta, \lambda, F_I, F_E, R_0$ and ζ are estimated using the quantitative implications of our model. For the first four of these parameters, θ, λ, F_I and F_E, we use the simulated method of moments (SMM). This approach chooses the parameter vector so as to minimize the distance between four moments as implied by our model and data, $\min \sum_{i=1}^{4} |\text{model}(i) - \text{data}(i)|$, where we index each moment by i. SMM iteratively searches repeatedly across sets of parameter values for θ, λ, F_I and F_E until the model's moments are as close as possible to the empirical moments (see Adda and Cooper, 2003, for further details). We also choose the fixed cost heterogeneity parameter, ξ, as 10% and verify that our results are not sensitive to this choice of parameter.

We use three moments from the microdata—firm entry rates, firm exit rates, and the average R&D/sales ratio of firms—together with the growth rate of the sector to identify these parameters. The entrant's labor share and exit rates are calculated across the five-year intervals of our Census Bureau data and then transformed into annualized net rates of 1.3% and 1.75%, respectively. We match the construction of these entry and exit rate moments in the model. The weighted average R&D-to-sales ratio is 6.56%, using log sales as weights and winsorizing the R&D/sales ratio at the 99th percentile to reduce outliers. The aggregate annual sales growth per worker is 1.23% for the sector across the 1975-2004 period.[18]

For the remaining parameters, ζ and R_0, we use the implied emission path of our model. BP (2010) reports that oil reserves are around 181.7 gigaton, but there is considerable uncertainty around this estimate as well as on coal reserves. Rogner (1997) argues that world coal supplies will be sufficient for several hundred years if used at the current rate. This motivates our

---

[18]The computational algorithm is described in the Online Appendix.

approach of computing these two parameters to match the level and the growth rate of emissions in 2008 (rather than using a specific value of total reserves).

4   Estimation Results

In this section, we provide the SMM estimates of the remaining parameters and discuss the fit of our model to non-targeted moments. Finally, we show how atmospheric carbon concentration, temperatures and aggregate output evolve given these parameters in a laissez-faire equilibrium (with no policy intervention) starting from the observed distribution of technology gaps.

4.1   Parameter Estimates and Goodness of Fit

Table 2 summarizes our parameter estimates. The estimate of $\theta = 0.958$ implies that a unit of skilled labor on a single product line generates an innovation with probability of approximately 24% per year. Our estimates of the innovation step size, $\lambda = 1.063$, implies a gross profit margin of 5.7%, which is reasonable. The model predicts a sizable fixed cost advantage for incumbent firms: their fixed cost of operation is equal to 5% of the entrants' fixed cost. Finally, our procedure of matching the recent past of emissions delivers the total energy resources as 13,549 gigatons of carbon equivalent (GtC), which is reasonable given the extent of known reserves.

TABLE 2. PARAMETER ESTIMATES

Parameter	Description	Value	Parameter	Description	Value
θ	Innovation productivity	0.958	F_E	Fixed cost of entry	0.040
λ	Innovation step size	1.063	R_0	Total energy resources	13549
F_I	Fixed cost of R&D	0.002	ζ	Extraction technology	0.016

Note: This table presents the parameter estimates from SMM.

Table 3 shows the values of the moments used for estimation, which generally match the data quite closely.

TABLE 3. MOMENT MATCHING

Moments	Model	Data	Moments	Model	Data
Entry Share	0.014	0.013	Aggregate Sales/Worker Growth	0.012	0.012
Exit Rate	0.032	0.018	Emissions Level in 2008	8.461	8.749
Average R&D/Sales	0.065	0.066	Growth of Emissions in 2008	0.023	0.024

Note: This table compares the model and the data across the moments used for estimation.
.

Our main method of evaluating the quantitative fit of our model is to look at a range of non-targeted moments, which are presented with the model implications in Tables $4A - 4C$. We choose the non-targeted moments to represent aspects of the firm size distribution and its growth properties, which are quite different from the moments we targeted in our estimation.

Our first non-targeted moment compares the size ratio of the median entrant to the median incumbent firm. Our targeted moments on entry/exit rates, overall sector growth, and R&D intensity do not directly impose strong constraints on this size distribution of firms. Table 4A contrasts the size ratios in the model and data with respect to employment, sales, and sales per employee, and shows that our model implications match the data very closely with respect to the latter two metrics, though not as well for employment.[19]

<div align="center">

TABLE 4A. ENTRANT SIZE RATIO TO INCUMBENTS

</div>

	Ratio of Median Sizes	
Size Measure:	**Model**	**Data**
Employment	0.18	0.03
Sales	0.18	0.20
Sales per Employee	1.13	1.05

Note: This table compares non-targeted moments in the model and the data.

We next compare growth distributions. We first calculate the unconditional growth rate of employment for each firm in the model and data defined as $(Emp_t - Emp_{t-1})/((Emp_t + Emp_{t-1})/2)$. As argued by Davis, Haltiwanger and Schuh (1996), this measure of growth has attractive properties like a symmetric treatment of positive and negative growth and bounded values that minimize outliers. We calculate growth over five-year intervals. We then calculate the probability that firms experience substantial movements in either positive or negative directions. Comparing these movements in the model to the data provides insights into how well the innovation step sizes and associated firm dynamics mirror the sector's true performance. Table 4B shows that the model matches the data quite well on this dimension. The model somewhat overpredicts employment declines of 25% or more, but it closely matches otherwise.

<div align="center">

TABLE 4B. COMPARISON OF GROWTH DISTRIBUTION

</div>

		Employment Growth Probability			
Change over 5-Years:	**Model**	**Data**	**Change over 5-Years:**	**Model**	**Data**
Decrease 75% or more	0.25	0.11	Increase 25% or more	0.26	0.31
Decrease 50% or more	0.30	0.15	Increase 50% or more	0.20	0.20
Decrease 25% or more	0.38	0.26	Increase 75% or more	0.17	0.14
			Increase 100% or more	0.10	0.11

Note: This table compares non-targeted moments in model and data.

Finally, Table 4C turns to a comparison of the model and the data for growth rates of firms conditional on different quartiles of the size distribution. For this exercise, we divide firms into quantiles based upon their initial size in each five-year period and then compute the growth rates using the above formula. The employment distribution implied by the model is a little less fine-grained than the data, as about 50% of our firms have one product and employment

---

[19]To pass Census Bureau disclosure restrictions, the empirical medians are "fuzzy" median estimates that use the average values over the 45th to 55th percentiles.

is partially proportional to product counts. Nevertheless, the model's implications match the general patterns in the data, and in particular, the model successfully generates the lack of growth in the top two quantiles compared to the bottom three. Overall, the model performs reasonably well for these diverse tests, providing confidence in its ability to capture production and innovation dynamics in the energy sector.

TABLE 4C. COMPARISON OF GROWTH OVER SIZE DISTRIBUTION

| | | | 5-Year Conditional Growth Rates | | |
Quantile of Sizes:	Model	Data	Quantile of Sizes:	Model	Data
Smallest	29%	31%	3rd	26%	11%
2nd	38%	14%	4th	-8%	-1%
			Largest	-3%	-10%

Note: This table compares non-targeted moments in model and data.

4.2 Climate Dynamics in the Laissez-faire Economy

We next describe the implied future equilibrium and atmospheric carbon paths of the model under laissez-faire (meaning no carbon taxes and research subsidies). Given the initial distribution of technology gaps, dirty innovation is more profitable and with no policy intervention, most R&D is initially targeted to the dirty technology as shown in Figure 4.
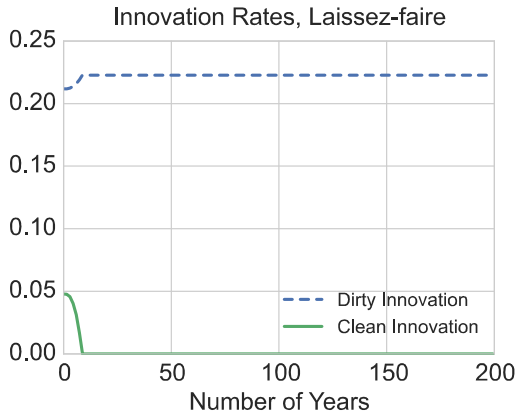

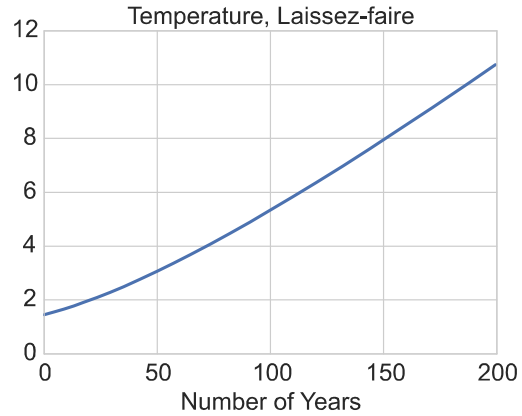Figure 4. Innovation rates under laissez-faire.

Figure 5. The time path of temperature increases under laissez-faire.

Moreover, at these innovation rates, technology gaps and the profitability of dirty technologies increase relative to those of clean technologies, and clean R&D rapidly converges to zero. Consequently, in the long run clean technologies disappear completely and dirty technologies take over the entire economy.[20] The obvious implication of this time path of innovations is a

---

[20]Note that this rapid disappearance of clean technologies does not conflict with the initial conditions shown in Figure 3, which had 9% of product lines being led by clean technology. This is because the current stock of technologies is not generated by a laissez-faire equilibrium, but by one in which clean technologies receive R&D subsidies as described in Section 5 and other protections (e.g., through pollution standards).

steady increase in dirty energy production and carbon emissions. Figure 5 shows an increase in temperature of an additional 11°C in the next 200 years.[21]

5 Policy Analysis

In this section, we characterize the policies that maximize discounted welfare given our estimated parameters and then consider various counterfactual policy experiments.

5.1 Optimal Policy

We start with optimal policy.[22] Throughout, we do not allow the social planner to correct for monopoly distortions, thus limiting ourselves to the policy instruments discussed above—carbon taxes and subsidies to clean research.[23] In fact, our theoretical analysis makes it clear that what is relevant is the differential tax and subsidy rates for clean vs. dirty energy, motivating us to focus on taxes on dirty production, which we refer to as "carbon taxes," and subsidies to clean innovation. Finally, for computational reasons, we model taxes and subsidies as quartic functions of calendar time. The resulting optimal policies are presented in Figure 6 (with the research subsidy shown on the left axis and the carbon tax on the right axis).
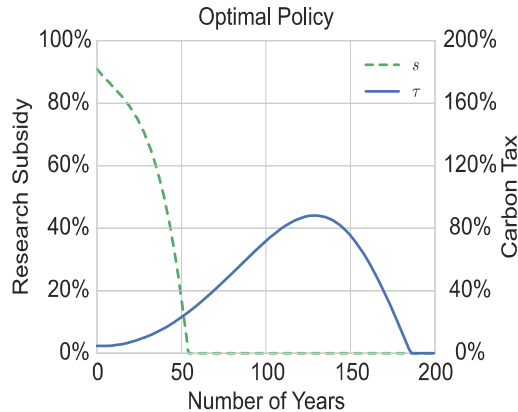

Figure 6. Optimal policies (carbon taxes and research subsidies) under baseline parameters.

Figure 6 shows a very high level of research subsidy, especially during the first few decades. The intuition for why optimal policy relies so much on subsidies to clean research is instructive. The social planner would like to divert R&D from carbon-intensive dirty technologies towards clean technologies. She can do so by choosing a sufficiently high carbon tax rate today and

---

[21] We compute temperature changes as $\Delta temperature = \lambda \left(\ln S_t - \ln \bar{S} \right) / \ln 2$.

[22] Because of the non-linear dynamics of atmospheric carbon concentration, optimal policy is not necessarily time-consistent. We ignore this problem by assuming that the social planner is able to commit to the future sequence of taxes and subsidies.

[23] As mentioned above, in the one-sector version of our model (either with only dirty or only clean technology), taxes or subsidies to research would only affect relative wages of skilled workers (employed in the research sector), and crucially not the aggregate rate of innovation. For this reason, subsidies to clean research or taxes on dirty research are identical in our model.

in the future, as this would reduce the profitability of production using dirty technologies and secure both a switch to clean production and, on the basis of this, to research directed at clean technologies. However, this is socially costly because given the current state of technology, switching most production to clean technology has a high consumption cost (because the marginal costs of production of clean technologies are initially significantly higher than those of dirty technologies). Hence, the social planner prefers to use the carbon tax to only deal with the carbon emission externality and to rely on the research subsidy to redirect R&D towards clean technologies. Figure 7 shows that the research subsidy is indeed sufficient to rapidly switch innovation from the dirty to clean technology.
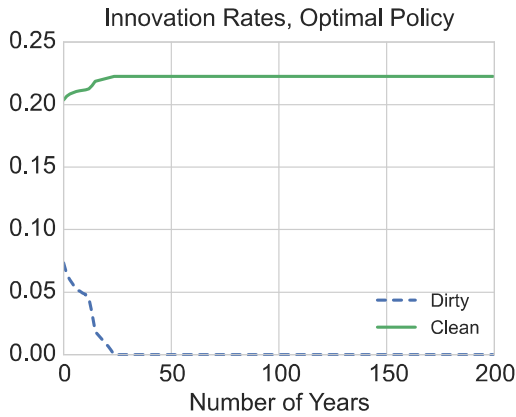

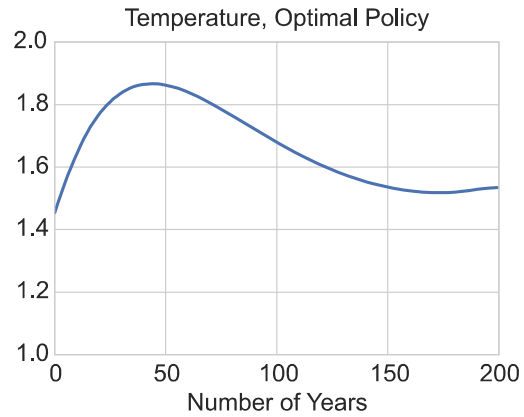Figure 7. Innovation rates under optimal policies.   Figure 8. The time path of temperature increases under optimal policies.

Figure 6 also shows that while the optimal research subsidy is front-loaded, the optimal carbon tax is hump-shaped. It starts out very low and increases gradually over the first 130 years to almost 90%, and then declines back down to zero by about year 180. This is also intuitive: the research subsidy is front-loaded because the social planner would like to switch research towards the clean technology as soon as possible. Once this is achieved and the change in the distribution of technology gaps makes clean research sufficiently profitable, the research subsidy is phased out. Given the research subsidy, which is highly effective in redirecting technological change, the planner initially sets a low carbon tax so as not to excessively distort early consumption. The carbon tax is raised over time, which reflects two forces; (i) future distortions are less costly to the social planner because of discounting; (ii) as clean technologies improve, the distortionary effects of the carbon tax are reduced. Finally, as most production switches to the clean technology, there is less need for the carbon tax, and it is phased out.[24]

---

[24] The path of optimal research subsidy and carbon tax here are fairly similar to that in Acemoglu et al. (2012), despite the fact that the details of the two models and the exact distortions that the carbon tax creates are quite different. In particular, the different production structure here implies that a carbon tax creates no

Figure 8 depicts the implied path of temperature under the optimal policy. The combination of the research subsidy and the carbon tax keeps the temperature increase below 2 degrees, and in fact, the temperature starts reverting down after about 40 years. This reflects both the switch of technological change to clean technologies and the role of the carbon tax in discouraging the use of existing dirty technologies.

5.2 Counterfactual Policy Analysis

We next investigate the welfare and climatic implications of a range of alternative policies. We first focus on two counterfactuals. The first concerns the choice of an optimal policy relying only on a carbon tax (i.e., no research subsidy), and the second involves delaying intervention for 50 years and then choosing the optimal policy from that point onwards. The optimal policies are presented in Figure 9.
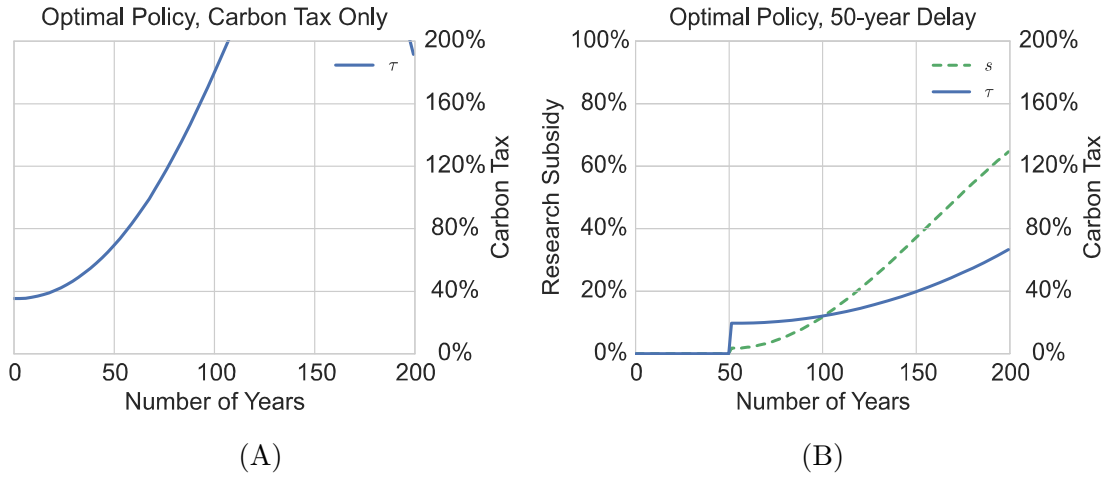

Figure 9. Constrained optimal policies following a 50-year delay
and the restriction to only carbon taxes

Panel A shows that optimal policy becomes more aggressive when the only policy tool is the carbon tax. The carbon tax in this case starts out higher than in the baseline and continues to increase to above 300%, though it still has a hump-shaped form and declines after about 150 years, coming down to below 200% around year 200 as is visible at the end of the 200-year window shown in the figure. The much higher level of the carbon tax in this case is because this policy tool is being used not just to reduce current emissions but also to redirect innovation towards clean technologies. An interesting implication of this constrained policy

distortion unless it induces some intermediates to switch to clean technology. This means that when there are few intermediates remaining with the dirty technology, the benefits from a carbon tax are low, but so are the costs, as we show analytically in the Online Appendix. Nevertheless, along the path induced by that optimal policy, the fraction of product lines using the dirty technology falls below 1% in 100 years, thus significantly reducing the role of carbon taxes thereafter and accounting for the decline of the carbon tax to zero in about 180 years.

is that aggregate temperature increases less at long horizons because of the more aggressive carbon tax, but this is at the expense of slower consumption growth, especially early on. As a result, the welfare cost of just relying on carbon tax for optimal policy (in terms of an equivalent loss of initial consumption) is 1.9%.

Somewhat paradoxically, delaying the start of optimal policies by 50 years leads to less aggressive policies from that point onwards. This is because the intervening interval has generated a bigger technology gap between the clean and dirty sectors, making a rapid switch from dirty to clean technologies thereafter undesirable (the switch now takes place in about 300 years). Because the economy now generates greater consumption early on, the welfare loss from this 50-year delay is a relatively modest 1.7%.[25]

A related counterfactual is to focus on time-invariant policies. Under this restriction, the optimal research subsidy and carbon tax are 63% and 13%, respectively. The welfare loss from using time-invariant policies is also relatively modest, just 1%, which in part reflects the fact that once the switch to clean technology has taken place, keeping the research subsidy and the carbon tax high only have moderate costs.

Finally, we also evaluate climatic and welfare implications of maintaining current US policies (here interpreted for the whole world) relative to adopting the optimal policy. For this purpose, we have tried to estimate the carbon taxes implied by US policies and the current subsidies to clean innovation (relative to dirty R&D) in our sample of firms. There is much uncertainty about what the carbon tax in the United States will be moving forward. A cap-and-trade program is likely to be implemented, but it is unclear what the implied carbon tax rate will be. Greenstone et al. (2013) estimate a social cost of carbon equal to about \$21 in 2010, expressed in 2007 dollars, and this number is currently being used for cost-benefit analysis by US agencies. Because future emissions are expected to become more harmful, this social cost estimate is forecasted to increase to \$45 in 2050 (in \$2007). We therefore use two values for the "business-as-usual" carbon tax, 0% consistent with the current situation, and 23% (consistent with a \$45 social cost of carbon in 2050).[26] We estimate the current clean research subsidy from our sample as follows: over our full 30-year period, 49% of all R&D expenditures by our clean firms are federally funded, while the same number is 11% for our dirty firms. This implies a 43% $((1 - 0.49) / (1 - 0.11) \simeq 1 - 0.43)$ subsidy for clean R&D relative to dirty R&D.

Because a 43% research subsidy for clean is insufficient to redirect technological change towards clean with no carbon tax, the scenario with a zero carbon tax involves rapid increases

---

[25] If we increase γ to five times its baseline value (as we do in our robustness checks below) so that the damages from not reducing carbon emissions become much higher, then policies following a 50-year delay are indeed more aggressive than the baseline and the cost of delaying intervention is significant (about 12.5%).

[26] In particular, US carbon emissions are 1.58 billion tons in 2002. One metric ton of carbon is equivalent to 3.667 units of carbon dioxide. Our dirty firms have sales of approximately one trillion dollars in this year. The \$45 social cost is \$39 in 2002 terms. These numbers imply a real tax rate in 2050 of about 23% $((39 \times 3.667 \times 1.58 \times 10^9)/10^{12} \simeq 0.23)$. This carbon tax rate is much less than currently used in Sweden (see Golosov et al. 2014) and also less than the numbers suggested by Nordhaus (2008).

in temperature in the first several hundred years and leads to close to 100% welfare costs (regardless of the discount rate). Interestingly, however, even with this less than optimal subsidy to clean research, it turns out that the temperature increase can be contained provided that there is a moderate carbon tax of 23%. In fact, because this (suboptimally low) combination of research subsidy and carbon tax is still sufficient to induce a rapid switch to clean technology, the welfare loss is only 1.2%.

6   Robustness and Extensions

In this section, we investigate how our estimation, optimal policy, and counterfactual results are affected by a range of different approaches, modeling assumptions, and variations on parameter estimates. We economize on space by only reporting the implied optimal policies, even when the variation in question involves reestimation. Overall, we find that our main conclusions, especially those concerning the form of optimal policies, are fairly robust across these variations.

6.1   Three-Step Policy

The first robustness exercise considers policies that take a simple "step function form" with three endogenously determined switch points rather than optimal policies modeled as a quartic in time.

The optimal policies in this case are depicted in Figure 10, which shows a similar pattern to Figure 6, except that the carbon tax now remains high for a longer interval, coming down only after about 440 years. We therefore conclude that the specific computation restrictions we imposed in our baseline analysis are not particularly important for our main conclusions.
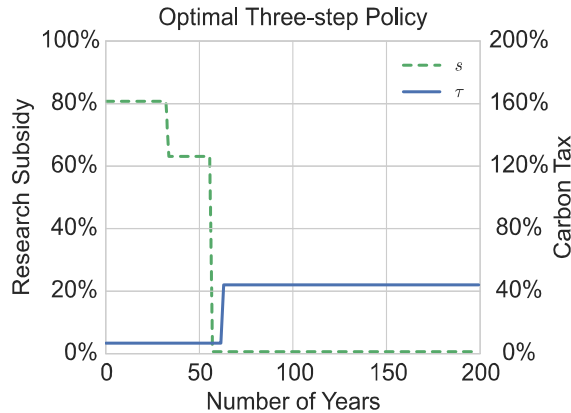

Figure10. Optimal policies restricted to a three-step form.

6.2   Lower Social Discount Rate

An important debate in the optimal climate policy literature concerns the social discount rate. Here we follow Stern (2007) and investigate the implications of lower social discount rates, in

particular, $\rho = 0.5\%$ and 0.1% (while still keeping the private discount rate at 1% so that the implied interest rate is not counterfactually low). Figure 11 shows that with 0.5%, the qualitative features of the optimal policy are very similar to our baseline, though the carbon tax now increases to somewhat higher levels.

With the much lower social discount rate of 0.1%, the carbon tax rises much more rapidly to much higher levels so as to cut emissions immediately (since these emissions create persistent costs in the future, which the social planner now cares much more about). But even in this case, the carbon tax has a hump-shaped path, and starts coming down after 150 periods. Overall, we read the patterns in Figure 11 as suggesting that our qualitative, and to some degree quantitative, conclusions are fairly robust to reasonable variations in social discount rate.
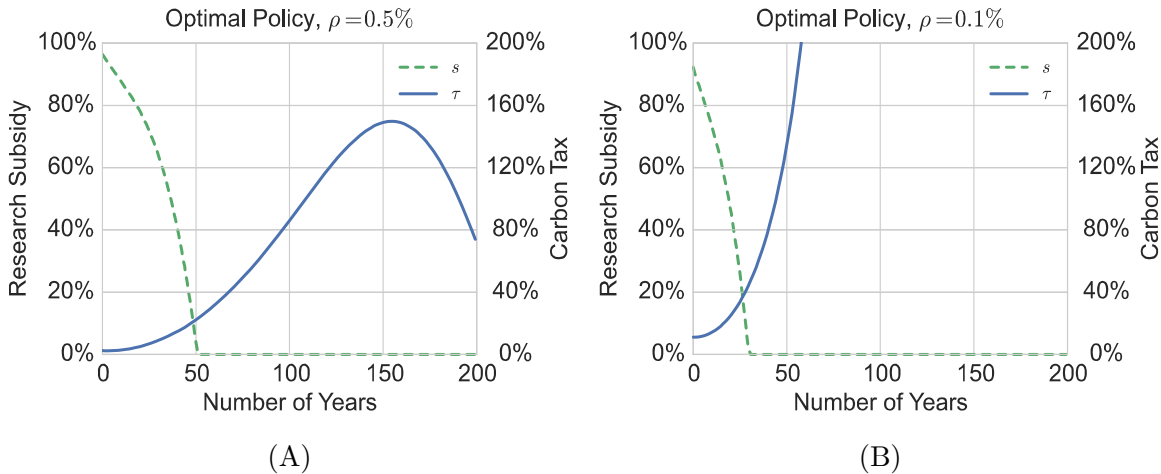

Figure 11. Optimal policies under a social discount rate of 0.5% and 0.1%.

6.3 Alternative Specifications of Carbon Emissions

Our model considers emissions that are proportional to the use of dirty output. An alternative is to assume that emissions are proportional to the amount of exhaustible resource used in dirty production as in Golosov et al. (2014). As noted in footnote 10, this amounts to assuming that even technological changes in the dirty sector reduce emissions per unit of energy output. As in Golosov et al. (2014), this would not generate enough emissions to make the climate change implications sufficiently costly. We therefore follow their paper in this case and assume that, in addition to having emissions proportional to the use of the exhaustible resource, there is also technological change reducing the cost of extracting the exhaustible resource by 2% a year. Figure 12 shows that the qualitative features of the optimal policy are similar to our baseline in this case, with the only exception that the carbon tax now increases to significantly higher levels and comes down only after about 500 years.

Figure 13 shows that if we eliminate the exhaustible resource entirely (so that we have

$\nu = 0$ in the above model and the production function of the dirty sector in (2) simplifies to $y_{i,t}^d = q_{i,t}^d l_{i,t}^d$), the pattern of the optimal policy in this case is very similar to our baseline.
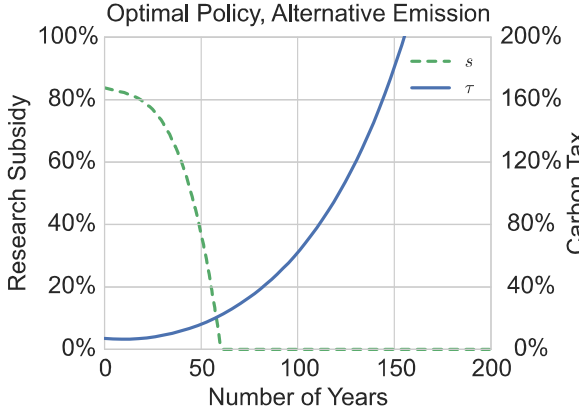

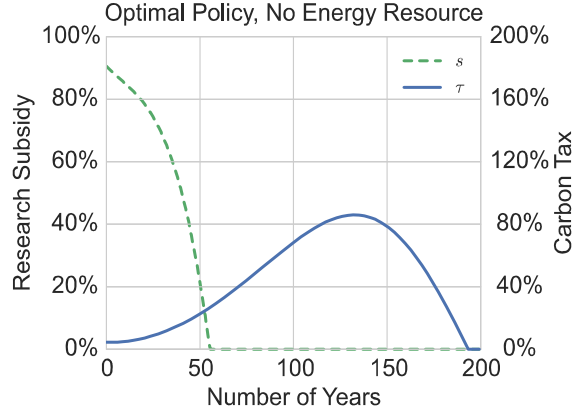Figure 12. Optimal policies when emissions are caused by the use of the exhaustible resource.

Figure 13. Optimal policies when dirty production does not use the exhaustible resource.

6.4 Alternative Damage Elasticity γ

As noted above, actual damages from atmospheric carbon may be greater than the estimates commonly used in the economics literature. We now show the sensitivity of our results to higher values of these damages, captured by the parameter γ, in our model. We consider two cases: first, where γ is twice as large as our baseline value of $\gamma = 5.3 \times 10^{-5}$, and second, where γ is five times as large as the baseline.

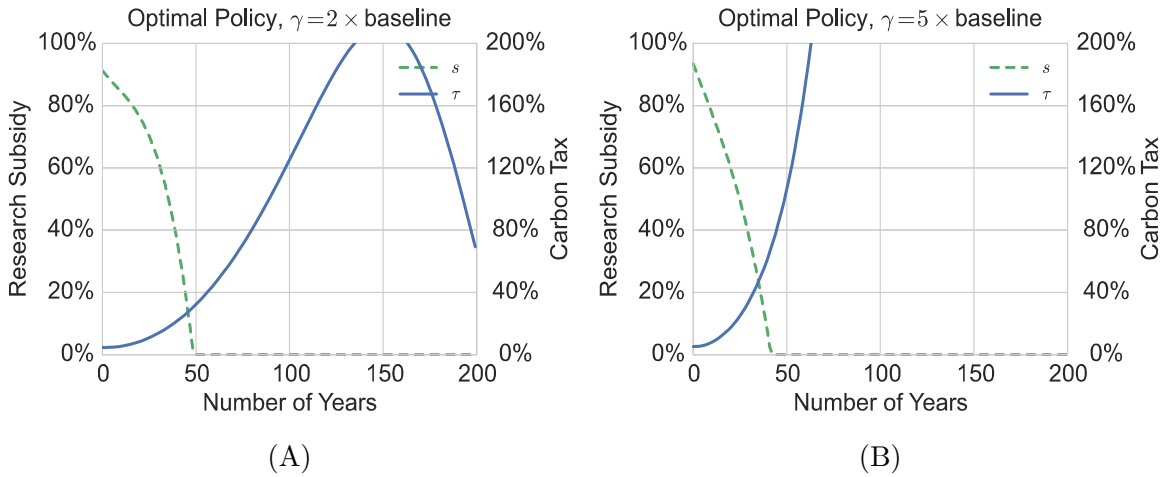(A)   (B)

Figure 14. Optimal policies under alternative economic damage scenarios.

34

Optimal policies are also similar when γ is twice as large as the baseline, with the only difference being that the carbon tax now increases to a higher level (about 200%) within the first 150 years, before declining towards zero again. When γ is raised to five times the baseline, the qualitative pattern is also similar and involves a front-loaded research subsidy and a hump-shaped path for the carbon tax, though the carbon tax now increases to much higher levels before starting to decline. We conclude that the heavy reliance of the optimal policy on research subsidies for the clean sector is robust to a fairly wide range of damage elasticities.

6.5 Varying the Distortions from Research Subsidies χ

We next investigate the robustness of our results to scenarios in which research subsidies create no distortions ($\chi = 0\%$) or create twice as large distortions as in the baseline model ($\chi = 20\%$). Without any additional distortions from the use of the research subsidy, optimal policy relies more heavily on this instrument, and the carbon tax increases less rapidly (and because it ramps up slowly, it also starts declining later, after about 540 years). With significantly higher distortions from the use of the research subsidy (20%), the form of the optimal policy is remarkably similar to the baseline, with only slightly higher carbon tax rates. Overall, these results imply that one of our main conclusions, the heavy reliance of optimal policy on research subsidies, is quite robust to reasonable distortions from such subsidies.
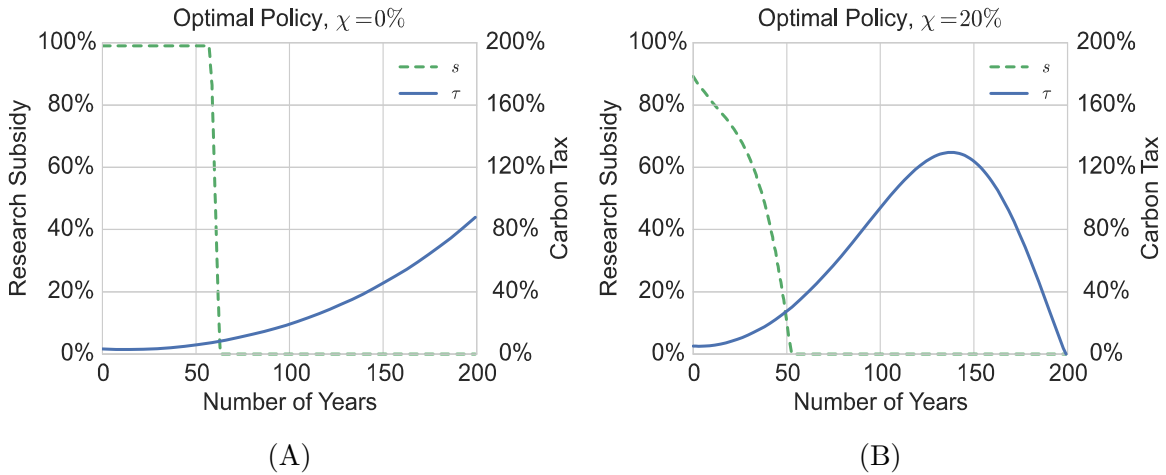

Figure 15. Optimal policies under different assumptions on the research subsidy distortions.

6.6 Alternative R&D Elasticities η

Our baseline results are for $\eta = 0.50$, which averaged across cross-sectional and first-difference estimates. In Figure 16, we reestimate the model using first a value of η in the ballpark of the cross-sectional estimates ($\eta = 0.65$) and then for a value corresponding to the first-difference estimates ($\eta = 0.35$), and investigate the implications of this for optimal policy.

35

The resulting optimal policies are again similar to the baseline, except that with $\eta = 0.35$, because research more heavily relies on the past stock of knowledge, redirecting technological change towards the clean sector is somewhat less effective, and the social planner makes heavier use of the carbon tax (which increases to 530% in the first 200 years, coming down only after about 390 years). When $\eta = 0.65$, the optimal policy is very similar to the baseline.
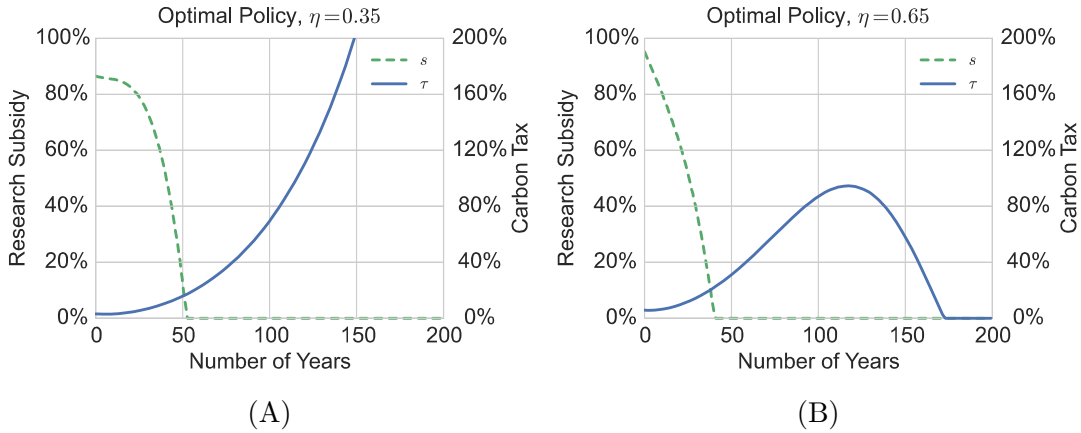

(A)   (B)

Figure 16. Optimal policies under alternative R&D elasticities.

6.7   Alternative Leapfrogging Probabilities α

We also investigate the implications of lower and higher leapfrogging rates ($\alpha = 0.03$ and 0.05). Figure 17 shows that the results are very similar to the baseline, with the only noteworthy difference being that with $\alpha = 0.03$, the slower rate of leapfrogging increases the reliance on the carbon tax slightly, leading to a somewhat higher peak (around 140%) for the optimal carbon tax schedule.
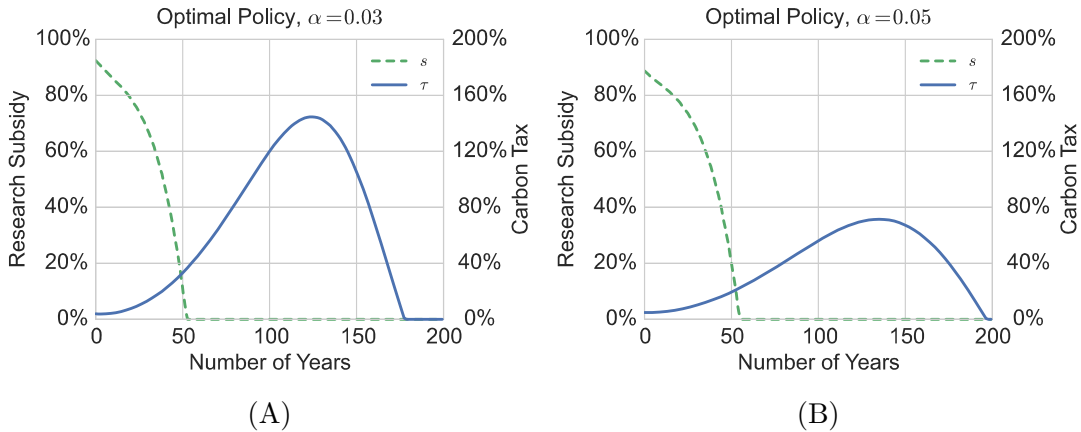

(A)   (B)

Figure 17. Optimal policies under alternative leapfrogging probabilities.

36

6.8 Alternative Initial Technology Distribution

Finally, we consider a variation on the initial technology distribution. We compute an alternative initial technology gap distribution with three modifications: [1] we weight patents by their normalized citation counts; [2] we consider four-digit (SIC4) industries, and [3] we focus on the manufacturing and energy sectors. There are 332 SIC4 industries that are of sufficient size to pass disclosure restrictions, and 9.4% are led by the clean technologies.[27] Figure 18 shows that the optimal policy is again very similar to the baseline.
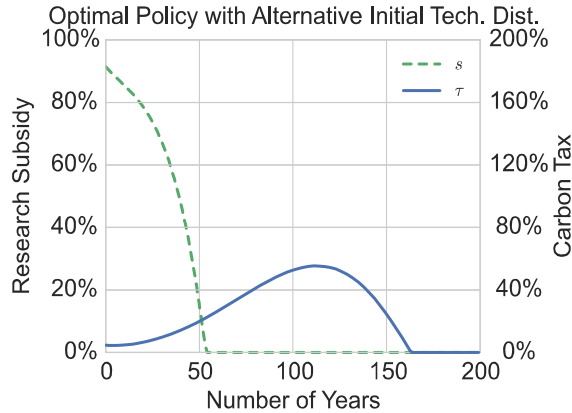

Figure 18. Optimal policies under alternative initial technology distributions.

Overall, the robustness exercises presented in this section show that the qualitative, and even quantitative, patterns of optimal policy, including the heavy reliance on research subsidies, are fairly robust.

7 Conclusion

One of the central challenges facing the world economy is reducing carbon emissions, which appears to be feasible only if a successful transition to clean technology is induced. This paper has investigated the nature of a transition to clean technology theoretically and empirically. We developed a microeconomic model where clean and dirty technologies compete in production and innovation. If dirty technologies are more advanced to start with, the potential transition to clean technology can be difficult both because clean research must climb several steps to catch up with dirty technology and because this gap discourages research effort directed towards clean technologies. We estimated our key model parameters from firm-level microdata in the US energy sector, using regression analysis and SMM. Our model performs fairly well in matching a range of patterns in the data that were not directly targeted in the estimation, giving us confidence that it is useful for the analysis of the transition to clean technology in the US energy sector.

---

[27] Additional moments of this distribution are provided in the Online Appendix.

Theoretically, carbon taxes and research subsidies encourage production and innovation in clean technologies. The key question we investigate using our estimated quantitative model is the structure and time path of optimal policies, how rapidly they will be able to secure a transition to clean technology and slow down the potential increases in temperatures, and what the costs of alternative, non-optimal policies are.

A natural intuition would be that only carbon taxes should be used because the key externalities in this model are created by carbon (in the absence of these carbon externalities, the social planner would have no reason to interfere with or subsidize research). In contrast to this intuition, we find that optimal policy heavily relies on research subsidies, and this result is very robust across a range of variations and for different damages and social discount rates. We also use the model to evaluate the welfare consequences of a range of alternative policy structures. For example, just relying on carbon taxes or delaying intervention both have significant welfare costs.

Though, to the best of our knowledge, this is the first attempt to develop a microeconomic model of the transition to clean technology and to quantitatively characterize optimal policy in such a setup, our paper has inevitably left several questions unanswered and taken a number of shortcuts, all of which constitute interesting areas for future research and investigation. We list some of these that we view as particularly important here. First, we have abstracted both from cross-country variation in policies and from the endogenous speed of clean technology transfer across countries, which is likely to be central for the future of climate change and for the structure of optimal policy. Second and related, we have also abstracted from game-theoretic interactions in emissions, policies and technology choice across several countries in the global economy, which become important when multiple countries individually choose their policies (e.g., Harstad, 2012; Dutta and Radner, 2006). Third, we have followed the literature in this area in ignoring non-linear threshold effects in the impact of atmospheric carbon on economic efficiency. Incorporating such non-linearities, together with an explicit approach to uncertainty along the lines of Weitzman (2009), would be an important area for future research. Fourth, as noted above, our optimal policies are characterized under the assumptions of commitment to the policy sequence by the social planner. A major next step is to characterize time-consistent optimal policy in the absence of perfect commitment. Fifth, another interesting area is to investigate the interactions between international trade, technology and emissions (see Hemous, 2012). Finally, our framework can also be augmented by considering a richer set of possible technological improvements, including those that enable expansions in the stock of exhaustible resources and technologies, such as carbon sequestration, that reduce the climatic damage from dirty technologies.

References

Acemoglu, Daron. 1998. "Why do New Technologies Complement Skills? Directed Technical Change And Wage Inequality," *Quarterly Journal of Economics* 113: 1055-1090.

Acemoglu, Daron. 2002. "Directed Technical Change." *Review of Economic Studies* 69: 781-810.

Acemoglu, Daron and Ufuk Akcigit. 2012. "Intellectual Property Rights Policy, Competition and Innovation." *Journal of the European Economic Association* 10: 1-42.

Acemoglu, Daron, Philippe Aghion, Leonardo Bursztyn, and David Hemous. 2012. "The Environment and Directed Technical Change." *American Economic Review* 102: 131-166.

Acemoglu, Daron, Ufuk Akcigit, Nicholas Bloom, and William Kerr. 2013. "Innovation, Reallocation, and Growth." NBER Working Paper 18993.

Adao, Bernardino, Borghan Narajabad, and Ted Temzelides. 2012. "Renewable Technology Adoption and the Macroeconomy." Rice University Working Paper.

Adda, Jerome, and Russell Cooper. 2003. *Dynamic Economics: Quantitative Methods and Applications*, MIT Press.

Aghion, Philippe, Antoin Dechezlepretre, David Hemous, Ralf Martin, and John Van Reenen. 2014. "Carbon Taxes, Path Dependency and Directed Technical Change: Evidence from the Auto Industry." *Journal of Political Economy*, forthcoming.

Aghion, Philippe, Christopher Harris, Peter Howitt, and John Vickers. 2001. "Competition, Imitation and Growth with Step-by-Step Innovation." *Review of Economic Studies* 68: 467-492.

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica* 110: 323-351.

Akcigit, Ufuk, and William Kerr. 2010. "Growth through Heterogeneous Innovations." NBER Working Paper 16443.

Archer, David. 2005. "Fate of Fossil Fuel CO2 in Geologic Time." *Journal of Geophysical Research* 110: C09S05.

Boden, Tom A., Gregg Marland, and Robert J. Andres (2010). "Global, Regional, and National Fossil-Fuel CO2 Emissions." Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, TN.

Bovenberg, Lans, and Sjak Smulders. 1995. "Environmental Quality and Pollution-Augmenting Technological Change in a Two-Sector Endogenous Growth Model." *Journal of Public Economics* 57: 369-391.

Bovenberg, Lans, and Sjak Smulders. 1996. "Transitional Impacts of Environmental Policy in an Endogenous Growth Model." *International Economic Review* 37: 861-893.

BP (British Petroleum). 2010. "BP Statistical Review of World Energy," available at http://bp.com/statisticalreview.

Davis, Steven, John Haltiwanger, and Scott Schuh. 1996. *Job Creation and Destruction,* MIT Press.

DOE (Department of Energy). 2012. "The Federal Budget, Fiscal Year 2012," available at www.whitehouse.gov/omb/factsheet_department_energy.

Dutta, Prajit K., and Roy Radner. 2006. "A Game-Theoretic Approach to GlobalWarming," *Advances in Mathematical Economics* 8: 135-53.

Emanuel, Kerry. 2005. "Increased Destructiveness of Tropical Cyclones Over the Past 30 Years." *Nature* 436: 686-688.

Golosov, Mikhail, John Hassler, Per Krusell, and Aleh Tsyvinski. 2014. "Optimal Taxes on Fossil Fuel in General Equilibrium." *Econometrica* 82: 41-88.

Goulder, Lawrence, and Koshy Mathai. 2000. "Optimal CO2 Abatement in the Presence of Induced Technological Change." *Journal of Environmental Economics and Management* 39: 1-38.

Goulder, Lawrence, and Stephen Schneider. 1999. "Induced Technical Change and the Attractiveness of CO2 Abatement Policies." *Resource and Energy Economics* 21: 211-253.

Greenstone, Michael, Elizabeth Kopits, and Anne Wolverton. 2013. "Developing a Social Cost of Carbon for US Regulatory Analysis: A Methodology and Interpretation." *Review of Environmental Economics and Policy* 7(1): 23-46.

Grimaud, Andre, Gilles Lafforgue, and Bertrand Magné. 2011. "Climate Change, Mitigation Options and Directed Technical Change: A Decentralized Equilibrium Analysis." *Resource and Energy Economics* 33: 938-962.

Grossman, Gene, and Elhanan Helpman. 1991. *Innovation and Growth in the Global Economy,* MIT Press.

Harris, Christopher, and John Vickers. 1995. "Innovation and Natural Resources: A Dynamic Game with Uncertainty." *RAND Journal of Economics* 26: 418-430.

Harstad, Bard. 2012. "Climate Contracts: A Game of Emissions, Investments, Negotiations, and Renegotiations." *Review of Economic Studies* 79(4): 1527-1557.

Hartley, Peter, Kenneth Medlock, Ted Temzelides, and Xinya Zhang. 2011. "Energy Sector Innovation and Growth: An Optimal Energy Crisis." Rice University Working Paper.

Hassler, John, and Per Krusell. 2012. "Economics and Climate Change: Integrated Assessment in a Multi-Region World." *Journal of the European Economic Association* 10(5): 974-1000.

Hassler, John, Per Krusell, and Conny Olovsson. 2011. "Energy-Saving Technical Change." NBER Working Paper 18456.

Hemous, David. 2012. "Environmental Policy and Directed Technical Change in a Global Economy: The Dynamic Impact of Unilateral Environmental Policies?", INSEAD Mimeo.

IPCC (Intergovernmental Panel on Climate Change). 2007. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of*

the Intergovernmental Panel on Climate Change, Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jaffe, Adam, David Popp, and Richard Newell. 2010. "Energy, the Environment, and Technological Change." in *Handbook of the Economics of Innovation: Vol. 2*, Bronwyn Hall and Nathan Rosenberg, eds., Academic Press/Elsevier, 873-937.

Klette, Tor Jakob, and Samuel Kortum. 2004. "Innovating Firms and Aggregate Innovation." *Journal of Political Economy* 112: 986-1018.

Krusell, Per, and Anthony Smith. 2009. "Macroeconomics and Global Climate Change: Transition for a Many-Region Economy." Working Paper.

Landsea, Christopher. 2005. "Atlantic Hurricanes and Global Warming." *Nature* 438: E11-E12.

Lentz, Rasmus, and Dale Mortensen. 2008. "An Empirical Model of Growth through Product Innovation." *Econometrica* 76: 1317-1373.

MacCracken, Christopher, James Edmonds, Son Kim, and Ronald Sands. 1999. "The Economics of the Kyoto Protocol." *The Energy Journal* 20: 25-72.

Mendelsohn, Robert, William Nordhaus, and Daigee Shaw. 1994. "The Impact of Global Warming on Agriculture: A Ricardian Analysis." *American Economic Review* 84: 753-771.

Mueller, Nicholas, Robert Mendelsohn, and William Nordhaus. 2011. "Environmental Accounting for Pollution in the United States Economy." *American Economic Review* 101: 1649-1675.

Newell, Richard, Adam Jaffe, and Robert Stavins. 1999. "The Induced Innovation Hypothesis and Energy-Saving Technological Change." *Quarterly Journal of Economics* 114: 941-975.

Nicholls, Robert, and Jason Lowe. 2006. "Climate Stabilisation and Impacts of Sea-level Rise." in *Avoiding Dangerous Climate Change*, H.J. Schellnhuber, ed., 195-201.

Nordhaus, William. 1994. *Managing the Global Commons: The Economics of Climate Change*, MIT Press.

Nordhaus, William. 2007. "To Tax or Not to Tax: The Case for a Carbon Tax." *Review of Environmental Economics and Policy*, 26-44.

Nordhaus, William. 2008. *A Question of Balance: Weighing the Options on Global Warming Policies*, Yale University Press.

Nordhaus, William, and Joseph Boyer. 2000. *Warming the World: Economic Modeling of Global Warming,* MIT Press.

Pizer, William. 1999. "The Optimal Choice of Climate Change Policy in the Presence of Uncertainty." *Resource and Energy Economics* 21: 255-287.

Popp, David. 2002. "Induced Innovation and Energy Prices." *American Economic Review* 92: 160-180.

Popp, David. 2004. "ENTICE: Endogenous Technological Change in the DICE Model of

Global Warming." *Journal of Environmental Economics and Management* 48: 742-768.

Popp, David. 2006. "Innovation in Climate Policy Models: Implementing Lessons from the Economics of R&D." *Energy Economics* 28: 596-609.

Popp, David, and Richard Newell. 2012. "Where Does Energy R&D Come From? Examining Crowding Out From Energy R&D." *Energy Economics* 3: 980-991.

Rogner, Holger H. 1997. "An Assessment of World Hydrocarbon Resources," *Annual Review of Energy and the Environment* 22: 217–262.

Stern, Nicholas. 2007. *The Economics of Climate Change: The Stern Review*, Cambridge University Press.

Stott, Peter, D.A. Stone, and M.R. Allen. 2004. "Human Contribution to the European Heatwave of 2003." *Nature* 432: 610-614.

Van der Zwaan, Robert, Reyer Gerlagh, G. Klaassen, and L. Schrattenholzer. 2002. "Endogenous Technical Change in Climate Change Modelling." *Energy Economics* 24: 1-19.

Weitzman, Martin. 2009. "On Modeling and Interpreting the Economics of Catastrophic Climate Change." *Review of Economics and Statistics* 91: 1-19.

Online Appendix

A-1 Proofs and Additional Theoretical Results

A-1.1 Proof of Lemma 1

Proof of Lemma 1. Substituting (15) into (14), we obtain

$$
r_t \left(\sum_{i=1}^u v_{n_i,t}^j \right) Y_t - \left(\sum_{i=1}^u \dot{v}_{n_i,t}^j \right) Y_t - \left(\sum_{i=1}^u v_{n_i,t}^j \right) \dot{Y}_t
$$
$$
= \sum_{i=1}^u \left[\begin{array}{c} \pi_{n_i}^j + z^j \left(-v_{n_i,t}^j Y_t \right) + \\ z^{-j} \left(1 - \alpha + \mathbb{I}_{(n_i^j \le 0)} \alpha \right) \left(-v_{n_i,t}^j Y_t + v_{n-1,t}^j Y_t \right) + \mathbb{I}_{(n_i^j>0)} z^{-j} \alpha \left(-v_{n_i,t}^j Y_t + v_{-1,t}^j Y_t \right) \end{array} \right]
$$
$$
+ \int \max_{x^j \ge 0} \left[u^j x^j \mathbb{E}_n v_{n_i,t}^j Y_t - \left(1 - s_I^j \right) u^j w^s \left((x^j)^{\frac{1}{\eta}} \theta^{-\frac{1}{\eta}} + \mathbb{I}_{(x^j>0)} F_I \right) \right] dF_I
$$

This can further be simplified to

$$
r_t \sum_{i=1}^u v_{n_i,t}^j - \sum_{i=1}^u \dot{v}_{n_i,t}^j - \sum_{i=1}^u v_{n_i,t}^j g_{Y,t}
$$
$$
= \sum_{i=1}^u \left[\begin{array}{c} \tilde{\pi}_{n_i}^j - z^j v_{n_i,t}^j + \\ z^{-j} \left(1 - \alpha + \mathbb{I}_{(n_i^j \le 0)} \alpha \right) \left(v_{n-1,t}^j - v_{n_i,t}^j \right) + \mathbb{I}_{(n_i^j>0)} z^{-j} \alpha \left(v_{-1,t}^j - v_{n_i,t}^j \right) \end{array} \right]
$$
$$
+ \int \max_{x^j \ge 0} \left[u^j x^j \bar{v}_t^j - \left(1 - s_I^j \right) u^j \tilde{w}^s \left((x^j)^{\frac{1}{\eta}} \theta^{-\frac{1}{\eta}} + \mathbb{I}_{(x^j>0)} F_I \right) \right] dF_I.
$$

where

$$
\tilde{w}^s \equiv \frac{w^s}{Y} \quad \text{and} \quad \tilde{\pi} = \frac{\pi}{Y}.
$$

Defining $\tilde{r}_t \equiv r_t - g_{Y,t}$ delivers the desired result. ∎

A-1.2 The Evolution of Quality Distributions

The quality indices in dirty sectors evolve according to following flow equations:

$$
\dot{Q}_{n>1,t}^d = z_t^d \lambda Q_{n-1,t}^d + (1-\alpha) z_t^c Q_{n+1,t}^d - z_t Q_{n,t}^d
$$
$$
\dot{Q}_{1,t}^d = z_t^d \lambda Q_{0,t}^d + (1-\alpha) z_t^c Q_{2,t}^d + \alpha z_t^d \lambda \sum_{n<0} Q_{n,t}^c - z_t Q_{1,t}^d
$$
$$
\dot{Q}_{0,t}^d = (1-\alpha) z_t^d \lambda Q_{-1,t}^d + (1-\alpha) z_t^c Q_{1,t}^d - z_t Q_{0,t}^d
$$
$$
\dot{Q}_{-1,t}^d = (1-\alpha) z_t^d \lambda Q_{-2,t}^d + z_t^c Q_{0,t}^d + \alpha z_t^c \sum_{n>0} Q_{n,t}^d - z_t Q_{-1,t}^d
$$
$$
\dot{Q}_{n<-1,t}^d = (1-\alpha) z_t^d \lambda Q_{n-1,t}^d + z_t^c Q_{n+1,t}^d - z_t Q_{n,t}^d
$$

The first line expresses the change in the quality index of sectors with $n > 1$. When there is a new dirty innovation in sector with $n - 1$ at the rate z_t^d, on average its quality is improved

A-1

by a multiplicative factor λ. Likewise, when there is a new incremental clean innovation at the rate $(1-\alpha)z_t^c$ in sectors with $n+1$, they enter into state n. Finally, there is an outflow from state n through clean or dirty innovations that occur at the rate z_t. The other lines represent the flow equations for state $n-1$, $n=0$ and $n<0$ respectively. Similar flow equations can be expressed for clean technologies as follows:

$$
\begin{aligned}
\dot{Q}_{n>1,t}^c &= z_t^d Q_{n-1,t}^c + (1-\alpha)z_t^c \lambda Q_{n+1,t}^c - z_t Q_{n,t}^c \\
\dot{Q}_{1,t}^c &= z_t^d Q_{0,t}^c + (1-\alpha)z_t^c \lambda Q_{2,t}^c + \alpha z_t^d \sum_{n<0} Q_{n,t}^c - z_t Q_{1,t}^c \\
\dot{Q}_{0,t}^c &= (1-\alpha)z_t^d Q_{-1,t}^c + (1-\alpha)z_t^c \lambda Q_{1,t}^c - z_t Q_{0,t}^c \\
\dot{Q}_{-1,t}^c &= (1-\alpha)z_t^d Q_{-2,t}^c + z_t^c \lambda Q_{0,t}^c + \alpha z_t^c \lambda \sum_{n>0} Q_{n,t}^d - z_t Q_{-1,t}^c \\
\dot{Q}_{n<-1,t}^c &= (1-\alpha)z_t^d Q_{n-1,t}^c + z_t^c \lambda Q_{n+1,t}^c - z_t Q_{n,t}^c.
\end{aligned}
$$

A-1.3 Optimal Constant Policy

To develop an intuition about the nature of optimal policy, we now consider a simplified optimal policy problem. In particular, we abstract from dynamics and from the exhaustible resource, and assume that the dirty output enters linearly into the production function, so that

$$
\ln Y = \int_0^1 \ln y_i \, di - \gamma \int_0^1 y_i^d \, di. \tag{A-1}
$$

This implies that the objective of the social planners to maximize net output. In the absence of the exhaustible resource, the unskilled labor market clearing condition becomes

$$
1 = \int_0^1 l_i^c \, di + \int_0^1 l_i^d \, di.
$$

Define

$$
x_i \equiv \frac{q_i^d}{q_i^c}.
$$

Let us rank sectors from 0 to 1 with ascending x_i, and also for ease of exposition, let us make the following assumptions:

1. x_i's distribution F is continuous and differentiable with density f,

2. all sectors charge a markup of λ.

Production is linear and satisfies $y_i = q_i^j l_i^j$, depending on which $j \in \{d, c\}$ is used. Furthermore, let us also normalize $q_i^d = 1$. We can order the product lines by x_i and define the cutoff to be μ, which will be a function of the tax τ, so that we can rewrite (A-1) as

$$
\ln Y = -\int_\mu^1 \ln x_i \, di + \underbrace{\int_0^\mu \ln l_i^d \, di + \int_\mu^1 \ln l_i^c \, di}_{\equiv L} - \gamma Y^d. \tag{A-2}
$$

A-2

Note that the marginal product line can be defined as

$$x^* = 1 + \tau \text{ and } \mu = 1 - F(1 + \tau). \tag{A-3}$$

In addition, noting that all dirty sectors will use the same labor and all clean sectors will use the same labor, and taking into account the impact of the carbon tax on dirty labor demand, we have

$$l^d = \frac{1}{\mu + (1 + \tau)(1 - \mu)} \text{ and } l^c = \frac{1 + \tau}{\mu + (1 + \tau)(1 - \mu)}.$$

Now we can show that

$$
\begin{aligned}
L &= \mu \ln \frac{1}{\mu + (1 + \tau)(1 - \mu)} + (1 - \mu) \ln \frac{1 + \tau}{\mu + (1 + \tau)(1 - \mu)} \\
&= (1 - \mu) \ln(1 + \tau) - \ln[\mu + (1 + \tau)(1 - \mu)]
\end{aligned}
$$

Likewise

$$Y^d = \frac{\mu}{\mu + (1 + \tau)(1 - \mu)}.$$

Let us assume that

$$F(x) = 1 - \frac{1}{x}.$$

Then from (A-3) we get the fraction of dirty sectors as

$$\mu = \frac{1}{1 + \tau}$$

and the total sum of the dirty output in (23)

$$Y^d = \frac{1}{1 + (1 + \tau)\tau}$$

Now the objective function can be expressed as

$$\max_\tau \left\{ -\int_{\frac{1}{1+\tau}}^1 \ln x_i di + \frac{\tau}{1 + \tau} \ln(1 + \tau) - \ln\left[\frac{1}{1 + \tau} + \tau\right] - \frac{\gamma}{1 + (1 + \tau)\tau} \right\}.$$

The first-order condition of the maximization problem is

$$\frac{\tau}{(1 + \tau)^2} - \frac{(1 + \tau)^2 - 1}{(1 + \tau)[1 + (1 + \tau)\tau]} + \gamma \frac{1 + 2\tau}{[1 + (1 + \tau)\tau]^2} = 0$$

This can be expressed as

$$\gamma = \frac{\tau[1 + (1 + \tau)\tau]}{(1 + \tau)^2} \equiv \Psi(\tau)$$

Note that $\Psi(0) = 0$ and $\lim_{\tau \to \infty} \Psi(\tau) = \infty$. Note also that $\Psi(\tau)$ is monotonically increasing in τ. Hence from intermediate value theorem the optimal tax rate $\tau^*(\gamma)$ exists and it is unique. Moreover, the optimal tax rate is increasing in the damage parameter γ.

A-3

A-2　Computational and Estimation Algorithms

A-2.1　Computational Algorithm

Our theoretical analysis shows that the key microeconomic decisions are independent of climate dynamics. We therefore solve for value functions, innovation rates, and distributions first, then use those to find the time path for the carbon stock, temperature, and other variables of interest.

The solution algorithm for this model involves finding the transition dynamics as the fixed point of a forward-backward iteration process, as in Conesa and Krueger (1998). To compute this fixed point, we first update the product line distribution (μ) in the forward direction and the value functions (V^j) in the backward direction, using the long-term (clean) steady state as the terminal condition.

For the estimation, we simulate a large panel of 16,384 firms, each with an evolving portfolio of product lines with various technological leads (n). After a lengthy burn-in time, we then sample the targeted statistics in order to find the best fit parameters. We use standard optimization routines to compute optimal policies. When focusing on constant policies, we use a straightforward grid search to find the optimum. In the time-varying case, we parameterize policies using a quartic carbon tax and a quartic research subsidy. We then search over this space of functions using a combination of a simple simulated annealing algorithm (Kirkpatrick et al., 1983) and a Nelder-Mead (simplex) algorithm (Nelder and Mead, 1965).

To solve for the fixed point of the sequence of value functions, we first discretize time into $N = 2048$ steps and set a terminal period $T = 2000$. Due to the symmetry between technology types inherent in this model, when a single type of technology is dominant—in the sense that the technology gap distribution is heavily skewed to either clean or dirty technology—one can analytically characterize value functions $v_{n,\infty}^j$ and innovation rates x_∞^j and z_∞^j. We use these values as terminal conditions, though we do not know in advance which technology (clean or dirty) will be dominant. In addition, we set large upper (100) and lower (-100) bounds on the step gap distribution space. The algorithm proceeds as follows:

1. Posit an initial guess for the value function at time zero of the form

$$v_{n,t}^j(0) = \frac{\pi_n^j}{\rho + \bar{z}} \quad \forall t,$$

   where $\bar{z}$ represents an estimated rate of creative destruction (we use $\bar{z} = 0.15$). Instantiate the technology gap distribution using the patent data with

$$\mu_{n,t}(0) = \mu_{n,0} \quad \forall t.$$

2. Iterate forward in time from $t = 0$ to $t = T$ by finding innovation rates x_t^j and z_t^j given value function and product distributions guesses at time $t + 1$, $v_{n,t+1}^j(k)$ and $\mu_{n,t+1}(k)$.

Using these innovation rates, update the time $t + 1$ product distribution $\mu_{n,t+1}(k + 1)$ using discrete time versions of the flow equations in (27).

3. Find the implied dominant technology at the terminal period by determining which technology type has a higher aggregate innovation rate as some late stage period $T - T_P$ (we use $T_P = 200$). Use the corresponding terminal value functions to update $v_{n,T}^j(k + 1) = v_{n,\infty}^j$.

4. Iterate backward in time from $t = T$ to $t = 0$ by updating value function $v_{n,t-1}^j(k + 1)$ using $v_{n,t}^j(k)$ and $\mu_{n,t}(k)$ according to a discretized version of the HJB equation in Lemma 1, re-solving for innovation rates x_t^j and z_t^j in the process.

5. Repeat steps 2-4 until the convergence criterion

$$\max_{n,t} \left| v_{n,t}^j(k + 1) - v_{n,t}^j(k) \right| < \varepsilon$$

is met. We use $\varepsilon = 10^{-6}$.

In order to avoid any instability, particularly when one is close to a threshold where the asymptotically dominant technology switches over, we also introduce heterogeneity in incumbent fixed costs as explained in the text.[28]

Using up-to-date computer hardware, the equilibrium solver takes anywhere from five seconds to two minutes, depending on the speed of convergence. The code is written mostly in Python, with core routines written in C/C++.

Finally, in our of policy analysis, we approximate optimal research subsidies and carbon taxes as quartic functions of time. In Section 6, we verify that the qualitative patterns are similar when we approximate these policies by step functions.

A-2.2 Estimation Algorithm

To find the moments used in the SMM estimation,[29] we simulate a panel of $16,384$ firms using equilibrium variables from the above model solving routine. Each firm has a portfolio of product lines with various technology gap values. We cap the maximum number of product lines a firm can have at 200. In order to determine the sales and R&D activity of the firm, we must keep track of both the number of product lines it is currently operating in, as well as the knowledge stock of the firm, which can in general differ. We simulate this panel of firms for 5 years to replicate the data generating process, and discretize time to have 100 subperiods per year, so that the simulations have 500 periods.

---

[28] A similar heterogeneity can be introduced in entrant fixed costs, but because entrants are never in the region where this heterogeneity matters, this makes no difference to our numerical procedures.

[29] McFadden (1989), Pakes and Pollard (1989) and Gourieroux and Monfort (1996) characterize the statistical properties of the SMM estimator.

A-3   Further Details on Data and Sample Construction

Our patent database was first developed by the NBER and was subsequently extended by HBS Research to include patenting in recent years. Each patent record provides information about the invention and the inventors submitting the application. Hall et al. (2001) provide extensive details about these data, and Griliches (1990) surveys the use of patents as economic indicators of technology advancement. Our focus is on industrial assignees with inventors in United States. In a representative year, 1997, this group comprised about 77,000 patents (about 40% of the total USPTO patent count in 1997). We focus on technology codes, that number of 150,000, given the broad nature of patent classes (about 450 groups). This is important as energy-sector patents are spread out over multiple patent classes—two examples related to solar energy are "Power Plants/utilizing natural heat/Solar" and "Stoves and Furnaces/Solar heat collector". Moreover, we use patenting technologies to classify firms as being primarily clean- or dirty-energy firms. This separation can only be done at the technology level as the patent class level includes both types (e.g., "Power Plants" includes technologies for coal-powered plants too). As a representative year, 1997, our energy-related patents comprised 7.6% of the total US patent count.

Our classifications follow closely Popp (2002) and Popp and Newell (2012), and we report results for our key parameters that just use their classification system. When seeking to extend their work, we use three steps. We first utilize resources from the OECD's work to identify environment-related technologies. OECD (2011) lists such technologies using the International Patent Classification (IPC) scheme, which some observers believe is better designed to identify and group environment-related technologies than the USPTO classification framework. We use concordances between the IPC and USPTO framework to identify additional technologies to be included. We next use information on energy-related R&D data from the NSF R&D Survey. For the firms identified in this survey to be conducting energy-related R&D, we list their patent technology codes and frequency. We then manually search the 600 most-frequent codes to identify if they are energy related. In a related final step, we also manually search the USPTO database using key words like "Coal" and "Solar" to determine relevant technologies. This identification process constructs a pool of patents related to the energy sector.[30]

We match the patent data to these operating data using firm-name and geographic-location matching algorithms. The basic concept in these algorithms is to identify Census Bureau firms that have similar names to USPTO patent assignees and that have establishments in the same geographic area as where inventors of the patents are located.[31] This linkage also accomplishes

---

[30]Energy-related patents account for 5%-15% of US patents over our sample period, with a declining share in recent years; in absolute terms, patent counts for the energy sector are stable or growing throughout the period. The declining share is partly due to the sector not growing as fast as others, and partly due to external changes over time to allow for patents to be made in sectors that traditionally did not patent, especially software patents.

[31]The algorithms are described in detail in an internal Census Bureau report by Ghosh and Kerr (2010). This patent matching builds upon the prior work of Balasubramanian and Sivadasan (2011) and Kerr and Fu (2008).

a related step of aggregating patent assignees to firms, as some firms file patents through multiple patent assignee codes. This aggregation is due to the Census Bureau's establishment-firm hierarchy, as we observe establishment-level names within multi-unit firms that help identify subsidiaries and major corporate restructurings like mergers and acquisitions, and through the name-matching process that consolidates slight name variants over patent assignees.

We focus our sample on the years in which Economic Censuses are conducted, specifically every five years starting in 1977 and ending in 2002. We adopt this focus for several reasons: [1] the operating data are often best measured around these years due to heightened Census Bureau resources, [2] some specific variables from the Economic Censuses are only available at those five-year marks, and [3] our innovation data are most appropriately considered over short time periods. The third rationale is due, for example, to lumpiness in firm applications for patents; our R&D expenditures data are also often collected biannually. We thus measure variables using the average of observed values for firms in five-year windows surrounding these Economic Census years (e.g., 1985-1989 for the 1987-centered period).

The R&D Survey is the US government's primary instrument for surveying the R&D expenditures and innovative efforts of US firms. This is an annual or biannual survey conducted jointly by the Census Bureau and NSF (it is biannual for most of our sample period). The survey includes with certainty all public and private firms, as well as foreign-owned firms, undertaking over a minimum threshold of R&D expenditure in the United States. For most of our sample period, this expenditure threshold is one million dollars of R&D within the US. The survey frame also sub-samples firms conducting less than the certain expenditure threshold. These micro-records begin in 1972 and provide the most detailed statistics available on firm-level R&D efforts. In 1997, 3,741 firms reported positive R&D expenditures that sum to $158 billion. Foster and Grim (2010) and Akcigit and Kerr (2010) discuss these data in greater detail.

When calculating our innovation production function for the sector (e.g., the η and α parameters), we only utilize firm observations for which we always observe reported data on R&D expenditures or scientist counts—meaning that these calculations use only firms that conduct more than the minimum threshold of one million dollars in R&D or are sub-sampled completely. While this might present an issue for sectors like consumer internet start-ups, this is not very restrictive for the supply side of the energy sector given the large amounts of R&D expenditures required by many start-ups. For our broader moments on firm dynamics, this minimum threshold creates a challenge, however, for the consistent calculation of the entry margin and growth rates. Our model requires that firms be innovative from the start of their lives, and thus an innovative firm that falls below threshold value in its first period would be inappropriately dropped if we restricted the sample only to firms for which we always observe R&D expenditure. By contrast, the patent data are universally observed. To ensure a complete distribution, we thus use patents to impute R&D values for firms that are less than the certainty

threshold and not sub-sampled.[32] As the R&D expenditures in these sub-sampled cases are low (by definition), this imputation choice versus treating unsurveyed R&D expenditures as zero expenditures conditional on patenting is not very important. The firm does not need to conduct R&D or patent in every year of the initial five-year window, but the firm must do one of the two activities at least once.

The main text outlines how we focus on firms that are initially innovative, including those that transition out of innovation, but do not allow non-innovative incumbents to enter. Note that it would have been impossible to build a consistent sample that would also include incumbents switching into innovation. To see why, consider keeping all of the past records for incumbent firms that start conducting R&D in 1987. In the prior periods, this approach would induce a mismeasurement of exit propensities and growth dynamics because a portion of the sample will include firms conditioned on survival until 1987.

Thus, our compiled dataset includes innovative firms in the energy sector from 1975-2004. A record in our dataset is a firm-period observation that aggregates over the firm's different establishments.[33] In addition to the statistics reported in the main text, our sample accounts for over a trillion dollars in sales, 3.9 million employees, and 25 billion dollars in R&D expenditures, and the firms obtain 56,000 patents during 1995-1999.

Several additional points are worth noting about the sample and our data approach. First, we generally include technologies that are designed to make fossil fuels cleaner in the dirty-energy group. In this one regard, we deviate from the classifications developed by Popp and Newell (2012) where coal liquefaction and gasification are included in alternative energy, for example. When we directly use Popp and Newell's (2012) technology scheme as a robustness check, we classify the technologies as in their original work. Second, we have not built our sample selection or grouping procedures around technologies related to pollution abatement. We retain all patents for included firms, and thus they are part of our overall technology description, but we only use energy-directed patents to classify patents and firms into dirty-or clean-energy groups. Finally, we also use the more detailed information the R&D Survey collects from selected major R&D producers. We specifically utilize information collected from about 100 firms on R&D expenditures related to specific energy applications like coal or solar energy. We earlier identified one application of this extra information in that we manually searched the major patenting technology codes from these R&D entities to identify energy-sector patenting groups that we should be including. A second application is to verify our data development procedures, for example by assigning firms based upon the types of R&D they conduct rather than observed patents. This group from the R&D Survey also confirms the bimodal nature of our firm groupings. While the group of firms asked these questions is

---

[32] In a small number of cases where we have scientists counts from the R&D Survey but lack R&D expenditures, we similarly use the scientist counts to impute R&D values for firms below the certainty threshold.

[33] We exclude approximately 50 non-profit research centers and similar entities to match our model's focus on profit-seeking firms. Our estimations are robust to retaining these entities.

too small and selected to serve as the backbone of our sample, these checks are comforting. While Census Bureau disclosure prevents us from listing firms, we overlap well with Popp and Newell's (2012) listed producers as one example.

Tables A1 and A2 provide additional results mentioned in the main text:

TABLE A1. POISSON ESTIMATES FOR η PARAMETER

Technique, Firm Size Measure u_f:	R&D Input Measure H_f	
	R&D Expenditure	Scientist Counts
Random Effects, SIC4 Counts	0.326 (0.122)	0.361 (0.079)
Fixed Effects, SIC4 Counts	0.321 (0.106)	0.357 (0.089)
Random Effects, Establishments	0.567 (0.108)	0.584 (0.064)
Fixed Effects, Establishments	0.565 (0.103)	0.583 (0.076)

Notes: Table presents fixed and random effects Poisson estimates similar to Table 1.

TABLE A2. INITIAL CONDITION DISTRIBUTIONS SIC3

Metric:	Clean Energy	Dirty Energy
Mean Patent Total	260	1029
Standard Deviation	515	1500
Share: [0,20]	37%	0%
Share: [21,100]	25%	6%
Share: [101,500]	22%	50%
Share: [500+]	16%	44%

Table $A3$ summarizes the alternative initial technology distribution considered in the robustness analysis. The average gap to the frontier for dirty-patent stocks in the 9% of cases where clean patents have the lead is 463 patents, or in relative terms, 33% of the total patenting in that line to date. The average gap to the frontier for clean-patent stocks in the 91% of cases where dirty patents have the lead is 624 patents and 82% in relative terms. The distribution graph has a broadly similar shape as Figure 3. The fraction of product lines with a non-zero gap in terms of step sizes is 82%. Clean energy leads by one or more step sizes in 7% of cases. Dirty energy has a lead of 20 and 50 step sizes or more in 8% and 2% of technologies, respectively.

TABLE A3. INITIAL CONDITION DISTRIBUTIONS SIC4

Metric:	Clean Energy	Dirty Energy
Mean Patent Total	140	663
Standard Deviation	401	1242
Share: [0,20]	53%	2%
Share: [21,100]	23%	18%
Share: [101,500]	17%	48%
Share: [500+]	6%	33%

Additional References

Balasubramanian, Natarajan, and Jagadeesh Sivadasan. 2011. "What Happens When Firms Patent? New Evidence from US Economic Census Data." *Review of Economics and Statistics* 93: 126-146.

Conesa, Juan Carlos, and Dirk Krueger. 1998. "Social Security with Heterogeneous Agents." *Review of Economic Dynamics*, 2: 757-795.

Foster, Lucia, and Cheryl Grim. 2010. "Characteristics of the Top R&D Performing Firms in the US: Evidence from the Survey of Industrial R&D." CES Working Paper 10-33.

Ghosh, Kaushik, and William Kerr. 2010. "Patent-SSEL Matching." Census Bureau Technical Paper.

Gouriéroux, Christian, and Alain Monfort. 1996. *Simulation-Based Econometric Methods*, Oxford University Press.

Griliches, Zvi. 1990. "Patent Statistics as Economic Indicators: A Survey." *Journal of Economic Literature* 28: 1661-1707.

Hall, Bronwyn, Adam Jaffe, and Manuel Trajtenberg. 2001. "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." NBER Working Paper 8498.

Kerr, William, and Shihe Fu. 2008. "The Survey of Industrial R&D—Patent Database Link Project." *Journal of Technology Transfer* 33: 173-186.

Kirkpatrick, S., C.D. Gelatt, and M.P. Vecchi. 1983. "Optimization by Simulated Annealing." *Science* 220: 671-680.

McFadden, Daniel. 1989. "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration." *Econometrica* 57: 995-1026.

Nelder, John, and Roger Mead. 1965. "A Simplex Method for Function Minimization." *The Computer Journal* 7: 308-313.

OECD Environment Directorate. 2011. "Patent Search Strategies for the Identification of Selected Environment-Related Technologies." www.oecd.org/environment/innovation.

Pakes, Ariel, and David Pollard. 1989. "Simulation and the Asymptotics of Optimization Estimators." *Econometrica* 57: 1027-1057.

Popp, David. 2002. "Induced Innovation and Energy Prices." *American Economic Review* 92: 160-180.

Popp, David, and Richard Newell. 2012. "Where Does Energy R&D Come From? Examining Crowding Out From Energy R&D." *Energy Economics* 3: 980-991.