Motivation Crowding-Out and Green-Paradox-Like Outcomes

Simone Marsiglio* Marco Tolotti†

Forthcoming in Journal of Public Economic Theory

Abstract

We analyze the effectiveness of environmental policy in a framework in which households' utility is determined by both private and social components, representing their extrinsic and intrinsic motivations to undertake green actions, respectively. Environmental policy, in the form of a subsidy aiming to incentivize the adoption of a green technology, on the one hand, directly increases households' extrinsic motivation, while, on the other hand, indirectly decreases their intrinsic motivation. We show that, provided that the indirect effect dominates, the policy leads to crowding-out of intrinsic motivation which ultimately undermines the effectiveness of the policy itself. Specifically, despite its positive effect on environmental outcomes in the short run, the policy will lead to a deterioration in long run environmental outcomes, giving rise to a reverse green-paradox-like outcome. Moreover, even in the case in which the direct effect dominates, provided that the indirect effect is large enough, the policy will generate a deterioration in short run environmental outcomes. These results clearly suggest that the optimal design of environmental policy is particularly complicated since it requires to take into account also its effects on intrinsic motivation.

Keywords: Environmental Policy, Green Paradox, Intrinsic and Extrinsic Motivations, Social Effects
JEL Classification: C60, D70, Q50

1 Introduction

The optimal design of economic policy put in place to achieve desired goals is a complex task: it often happens that the ex-ante effects are dramatically different from the ex-post results. The problem is even more severe when dealing with environmental policy, since because of uncertainty, irreversibility and transboundary externality, understanding a priori how the natural ecosystem will be affected by specific policy tools is often not possible. The most famous example of how problematic the design of environmental policy might be is represented by the so-called green paradox (Sinn (2008)). Indeed, the green paradox suggests that since short run and long run effects may be different and economic agents optimally respond to policy changes, environmental policy may eventually achieve an improvement in environmental outcomes in the long run but only at the cost of deteriorating environmental outcomes in the short run (see Sinn (2008) and Jensen et al. (2015)). In its simplest and original form the argument underlying the green paradox is straightforward and extremely intuitive. Suppose that policymakers in an attempt to reduce emissions set a carbon tax on polluting activities which will rise over time consistently with the expected increase in the environmental damages; owners or producers of these activities will optimally respond to the tax by anticipating their activities in order to lower their tax exposure; even if lowering long run emissions, this clearly will rise short term emissions, suggesting thus that green policies might end up being detrimental for the environment. The traditional discussion of the green paradox involves extraction of a non-renewable resource (i.e., oil);

*University of Pisa, Department of Economics and Management, via Cosimo Ridolfi 10, 56124 Pisa, Italy. Contact: simone.marsiglio@unipi.it.

†Università Ca' Foscari Venezia, Department of Management, Cannaregio 873, 30121 Venice, Italy. Contact: tolotti@unive.it

in this context it is evident how carbon taxes, by affecting the timing of optimal choices, might encourage producers to anticipate extraction to maximize the present value of profits. More recently, the term green paradox has been used to refer to the unintended effects of any type of environmental or climatic policy (see Jensen et al. (2015)). Following Sinn's (2008) seminal paper, a growing share of literature has tried to identify alternative circumstances under which green-paradox outcomes might occur, ranging from carbon price dynamics, technology spillovers, implementation delays, and international trade effects (see van der Werf & Di Maria (2012) or Cairns (2014) for some recent surveys). The aim of this paper is to contribute to this literature by proposing a novel mechanism able to eventually give rise to green-paradox-like outcomes, related to social effects and crowding-out of intrinsic motivation.

Consistent with the most recent interpretation of the term green paradox, we focus on environmental policy not necessarily meant as a carbon tax on extraction activities. Specifically, we consider a green policy which is introduced in order to provide individual households with some additional incentive to adopt a new technology allowing to reduce ecological footprints.[1] The policy is implemented through a subsidy, which in line with the expected reduction in environmental damage and balanced budget considerations, will fall over time; this implies that the size of the incentive effectively received by single individuals adopting the green technology changes according to the timing of their adoption decision. We show that if social effects and intrinsic motivation matter in determining single individual's optimal behavior and individuals' intrinsic motivation is negatively affected by the subsidy, it might happen that the green policy generates results very different from those hoped for by policymakers and an outcome similar to what hypothesized by the green paradox might occur. However, in our setting the short and long run effects of the green policy are reversed with respect to those traditionally discussed in the green paradox literature. Indeed, we show that the green policy might improve environmental outcomes in the short run but at the cost of deteriorating them in the long run, characterizing thus a "reverse" green paradox outcome. This result suggests that understanding the effectiveness of environmental policy might even be more difficult than what traditionally believed, since even if short run effects are consistent with the desired goals, what will happen in the long run cannot be known a priori. From a sustainability point of view, which necessarily requires to take into account also long run outcomes, this is particularly problematic since it might make not possible at all to assess the long run consequences of alternative environmental policies.

The type of environmental policies considered in our analysis refers to the introduction of subsidies aiming to push single individuals to adopt a green technology reducing thus carbon emissions at national or regional level. Specific real world examples include subsidies to the use of renewable energy both at household and firm levels (i.e., solar and wind power), or the purchase of hybrid low-polluting vehicles (i.e., LPG and electric cars). These green policy schemes are traditionally designed to provide the new adopters of the promoted goods with an incentive which falls over time as the number of adopters increases. Typical examples of this type of scheme are feed-in-tariffs which upon achievement of a certain predetermined adoption target drastically fall reducing the benefits obtained by later adopters[2]. This tends to generate a boom effect in the adoption of the green good which definitely decreases emissions in the short run, but understanding what the long run effect will be is not possible a priori. Our argument is indeed that the boom in the adoption of green goods is driven by the direct effect of the subsidy which provides single individuals with additional extrinsic motivation to adopt; however, this is accompanied by an indirect effect which may weaken the social norm determining the desirability to undertake green actions, undermining thus the intrinsic motivation of single individuals. Provided that this latter effect is strong enough for extrinsic motivation to crowd–out

[1] The drivers of households' preferences towards green technologies has received growing attention over the last decade (see, among others, Sundt & Rehdanz (2015) and Cardella et al. (2017)). Different from extant literature, in this paper we focus on the role of social effects and intrinsic motivation as a determinant of households' behavior and its implications in terms of green-paradox-like outcomes.

[2] An emblematic case is represented by feed-in-tariffs to promote the use of renewable resources in Italy (see Section 5 for more details). Similar feed-in-tariffs schemes have been introduced in several EU countries as well, including France, Germany, Greece and Spain.

intrinsic motivation, it might happen that in the long run the overall adoption rate will be lower than what it would be in the absence of environmental policy. Quantifying these effects is clearly not simple at all, but the eventual possibility that environmental policy might generate paradoxically negative long run environmental consequences suggests that more caution is needed in the design of environmental policy, and in particular eventual effects induced on intrinsic motivation need to be carefully taken into account.

Specifically, we focus on a social interactions framework based upon a random utility setting (Brock & Durlauf (2001); Blume & Durlauf (2003)). Households are heterogeneous in their individual valuation of a green technology and determine their technology adoption decision (i.e., whether to adopt or not to adopt) by maximizing their individual utility function. Such an utility function depends on private and social components: the private term measures the utility associated with monetary incentives representing thus individuals' extrinsic motivation to adopt; the social term measures the utility related to adhering to a social norm stressing the desirability to undertake green actions, and thus it represents individuals' intrinsic motivation (Steg et al. (2015)). In such a setting, with no public intervention, since single individuals care for the environment in the long run everyone will naturally adopt the green technology minimizing society's ecological footprints. However the convergence to the long run equilibrium might be particularly slow and thus policymakers might wish to intervene in order to improve short run dynamics. By maintaining a balanced budget at any point in time and allocating a fixed amount of resources to environmental policy, policymakers introduce a subsidy to provide single individuals with additional incentives to speed up the adoption rate. In theory, this should clearly be beneficial, since by increasing individuals' extrinsic motivation the direct effect of the subsidy is to increase the speed of convergence towards the long run equilibrium. However, in practice, this prediction does not account for the eventual indirect effect induced on the social norm and thus on individuals' intrinsic motivation (Rege (2004); Benabou & Tirole (2003)). Several studies discuss how social norms play an essential role in determining the effectiveness of environmental policy and how public policy may have hidden costs related to its effects on intrinsic motivation (Nyborg (2003); Frey & Oberholzer–Gee (1997)). We show that provided that the subsidy actually weakens the social norm undermining individuals' intrinsic motivation and such an effect is large enough, it might even happen that, despite the initial policy-induced boom, in the long run the adoption rate will be lower and thus ecological footprints higher than in a setting with no public intervention. This suggests that there might be a trade off between the short run and long run effects of environmental policy, and thus understanding its true consequences is not simple at all.

This paper proceeds as follows. Section 2 introduces the unregulated economy in which there is no public intervention. Section 3 describes the regulated economy in which public intervention takes the form of a subsidy, which generates only a direct effect on individuals' extrinsic motivation. Section 4 discusses how the presence of a subsidy-induced indirect effect affecting intrinsic motivation may alter the effectiveness of environmental policy, leading to a reverse green-paradox-like outcome. Section 5 presents a calibration based on the case of feed-in-tariffs in Italy lending support to our theoretical conclusions. Section 6 introduces an extension of our baseline model to account for a pollution externality affecting households' utility function. Section 7 generalizes our results by deriving general (sufficient) conditions for environmental policy to deteriorate long run environmental outcomes. Section 8 presents concluding remarks and proposes directions for future research. Appendix A discusses the numerical procedure employed in our calibration.

2 The Laissez-Faire Outcome

We consider a very simple economy populated by a large number, N, of heterogeneous households, indexed by $i = 1, \ldots, N$. Their daily-life activities generate emissions and in particular each household generates a constant amount of pollution equal to $e > 0$. A new green technology with the ability to reduce the emissions generated by households' activities by a factor $0 < \psi < 1$ becomes available on the market, such that emissions generated by the adopters of such a new technology will be equal to ψe. This new technology may be thought of as a photovoltaic power plant, which by producing renewable and clean energy allow households to reduce their individual ecological footprints even without implementing any behavioral change.

In this setting, at any moment in time t, where $t \in [0, \infty]$, each single household may decide to purchase the new technology in an attempt to maximize his utility. The new technology is a durable non-perishable good such that once households have purchased it they will not need to make any further decision.[3] Households are heterogeneous in their individual valuation of the technology. The utility function of each household i, who at time t has not purchased the new technology yet, is associated with the choice $\omega_{i,t} \in \{0, 1\}$ such that $\omega_{i,t} = 1$ ($\omega_{i,t} = 0$) denotes that the household adopts (does not adopt) the technology. The utility related to such an adoption decision depends on two elements: a private component which is household-specific and a social component which is common to all households. The private component is driven by the net benefit from adoption given by $b + \varepsilon_i - c$. The benefit $b + \varepsilon_i$ is the sum of a deterministic term common to all households, $b \geq 0$, and a random household-specific term determining the effective valuation that each household i attaches to the technology, ε_i. The cost is common to all households and it is equal to $c \geq 0$. For the sake of simplicity, we assume that the private utility component, including households' own valuation of the technology, is constant over time. The social component of the utility is related to a social externality quantifying the importance of a social norm stressing the desirability of undertaking green actions. Mathematically, it is given by Jx_t^N, where the size of such a social externality is measured by $J > 0$, while $x_t^N = \frac{1}{N}\sum_j \omega_{j,t}$ represents the time-varying fraction of green households (i.e., households adopting the green technology); for the sake of expositional simplicity, we will refer to x_t^N as the *"green share"*. This means that the effective magnitude of the social externality in the utility function depends on the number of households adhering to the norm, that is on the number of households actively adopting the green technology: the eventual adoption by other households reinforces the social component positively affecting individual utility. Formally, we model the individual household's utility as in random utility models (see Brock & Durlauf (2001) and Barucci & Tolotti (2012)). Eventually, the adoption decision is based on the following utility function:

$$u_{i,t}(\omega_{i,t}) = \omega_{i,t}\left(b + \varepsilon_i - c + Jx_t^N\right). \tag{1}$$

If the agent i does not adopt the green technology ($\omega_{i,t} = 0$) his utility is simply null, $u_{i,t}(0) = 0$.[4] If the agent does adopt it ($\omega_{i,t} = 1$) his utility is equal to $u_{i,t}(1) = b + \varepsilon_i - c + Jx_t^N$. Clearly, an individual will adopt the technology whenever the sum of the private and social utility components is positive (i.e., $b + \varepsilon_i - c + Jx_t^N > 0$), otherwise he will not (i.e., $b + \varepsilon_i - c + Jx_t^N \leq 0$). Note that the private and the social components represent alternative determinants of the adoption decision. The private component captures the individual benefit and cost for each household: the benefit can be thought of as the (non-renewable) energy savings generated by the new technology which can largely vary from household to household according to a number of factors (as dwelling size, number of members, locations); the cost can be thought of as the price to pay for installing the new technology which is the same across different households. The social component captures the individually-perceived benefit associated with warm-glow behavior: this can be thought of as the desire to help the planet which may become stronger as more people

[3]Since the technology we consider consists of a durable good we rely on the framework of innovation diffusion based on Bass (1969) where agents may only purchase the good once and for all. We could consider a setup in which households may opt for *multiple repurchasing* during their lifetime, but this context would be more appropriate to model non-durable goods (see Peres et. al. (2010) for a recent survey of different frameworks and models in the field of diffusion of innovation). Alternatively, if we interpret the cost parameter c as the expected present value associated with eventual repurchases of the goods, our model accounts also for the possibility of multiple repurchases.

[4]For normalization issues, we assume that the utility of not adopting, which determines the threshold value for adoption decisions, is null but assuming any other exogenous value for such a threshold will not change any of our conclusions. The results may be different (and more interesting) if this value is endogenously determined. We will present in Section 6 an extension of our baseline model in which the adoption threshold is endogenous since it depends on the emissions level which in turn depends on the green share. We will show that results are qualitatively identical to those of our baseline model, thus it seems convenient to present our arguments in the simplest possible way first.

contribute to fix environmental problems. Therefore, while the private component measures the utility associated with individual-specific monetary incentives representing thus individuals' extrinsic motivation to adopt the new technology, the social component measures the utility associated with non-monetary moral incentives representing individuals' intrinsic motivation to adopt (see Steg et al. (2015) for a recent survey on the relation between motivations and norms in an environmental setting). We assume that the net benefit of adopting the green technology is on average positive, that is $b > c$, but the effective value of the net benefit is determined at household level by his specific valuation of the technology. This is driven by the one-time (i.e., at time $t = 0$) realization of ε_i which are i.i.d. random variables drawn from a common cumulative distribution function, η, which ultimately determines the specific valuation type of each single household, which does not change over time.

Similar to Perino & Requate (2012), the choice of individual households determines at aggregate level the evolution of emissions, E_t^N, as follows:

$$E_t^N = e\left(N - \sum_i \omega_{i,t}\right) + \psi e \sum_i \omega_{i,t}, \tag{2}$$

where the first term represents the emissions generated by non-adopters of the new technology while the second term those generated by its adopters. Clearly, a higher number of adopters of the green technology will decrease aggregate emissions, allowing for a reduction in ecological footprints to effectively occur.

As discussed in the literature on social interactions (see Blume & Durlauf (2003); Marsiglio & Tolotti (2018)), the households' optimization in (1) gives rise to a dynamic probabilistic choice model in which each agent is characterized by a time-varying likelihood to adopt the new technology. In particular, by considering the net benefit $b - c$, the social externality J, and the green share x_t^N, at any time t household i will decide to adopt with probability:[5]

$$\mathbb{P}(\omega_{i,t+\Delta t} = 1 |\, \omega_{i,t}, x_t^N) = \eta\left(b - c + Jx_t^N\right). \tag{3}$$

This implies that the dynamic evolution of emissions in per capita terms, $e_t^N = \frac{E_t}{N}$, is given by the following expression:

$$e_t^N = e\left(1 - x_t^N\right) + \psi e x_t^N = e - e(1-\psi)x_t^N \tag{4}$$

The Markovian dynamics induced by (3) are cumbersome to study in the finite dimensional population model; nevertheless, it is possible to describe in closed-form the (deterministic) dynamics emerging from the asymptotic system characterized by an infinite number of households (see Marsiglio & Tolotti (2018), for a comparison of the deterministic and stochastic versions of a similar random utility model). For the sake of simplicity we assume that the distribution of household types is uniform, and this allows us to describe the dynamics of the green share through a simple differential equation. Indeed, by assuming that the valuation types are uniformly distributed on the unit interval, it possible to show[6] that, whenever $N \to \infty$, the sequence of stochastic processes $\{x_t^N\}_{t\geq 0}$ converges to x_t obtained as the solution of the following differential equation:

$$\dot{x}_t = (1 - x_t)\left[(b - c + Jx_t) \wedge 1\right], \tag{5}$$

where $a \wedge b = \min\{a, b\}$.[7] Equation (5) describes in a very simple form the evolution of the share of households adopting the green technology, which in turn determines the evolution of per-capita emissions as follows:

$$e_t = e - e(1-\psi)x_t. \tag{6}$$

[5] Mathematically speaking, each agent at random times needs to decide whether to adopt the new technology in order to maximize his utility function. With a probability as in (3), at a given point in time he may either choose not to adopt and if so he will face the same optimization problem at the next random time; on the contrary, in case of adoption, he will not need to make any further decision (see Bass (1969)).

[6] We will provide a proof of this result, stated for a more general class of models, in Section 7.

[7] Dealing with a uniform distribution with support on $[0, 1]$, we have that $\eta(z) = 1$, for $z \geq 1$ and $\eta(z) = 0$ for $z < 0$. This latter situation is automatically satisfied in our case, since $b - c + Jx_t \geq 0$ for all $x \in [0, 1]$. The former has to be considered and this is the reason why we put a *cup* at the level 1.

Since (5) describes a logistic-type differential equation, the technology adoption rate follows a typical S-shaped dynamics as discussed in literature of innovation diffusion (see Bass (1969); Colapinto et.al. (2014)) and its closed form evolution path can be derived. Indeed, provided that $b-c+J \leq 1$ and $x_0 = 0$, which we will assume to hold true in what follows,[8] the time evolution of the green share can be explicitly computed as follows:

$$x_t = \frac{1 - e^{-(b-c+J)t}}{1 + \frac{J}{b-c} e^{-(b-c+J)t}}. \tag{7}$$

This shows that for the relevant parameter values the green share will converge to 1 in the long run, meaning that the entire population in the long run will adopt the green technology minimizing thus emissions. We summarize this result in the following proposition.

Proposition 1. *The green share, x_t, in the long run will converge to its unique (asymptotically) stable equilibrium $\overline{x}^* = 1$, and thus emissions, e_t, will reach their minimal level $\overline{e}^* = \psi e$.*

Proposition 1 suggests that even in the absence of environmental policy the share of individuals who will adopt the green technology is equal to one, and thus emissions will reach their minimum $\overline{e}^* = \psi e$. Therefore, if policymakers care only about long run outcomes there is no need to intervene in order to try modifying households' behavior and environmental outcomes. However, the transition to such a long run outcome might even be very long, thus in such a case policymakers may wish to intervene in order to try speeding up the adoption rate and reducing short run emissions. Indeed, the speed of convergence to the long run equilibrium is equal to $\lambda = b-c+J$, thus according to the parameter values adoption in the short run might be either fast or slow. In particular, the short run adoption rate will be particularly slow whenever the net benefit, $b-c$, and/or the social externality, J, are relatively small. The time that might be required to

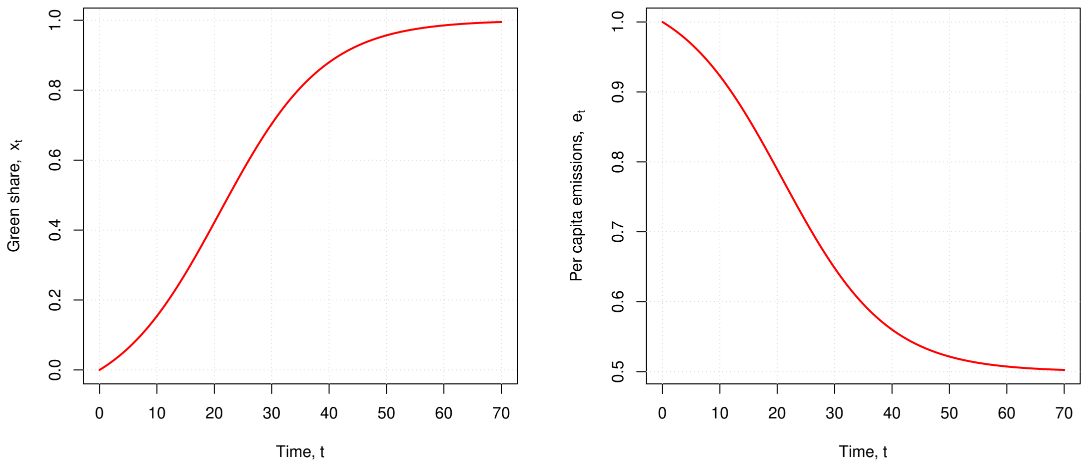

Figure 1: Time evolution of the green share (left) and per capita emissions (right) in the unregulated economy.

effectively converge to the long run equilibrium in the case of a small λ is shown in Figure 1, which represents the time evolution of the green share (left) and emissions (right) for the following parametrization: $b = 0.2$, $c = 0.19$, $J = 0.1$, $e = 1$ and $\psi = 0.5$, implying thus that $\lambda = 0.2$. It might take about 70 years for the green technology to be adopted by the entire household population and thus for society's ecological footprints to be minimized.

[8]Note that the condition $b-c+J < 1$ ensures that $[(b-c+Jx_t) \wedge 1] = (b-c+J)$, meaning that the evolution of the green share reads as follows: $\dot{x}_t = (1-x_t)(b-c+Jx_t)$.

3 The Regulated Outcome: Direct Effect

Let us now analyze how the presence of environmental policy may affect the results described above. Indeed, whenever the speed of convergence to the long run equilibrium is low (i.e., when λ is small), policymakers may be interested in rising (through specific public policy) the green share in the short run in order to lower short run emissions. This scenario may represent a typical situation in which awareness of the benefits associated with green technologies spreads slowly across the population, while policymakers moved by political incentives wish to promote a faster diffusion of the technology. In order to maintain the framework as simple as possible, we assume that policymakers allocate to this goal a certain fixed amount τ (financed through non-distortionary income taxation),[9] which is entirely used to provide agents with some (additional) incentive to adopt the green technology, and such an incentive takes the form of a subsidy s_t. Note that, by increasing the net benefit from adoption, the subsidy directly provides individuals with additional extrinsic motivation to adopt. By taking into account the subsidy which increases the private utility component, the dynamics of the green share becomes:

$$\dot{x}_t = (1 - x_t)\left[(b + s_t - c + Jx_t) \wedge 1\right]. \tag{8}$$

Clearly, since the budget available to finance environmental policy is fixed, the size of the subsidy received by each individual agent cannot be constant, but it tends to change with the number of agents adopting the new technology, x_t. Specifically, we assume that in order to maintain a balanced budget at any point in time, the subsidy is determined as follows:

$$s_t = \frac{\tau}{1 + x_t}. \tag{9}$$

This expression suggests that when $x_t = 0$ the size of the subsidy is maximal and equal to $s_t = \tau$, but as x_t increases this tends to fall approaching $s_t = \frac{\tau}{2}$ when $x_t = 1$. The implications of such an incentive scheme are intuitive and consistent with several real world green experiences (i.e., feed-in-tariffs): the subsidy is larger when the target population is larger, while it gets smaller as soon as the policy becomes effective enough to reduce the size of the target population. The decreasing size of the subsidy is also consistent with expectations about environmental damages: since gradually even with no subsidy more and more households will adopt the green technology ecological footprints will tend to decrease over time and thus the need for green incentives will be less stringent in the future. By plugging (9) into (8) we obtain the law of motion of the green share in general equilibrium under public policy:

$$\dot{x}_t = (1 - x_t)\left[\left(b + \frac{\tau}{1 + x_t} - c + Jx_t\right) \wedge 1\right], \tag{10}$$

while per-capita emissions are still given by (6). Despite the nature of the time-varying incentive provided by this type of environmental policy, the policy (at least a priori) seems effective since allowing to increase the speed of convergence to the unique equilibrium $(\overline{x}^*, \overline{e}^*)$. Indeed, the result in Proposition 1 still holds true and the only effect of the policy is to modify the speed of convergence which is now given by $\lambda = b - c + J + \frac{\tau}{2}$, meaning that the speed of convergence increases with τ. This allows to state the following result.

Proposition 2. *Environmental policy, even if not modifying the long run outcome, affects the transitional dynamics. Specifically, a stronger environmental policy tool (a higher τ) increases the speed of convergence to the unique equilibrium $(\overline{x}^*, \overline{e}^*)$.*

[9] Since the main goal of our analysis is positive rather than normative, it seems convenient to assume that the resources allocated to environmental policy are predetermined (see Khalil et al. (2019) for a discussion of public spending decisions under a fixed budget). This is consistent with real-world experiences (i.e., feed-in-tariffs) in which during the annual budgeting process, the overall available resources are allotted to different spending needs in an arbitrary way driven to a large extent by political negotiations.

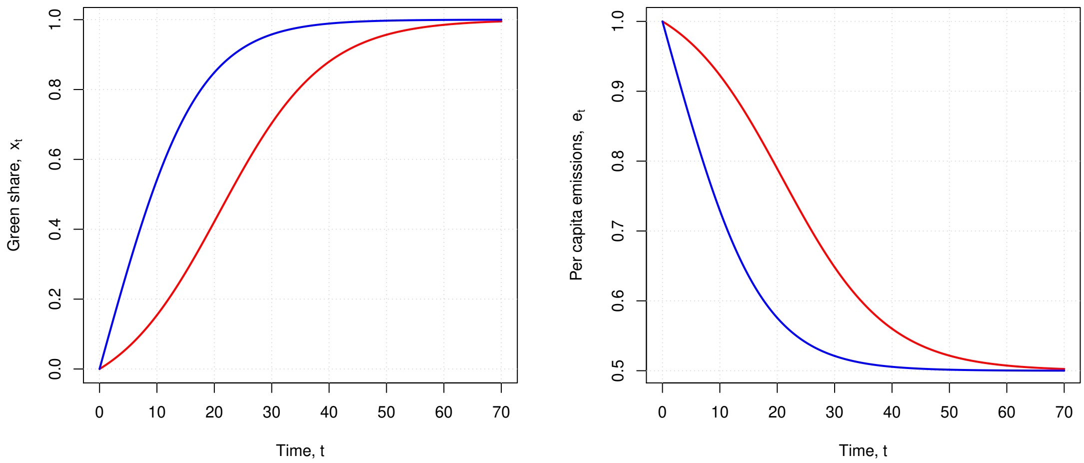

Figure 2: Time evolution of the green share (left) and per capita emissions (right) without (red) and with (blue) environmental policy.

Proposition 2 suggests that the environmental policy, as expected, is desirable since allowing to speed up the adoption rate, reducing environmental degradation. Even if the policy does not improve environmental outcomes in the long run it improves them at all points in time during the transition to the long run equilibrium. The effectiveness of environmental policy is shown in Figure 2, in which we compare the laissez-faire outcome (in red) and the regulated outcome (in blue); the parameter values are the same employed in the previous section with $\tau = 0.05$. It is clear that in the latter case the speed of convergence is faster and thus the short run outcome is effectively improved thanks to the green policy; in particular, in the regulated economy it takes about 40 years for the entire population to adopt the green technology and thus for emissions to reach their minimum level, which is much less than the 70 years required in the unregulated economy. Since the speed of convergence rises with τ, the stronger the environmental policy the faster the convergence to the long run equilibrium, and thus the time effectively needed to minimize ecological footprints can be reduced further.

4 The Regulated Outcome: Indirect Effect

A potential problem which we have not analyzed yet and which is generally not considered at all in policy design is associated with the eventual social effects of public policy (see Rege (2004)). Several studies discuss that social effects play a fundamental role in determining the effectiveness of environmental policy: these effects allow to explain why environmental policy may generate virtuous or perverse mechanisms, and thus should be fully taken into account as a possible solution to environmental problems (Nyborg (2003), Nyborg et al. (2016)). Consistent with this view we now analyze what a change in the social norm induced by environmental policy might imply for environmental outcomes (Brekke et al. (2003), Nyborg (2003)). Recall that in our setup the social norm determines at single individual level the desirability to undertake green actions independently of individual-specific monetary incentives, and thus it represents individuals' intrinsic motivation. Therefore, what we now analyze is whether environmental policy by affecting the social norm and intrinsic motivation may give rise to results different from those hypothesized and hoped for by policymakers.

Specifically, we assume that the environmental policy indirectly affects J, quantifying the social externality. By providing additional monetary incentives to adopt the green technology, the subsidy might distort agents' perception about the desirability to eventually adhere to the social norm and thus to undertake green actions, which might induce agents to perceive a less stringent need to help the environment. Such an effect is consistent with the view that *"economic incentives may [...] have adverse effects, through reducing individuals' perceived moral responsibility"* (Nyborg (2003)). This in turn might affect not only the short run transitional dynamics but also the long run equilibrium. To take these effects into account, we propose a more general framework where J is now a function of the subsidy itself. From now on, J is replaced by $J(1-\epsilon x_t s_t)$, where $\epsilon \in \mathbb{R}$ measures the size of this social effect, x_t the size of the population reached by the subsidy and, finally, s_t the subsidy. This implies that the dynamics of the green share is given by:

$$\dot{x}_t = (1 - x_t)\,\eta\,(b + s_t - c + J\,(1-\epsilon x_t s_t)\,x_t)\,, \tag{11}$$

while per-capita emissions are still given by (6). Note that the subsidy has two different effects on single individual's utility function and thus on the adoption rate. On the one hand, the policy provides single individuals with a higher net benefit from adopting, which by increasing the private component of utility tends to increase adoption (i.e., it is a source of extrinsic motivation); we refer to this term ($+s_t$) as the *"direct effect"*. On the other hand, the policy by affecting the social norm alters the size of the social externality which by determining the social component of utility tends to increase/decrease adoption (i.e., it affects intrinsic motivation); we refer to this term ($-J\epsilon x_t^2 s_t$) as the *"indirect effect"*. The overall effect of the subsidy on the adoption rate clearly depends on whether the direct and indirect effects go in the same direction and, if not, on which effect dominates. Specifically, whenever $\epsilon = 0$, the subsidy does not generate any indirect effect and thus we are back exactly to the framework discussed in the previous section. Whenever $\epsilon \neq 0$, the size of the indirect effect increases with the green share: the larger the share of households adopting the green technology the larger the effects induced on the social utility component. In particular, the direction of such an effect is determined by the sign of ϵ: if $\epsilon < 0$ the indirect effect is positive and goes hand-in-hand with the direct effect; if $\epsilon > 0$ the indirect effect is negative and goes in the opposite direction of the direct effect. In this latter case it may be possible that extrinsic motivation crowds-out intrinsic motivation (Benabou & Tirole (2003); Frey & Oberholzer–Gee (1997)) such that the negative indirect effect dominates the direct effect and thus in the regulated outcome the adoption rate ends up being lower and so emissions higher than in the lasseiz-faire outcome, generating thus a green-paradox-like outcome.

Consistent with Benabou & Tirole (2003), we assume that $\epsilon > 0$, meaning that the indirect effect induced by the subsidy is negative, partly reducing utility and adoption. This is due to the fact that the moral responsibility driving individual intrinsic motivation decreases with the effectiveness of the incentive mechanism (see Brekke et al. (2003) for a discussion of the effects on moral responsibility induced by public policy): households prefer to live in a healthier environment but adopting the green technology is costly, thus when the green share is low the single individual is compelled by his own moral responsibility to act in order to reduce ecological footprints; when the green share increases, the single individual's marginal contribution to environmental improvements falls, weakening his own moral responsibility through a free-riding effect. By recalling that $s_t = \frac{\tau}{1+x_t}$, the dynamics of the green share becomes:

$$\dot{x}_t = (1 - x_t)\,\eta\left(b + \frac{\tau}{1 + x_t} - c + J\left(1 - \frac{\epsilon\tau x_t}{1 + x_t}\right)x_t\right), \tag{12}$$

where:

$$\eta(z) = \begin{cases} 0 & if \quad z < 0 \\ z & if \quad 0 \leq z < 1 \\ 1 & if \quad z \geq 1 \end{cases}$$

From the above equation, the direct and indirect effects of the subsidy are given by $\frac{\tau}{1+x_t}$ and $-\frac{J\epsilon\tau\, x_t^2}{1+x_t}$, respectively. Note that, in the classical static setup proposed by Benabou & Tirole (2003) and Frey &

Oberholzer–Gee (1997), no matter the parametrization of the model, one of the two effects always dominates the other. Differently, in our framework, since the size of the indirect effect changes with the adoption rate, it is not possible to undoubtedly state which effects dominates and thus whether crowding-out of intrinsic motivation effectively occurs. We now analyze under which nontrivial conditions the indirect effect may effectively dominate the direct effect, such that extrinsic motivation may crowd-out intrinsic motivation eventually deteriorating environmental outcomes.

In order to understand whether such an outcome is actually possible, we need to focus on both the long run and short run behavior of the green share in (12). Let us start by focusing on the long run equilibrium, which in this case might not be unique. Indeed, $\overline{x}^* = 1$ is still an equilibrium exactly as before, but its stability properties depend upon whether some other equilibrium might exist or not. If another equilibrium $0 < \overline{x} < 1$ exists (i.e., $\eta(z) = 0$ for some admissible values), this equilibrium will be (asymptotically) stable. On the opposite, if such an equilibrium $\bar{x}$ does not exist, then $\overline{x}^* = 1$ will be asymptotically stable. It is possible to show[10] that, as soon as:

$$\max\left\{\frac{1}{\epsilon};\ \frac{2(b-c+J)}{J\epsilon-1}\right\} < \tau < 1-(b-c), \tag{13}$$

the green share converges to:

$$\bar{x} = \frac{(b-c+J)+\sqrt{(b-c+J)^2+4J(b-c+\tau)(\epsilon\tau-1)}}{2J(\epsilon\tau-1)}. \tag{14}$$

Whenever the technical condition (13) is met, $\bar{x}$ turns out to be well defined, that is real and belonging to the open set $(0,1)$, such that emissions converge to $\bar{e} = e - e(1-\psi)\bar{x} > \bar{e}^*$. If τ does not satisfy the above condition, the long run equilibrium $(\bar{x}^*, \bar{e}^*)$ will be naturally achieved. However, also in this case, even if the green policy does not affect the long run outcome, it may lead to a deterioration of short run outcomes. In order to see this, we need to focus on the transitional dynamics. In this respect, it is straightforward to show that the speed of convergence to $\overline{x}^*$, $\lambda = b - c + J + \frac{\tau}{2}(1 - J\epsilon)$, is even smaller than the speed of convergence in the unregulated economy provided that $\epsilon > 1/J$. We can summarize these results in the following proposition.

Proposition 3. *If negatively affecting intrinsic motivation, environmental policy might deteriorate environmental outcomes. Specifically, if the indirect effect dominates then environmental policy deteriorates long run outcomes, since at the (asymptotically) stable equilibrium $(\overline{x}, \overline{e})$ emissions are larger than in the lasseiz-faire equilibrium (i.e., $\overline{e} > \overline{e}^*$). If the direct effect dominates but the indirect effect is large enough then environmental policy affects the short run transitional dynamics, reducing the speed of convergence to the (asymptotically) stable equilibrium $(\overline{x}^*, \overline{e}^*)$.*

Proposition 3 suggests that reverse green-paradox-like outcomes are possible only whenever the indirect effect dominates the direct effect, and in such a case a stronger policy-induced social effect (a higher ϵ) decreases the long run equilibrium value $\overline{x}$ ultimately increasing emissions to $\overline{e}$. If instead the direct effect dominates the indirect effect reverse green-paradox-like outcomes cannot occur but also in such a framework it is possible that green policy deteriorates (short run) environmental outcomes. This suggests that whenever green policies negatively affect individuals' intrinsic motivation it is very difficult to predict their effectiveness since both short run and long run outcomes might be perversely affected.

The result is illustrated in Figure 3 where we compare the laissez-fare outcome (red) with the outcome in the regulated economy, both in the cases in which environmental policy does not affect (blue) and does affect (pale blue) intrinsic motivation. The parameter values are the same employed earlier with two different

[10] A proof of this result is postponed to Section 7 where we discuss under which more general conditions a green-paradox-like outcome might take place.

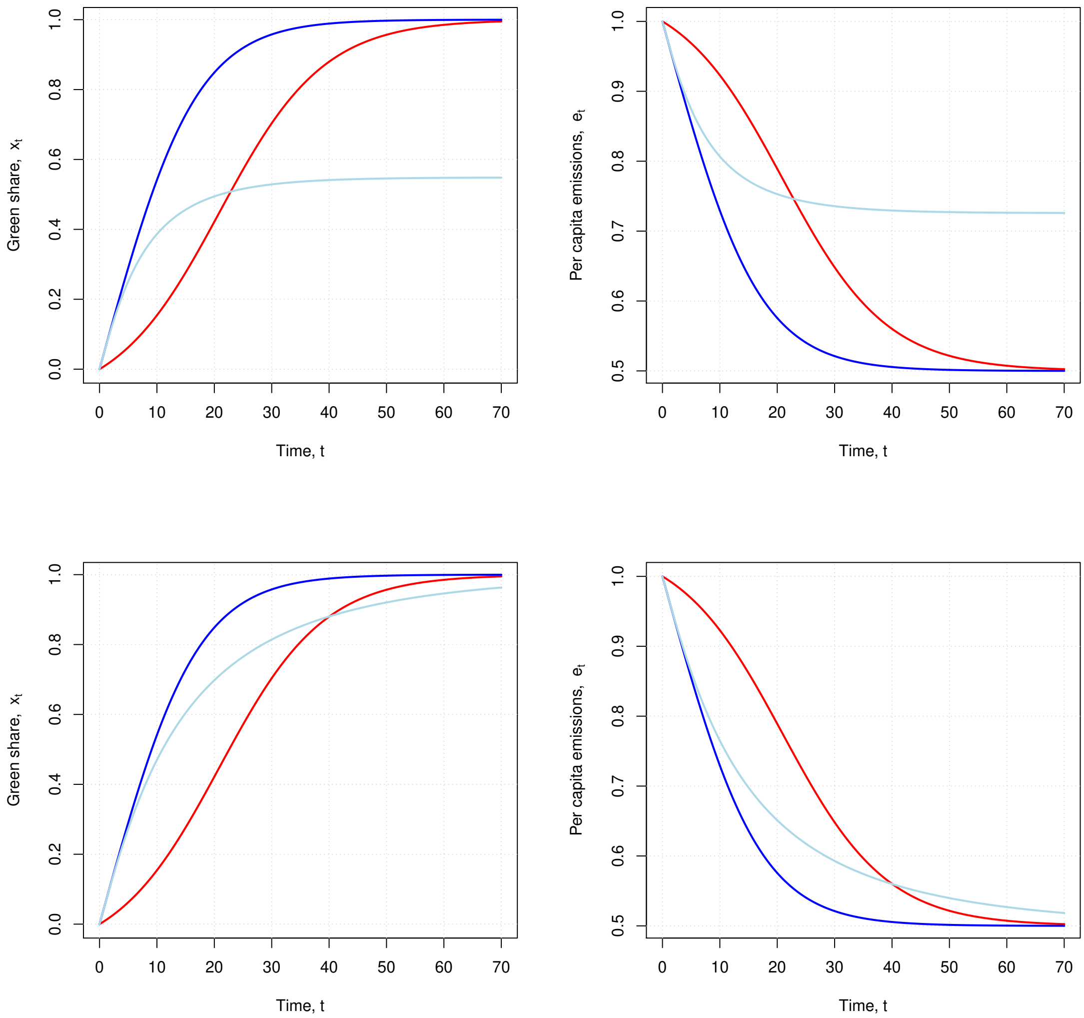

Figure 3: Time evolution of the green share (left) and per capita emissions (right) without (red) and with environmental policy, in the case in which environmental policy affects intrinsic motivation (pale blue) or does not (blue).

configurations of the policy-induced social effect parameter, $\epsilon = 100$ (top) and $\epsilon = 40$ (bottom).[11] Note that in the former case the parameters values satisfy the condition for $\bar{x} < 1$ to exist (the indirect effect dominates), while in the latter the condition is not met and the prevailing equilibrium is $\bar{x}^* = 1$ (the direct effect dominates). By comparing the blue and pale blue curves in the top panels, we can see that initially environmental policy generates the same effects on the green share and emissions independently of the implications on intrinsic motivation; however, after a while the consequences of the indirect effect become clear. The effect induced on intrinsic motivation leads to an outcome which is even worse than what would be achieved in the unregulated economy. Indeed, by comparing the pale blue and the red curves, we can see that after the initial boom in the adoption rate and fall in emissions generated by environmental policy, since the social effect is larger than the direct effect in the long run the adoption rate ends up being lower and emissions higher than in the unregulated economy, suggesting thus that understanding a priori the long run effects of environmental policy is not simple at all. Specifically, after the initial 20 years period in which the

[11]The values of ϵ seem huge. This is due to the fact that in all equations, it is multiplied by τ, so that the product $\epsilon\tau$ (the proper coefficient for the indirect effect) is worth 5 in the first parametrization and 2 in the second.

environmental policy seems effective, its long run implications for environmental outcomes become strongly negative. In the bottom panels we can see that even if the indirect effect is smaller than the direct effect environmental policy does not undoubtedly improve environmental outcomes. Indeed, even if the long run equilibrium is still $(\overline{x}^*, \overline{e}^*)$ the speed of convergence is non-monotonically affected: in earlier times the policy boosts adoption but later it slows it down, implying that emissions fall rapidly in earlier times and slowly later on (note that the pale blue emissions curve lies above the red emissions curve after year 40).

Our results show that whenever environmental policy affects intrinsic motivation, green policy may end up harming the environment, either in the short or in the long run. In such a case it is natural to wonder which options may be available to policymakers to avoid such a deterioration in environmental outcomes. A crucial role in determining the eventual negative implications of environmental policy is played by the size of the policy-induced social effect, ϵ. In fact, while J plays a *positive* role in boosting the adoption rate, when combined with ϵ as in (11), it could *negatively* affect adoptions due to a second order effect.[12] If policymakers can design some appropriate policy to prevent such a second order effect, they will be able to restore the equilibrium $(\bar{x}^*, \bar{e}^*)$ and at the same time increase the speed of convergence to it. Reducing ϵ requires to decouple extrinsic and intrinsic motivation: households' perception of their moral responsibility to undertake green actions should become little related to the monetary incentives associated with environmental policy itself. This can eventually be achieved only through education policies aiming to promote self-accountability at micro level for aggregate macro outcomes. Clearly, effectively designing such policies is easier to say than done, thus in reality we will always incur into the potential risk of undesirable policy-induced environmental effects. These results clearly suggest that such policy-induced effects on intrinsic motivation might have important implications on the effectiveness of environmental policy and thus should be carefully taken into account in the design of public policy.

5 Photovoltaics and Feed-in-Tariffs in Italy

We now present a calibration based on a specific case of environmental policy to show that our conclusions are consistent with real world experiences. Specifically, we focus on the introduction of feed-in-tariffs in Italy to promote the installation of photovoltaics in order to increase the contribution of renewable sources to the overall production of electricity at national level (see Gestore Servizi Energetici (2017)). Renewable sources have traditionally played an important role in the production of electricity in Italy, where hydroelectricity has been for decades the main source of renewable energy (in early 2000s hydroelectricity accounted for more than 80% of the total electricity produced via renewable sources). In order to favor a diversification in its portfolio of renewable energy, the Italian government has introduced in mid 2000s a system of feed-in-tariffs targeting in particular photovoltaics. The so-called "Conto Energia" scheme was firstly introduced in 2005 to incentivize investments in the installation of photovoltaic plants from both households and businesses by providing them with generous tax breaks and investment rebates, and given its success other schemes were introduced in the following years. Between 2005 and 2013 five different Conto Energia schemes were introduced. Over time, the underlying incentive mechanism became less and less strong with the duration and size of tax breaks and rebates substantially falling from one scheme to the next. The nature of such a scheme is consistent with the type of green policy we discussed earlier and thus provides us with an interesting opportunity to assess the validity of our conclusions. Note that since in our setting there exists an inverse relation between adoption and emissions (see equation 6), focusing on either the former or the latter in order to test our theoretical model will be completely equivalent. Therefore, due to data limitations which do not allow to disentangle emissions according to its different sources, in the following we will simply focus on adoption.

In early 2000s photovoltaics were still in their infancy and prior to the incentive mechanism, in Italy

[12] We postpone to section 7 (see Example 1) a detailed analysis of the mathematical implications of such a model.

Year	Photovoltaics (GWh)	Total (GWh)	%
2005	31	303,672	0.0102
2006	35	314,090	0.0111
2007	39	313,888	0.0124
2008	193	319,130	0.0605
2009	676	292,642	0.2310
2010	1,906	302,062	0.6310
2011	10,795	302,570	3.5678
2012	18,861	299,276	6.3022
2013	21,588	289,803	7.4492
2014	22,306	279,829	7.9713
2015	22,942	282,994	8.1069

Table 1: Data for energy produced via photovoltaics and total energy produced in Italy in the period 2005-2015.

as in several other European countries, photovoltaics were to a large extent unknown to the public and the amount of electricity produced via solar power was almost negligible. Following the introduction of the Conto Energia schemes from 2005 the installation of photovoltatics has boomed increasing the photovoltaics energy power by 22,800 GWh over an eight-years period (2008–2015). The share of total energy produced, expressed in GWh, via photovoltaics has substantially increased over time reaching about 8.11% in 2015, as reported in Table 1. In order to understand what lies at the basis of this sharp increase in the contribution of photovoltaics to energy production, we now fit this curve through our model, given by (12). The best fit[13] is represented by the dashed curve in Figure 4, which is consistent with the following parameter values: $b - c = 1 \cdot 10^{-6}$, $\tau = 4 \cdot 10^{-6}$, $J = 1.40$ and $\tilde{\epsilon} := \epsilon\tau = 13.02$.[14] As we can see, the calibrated model fits particularly well the data and it allows to identify the role played by the different utility components. While both the small private $(b - c)$ and the policy-induced incentive (τ) terms jointly contribute to explain the flat pattern of the curve over the initial period (2005–2008), the large social externality term (J) explains its sudden increase in the middle period (2009–2012) and the huge policy-induced social effect (ϵ) is essential to explain its flattening out in the final period (2013–2015). Note that each of these elements is crucial to effectively mimic the dynamics of the electricity share produced by photovoltaics, and a critical role is played by the social utility components captured by the parameters J and ϵ. In particular, if this latter term is null (that is, public policy does not affect intrinsic motivation such that the indirect effect is null) it will be not possible to replicate the observed evolution of the electricity share produced by photovoltaics over the 2013–2015 period.

Our calibration results are clearly consistent with our theory stating that, since extrinsic motivation may crowd-out intrinsic motivation due to the social effects induced by public policy, green policy may end up harming the environment. However, there exists another alternative interpretation of the observed adoption rate. It is possible that households' valuation of the green technology decreases over time either because of the development of alternative better technologies or because of the introduction of competing environmental subsidies. To the best of our knowledge, in the specific case under consideration and over that specific time period, neither alternative technologies allowing to produce clean electricity have been developed nor policymakers have introduced other forms of feed-in-tariffs or subsidies to incentivize some different form of renewable electricity production. Thus, the substitution motives potentially reducing individuals' valuation of photovoltaics can be ruled out.

[13]The calibration is performed using a *rough square errors minimization* over a grid of predetermined values for the parameters. More details on the numerical procedure are discussed in appendix A.

[14]Since in all equations ϵ is multiplied by τ, it seems convenient to estimate directly the product $\varepsilon = \epsilon\tau$.

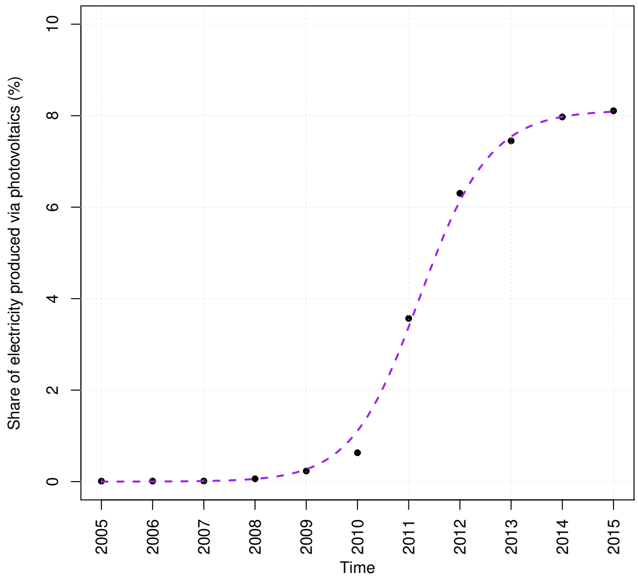

Figure 4: Percentage of total electricity produced by photovoltaics in Italy: data (dotted points) ranging in the period 2005–2015 and predicted values (dashed curve).

A formal empirical investigation of the role of environmental policy in affecting social norms and intrinsic motivation is outside the scope of our paper, but such a simple calibration lends support to our conclusion that environmental policy through its social effects may lead to crowding-out of intrinsic motivation which may ultimately undermine the overall effectiveness of the implemented policy. This result is consistent with Frey & Oberholzer–Gee (1997), who show that since extrinsic motivation crowds-out intrinsic motivation, incentives introduced to compensate individuals for accepting the development of noxious facilities may reduce local support rather than increasing it. Different from them, our analysis shows that similar results may arise even if the incentives are introduced to promote the development of a locally desirable project.

6 Pollution Externality

The setup we have analyzed thus far assumes that households do not account for environmental quality when deciding whether to adopt the green technology. We now extend our baseline framework to allow for a pollution externality, by assuming that the level of emissions directly affects households' utility and therefore it plays a crucial role in determining households' decisions (see for example Gradus and Smulders (1993), or more recently Bosi et al. (2019)). We will show that our main results hold true also in such a richer context, confirming thus that our conclusions can be safely applied also in a more traditional environmental economics framework with a pollution externality. Specifically, we assume that the adoption decision is based on the following utility function:

$$u_{i,t}(\omega_i) = \omega_i \left[b + \epsilon_i - c + Jx_t^N + V(e_t^N)\right], \tag{15}$$

where $V(\cdot)$ is an increasing function of per-capita emissions as defined in (4). Since households care for the environment, the higher the level of emissions, the higher their desire to contribute to the green cause and so the higher the perceived value of the new technology and the higher the likelihood of adoption. Indeed, as discussed in our baseline model, household i will decide to adopt the green technology ($\omega_{i,t} = 1$)

if $b + \epsilon_i - c + Jx_t^N + V(e_t^N) \geq 0$.[15] For the sake of simplicity, we assume $V(\cdot)$ to be linear thus we postulate $V(e_t^N) = e_t^N = e - e(1-\psi)x_t^N$. By plugging this into (15), the utility function reads as follows:

$$u_{i,t}(\omega_i) = \omega_i \left[b - c + e + (J - e(1-\psi))x_t^N - \epsilon_i\right], \tag{16}$$

from which we can conclude that the form of the utility function is still as in our baseline model (and thus the asymptotic results presented in Section 7 will still apply also in this extended setup), meaning that it is as if we are adjusting some parameters to account for the emissions level. Indeed, by defining such new (adjusted) parameters as $\tilde{b} = b + e$ and $\tilde{J} = J - e(1-\psi)$, we can rewrite the green share in equation (5) as follows:

$$\dot{x}_t = (1 - x_t) \left[(\tilde{b} - c + \tilde{J}x_t) \wedge 1\right]. \tag{17}$$

To discuss the possible equilibrium outcomes, we follow the same approach discussed earlier for our baseline model. Intuitively, it turns out that exactly the same results apply provided that $\tilde{J}$ is positive. However, now $\tilde{J}$ could also be negative thus we need to discuss what may happen in this new scenario, and in particular we need to understand whether a new equilibrium $\tilde{x} < 1$ may appear. If this is the case, the new equilibrium must take the following form $\tilde{x} = \frac{\tilde{b}-c}{-\tilde{J}}$. For this to be well defined ($\tilde{x} < 1$) we will need that $\tilde{b} - c < -\tilde{J}$, which is equivalent to $b - c < -J - \psi$, but since by assumption $b - c > 0$ and $-J - \psi < 0$ this situation is not possible, meaning that also in the $\tilde{J} < 0$ case, there is no other equilibrium different from $(\bar{x}^* = 1, \bar{e}^* = \psi e)$. Therefore, also in the presence of a pollution externality Proposition 1 still applies.

In such an extended model the green share in the regulated outcome (with a direct effect only), given by equation (10) in our baseline model, now reads as follows:

$$\dot{x}_t = (1 - x_t) \left[\left(\tilde{b} + \frac{\tau}{1 + x_t} - c + \tilde{J}x_t\right) \wedge 1\right]. \tag{18}$$

By applying the same arguments presented above it is possible to show that in the presence of a pollution externality, the equilibrium is still given by $(\bar{x}^*, \bar{e}^*)$ and thus Proposition 2 still holds true by adjusting the parameter values to account for the effects of emissions.

The green share in the regulated outcome (with both direct and indirect effects), given by equation (12) in our baseline model, now becomes:

$$\dot{x}_t = (1 - x_t)\,\eta \left(\tilde{b} + \frac{\tau}{1 + x_t} - c + \tilde{J}\left(1 - \frac{\epsilon\tau x_t}{1 + x_t}\right) x_t\right). \tag{19}$$

Following the same approach as in our baseline model, we need to generalize the sufficient condition in (13) for a non-trivial equilibrium $\bar{x} \in (0, 1)$ to exist. It can be shown that the same condition is still sufficient (by substituting the original with the adjusted parameters) provided that also $\tilde{J} = J - e(1-\psi) > 0$ is met, which requires that the emissions generated by households' activities adjusted for the adoption of the green technology are not too large. If this were not the case, that is net (of adoption) emissions were particularly large, the incentive to purchase the new technology would be particularly strong leading to full adoption at equilibrium $(\bar{x}^*, \bar{e}^*)$. Therefore, the sufficient condition for a non-trivial equilibrium $\bar{x}$ to exist is the following:

$$\max\left\{\frac{1}{\epsilon}; \frac{2(b - c + J + e\psi)}{(J - e(1-\psi))\epsilon - 1}\right\} < \tau < 1 - (b - c + e); \quad J > e(1-\psi), \tag{20}$$

and whenever (20) is met, the equilibrium achieved in the long run will be $(\bar{x}, \bar{e})$ associated with partial adoption and emissions larger than in the lasseiz-faire equilibrium (i.e., $\bar{x} < \bar{x}^*$ and $\bar{e} > \bar{e}^*$). In such a case

[15] From another point of view, household i will adopt the new technology if $b + \epsilon_i - c + Jx_t^N \geq -V(e_t^N)$, meaning that in our extended framework the adoption threshold ($u_{i,t}(0) = 0$ in our baseline model) is now endogenous and depending on emissions, which from (4) depend on the green share. Since $V' > 0$, as more household adopt the new technology emissions decrease and the adoption threshold increases reducing new adoptions, since intuitively when the environment is healthy the individual incentive to undertake green actions is lower.

exactly the same results discussed in Proposition 3 still hold true. This shows that the introduction of a pollution externality in the utility function does not modify our main conclusions and it is still possible (provided that some additional parameter restriction is verified) that green policies end up harming the environment either in the short or in the long run.

7 A General Class of Models for the Green Paradox

The results discussed thus far are based upon some specific functional forms which have allowed us to present our argument in the simplest and most intuitive way. We now show that such explicit functional forms are not needed to give rise to our main results, which instead do hold true under rather general assumptions about the implications of the environmental policy on the private and social components of individuals' utility function. To this aim, assume that the individual utility structure is given by:

$$u_i(\omega_{i,t}) = \omega_{i,t} \left[F(x_t^N) - \varepsilon_i\right], \tag{21}$$

for a differentiable function F, capturing both private and social utility components (even those associated with the direct and indirect effects of environmental policy, and those associated with pollution externalities). Recall that ε_i, for $i = 1, \ldots, N$, are i.i.d. random terms with distribution η. The probabilistic description of the model, in line with (21), is now given by:

$$\mathbb{P}(\omega_{i,t+\Delta t} = 1 | \omega_{i,t}, x_t^N) = \eta(F(x_t^N)). \tag{22}$$

The infinitesimal generator $\mathcal{L}_N$ of the continuous time Markov chain x_t^N related to (22) has the following first order approximation:

$$\mathcal{L}_N f(x) = (1 - x_t^N)\, \eta(F(x_t^N))\, f'(x) + o\left(\frac{1}{N}\right). \tag{23}$$

Since all the derivatives are uniformly bounded, it can be proved that

$$\lim_{N\to\infty} \sup_{x\in\mathbb{R}} |\mathcal{L}_N f(x) - \mathcal{L} f(x)| = 0 \tag{24}$$

where $\mathcal{L} f(x) = (1 - x)\, \eta(F(x))\, f'(x)$. By considering $f \equiv x$, it is straightforward to show that $\mathcal{L}$ is the infinitesimal generator of the process x_t, where

$$\dot{x}_t = (1 - x_t)\, \eta\left(F(x_t)\right). \tag{25}$$

By virtue of Theorem 1.6.1 in Ethier and Kurtz (1986), x_t^N converges to x_t. Accordingly, emissions per capita terms, e_t^N, will converge to e_t, where $e_t = e - e(1 - \psi)x_t$. Summarizing:

Theorem 1. *Consider N agents taking decisions according to (21), where $F : \mathbb{R} \to \mathbb{R}$ is a differentiable function and where ε_i, for $i = 1, \ldots, N$, are i.i.d. random terms with distribution η. Define the stochastic process $\{x_t^N\}_{t\geq 0}$, such that, for all t, $x_t^N = \frac{1}{N}\sum_{i=1}^N \omega_{i,t}$. Then, when N goes to infinity, x_t^N converges to the (deterministic) process x_t solving (25), whereas e_t^N converges to $e_t = e - e(1 - \psi)x_t$.*

Note that, by specifying F to meet the assumptions of the models discussed in the previous sections, we can easily obtain equations (5), (10) and (12) as a corollary of the above theorem. Now, having green-paradox-like outcomes in mind, we concentrate on the variable x_t, proposing general conditions on F under which the following condition is verified:

Condition 1. *There exists a stable equilibrium $\bar{x} \in (x_0, 1)$ for (25).*

A necessary and sufficient condition for Condition 1 to hold is:

$$F(\bar{x}) = 0 \text{ for some } \bar{x} \in (x_0, 1); \quad 0 < F(x) < 1 \text{ for all } x \in [x_0, \bar{x}). \tag{26}$$

In the case $F(\bar{x}) = 0$ admits a unique solution $\bar{x} < 1$, it is immediately derived that it is the unique (different from $\bar{x}^* = 1$) asymptotically stable equilibrium of the dynamics starting at $x_0 < \bar{x}$. If a multiplicity of such points exist, the minimum is the stable equilibrium. Condition $0 < F(x) < 1$ for all $x \in [x_0, \bar{x})$ is needed to ensure that the dynamics starts with a positive slope and that F does not exit the support of η (otherwise, $\dot{x}_t = (1 - x_t)$ would converge to $\bar{x}^* = 1$). As a matter of fact, we necessarily need the following condition on F in order for Condition 1 to be met: there exists a value $\tilde{x} \in (x_0, 1)$ such that

$$\frac{dF(\tilde{x})}{dx} < 0. \tag{27}$$

Basically, in order for F to be zero somewhere it needs to turn from positive to negative, thus there must exist (at least a) *turning point* of the *perception of social utility.* This phenomenon, described in the literature on social norms (see Nyborg (2003)) and due to the crowding-out of intrinsic motivation in our setting, is precluded under the classical paradigm of social utility á-la Brock & Durlauf (2001) where social interactions always give rise to a positive externality. Indeed, that framework would prescribe $\frac{dF}{dx} > 0$ for all x, suggesting that the larger x the higher the incentive for single agents to choose $\omega_i = 1$. Therefore, Condition 1 can be obtained only if we relax the basic hypothesis behind classical social interaction models, which is exactly what we are doing by introducing an indirect effect associated with environmental policy.

Coming back to the study of sufficient conditions, (26) turns out to be difficult to be tested in practice. A slightly more restrictive condition is:

$$F(1) < 0 < F(x_0); \quad \max_{x \in [0,1]} F(x) < 1 \tag{28}$$

$F(x_0) > 0$ ensures that at time zero the diffusion effectively takes off. By continuity of F, as soon as $F(1) < 0$, it must exist (at least one) $\bar{x} \in (x_0, 1)$ at which $F(\bar{x}) = 0$. The second condition ensures that F does not exceed 1. We summarize this straightforward analysis into the following statement.

Corollary 1. *Under the assumptions of Theorem 1, assume that* (28) *holds true, then Condition 1 is satisfied. Moreover, in this case, necessarily F is such that:*

$$\frac{dF(\tilde{x})}{dx} < 0$$

for some $\tilde{x} \in [0, 1]$. Therefore, the monotonicity paradigm underlying classical random utility models with social interactions is not in place.

Corollary 1 provides us with sufficient conditions for green-paradox-like outcomes to occur in more general settings than those discussed earlier. Indeed, it is straightforward to show that our analysis in Section 4 represents a special case of the above corollary, as illustrated in the following example.

Example 1. *Consider the model described by* (12)*; because of an indirect effect, the social externality may turn from positive to negative in such a way that a non-trivial equilibrium emerges. For sake of clarity we recall that, in this case, F is given by:*

$$\begin{aligned} F(x) &= b - c + \frac{\tau}{1+x} + J\Big(1 - \frac{\epsilon\,\tau x}{1+x}\Big)x \\ &= \frac{J(1-\epsilon\,\tau)\,x^2 + (b - c + J)\,x + (b - c + \tau)}{1+x}. \end{aligned} \tag{29}$$

It is not difficult to see that, in order to have a positive root for the quadratic polynomial at the numerator in (29)*, necessarily* $(1 - \epsilon\,\tau) < 0$*, so that the quadratic function itself is concave. To find conditions ensuring*

(28)*, we need to discuss the values of* F *at* $x = 0$*,* $x = 1$ *and* $x \in (0, 1)$*. Indeed, when* $x_0 = 0$*,* $F(0) = b - c + \tau$*. On the other hand,* $F(1) = \frac{J(1-\epsilon\,\tau)+(b-c+J)+(b-c+\tau)}{2}$*. To ensure that* $F(x) < 1$ *for all* $x \in [0, 1]$*, notice that on this interval*

$$F(x) = \frac{J(1-\epsilon\,\tau)\,x^2 + (b-c+J)\,x + (b-c+\tau)}{1+x} \leq J(1-\epsilon\,\tau)\,x^2 + (b-c+J)\,x + (b-c+\tau).$$

Since the maximum for the quadratic function on the RHS of the previous inequality is reached when $x = \frac{b-c+J}{2J(1-\epsilon\,\tau)} < 0$*, and being the quadratic function concave, it is maximized (when restricting on the unitary interval) at* $x = 0$*, so that* $F(x) \leq F(0) = b - c + \tau$*.*

We thus obtain the following sufficient conditions for a stable equilibrium $\bar{x} \in (0, 1)$ *to exist:*[16]

$$\begin{cases} \tau > \frac{1}{\epsilon} \\ \tau > \frac{2(b-c+J)}{J\epsilon - 1} \\ \tau < 1 - (b - c) \end{cases}$$

which are exactly the conditions given by (13). Since $\bar{x}$ *in* (14) *is the unique positive solution to* $F(0) = 0$*, it is also the only (different from* $\bar{x}^* = 1$*) stable equilibrium for* (12)*. Finally, note that, as far as the calibrated model in Section 5 is concerned, parameters satisfy all the above assumptions.*

To the best of our knowledge, in the social interactions literature none, not even in other contexts, has ever analyzed a dynamic model where the social utility turns from positive to negative due to the crowding-out of intrinsic motivation. Thus, apart from the green-paradox implications, our paper presents an interesting contribution to the classical social interaction literature by considering a possibly important extension.

8 Conclusions

The design of environmental policy is very complex and it is well known that public policies might give rise to effects very different from those hypothesized and hoped for by policymakers. The green paradox, in its broader interpretation, suggests that green policy under certain specific conditions might result in a deterioration of environmental outcomes. Several mechanisms explaining why this might be the case have been proposed thus far, but none takes into account social effects and crowding-out of intrinsic motivation. In this paper we show that policy-induced social effects altering individual intrinsic motivation might be an alternative source of outcomes similar to those conjectured by the green paradox. However, we show that this novel channel might eventually give rise to a reserve green paradox, in the sense that environmental policy might improve the short run but deteriorate the long run environmental outcomes. Whether this effectively happens crucially depends on how the two components of individual utility, namely private and social utility, change with environmental policy. A green subsidy will undoubtedly increase the private component through its direct effect on extrinsic motivation, but the social component might even decrease if the policy indirectly weakens individual intrinsic motivation. In such a case, if the negative indirect effect outweighs the positive direct effect, the policy will have negative long run consequences. Moreover, even in the case in which the direct effect dominates, provided that the negative indirect effect is large enough, the policy will have negative short run consequences. Since quantifying social effects is generally not possible, our results suggest that the design of environmental policy is even more complicated than what traditionally believed since requiring also to take into account the eventual effects induced on intrinsic motivation. Our results are consistent with real world experiences, as shown in a calibration based on the

[16]Note that this conditions, although more restrictive than (26), have the huge advantage to be independent of the initial condition: indeed, they only depend on parameter values. Moreover, in case of a quadratic function as in (29), they are also necessary for $\bar{x} \in (0, 1)$ to exist.

Italian photovoltaics case over the 2000–2015 period, suggesting that our theoretical conclusions match well empirical observations.

This paper represents a first attempt to analyze the social effects of environmental policy and their implications for environmental outcomes. The approach has thus been quite simplistic and several issues have not been taken into account at all. First, we have assumed that environmental policy affects individuals' intrinsic motivation which alters the magnitude of the social component in individuals' utility function; however, we have not formally modeled how this change in intrinsic motivation occurs, while understanding the underlying mechanisms and dynamics might provide us with some further insights on the possible problems involved in the design of public policy. Second, we have focused on a partial equilibrium approach to analyze the adoption decision of a green technology; however, taking a general equilibrium perspective might be useful to introduce adoption into a consumption and investment decision setting shedding some more light on households' behavior. Third, we have have not discussed the determinants of emissions and the financing sources for public policy; however, in a traditional macroeconomic-environmental context both emissions and taxes may be related to production allowing to analyze the feedback effects among production, emissions and public policy. Extending the analysis along these directions is left for future research.

A Numerical Procedure

In this appendix we briefly illustrate the methodology used for the calibration of the model discussed in Section 5. The aim of the numerical procedure is to identify the values of the triplet of parameters $\theta = (\epsilon, J, \tau)$, defined over a prespecified grid Θ, that best fits the data in Table 1. In particular, we numerically solve the problem of minimizing a *fit function* $f(\theta)$ as follows:

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} f(\theta), \quad f(\theta) = \sum_{k=1}^{11} \big(Data(k) - Curve_\theta(k)\big)^2,$$

where $Data(k)$, $k = 1, \ldots, 11$ is the percentage of energy produced via photovoltaics as reported in the last column of Table 1 and $Curve_\theta(k)$ is the evaluation of the simulated trajectory using the triplet θ. To this aim, we have run a Matlab code[17] to evaluate $f(\theta)$ for $\theta \in \Theta$. We implemented a double procedure:

1. Preprocessing procedure over a coarse grid of parameters.

   In Table 2 we report the range of the parameters and the increment used to build the grid.

Parameter	Lower value	Upper value	Increment
ϵ	5	15	0.5
J	0.6	1.6	0.1
τ	$1 \cdot 10^{-6}$	0.001	$5 \cdot 10^{-5}$

Table 2: Range of parameters for the preprocessing phase of calibration.

   Some remarks about the chosen range of the parameters. Concerning ϵ, as already stressed in Section 5, it is convenient to calibrate the product $\tilde{\epsilon} = \epsilon\tau$. $J \approx 1$ is a typical value of the parameter related to social interactions. The magnitude of the incentive τ has to be, for our purposes, of the order of 10^{-5}.

   The result of this preprocessing phase gives as estimated parameters: $\tilde{\epsilon} = 13$, $J = 1.3$, $\tau = 1 \cdot 10^{-6}$. We use this parameters to create a finer grid and to obtain a more precise (final) calibration of the triplet of parameters.

[17] We used the software release Matlab R2017b with a i7-6700 CPU machine. The Matlab source code is available upon request from the authors.

2. Final procedure over a finer grid of pre-selected values of the parameters.

   This second phase uses a finer grid as in Table 3; it returns the values of parameters $\tilde{\epsilon} = 13.02$, $J = 1.40$, $\tau = 4 \cdot 10^{-6}$, reported in Section 5. The final calibration error is $f(\hat{\theta}) \approx 6.2252 \cdot 10^{-5}$.

Parameter	Lower value	Upper value	Increment
ϵ	12.5	13.5	0.02
J	1.2	1.4	0.002
τ	$1 \cdot 10^{-6}$	$1 \cdot 10^{-5}$	$2 \cdot 10^{-7}$

Table 3: Range of parameters for the final phase of calibration.

Acknowledgements

We are indebted to two anonymous referees for their constructive comments allowing us to substantially improve our paper. We also wish to thank the seminar participants at the University of Turin and Aarhus University for their helpful comments and suggestions. Marco Tolotti acknowledges Università Ca' Foscari Venezia for the financial support under the grant *"Interactions in complex economic systems: contagion, innovation and crises"*.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study

References

Barucci, E., Tolotti, M. (2012). Social interaction and conformism in a random utility model, Journal of Economics Dynamics & Control 36, 1855–1866

Bass, F.M. (1969). A new product growth for model consumer durables, Management Science 15, 215-–227

Benabou, R., Tirole, J. (2003). Intrinsic and extrinsic motivation, Review of Economic Studies 70, 489—520

Blume, L., Durlauf, S. (2003). Equilibrium concepts for social interaction models, International Game Theory Review 5, 193–209

Bosi, S., Desmarchelier, D., Ragot, L. (2019). Pollution effects on preferences: a unified approach, Journal of Public Economic Theory 21, 371–399

Brekke, K.A., Kvemdokk, S., Nyborg, K. (2003). An economic model of moral motivation, Journal of Public Economics 87, 1967—1983

Brock, W., Durlauf, S. (2001). Discrete choice with social interactions, Review of Economic studies 68, 235–260

Cardella, E.,Ewing, B.T., Williams, R.B. (2017). Price volatility and residential electricity decisions: experimental evidence on the convergence of energy generating source, Energy Economics 62, 428–437

Colapinto, C., Sartori, E., Tolotti, M. (2014). Awareness, persuasion, and adoption: enriching the Bass model, Physica A 395, 1–10

Cairns, R.D. (2014). The green paradox of the economics of exhaustible resources, Energy Policy 65, 78-–85

Ethier, S.N., Kurtz, T.G. (1986). Markov processes: characterization and convergence (John Wiley and Sons)

Frey, B. Oberholzer–Gee, F. (1997). The cost of price incentives: an empirical analysis of motivation crowding–out, American Economic Review 87, 746-–755

Gestore Servizi Energetici (2017). Rapporto Statistico - Energia da Fonti Rinnovabili in Italia - Anno 2015, available at: http://www.gse.it/it/Statistiche/RapportiStatistici/Pagine/default.aspx

Gradus, R., Smulders, S. (1993). The trade-off between environmental care and long-term growth: pollution in three prototype growth models, Journal of Economics 58, 25–51

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

Khalil, F., Kim, D., Lawarree, J. (2019). Use it or lose it, Journal of Public Economic Theory 21, 991–1016

Marsiglio, S., Tolotti, M. (2018). Endogenous growth and technological progress with innovation driven by social interactions, Economic Theory, 65, pp. 293-328

Nyborg, K. (2003). The impact of public policy on social and moral norms: some examples, Journal of Consumer Policy 26, 259-–277

Nyborg, K., Anderies, J.M., Dannenberg, A., Lindahl, T. Schill, C., Schlüter, M., Adger, W.N., Arrow, K.J., Barrett, S., Carpenter, S., F.S., Chapin III, Crepin, A.S., Daily, G., Ehrlich, P., Folke, C., Jager, W., Kautsky, N., Levin, S.A., Madsen, O.J., Polasky, S., Scheffer, M., Walker, B., Weber, E.U., Wilen, J., Xepapadeas, A.,de Zeeuw, A. (2016). Social norms as solutions, Science 354, 42–43

Peres, R., Muller, E., Mahajan, V. (2010). Innovation diffusion and new product growth models: a critical review and research directions, International journal of research in marketing, 27, 91–106

Perino, G., Requate, T. (2012). Does more stringent environmental regulation induce or reduce technology adoption? When the rate of technology adoption is inverted U-shaped, Journal of Environmental Economics and Management 64, 456–467

Rege, M. (2004). Social norms and private provision of public goods, Journal of Public Economic Theory 6, 65—77

Sinn, H.W. (2008). Public policies against global warming, International Tax and Public Finance 15, 360—394

Steg, L., Lindenberg, S., Keizer, K. (2005). Intrinsic motivation, norms and environmental behaviour: the dynamics of overarching goals, International Review of Environmental and Resource Economics 9, 179—207

Sundt, S,, Rehdanz, K. (2015). Consumers' willingness to pay for green electricity: a meta-analysis of the literature, Energy Economics 51, 1–8

van der Werf, E., Di Maria, C. (2012). Imperfect environmental policy and polluting emissions: the green paradox and beyond, International Review of Environmental and Resource Economics 6, 153-–194