THE ADOPTION OF CLEANER HOUSEHOLD TECHNOLOGIES: WHEN IS BACKFIRE WELFARE-IMPROVING?

Peter Kennedy
Department of Economics
University of Victoria

29 April 2019

ABSTRACT

I examine a setting in which households with "green preferences" choose between two available technologies on the basis of their costs, and on the basis of their associated emissions. The green preferences give rise to a reciprocal externality among the households, the correction of which requires policy intervention. I show that in the absence of corrective policy, the adoption of a cleaner technology can be welfare-improving even when it induces an increase in emissions (backfire). More generally, a reduction in emissions is neither necessary nor sufficient for the equilibrium adoption of a cleaner technology to be welfare-improving. I also show that mandated adoption of a cleaner technology – when households have otherwise chosen not to adopt it – is never welfare-improving if it induces backfire. Conversely, there exist conditions under which equilibrium adoption of the cleaner technology is welfare-reducing, due to the associated backfire. Under these conditions, policy intervention calls for a prohibition on adoption of the cleaner technology. I then characterize the first-best solution, and show that emissions can be higher after the optimal adoption of a cleaner technology; that is, backfire can be optimal. Problematically, implementation of the first-best solution may not be possible with a corrective tax on emissions because the right Pigouvian tax can induce the wrong technology choice in the corrected equilibrium. Under those circumstances, a finely-tuned supplementary policy involving a subsidy or tax on the technology itself is needed to achieve first-best.

1. INTRODUCTION

Does the adoption of a cleaner technology necessarily lead to lower emissions? This question has been widely studied in the existing literature, primarily in the context of energy efficiency improvements, and that literature has identified a variety of mechanisms through which energy use – and hence, emissions from energy use – can "rebound" after the adoption of more energy-efficient devices and production technologies.

The most obvious of those rebound mechanisms operates through price. The effective price of energy-related services, like space-conditioning, lighting and transportation, falls when energy efficiency rises. This fall in effective price leads to an increase in the use of those services, thereby offsetting – at least to some extent – the direct effect on energy use via the energy-efficiency improvement. This offsetting "rebound" can in principle be large enough to cause overall energy use to rise, an outcome that has come to be called "backfire". The same phenomenon can arise more broadly in any context where the adoption of a cleaner technology reduces the marginal cost of producing a good or household service, leading to an increase in output that more than offsets the lower emissions-intensity of the cleaner technology.

In the context of energy-efficiency improvements that are mandated by policy for the express purpose of reducing energy use, the term "backfire" – with its negative connotation – is perhaps appropriate. However, in a broader sense the term can be somewhat misleading because backfire can be an *optimal* outcome. In particular, if the benefits of higher consumption under a cleaner technology outweigh the costs of an induced increase in emissions, then backfire is welfare-improving. This basic point is the over-arching theme of this paper.

The vast majority of the existing literature on rebound focuses solely on the sign and size of the rebound effect, either empirically or in the context of theoretical models.[1] There is surprisingly little analysis of its welfare effects. While a number of papers examine rebound in the context of choice-theoretic models that are in principle well-

[1] For reviews of the literature, see Greening et al. (2000), Sorrell et al. (2009), and Gillingham et al. (2016).

suited to welfare analysis [2], that analysis typically excludes any explicit consideration of external costs. Yet it is these costs that typically motivate policy intervention in practice.

A notable exception to this pattern in the literature is Chan and Gillingham (2014). Their work is the first paper, to my knowledge, that explicitly includes external costs in a formal model of rebound. They examine the welfare effects of an energy-efficiency improvement, and the associated rebound, in the context of a social welfare function that is linearly-separable in the utility of a representative consumer and the external cost of that consumer's consumption of energy services. They show that the welfare effect of an energy-efficiency improvement with rebound is ambiguous, and they point out that an "energy efficiency policy should not be dismissed simply because it results in a large rebound". This is an important message for policy-makers. My work here elaborates upon that message and extends it beyond the issue of energy efficiency.

My analysis employs a model that is more rudimentary than that of Chan and Gillingham. It lacks the real-world complexity of multiple energy services and multiple fuel types that they build into their analysis. My model has only a "black-box" linkage between consumption and emissions. However, this simplicity allows me to tractably introduce three important elements of the rebound issue that are absent from the Chan-Gillingham model: technology-adoption costs; non-marginal changes in technology; and the direct incorporation of external costs via "green preferences" among households.[3]

Technology-adoption costs are a surprisingly uncommon feature of rebound studies in the existing literature.[4] Yet those costs can have an important impact on individual decisions to adopt a new technology, and on post-adoption consumption behaviour. Moreover, mandated adoption of a new technology via regulation can impose costs on consumers that they would not otherwise choose to accept. All of this means that

---

[2] See Binswanger (2000), Berkhout et al. (2000), Thomas and Azevedo (2013), and Borenstein (2015).
[3] Chan and Gillingham do not exclude these elements by neglect. They take care to point out that their analysis does not incorporate some aspects of the rebound problem, including adoption costs and behavioural distortions, but they exclude them from their model in order to keep it tractable. Similarly, my model omits other potentially important elements of the problem that have been identified in the literature.
[4] Exceptions include Mizobuchi (2008), Nässén and Holmberg (2009), and Borenstein (2015) who include adoption costs in the context of partial equilibrium models. Allan et al. (2007), Barker et al. (2007, 2009), Turner (2009), and Chang et al. (2018) incorporate capital costs into general equilibrium models. Fullerton and Tan (2019) focus explicitly on the role of capital costs, and provide the most thorough investigation to date of their role in a general equilibrium model. However, none of this work examines the welfare implications of costly adoption in a model that also includes external costs.

a thorough welfare assessment of rebound should consider these costs explicitly. In this paper I allow the cleaner technology to have both a fixed adoption-cost, and a marginal cost of use different from that of the technology it replaces.

My model is simple enough to be solved analytically, and this means that I can examine non-marginal technology changes. This turns out to be important because all of my results indicate that the direction of welfare change after the adoption of a cleaner technology depends critically on the magnitude of the technology change itself.

The most novel feature of my model is the direct incorporation of external costs via "green preferences" among households. Households care directly about the damage their emissions create, and this has two important implications: the relative cleanliness of available technologies matters for their decision-making; and emissions are the source of a reciprocal externality among those households. The presence of this reciprocal externality necessitates the use of a game-theoretic approach to the analysis.

A key advantage of incorporating the externality-driven distortion directly into household choices is that it allows a policy intervention like mandated technology adoption to be assessed in the context of a model that explicitly incorporates the welfare-based rationale for that policy. This allows me to characterize optimal policy in terms of the preferences of the regulated households themselves. Policy is not an add-on to the model; it is motivated by the welfare of the households that populate the model itself.

The welfare analysis yields a number of interesting results. First, I show that a reduction in post-adoption emissions is neither necessary nor sufficient for the equilibrium adoption of a cleaner technology to raise welfare. In particular, equilibrium backfire can be welfare-improving. Second, mandated backfire is never welfare-improving. That is, if a policy-maker mandates adoption of a cleaner technology because households have chosen not to adopt it, and that adoption causes emissions to rise, then the mandatory adoption policy reduces welfare. Mandatory adoption is only welfare-improving if it causes emissions to fall. Third, there exist conditions under which equilibrium adoption of the cleaner technology is welfare-reducing, due to the associated backfire. Under these conditions, policy intervention calls for a prohibition on adoption of the cleaner technology.

I also allow the policy-maker to use a corrective tax on emissions. I first derive the first-best solution for emissions and for technology choices, and show that there are conditions under which backfire occurs even in this first-best solution. That is, backfire can be optimal. I then ask whether or not Pigouvian taxes can implement the first-best solution, and show that the answer is sometimes “no”. In particular, there exist conditions under which the Pigouvian tax matched to the optimal technology may not induce the choice of that technology in the tax-corrected equilibrium. That is, the right Pigouvian tax can cause households to choose the wrong technology. Correcting that problem sometimes requires a subsidy on the adoption of the cleaner technology, and sometimes it requires a tax on the adoption of that technology. Surprisingly, there exist conditions under which the cleaner technology would be adopted in the corrected equilibrium, and where adoption of that technology would reduce emissions, but where the optimal policy is to prevent its adoption with a tax on the technology itself.

The rest of the paper is organized as follows. Section 2 presents the model. Section 3 derives equilibrium outcomes for emissions and technology choices when there are no corrective taxes, and derives conditions under which there is equilibrium backfire in that setting. Section 4 examines the welfare properties of the equilibrium without corrective taxes, and derives conditions under which mandated adoption of the cleaner technology is welfare-improving. Section 5 derives the first-best solution, and introduces the corrective tax on emissions. I derive conditions under which the corrective tax alone is not enough to implement the first-best solution, and characterize the supplementary adoption-subsidy/tax policy needed to achieve first-best. Section 6 concludes with a synopsis of the main results.

2. THE MODEL

Households produce a “dirty” good or service (such as space-conditioning or transportation) via the use of a polluting input (such as energy). The existing technology generates one unit of emissions for every unit of the dirty good consumed. An available cleaner technology reduces that emissions-intensity to $(1-\theta)$, where $\theta \in (0,1)$. Adoption of the cleaner technology involves a fixed cost for the household, denoted $k \geq 0$.

The cleaner technology may also have a marginal cost of use different from that of the existing technology. In particular, let $p > 0$ denote the marginal cost to the household of producing the dirty good using the existing technology, and let

(1) $$p_C = p + \pi$$

denote the marginal cost of producing the dirty good using the cleaner technology, where π could be positive or negative. One can think of p and p_C respectively as the effective price of the dirty good under the old and cleaner technologies. Both prices are assumed to be positive, so $\pi > -p$ is a restriction imposed throughout.

In the special case where energy is the only input used to produce the dirty good, and where energy-use per unit falls to $(1-\theta)$ under the cleaner technology, the new effective price of the dirty good is simply $p_C = p(1-\theta)$. In that case, $\pi = -\theta p < 0$.

More generally, the effective price of the dirty good could include elements unrelated to energy use or to any other correlate of emissions-intensity. In particular, use of the cleaner technology might involve higher or lower time costs, higher or lower servicing and maintenance costs, or the use of different inputs. As an example, consider a switch from a conventional gasoline-powered car to an electric car equipped with regenerative braking. This switch in technology for producing household transportation services is more than a simple case of improved fuel efficiency. The fuel type is different, the mechanical features are different – some simpler, some more complex – and more advance-planning is needed to ensure the car is charged when it is needed and that the charge lasts for the entire journey. Some of these differences make the electric car more costly to operate, while others make it cheaper to operate. On balance, the effective price of transportation could be higher or lower under the electric technology; that is, π could be positive or negative. I allow for both possibilities here.

There are *n* identical households, each with preferences represented by

(2) $$u(x, z, E) = xz - \gamma E$$

where *x* is consumption of the dirty good, *z* is consumption of an emissions-free good, *E* is aggregate emissions from the *n* households, and γ is a parameter that reflects the damage that a household suffers from those emissions. That damaging impact could be direct (as with a health effect) or it could be indirect (as with concern about damage to future generations). The distinction is ultimately unimportant here, and henceforth I will simply refer to γ as the "green preference parameter".[5]

While this preference structure in (2) is obviously restrictive, it does allow the entire model to be solved analytically while still yielding an interesting range of possible equilibria and first-best outcomes.

The aggregate emissions generated by households is

(3) $$E = X_0 + (1-\theta)X_C$$

where X_0 is aggregate consumption of the dirty good by households using the old technology, and X_C is aggregate consumption of the dirty good by households using the cleaner technology. We will see that under most conditions, either all households adopt the new technology or none do (because households are homogeneous), but we will also see that in a setting where an emissions tax is used, there are conditions under which a symmetric pure-strategy equilibrium does not exist, and where some households adopt the cleaner technology and others do not.

A household that retains the existing technology faces the budget constraint

(4) $$px + z = m$$

where *m* is household income. A household that adopts the cleaner technology faces the budget constraint

(5) $$(p+\pi)x + z = m - k$$

I delay the introduction of an emissions tax until Section 5.

3. EQUILIBRIUM EMISSIONS AND TECHNOLOGY CHOICES

Recall that the damage from emissions in this economy is incurred by the households themselves, as specified in (2) above. This means that the externality problem is a reciprocal one, and so it must be modeled as a game. There are two stages to this game. In the first stage, households choose a technology, and in the second stage they choose their consumption of the dirty good. The game is solved recursively for a sub-game perfect Nash equilibrium. The simple specification of the model makes this very

[5] This model is easily extended to incorporate emissions from another jurisdiction as an exogenous component of aggregate emissions. The policy problem can then be set up as a game between jurisdictions.

straightforward. In particular, the linearity of utility in E means that there is a dominant strategy in each stage of the game. The equilibrium strategies are characterized in the following sub-sections.

3.1 Consumption Choices

It is straightforward to show that if a household retains the old technology then its consumption of the dirty good is

$$(6) \qquad \hat{x}_0 = \frac{m-\gamma}{2p}$$

and if it adopts the cleaner technology then its consumption of the dirty good is

$$(7) \qquad \hat{x}_C = \frac{m-k-\gamma(1-\theta)}{2(p+\pi)}$$

In each case there is the possibility of a corner solution where consumption is bound at zero but I henceforth restrict the green preference parameter to rule out this possibility. In particular, I assume henceforth that

$$(8) \qquad \gamma < \gamma_{\max} \equiv m-k$$

3.2 Technology Choices

This stage of the game is equally straightforward to solve. A household adopts the cleaner technology if and only if its equilibrium payoff in the second stage – based on the consumption choices in (6) and (7) above – is higher under that technology than under the old technology. This means that a household will adopt the cleaner technology if and only if $\pi < \hat{\pi}_A(\theta)$, where

$$(9) \qquad \hat{\pi}_A(\theta) = \frac{p(\gamma\theta-k)(2(m-\gamma)+\gamma\theta-k)}{(m-\gamma)^2}$$

I will henceforth to this critical threshold as the "equilibrium adoption threshold". It is depicted in (θ,π) space in **Figure 1**. (The "hat" notation here and throughout indicates that the adoption choice is made in a setting where there are no policies in place to correct the externality associated with consumption).

The intuition behind the shape of the adoption threshold in Figure 1 is straightforward: adoption of the cleaner technology is most attractive for the household if it is much cleaner than the existing one (high θ) and/or cheaper to use than the existing one (negative π).

Note from Figure 1 that if there is no change in the marginal cost of use ($\pi = 0$) then there exists a critical value of θ above which adoption occurs:

(10) $$\hat{\theta}_A = \frac{k}{\gamma}$$

This critical value highlights the impact of the fixed cost on the adoption choice. A low fixed cost makes adoption more attractive, so a reduction in *k* shifts the entire adoption threshold up in (θ, π) space, enlarging the region in which adoption occurs. If $k = 0$ then threshold passes through the origin in Figure 1: the cleaner technology is adopted for any $\theta > 0$ if it is not more costly to use than the old one. Conversely, if $k > \gamma$ then the threshold passes through $\theta = 1$ in Figure 1: the cleaner technology is never adopted if it is more costly to use than the old one.

The role of the green preference parameter in the adoption threshold is more nuanced. A smaller value of γ (reflecting less concern about emissions) causes the adoption threshold to pivot clockwise around a fixed point in (θ, π) space, as illustrated in **Figure 2**. This pivot point is $\{\overline{\theta}, \overline{\pi}\}$, where

(11) $$\overline{\theta} = \frac{k}{m}$$

(12) $$\overline{\pi} = -\frac{(2m-k)pk}{m^2}$$

If $k = 0$ then this pivot point lies at the origin in Figure 2 but otherwise it must lie in the lower interior region, as illustrated in the figure as drawn. As γ falls towards zero, the adoption threshold pivots and straightens to become a flat line at $\pi = \overline{\pi}$. In this limiting case, households base their adoption decisions exclusively on the change in effective prices, because cleanliness *per se* is of no concern to them.

The pivot point identified in (11) and (12), and its relationship to γ, plays a key role in all aspects of this model. Its interpretation is best understood in the context of a key question relating to backfire: does adoption of the cleaner technology cause emissions to rise or fall? I turn to that question next.

3.3 Backfire

Emissions from a household under the old technology and under the cleaner technology are, respectively,

$$(13)\qquad \hat{e}_0 = \hat{x}_0 = \frac{m-\gamma}{2p}$$

and

$$(14)\qquad \hat{e}_C = (1-\theta)\hat{x}_C = (1-\theta)\left(\frac{m-k-\gamma(1-\theta)}{2(p+\pi)}\right)$$

The change in emissions due to adoption is then straightforward to calculate, and can be usefully expressed as the sum of four terms as follows:

$$(15)\qquad \Delta e = -\theta\left(\frac{m}{2(p+\pi)}\right) - \pi\left(\frac{m-\gamma}{2p(p+\pi)}\right) - k\left(\frac{1-\theta}{2(p+\pi)}\right) + \gamma\left(\frac{\theta(2-\theta)}{2(p+\pi)}\right)$$

These four terms correspond respectively to a direct cleanliness effect, a price effect, a fixed-cost effect, and a green-consumer effect. Consider each of these effects in turn.

The *direct cleanliness effect* (*DCE*) in (15) captures the impact on emissions that would occur if consumption remained unchanged after adoption of the cleaner technology. That scenario would arise if and only if $\pi = k = \gamma = 0$, in which case the change in emissions would simply be

$$(16)\qquad \Delta e_{DCE} = -\theta\left(\frac{m}{2p}\right)$$

The *DCE* is unambiguously negative.

The *price effect* (*PE*) in expression (15) arises when the new technology has a higher or lower marginal cost of use relative to the old one. If cost is higher ($\pi > 0$) then consumption of the dirty good falls (all else equal), and so the *PE* is negative. In that case it reinforces the *DCE*. Conversely, if cost is lower ($\pi < 0$) then consumption of the dirty good rises (all else equal), and so the *PE* is positive. In that case it offsets the *DCE* at least to some extent. In the standard terminology, a positive *PE* causes “rebound”.

The *fixed-cost effect* (*FCE*) in expression (15) is due to an income effect. The dirty good is a normal good here, and so the reduction in disposable income due to the fixed adoption cost leads to a fall in consumption of the dirty good. Thus, the *FCE* is negative in this model, and it therefore reinforces the *DCE*. Notice however, that the *FCE* becomes weaker as θ rises because the income effect on consumption has an increasingly small effect on emissions as the technology becomes cleaner. This will turn out to be important.

The final effect in expression (15) is the *green-consumer effect* (*GCE*). While the other three effects are already well-recognized in the literature, the *GCE* is new. It is also the most interesting effect in the context of this model because its sign is perhaps unexpected: it is positive. That is, a green consumer (one with $\gamma > 0$) will increase her consumption of the dirty good after adopting the cleaner technology by more than a non-green consumer would. Why? Under any given technology, the damaging effect of the dirty good motivates a green consumer to consume less of it than she would if it were clean; there is some disutility from consumption because the good is dirty. That disutility is reduced if she adopts the cleaner technology. She can therefore happily consume more of the dirty good because it is not as dirty as it was under the old technology. For example, the buyer of a new electric car might switch from commuting by bicycle to commuting by car instead because her new car is less polluting than her old gas-guzzling one that she previously used only on weekends. One might say that the cleaner technology reduces the "green guilt" from consumption, and thereby leads to higher consumption.

It is worth noting from (15) that the *GCE* rises as θ rises but at a diminishing rate. In particular, differentiating that last term in (15) with respect to θ tells us that it reaches a turning point at $\theta = 1$. This is a perfectly reasonable property, but it will play an important role in some of the results below, so it is important to flag it here.

The overall effect of the four terms in expression (15) depends on relative parameter values, and on the size of θ in particular. It will therefore prove useful to construct a "backfire threshold" that partitions the (θ, π) space into a region in which $\Delta e < 0$, and a region in which $\Delta e > 0$ (backfire). From (15) we can set $\Delta e = 0$ and solve for π to derive this backfire threshold, which can be expressed instructively as

(17) $$\hat{\pi}_B(\theta) = \hat{\pi}_A(\theta) - (\theta - \overline{\theta})\left(\frac{pm(m-k-\gamma(1-\theta))}{(m-\gamma)^2} \right)$$

where the first term, $\hat{\pi}_A(\theta)$, is the equilibrium adoption threshold from (9), and where $\overline{\theta}$ is the pivot point from (11). It is straightforward to show that this threshold is strictly concave in θ, and that $\Delta e > 0$ (backfire) occurs if and only if $\pi < \hat{\pi}_B(\theta)$.

This backfire threshold is depicted in (θ, π) space in **Figure 3**, drawn for a relatively high value of *k*, and a value of $\gamma < \frac{1}{2}\gamma_{\max}$, as defined in (8) above. (The significance of this will become clear in a moment). The figure highlights the fact that the backfire here relates to the change in uncorrected emissions, in the sense that there are no policies in place to correct the externality distortion in consumption choices.

To understand the properties of the backfire threshold, and why it partitions the space the way it does, consider a new technology that is no cleaner than the old one ($\theta = 0$) but whose marginal cost of use is lower, with a negative value of π somewhere below the intercept-point $\hat{\pi}_B^0$ in Figure 3, where

$$(18) \qquad \hat{\pi}_B^0 = -\frac{pk}{m-\gamma}$$

The *PE* from adopting this technology is strong enough to outweigh the *FCE*, and so adoption causes emissions to rise; recall (15) above, evaluated at $\theta = 0$. If we then hold π fixed at this value below $\hat{\pi}_B^0$, and allow θ to rise, the strength of the *PE* in (15) does not change (because π is unchanged), but the negative *DCE* and the positive *GCE* from (15) both begin to kick in. These competing effects are initially weak because θ is small, and so their net effect is also small. Moreover, the *FCE* begins to weaken as θ rises, because as noted earlier, the income effect on consumption has an increasingly small effect on emissions as the technology becomes cleaner. Thus, the overall change in emissions remains positive even as θ rises. However, as θ rises still further the negative *DCE* strengthens at a linear rate, while the positive *GCE* strengthens at a diminishing rate, as noted above, and so the overall change in emissions eventually turns negative, at the point where the backfire threshold is crossed. Of course, as $\theta \to 1$, emissions necessarily fall towards zero.

The backfire story becomes more somewhat complicated when we consider a setting with a higher value of γ. In particular, the simple monotonicity of Δe in θ that Figure 3 depicts does not hold for higher values of γ. If we raise γ above the value for which Figure 3 is drawn, the curvature of the backfire threshold tightens, and at values of γ above $\frac{1}{2}\gamma_{\max}$, the threshold becomes non-monotonic in θ, as illustrated by the threshold labeled in $\hat{\pi}_{BH}(\theta)$ in **Figure 4** (where the "H" subscript indicates a high value of γ). The threshold from Figure 3, now labeled $\hat{\pi}_{BL}(\theta)$ in Figure 4, is plotted alongside it (where the "L" subscript indicates a low value of γ).[6]

The increasing curvature of the backfire threshold, as γ rises, occurs relative to a fixed point, as illustrated in Figure 4. This fixed point is the same pivot point $\{\overline{\theta},\overline{\pi}\}$ identified in (11) and (12) above. Why? Recall that the equilibrium adoption threshold passes through this point for any value of γ. In the extreme case where $\gamma = 0$, the change in emissions induced by adoption of the new technology is irrelevant to the adoption decision. At any $\gamma > 0$, the change in emissions cannot be irrelevant, and so this change in emissions could have no effect on the adoption threshold if and only if it is exactly zero. A zero change in emissions is the defining property of the backfire threshold, so it must also pass through the pivot point. Moreover, since the adoption threshold passes through this point for any value of γ, it follows that the backfire threshold must also pass through this point for any value of γ.

Note too that the high-γ backfire threshold in Figure 4 reaches an interior turning point, at

(19) $$\hat{\theta}_{peak} = 1 - \frac{m-k}{2\gamma}$$

This means that the backfire threshold can be crossed twice as θ rises. To understand why, again consider a new technology that is no cleaner than the old one ($\theta = 0$) but whose marginal cost of use is lower than the old one, at a value of π between the intercept-points $\hat{\pi}^0_{BH}$ and $\hat{\pi}^0_{BL}$ in Figure 4. In contrast to the low-γ case, emissions in the high-γ case are now lower under the new technology (even though $\theta = 0$) because the *PE* is now outweighed by the *FCE*. Why? The higher value of γ weakens the *PE*

[6] If instead γ is reduced below the value for which Figure 3 is drawn, the backfire threshold gradually loses its curvature, and eventually becomes linear when $\gamma = 0$.

because higher "green guilt" dampens the price-induced increase in consumption; note the role of γ in the second term in (15).

If we now hold π fixed at a value between $\hat{\pi}^0_{BH}$ and $\hat{\pi}^0_{BL}$, and allow θ to rise, the *PE* is unchanged but the *FCE* shrinks, again because the income effect on consumption has an increasingly small effect on emissions as the technology becomes cleaner. The negative *DCE* also kicks in as θ rises but so too does the positive *GCE*, and this latter effect is now stronger than it was in the low-γ case. Moreover, the *DCE* is now initially weaker than it was in the low-γ case because a higher degree of green guilt leads to a relatively low consumption of the dirty good under the old technology. On balance, the emissions cut delivered by the new technology *shrinks* as θ rises, and is eventually reversed where the $\hat{\pi}_{BH}(\theta)$ threshold is crossed, and backfire ensues. This jump in emissions becomes increasingly large until it reaches a maximum at the dashed vertical threshold labeled "peak emissions" in Figure 4. Increasing θ beyond this point starts to bring emissions down again as the *DCE* becomes increasingly dominant, and eventually the changes in emissions turns negative again as the $\hat{\pi}_{BH}(\theta)$ threshold is crossed for the second time. Of course, as $\theta \to 1$, emissions necessarily fall towards zero.

In the cases illustrated in Figures 3 and 4, the backfire threshold lies wholly below the $\pi = 0$ axis. It need not. In particular, if γ is greater than

(20) $$\hat{\gamma}_B = \frac{m + ((2m-k)k)^{\frac{1}{2}}}{2}$$

then the backfire threshold crosses the axis, as illustrated in **Figure 5**, which means that backfire can occur (for middle values of θ) even when the effective price of dirty consumption is higher under the new technology. Thus, it is possible to have "negative rebound" due to a negative *PE*, but still get backfire due to a strong offsetting *GCE*.[7]

[7] At the opposite extreme, we can ask whether or not the backfire region ever shrinks to an empty set. It does not, unless consumption of the dirty good is bound at zero. It is straightforward to show that if consumption is positive, then there must exist a region in which $\theta \in (0,1)$ and $\pi > -p$, and where emissions rise after adoption of the new technology.

The Simple Energy Efficiency Case

Recall from the introduction that much of the analysis on rebound in the existing literature has focused on energy-efficiency improvements. The simplest possible interpretation of that case can be captured in this model by setting $\pi = -\theta p$. This means that the effective price of the dirty good is inversely proportional to its cleanliness. The backfire threshold specified in (17) still applies in this special setting, but there is now a tight restriction on which points in the (θ, π) space can represent feasible technologies. In particular, all possible technologies must now lie on a straight line through the origin with slope $-p$, labeled "EE Case" in **Figure 6** (drawn for the same high-γ scenario depicted in Figure 4). This set of feasible technologies is partitioned by the backfire threshold into those whose adoption will reduce emissions (above the threshold) and those whose adoption will raise emissions (below the threshold).

3.4 Equilibrium Backfire

To this point I have described only the conditions under which backfire occurs if the new technology is adopted. The next step is to identify the conditions under which backfire occurs when households *choose* to adopt the new technology. That is, when does backfire occur in equilibrium?

A graphical approach provides the clearest answer. **Figure 7** overlays the adoption threshold from Figure 1 on the backfire threshold from Figure 3, drawn for the same set of parameter values. Recall that the two thresholds must cross at the pivot point $\{\bar{\theta}, \bar{\pi}\}$ identified in (11) and (12) above. This guarantees that there exist four distinct regions in the (θ, π) space, labeled R1 – R4 in Figure 7.[8]

In region R1 (above both thresholds) households do not adopt the cleaner technology even though doing so would reduce emissions. In region R2 (between the thresholds to the right of $\bar{\theta}$) households do adopt the cleaner technology, and doing so

[8] Recall that the backfire threshold in Figure 3 (and Figure 7) is drawn for $\gamma < \frac{1}{2}\gamma_{\max}$. In the case where $\gamma > \frac{1}{2}\gamma_{\max}$, the same four regions arise, with the same properties. In this particular regard, the size of γ relative to $\frac{1}{2}\gamma_{\max}$ is irrelevant.

reduces emissions; there is no backfire. In region R3 (below both thresholds) households adopt the cleaner technology and emissions rise; there is equilibrium backfire.

In region R4 (between the thresholds to the left of $\bar{\theta}$) emissions would rise if households did adopt the cleaner technology, but they choose not to do so. This region is special relative to the other regions: it shrinks towards empty as *k* falls towards zero. In contrast, the other regions remain non-empty at $k = 0$.

Region 4 is special for another reason too. If a policy-maker were to intervene in this economy and *mandate* the adoption of the cleaner technology because households have not chosen to adopt it voluntarily, then region R4 is the only region in which this mandated adoption would result in backfire. This raises a natural question: could "mandated backfire" ever be welfare improving? That is, would a welfare-maximizing policy-maker ever mandate adoption of the cleaner technology if the economy is in region R4? I now turn to that question, and to the assessment of welfare more generally.

4. WELFARE ANALYSIS

The unpriced externality from emissions in this economy (when $\gamma > 0$) means that equilibrium choices will typically not yield Pareto-efficient outcomes, with respect to either emissions or technology choices. A Pigouvian tax on emissions can in principle correct the distortion with respect to emissions but I will not introduce that policy instrument yet. At this point I wish to characterize the conditions under which the adoption of the cleaner technology is welfare-improving when emissions remain uncorrected at their equilibrium level for whichever technology is in place.

It is straightforward to construct the utility for a representative household under universal adoption of the cleaner technology, and under universal retention of the old one. It is then possible to derive a threshold condition that partitions the (θ, π) space into a region where universal adoption is welfare-improving, and a region where it is welfare-reducing. This critical threshold for "welfare-improving adoption" (WIA) can be expressed instructively as

$$(21) \qquad \tilde{\pi}_A(\theta) = \hat{\pi}_B(\theta) - \left(\frac{\theta - \bar{\theta}}{\gamma - \tilde{\gamma}}\right)\left(\frac{pm(m-k-\gamma(1-\theta))}{(m-\gamma)(2n-1)}\right)$$

where the first term, $\hat{\pi}_B(\theta)$, is the backfire threshold from (17), $\bar{\theta}$ is the pivot-point value from (11), and

(22) $$\tilde{\gamma} = \frac{m}{(2n-1)\delta}$$

is a critical value of the green preference parameter on which a key result will hinge. Note that *n*, the number of households, appears here for the first time in the results because the magnitude of the uncorrected emissions externality rises with *n*.[9]

Consider the properties of this WIA threshold. It is straightforward to show that the second bracketed term in (21) is unambiguously positive when consumption of the dirty good is positive, so we can relate the WIA threshold to the backfire threshold on the basis of the first bracketed term in (21), henceforth called the "hinge term":

(23) $$h(\theta,\gamma) = \frac{\theta - \bar{\theta}}{\gamma - \tilde{\gamma}}$$

The properties of this hinge term determine the curvature of the WIA threshold, and the manner in which it partitions the (θ,π) space relative to the backfire threshold. There are two key cases of interest.

Case 1: $\gamma < \tilde{\gamma}$

It is straightforward to show that in this case, $\tilde{\pi}_A(\theta)$ is strictly convex in θ, and universal adoption is welfare-improving if and only if $\pi < \tilde{\pi}_A(\theta)$. This case is depicted in Figures 8A and 8B, drawn for the same set of parameter values as figure 7. **Figure 8A** highlights the relationship between welfare-improving adoption and the subsequent change in emissions. **Figure 8B** highlights the relationship between welfare-improving adoption and equilibrium adoption choices. Consider each in turn.

Figure 8A depicts the WIA threshold alongside the backfire threshold. Note when $\gamma < \tilde{\gamma}$, the hinge term from (23) is positive for $\theta < \bar{\theta}$, and negative for $\theta > \bar{\theta}$; thus, $\tilde{\pi}_A(\theta) < \hat{\pi}_B(\theta)$ for $\theta < \bar{\theta}$, and $\tilde{\pi}_A(\theta) > \hat{\pi}_B(\theta)$ for $\theta > \bar{\theta}$, as depicted in Figure 8A.

[9] Chan and Gillingham (2015) also account for the size of the population damaged by emissions in their analysis, but these damaged agents are entirely external to the agents making choices in the model. The externality in their model is not a reciprocal one.

Four key points can be gleaned from the figure. First, the thresholds cross at $\theta = \overline{\theta}$. Why? We have already seen that the backfire threshold, $\hat{\pi}_B(\theta)$, and the equilibrium adoption threshold, $\hat{\pi}_A(\theta)$, cross at $\theta = \overline{\theta}$; recall Figure 7. The only difference between the private net benefit from adoption (which underlies $\hat{\pi}_A(\theta)$) and the social net benefit from adoption (which underlies $\tilde{\pi}_A(\theta)$) is the valuation of the change in emissions. If emissions do not change, then there can be no difference between them. It follows that if $\hat{\pi}_B(\theta)$ and $\hat{\pi}_A(\theta)$ cross at $\theta = \overline{\theta}$, as they must, then $\hat{\pi}_B(\theta)$ and $\tilde{\pi}_A(\theta)$ must also cross at $\theta = \overline{\theta}$.

Second, adoption of the cleaner technology is welfare-improving if it is very clean (high θ) and/or has a lower marginal cost of use (low π), but not otherwise. This might seem unsurprising but we will see in the discussion of Case 2 below that this relationship is reversed when γ is large.

Third, there exists a region in which adoption is welfare-improving even though emissions rise after adoption. This is the cone-shaped region in the lower-middle of Figure 8A bordered by $\tilde{\pi}_A(\theta)$ and $\hat{\pi}_B(\theta)$, labeled Q1. Backfire is welfare-improving in this region because the reduction in marginal cost of use under the cleaner technology is large enough to more-than offset the damage from higher emissions.

Fourth, there exists a region in which adoption of the cleaner technology would cause emissions to fall but welfare would nonetheless decline because the overall cost of adoption is too high. This is the cone-shaped region in the upper-left corner of Figure 8A bordered by $\tilde{\pi}_A(\theta)$ and $\hat{\pi}_B(\theta)$, labeled Q2.

Together these last two points highlight the fact that a post-adoption decline in emissions is neither necessary nor sufficient for welfare-improving adoption of a cleaner technology.

A second set of results can be gleaned from Figure 8B. This depicts the WIA threshold alongside both the backfire threshold and the equilibrium adoption threshold. Two key points are noteworthy. First, there exists a region in which adoption of the cleaner technology would be welfare-improving but it is not adopted in equilibrium. This is the cone-shaped region in the upper-right of Figure 8B, bordered by $\tilde{\pi}_A(\theta)$ and $\hat{\pi}_A(\theta)$, labeled Q3. In this region there is a role for policy-intervention. In particular, mandatory-adoption would be welfare-improving here. Note that intervention of this type would necessarily cause emissions to fall in this region; Q3 is entirely above the backfire threshold. This is precisely why the intervention would be welfare-improving. The private calculus that underlies adoption choices in equilibrium does not account for the external effect of any change in emissions that arises from adopting the cleaner technology. This means that the net private benefit of adoption *undervalues* the true net social benefit when adoption causes emissions to fall.

The converse is also true: the net private benefit of adoption *overvalues* the true net social benefit when adoption causes emissions to rise. This private overvaluation creates a region in which adoption occurs in equilibrium even though it is welfare-reducing. This is the cone-shaped region in the lower-left corner of Figure 8B, bordered by $\tilde{\pi}_A(\theta)$ and $\hat{\pi}_A(\theta)$, labeled Q4. This is a region of equilibrium backfire, but the backfire here is welfare-reducing. In this region, policy intervention would require a *prohibition* on adoption of the cleaner technology.

The second noteworthy point to be gleaned from Figure 8B is that there does *not* exist a region in which a welfare-maximizing regulator would over-ride private decisions not to adopt the cleaner technology by mandating adoption if that adoption would cause emissions to rise. That is, "mandated backfire" cannot be welfare-improving. If households have failed to adopt the cleaner technology when doing so would be welfare-improving, it can only be due to the fact that they do not fully value an associated *reduction* in emissions, as arises in region Q3. The increase in emissions that arises in Q4 can never cause households to under-adopt in equilibrium.

Case 2: $\gamma > \tilde{\gamma}$

Recall that Figures 8A and 8B are drawn for the case where $\gamma < \tilde{\gamma}$. If γ rises, the WIA threshold in those figures pivots in a counter-clockwise direction around the pivot point $(\bar{\theta}, \bar{\pi})$. Why? We have also already seen that the equilibrium adoption threshold, $\hat{\pi}_A(\theta)$, pivots around this point as γ rises, and we know that there is no change in emissions at this pivot point because it also lies on the backfire threshold. If there is no change in emissions, then an increase or decrease in γ cannot cause a divergence between the private net benefit from adoption and the social net benefit from adoption. It follows that if $\hat{\pi}_A(\theta)$ pivots around $(\bar{\theta}, \bar{\pi})$, as it must, then so too must $\tilde{\pi}_A(\theta)$. As it does so, it becomes increasingly steep until at $\gamma = \tilde{\gamma}$ it is perfectly vertical. As γ rises further, above $\tilde{\gamma}$, the hinge term in (23) switches sign, and this distinguishes Case 2 from the previous case.

In Case 2, $\tilde{\pi}_A(\theta)$ is strictly concave in θ, and universal adoption is welfare-improving if and only if $\pi > \tilde{\pi}_A(\theta)$. Note that these properties are the mirror-image of those from the previous case. Case 2 is depicted in **Figure 9**, alongside the backfire threshold and the equilibrium adoption threshold drawn for the same set of parameter values, as depicted in Figure 7 earlier.[10] The key difference between this case and the case depicted in Figures 8A and 8B is that the WIA threshold now crosses the backfire threshold from below; the hinge term is now negative for $\theta < \bar{\theta}$, and positive for $\theta > \bar{\theta}$. Thus, $\tilde{\pi}_A(\theta) > \hat{\pi}_B(\theta)$ for $\theta < \bar{\theta}$, and $\tilde{\pi}_A(\theta) < \hat{\pi}_B(\theta)$ for $\theta > \bar{\theta}$. This in turn means that adoption of the cleaner technology is welfare-improving only if its marginal cost of use is not too much *lower* than that of the old technology.

Note that the direction of this result is completely opposite to that from the previous case, where adoption is welfare-improving only if marginal cost is not too much higher. Why? Recall that Case 2 is distinguished by the large size of the green-preference parameter. Recall also that the green-consumer effect (*GCE*) is positive: green preferences tend to make emissions rise after adoption. The larger is γ, the larger is this positive effect, and the greater is the social cost of any increase in emissions. At values of γ above $\tilde{\gamma}$ this effect is large enough that adoption can be welfare-improving only if the *GCE* is more-than-offset by a *PE* that pulls in the <u>opposite</u> direction, as arises when the cleaner technology has a higher marginal cost of use than the old one.

[10] Recall from Section 3 that if $\gamma > \frac{1}{2}\gamma_{\max}$, then $\hat{\pi}_B(\theta)$ has an interior turning point at $\hat{\theta}_{peak}$. Similarly, the WIA threshold has an interior turning point at $\tilde{\theta}_{peak} = 1 - n(m-k)/((2n-1)\gamma)$ if $\gamma > n\gamma_{\max}/(2n-1)$. It is straightforward to show that if this condition on γ is met then $\gamma > \frac{1}{2}\gamma_{\max}$. That is, if $\tilde{\pi}_A(\theta)$ has an interior turning point, then so too does $\hat{\pi}_B(\theta)$. It also straightforward to show that $\tilde{\theta}_{peak} < \hat{\theta}_{peak}$. This very-high-$\gamma$ case is qualitatively identical to the one depicted in Figure 9, and so I do not deal with it separately.

It may at first seem odd that higher marginal costs can be welfare-improving. However, it is important to remember that equilibrium consumption of the dirty good is excessive, due to the uncorrected externality. A higher cost of consumption acts in one respect like a corrective tax here: it reduces the gap between equilibrium consumption and first-best consumption. Moreover, equilibrium consumption always remains higher than first-best, no matter how high cost rises, and so any cost increase will always reduce the gap a little further. If the environmental benefits of that reduction are high enough (when $\gamma > \tilde{\gamma}$) then a cost increase will always increase welfare.

Despite the key difference between Case 1 and Case 2 in terms of where adoption is welfare-improving, the set of possible outcomes is essentially the same in terms of whether or not equilibrium technology choices and welfare-improving technology choices coincide, and whether or not backfire is welfare-improving. In particular, the four regions identified in Figure 9 as Q1 – Q4 correspond to same regions identified in Figures 8A and 8B for Case 1. Similarly, there does *not* exist a region in Figure 9 in which a welfare-maximizing regulator would over-ride private decisions and mandate adoption of the cleaner technology if adoption causes emissions to rise. That is, “mandated backfire” cannot be welfare-improving in either of the two cases.

5. THE FIRST-BEST SOLUTION AND THE PIGOUVIAN TAX

The non-alignment of equilibrium adoption and welfare-improving adoption identified in the previous section is hardly surprising given that the externality from emissions is left uncorrected. In this section I introduce a Pigouvian tax as a policy tool for correcting the emissions externality. I begin by deriving the first-best solution for emissions and technologies, and show that there are conditions under which backfire occurs even in this first-best solution. That is, backfire can be optimal. I then ask whether or not Pigouvian taxes can implement the first-best solution, and show that the answer is “not always”. In particular, there are conditions under which the first-best solution can be achieved only by pairing a Pigouvian tax on emissions with a subsidy – or sometimes, a tax – on adoption of the cleaner technology itself.

5.1 The First-Best Solution

The planning problem is solved in two stages. In the first-stage I derive the optimal consumption levels under each of the two available technologies, and construct the maximized payoff for a representative household in each case. In the second stage I identify the optimal technology according to which technology yields the highest maximized payoff.

It is important to note at the outset that in a more complicated model than the one I am using here, where households are heterogeneous or where the damage is strictly convex in emissions, the first-best solution may not require either universal adoption of the cleaner technology, nor universal retention of the old technology; a mix of technologies could be optimal. In contrast, a mixed optimum of this type cannot arise in a simple setting with homogeneous agents and linear damage. I therefore confine the discussion to a comparison of universal adoption of the cleaner technology, and universal retention of the old technology.

First-Best Consumption

It is straightforward to show that first-best consumption of the dirty good by a representative household under universal use of the old technology is

(24) $$x_0^* = \frac{m - n\gamma}{2p}$$

if $x_0^* > 0$, and zero otherwise. If $x_0^* > 0$ then the associated payoff for a household is

(25) $$v_0^* = \frac{(m - n\gamma)^2}{4p}$$

Note that for any value of $\gamma > 0$, x_0^* goes to zero at sufficiently large but finite *n*.

First-best consumption of the dirty good by a representative household under universal use of the cleaner technology is

(26) $$x_C^* = \frac{m - n\gamma(1-\theta)}{2(p+\pi)}$$

if $x_C^* > 0$, and zero otherwise. If $x_C^* > 0$ then the associated payoff for a household is

(27) $$v_C^* = \frac{(m-k-n\gamma(1-\theta))^2}{4(p+\pi)}$$

Note again that for any value of $\gamma > 0$, x_C^* goes to zero at sufficiently large but finite n unless the technology is perfectly clean ($\theta = 1$). While it is entirely plausible that optimality in practice might require no consumption at all of some dirty goods, I will confine attention here to interior solutions. Accordingly, I place the following restriction on n:

(28) $$n < n_{\max} \equiv \frac{m-k}{\gamma}$$

This restriction guarantees that both x_0^* and x_C^* are positive at any value of $\theta < 1$, and that (25) and (27) are the correct payoffs at the optimum.

The First-Best Technology

It is straightforward to show that $v_C^* > v_0^*$ if and only if $\pi < \pi_A^*(\theta)$, where

(29) $$\pi_A^*(\theta) = \tilde{\pi}_A(\theta) - \gamma^2(n-1)^2\left(\frac{\theta - \overline{\theta}}{\gamma - \tilde{\gamma}}\right)\Psi$$

where first term, $\tilde{\pi}_A(\theta)$, is the WIA threshold from (21), and where

(30) $$\Psi = \frac{pm(m-k-n\gamma+(m-n\gamma)(1-\theta)+n\gamma\theta)}{(2n-1)(m-\gamma)(m-n\gamma)^2}$$

is a strictly positive term when consumption of the dirty good is positive. I will henceforth refer to $\pi_A^*(\theta)$ as the “first-best adoption threshold”. It partitions the (θ, π) space into a region in which universal adoption of the cleaner technology is optimal (below the threshold), and a region in which universal retention of the old technology is optimal (above the threshold).

There are two key points to note about $\pi_A^*(\theta)$. First, if $\gamma = 0$ or $n = 1$ then there is no externality in need of correction, and so $\pi_A^*(\theta) = \tilde{\pi}_A(\theta)$. Moreover, both thresholds coincide with the equilibrium adoption threshold, $\hat{\pi}_A(\theta)$, in that case.

Second, when $\gamma > 0$ and $n > 1$, the two thresholds intersect at the now-familiar pivot point from (11). Why? The only underlying difference between the WIA threshold and the optimal adoption threshold is the change in emissions, and this is necessarily zero at the pivot point because the backfire threshold also passes through that point.

Third, the relationship between $\pi^*_A(\theta)$ and $\tilde{\pi}_A(\theta)$ depends on sign of the "hinge term" identified earlier in (23). This means that there are again two cases of interest, relating to the size the of the preference parameter.

Case 1: $\gamma < \tilde{\gamma}$

The two thresholds in this case are depicted in **Figure 10**, alongside the backfire threshold from Figures 8A and 8B, drawn for the same parameter values (and with *n* chosen to ensure that (28) is satisfied). Figure 10 highlights two regions of interest. In the region labeled F1, bordered by $\pi^*_A(\theta)$ and $\tilde{\pi}_A(\theta)$, adoption of the cleaner technology is not first-best but it is welfare-improving if emissions are determined by *uncorrected* equilibrium choices. Note that in this region, emissions fall after adoption in the uncorrected setting, and this is precisely why it is welfare-improving in that setting but not first-best. In the first-best solution, emissions are significantly lower than in the uncorrected equilibrium, under both technologies, so the change in emissions after adoption is much smaller, and so has much smaller benefit relative to the cost of adoption.

The second region of interest in Figure 10 is labeled F2, bordered by $\pi^*_A(\theta)$ and $\tilde{\pi}_A(\theta)$. In this region, adoption of the cleaner technology is first-best but it is not welfare-improving in a setting where emissions are determined by uncorrected equilibrium choices. In this region, emissions rise after adoption in the uncorrected setting, thereby exacerbating the excessive-emissions problem, and this is why adoption is not welfare-improving in that setting but it is first-best when the externality is corrected because emissions are much lower in that setting.

Case 2: $\gamma > \tilde{\gamma}$

As γ rises, the optimal-adoption threshold pivots counter-clockwise around the pivot point. Its key properties remain unchanged but its relationship to the WIA threshold changes because that threshold switches from being convex to concave as γ rises past $\tilde{\gamma}$, as discussed in Section 4 (in relation to Figure 9). The case where $\gamma > \tilde{\gamma}$ is illustrated in **Figure 11**. The optimal-adoption threshold is plotted alongside the WIA threshold and the backfire threshold from Figure 9. Again, there arise two regions where the optimal-adoption rule yields an outcome different from that under the welfare-improving adoption rule in the setting where the emissions externality is uncorrected. These are labeled F1 and F2 in Figure 11, and they correspond to same regions from Figure 10. The explanation for the divergence in outcomes is exactly the same as in that previous case.

First-Best Backfire

When does optimal adoption of the cleaner technology lead to higher optimal emissions? It is straightforward to show that emissions in the first-best solution are higher after adoption of the cleaner technology if and only if $\pi < \pi_B^*(\theta)$, where

(31) $$\pi_B^*(\theta) = \pi_A^*(\theta) + (\theta - \bar{\theta})\left(\frac{pm(m - k - n\gamma(1-\theta))}{(m - n\gamma)^2}\right)$$

I will henceforth to this expression as the “optimal-backfire threshold”. The second bracketed term in (31) is unambiguously positive when consumption of the dirty good is positive, so we can relate the optimal-backfire threshold to the optimal-adoption threshold in terms of θ relative to the pivot point, $\bar{\theta}$. This relationship is illustrated in **Figure 12**.

Figure 12 is drawn for the case where $n > \frac{1}{2}n_{\max}$, which means that the optimal-backfire threshold has an interior turning point. In the case where $n < \frac{1}{2}n_{\max}$, the threshold is negatively-sloped for all $\theta > 0$. In both cases there exists a region in which backfire is optimal. This region is labeled F3 in Figure 12, bordered by $\pi_A^*(\theta)$ and $\pi_B^*(\theta)$. In this region, the cleaner technology is the first-best technology choice, and adoption of that technology causes first-best emissions to rise. There is first-best backfire.

It is worth noting that first-best backfire can arise even when there is no reduction in the marginal cost of use under the new technology; green preferences alone can induce it. In particular, it can be shown that if $k < m/5$ and $\gamma > \gamma_B^*$, where

(32) $$\gamma_B^* = \frac{m + (2m - k^2)^{\frac{1}{2}}}{2n}$$

then there is a range of θ over which backfire occurs even when $\pi \geq 0$, as illustrated in **Figure 13**.

The first-best backfire threshold can be related in a straightforward way to the backfire threshold from Section 3, which I will henceforth call the "uncorrected backfire threshold" to make the distinction clear. In particular,

$$\text{(33)} \qquad \pi_B^*(\theta) = \hat{\pi}_B(\theta) + (\theta - \bar{\theta})\left(\frac{pm(m-k-n\gamma(1-\theta))}{(m-n\gamma)^2}\right)$$

The second bracketed terms is positive if consumption of the dirty good is positive, so $\pi_B^*(\theta) < \hat{\pi}_B(\theta)$ for $\theta < \bar{\theta}$, and $\pi_B^*(\theta) > \hat{\pi}_B(\theta)$ for $\theta > \bar{\theta}$. This relationship is illustrated in **Figure 14**, which overlays the uncorrected backfire threshold from Figure 11 on the first-best thresholds from Figure 12. Note that there exists a region, bordered by $\pi_B^*(\theta)$ and $\hat{\pi}_B(\theta)$, and labeled B1 in Figure 14, in which backfire could not occur in the uncorrected equilibrium but does occur in the first-best solution. Conversely, there exists a region, bordered by $\pi_B^*(\theta)$ and $\hat{\pi}_B(\theta)$, and labeled B2 in Figure 14, in which backfire can occur in the uncorrected equilibrium but does not occur in the first-best solution.

5.2 Implementation via Pigouvian Taxes

I now turn to the question of whether or not a corrective tax on emissions can implement the first-best outcome. I begin by deriving the Pigouvian taxes under each technology, and then ask whether or not those taxes are consistent with the corrected-equilibrium adoption choices they induce.

An Equilibrium with Universal Retention of the Old Technology

Suppose that all households use the old technology and face tax rate t_0 on emissions. All revenue raised by the tax is returned to households as a lump-sum. Each household therefore faces the following budget constraint:

$$\text{(34)} \qquad y + px + t_0 x = m + \frac{T_0}{n}$$

where $T_0 = t_0 X_0$, and X_0 is aggregate consumption of the dirty good.[11] It is straightforward to show that the symmetric pure-strategy equilibrium level of consumption for each household is

$$\hat{x}_0(t_0) = \frac{n(m-\gamma)}{2np + (n-1)t_0} \tag{35}$$

We can now ask what value of t_0 will implement the first-best level of consumption identified in (24). This Pigouvian tax, matched to universal use of the old technology, is given by

$$t_0^* = \frac{2np\gamma}{m - n\gamma} \tag{36}$$

We now need to determine whether or not this tax rate actually induces all households to retain the old technology. If not, the Pigouvian tax based on the old technology is not consistent with the equilibrium it induces. To answer this question, we need to compare the candidate equilibrium payoff – when use of the old technology is universal – with the payoff to a household that unilaterally deviates from the candidate equilibrium, and adopts the cleaner technology.

It is straightforward to show that the candidate equilibrium payoff is simply equal to v_0^* from (25) above because the first-best solution and the corrected-equilibrium level of consumption necessarily coincide under the Pigouvian tax. To calculate the payoff from unilateral adoption of the cleaner technology, we first need to determine the privately-optimal out-of-equilibrium consumption choice for a household that adopts the cleaner technology while facing t_0^*. We can then find the adoption payoff at this optimal consumption level.

Both the privately-optimal out-of-equilibrium consumption level and the associated adoption payoff are too messy to report here usefully. However, it straightforward to show that there exists a critical value of π, denoted $\hat{\pi}_0(\theta, t_0^*)$, such that unilateral adoption of the cleaner technology yields a lower payoff than the candidate

[11] The inclusion of the refunded tax revenue now means that there are no dominant strategies in the game. In the limit as $n \to \infty$, each household would see T as independent of its own consumption choice, and in that limiting case there would again be dominant strategies. However, one cannot impose this limiting case here and focus on the associated dominant strategy equilibrium because we know that optimal consumption of the dirty good falls to zero at finite *n*, so the limiting case is in fact a corner solution.

equilibrium payoff if and only if $\pi > \hat{\pi}_0(\theta, t_0^*)$. This threshold is illustrated in **Figure 15** in (θ, π) space, drawn for the same parameter values that underlie Figure 14. It is henceforth called the "universal retention threshold". It partitions the space into two regions: universal retention of the old technology is a corrected equilibrium in the region above this threshold; universal retention of the old technology is not a corrected equilibrium in the region below this threshold. The other threshold in this figure is derived next.

An Equilibrium with Universal Adoption of the Cleaner Technology

Now suppose all households adopt the cleaner technology and face tax rate t_C on emissions. Each household now faces the following budget constraint:

(37) $$y + (p + \pi)x + k + t_C(1-\theta)x = m + \frac{T_C}{n}$$

where $T_C = t_C(1-\theta)X_C$, and X_C is aggregate consumption of the dirty good. It is straightforward to show that the symmetric pure-strategy equilibrium level of consumption for each household is

(38) $$\hat{x}_C(t_C) = \frac{n(m-k-\gamma(1-\theta))}{2n(p+\pi)+(n-1)(1-\theta)t_C}$$

We can now ask what value of t_C will implement the first-best consumption level identified in (26). This Pigouvian tax, matched to universal use of the cleaner technology, is given by

(39) $$t_C^* = \frac{2n(p+\pi)\gamma}{m-k-n\gamma(1-\theta)}$$

We now need to determine whether or not this tax rate actually induces all households to adopt the cleaner technology. If not, then the Pigouvian tax under the cleaner technology is not consistent with the equilibrium it induces. To answer this question, we need to compare the candidate equilibrium payoff – when use of the cleaner technology is universal – with the payoff to a household that unilaterally deviates from the candidate equilibrium, and retains the old technology.

It is straightforward to show that the candidate equilibrium payoff is simply equal to v_C^* from (27) above, because the first-best solution and the corrected-equilibrium level of consumption necessarily coincide under the Pigouvian tax. To calculate the payoff from unilateral retention of the old technology, we first need to determine the privately-optimal out-of-equilibrium consumption choice for a household that retains the old technology while facing t_C^*. We can then find the retention payoff at this consumption level.

Again, both the privately-optimal out-of-equilibrium consumption level and the associated retention payoff are too messy to report here usefully. However, it straightforward to show that there exists a critical value of π, denoted $\hat{\pi}_C(\theta, t_C^*)$, such that unilateral retention of the old technology yields a lower payoff than the candidate equilibrium payoff if and only if $\pi < \hat{\pi}_C(\theta, t_C^*)$. This threshold is illustrated in **Figure 15** in (θ, π) space, drawn for the same parameter values that underlie Figure 14, and is henceforth called the "universal adoption threshold". It partitions the space into two regions: universal adoption of the cleaner technology is an equilibrium in the region below this threshold; universal adoption of the cleaner technology is not an equilibrium in the region above this threshold.

An Equilibrium with Partial Adoption of the Cleaner Technology

It is clear from Figure 15 that there exists a region, between the two thresholds labeled PA, in which neither universal adoption of the cleaner technology, nor universal retention of the old technology, is an equilibrium. In this region there exists only an equilibrium with partial adoption, in which a fraction $\psi(t) \in (0,1)$ of households adopt the cleaner technology, and where this fraction is an increasing function of the tax rate in place.[12] Partial adoption is not a first-best solution in this economy, so a supplementary policy instrument will be needed in this PA region. To characterize that instrument, we first need to relate this region of partial adoption to the first-best adoption threshold.

[12] There is an analytical solution for $\psi(t)$ but it is extremely complicated, and reporting it here serves no purpose.

The Tax-Induced Equilibrium vs. the First-Best Solution

Figure 16 overlays the first-best adoption threshold on the universal adoption threshold from Figure 15. It is straightforward to show that there always exists a non-empty range $\theta \in (\theta_L^C, \theta_H^C)$ – as illustrated in Figure 16 – over which the universal adoption threshold lies strictly below the first-best adoption threshold. In this lens-shaped region between the two thresholds in Figure 16, the first-best solution is universal adoption but this is not an equilibrium. In particular, when the tax rate on emissions is set to match the first-best technology (t_C^* in this case), that tax rate does not create the incentive needed to induce all households to adopt the cleaner technology; some households retain the old technology. In this region there is *under-adoption* of the cleaner technology relative to first-best.

Under-adoption is not the only possibility. **Figure 17** overlays the first-best adoption threshold on the universal retention threshold from Figure 15. It is straightforward to show that always exist a non-empty range $\theta \in (\theta_L^0, \theta_H^0)$ – as illustrated in Figure 17 – over which the universal retention threshold lies strictly above the first-best threshold. In this lens-shaped region between the two thresholds in Figure 17, the first-best solution is universal retention of the old technology but this is not an equilibrium. In particular, when the tax on emissions is set to match the first-best technology (t_0^* in this case), that tax does not create the incentive needed to induce all households to retain the old technology; some households adopt the cleaner technology. In this region there is *over-adoption* of the new technology relative to first-best.[13]

5.3 Supplementary Policy

There is one final question to be answered here: can the first-best solution be implemented if the Pigouvian tax is used in concert with a supplementary policy that targets the incentive to adopt directly? In particular, suppose households are paid a fixed

[13] There is one qualification needed here. Consider the special case where $\pi = 0$ and $\theta = \theta^+$, where $\theta^+ = k/n\gamma$ is the value of θ at which $\pi_A^*(\theta) = 0$. In this case, $t_0^* = t_C^*$, and welfare is exactly the same under universal adoption of the cleaner technology and universal retention of the old one, and at any level of partial adoption in between those extremes. In this very special case – which corresponds to a single point in the parameter space, where the optimal-adoption threshold crosses the axis in Figures 16 and 17 – an equilibrium with partial adoption of the cleaner technology is first-best.

subsidy if they adopt the cleaner technology, where the aggregate funding for this subsidy is deducted from the aggregate tax revenue returned to households. How would that subsidy be set?

In the under-adoption region in Figure 16, a subsidy $s^*(\theta) > 0$ can be paid to households who adopt the cleaner technology to ensure that all households adopt that technology when facing tax rate t_C^* on emissions. The analytical expression for this optimal subsidy is too cumbersome to report usefully here but its key properties are illustrated in the upper panel of **Figure 18**. First, it is zero-valued for $\theta \notin (\theta_L^C, \theta_H^C)$ because we know (from Figure 16) that no subsidy is needed to achieve first-best in that region. Second, its value is independent of π, but the range of θ over which it should be applied is <u>not</u> independent of π. For example, Figure 18 highlights the range of application when $\pi = 0$. This range corresponds to that part of the horizontal axis in Figure 16 that lies between the two thresholds. At higher values of π this range shifts to the right; at lower values of π this range shifts to the left.

A subsidy can also be used to correct for over-adoption of the cleaner technology but in that case the "subsidy" must be negative. In particular, in the over-adoption region from Figure 17, a tax $\tau^*(\theta) > 0$ must be charged on the cleaner technology to ensure that all households retain the old technology when facing tax rate t_0^* on emissions. Again, the analytical expression for this optimal tax is too cumbersome to report usefully here but its key properties are illustrated in the lower panel of Figure 18, which depicts the negative of $\tau^*(\theta)$. First, it is zero-valued for $\theta \notin (\theta_L^0, \theta_H^0)$ because we know (from Figure 17) that no subsidy is needed to achieve first-best in that range. Second, its value is independent of π, but the range of θ over which it should be applied is *not* independent of π. For example, Figure 18 highlights the range of application when $\pi = 0$. This range corresponds to that part of the horizontal axis in Figure 17 that lies between the two thresholds. (Note that the upper bound of this range must correspond to the lower bound of the range in the upper panel of Figure 18 where the subsidy should be positive). At higher values of π the range of θ over which the adoption tax should be applied shifts to the right; at lower values of π that range shifts to the left.

While these supplementary policies are easily derived in the context of this model, using them in practice would be very challenging. In particular, there is a very fine line between the conditions under which a subsidy should be used and the conditions under which a tax should be used because the policy problem involves a discrete choice between two very different outcomes. In a more realistic setting with heterogeneous agents the problem would loose that discreteness because first-best would typically involve non-universal adoption, but in that setting the policy problem acquires a new complication: the correct Pigouvian taxes are specific to individual households. In particular, the green preference parameter for any given household shows up in the tax rate that this household should face. Implementing such taxes would be almost impossible.

These practical issues suggest that the best approach to the policy problem may instead be to identify a set of simple policies that a regulator could reasonably use in practice, and then assess how these policies are likely to perform, given the underlying complexity of the setting in which they are used. A reasonable conjecture might be that highly discrete policies – like mandatory adoption policies, or large technology-adoption subsidies – have greater potential for being very wrong than do more continuous policies like an emissions tax. I leave further consideration of that conjecture to future work.

6. CONCLUSION

I have examined a setting in which households with green preferences choose between two available technologies on the basis of their costs, and on the basis of their associated emissions. These green preferences in turn give rise to a reciprocal externality among the households, the correction of which requires policy intervention. I have shown that in the absence of corrective policy, the adoption of a cleaner technology can be welfare-improving even when it induces backfire. More generally, a reduction in emissions is neither necessary nor sufficient for the equilibrium adoption of a cleaner technology to be welfare-improving. I have also shown that mandated adoption of a cleaner technology – when households have otherwise chosen not to adopt it – is never welfare-improving if it induces backfire. Conversely, there exist conditions under which equilibrium adoption of the cleaner technology is welfare-reducing, due to the associated backfire. Under these conditions, policy intervention calls for a prohibition on adoption of the cleaner technology.

I have also characterized the first-best solution in this setting, and shown that this solution can also involve higher emissions under the optimally-adopted cleaner technology; that is, backfire can be optimal. Problematically, implementation of the first-best solution may not be possible with a corrective tax on emissions because the right Pigouvian tax can induce the wrong technology choice in the corrected equilibrium. In those circumstances, a finely-tuned supplementary policy involving a subsidy or tax on the technology itself is in principle needed to achieve first-best.

REFERENCES

Allan, Grant, Nick **Hanley**, Peter **McGregor**, Kim **Swales**, and Karen **Turner** (2007), The Impact of Increased Efficiency in the Industrial Use of Energy: A Computable General Equilibrium Analysis for the United Kingdom, *Energy Economics*, 29(4), 779-798.

Barker, Terry, Paul **Ekins**, and Timothy **Foxon** (2007), The Macroeconomic Rebound Effect and the UK Economy, *Energy Policy*, 35 (10), 4935-4946.

Barker, Terry, Athanasios **Dagoumas**, and Jonathan **Rubin** (2009), The Macroeconomic Rebound Rffect and the World Economy, *Energy Policy*, 2 (4): 411-427.

Berkhout, Peter H.G., Jos C. **Muskens**, and Jan W. **Velthuijsen** (2000), Defining the Rebound Effect, *Energy Policy*, 28(6-7), 425-432.

Binswanger, Mathias (2001), Technological Progress and Sustainable Development: What about the Rebound Effect?, *Ecological Economics*, 36(1), 119-132.

Borenstein, Severin (2015), A Microeconomic Framework for Evaluating Energy Efficiency Rebound and Some Implications, *Energy Journal*, 36 (1), 1-21.

Chan, Nathan W., and Kenneth **Gillingham** (2015), The Microeconomic Theory of the Rebound Effect and its Welfare Implications. *Journal of the Association of Environmental and Resource Economists*, 2(1), 133-159.

Chang, Juin-Jen, Wei-Neng **Wang**, Jhy-Yuan **Shieh** (2018), Environmental Rebounds/Backfires: Macroeconomic Implications for the Promotion of Environmentally-Friendly Products, *Journal of the Environmental Economics and Management*, 88, 35-68.

Fullerton, Don and Chi L. **Tan** (2019), Costs of Energy Efficiency Mandates Can Reverse the Sign of Rebound, *https://works.bepress.com/don_fullerton/81*

Gillingham, Kenneth, David **Rapson**, and Gernot **Wagner** (2016), The Rebound Effect and Energy Efficiency Policy, *Review of Environmental Economics and Policy*, 10 (1), 68-88.

Greening, Lorna A., David L. **Greene**, and Carmen **Difiglio** (2000), Energy Efficiency and Consumption – the Rebound Effect – a Survey, *Energy Policy*, 28 (6-7), 389-401.

Mizobuchi, Kenichi (2000), An Empirical Study on the Rebound Effect Considering Capital Costs, *Energy Economics*, 30(5), 2486-2517.

Nässén, Jonas, and John **Holmberg** (2009), Quantifying the Rebound Effects of Energy Efficiency Improvements and Energy Conserving Behavior in Sweden, *Energy Efficiency*, 2(3), 221-231.

Sorrell, Steve, John **Dimitropoulos**, and Matt **Sommerville**, (2009), Empirical Estimates of the Direct Rebound Effect: A Review, *Energy Policy*, 37 (4), 1356-1371.

Thomas, Brinda A. and Inês L. **Azevedo**, (2013), Estimating Direct and Indirect Rebound Effects for U.S. Households with Input-Output Analysis Part 1: Theoretical Framework, *Ecological Economics*, 86, 199-210.

Turner, Karen (2009), Negative Rebound and Disinvestment Effects in Response to an Improvement in the UK Economy, *Energy Economics*, 31(5), 648-666.

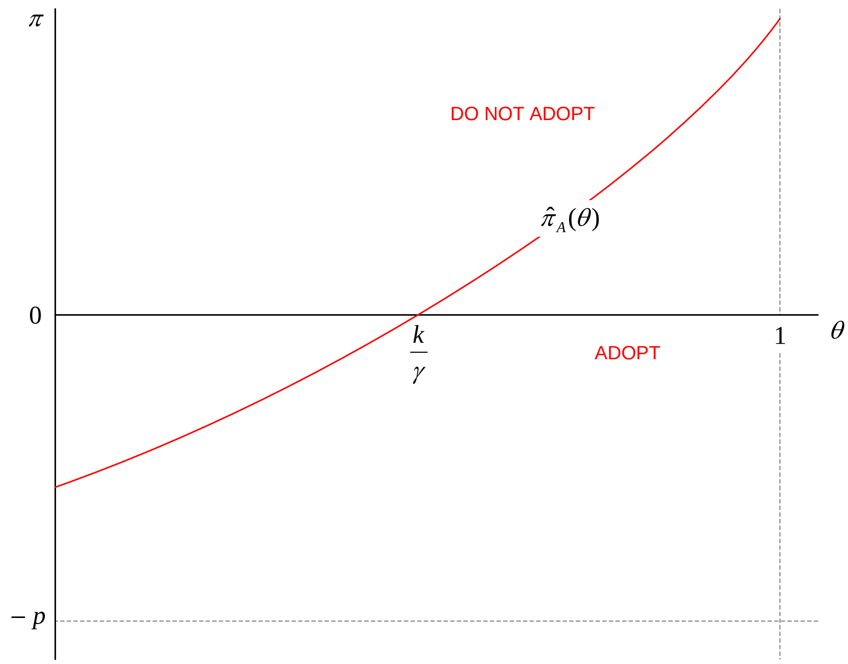


FIGURE 1

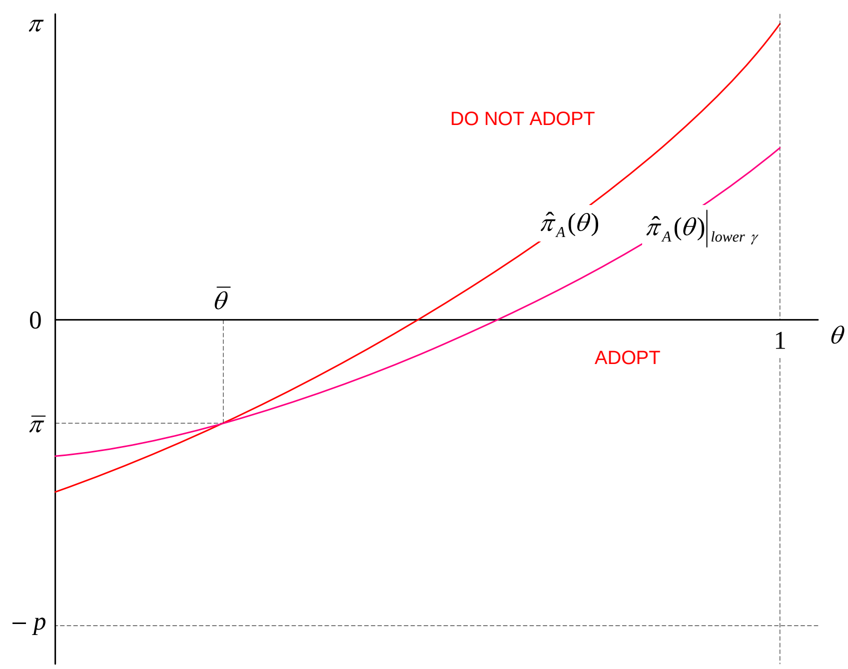


FIGURE 2

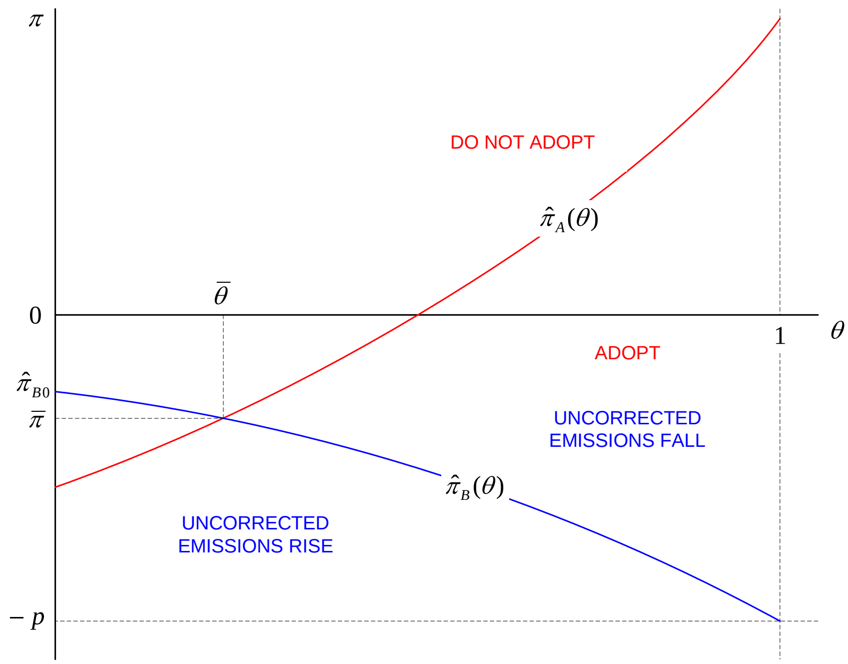


FIGURE 3

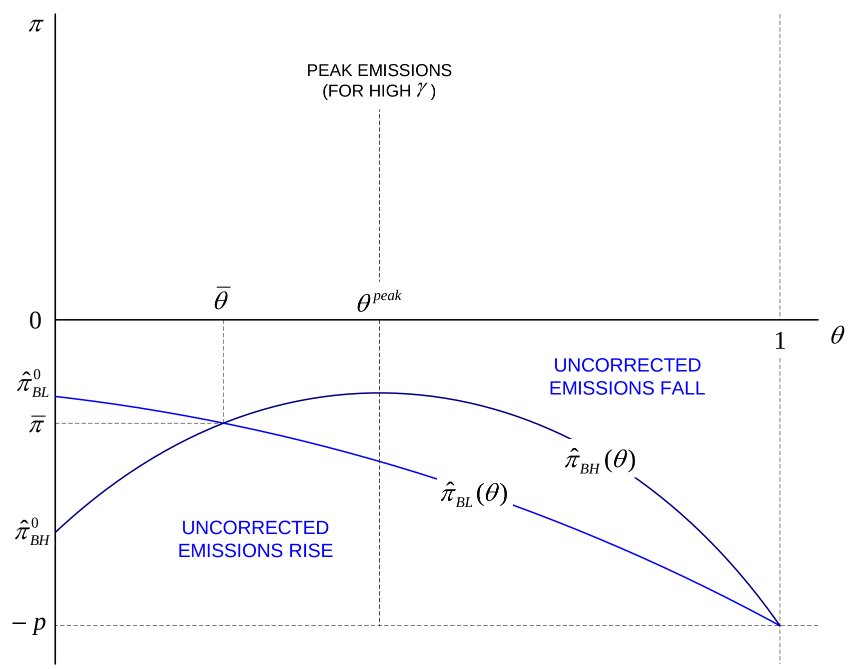


FIGURE 4

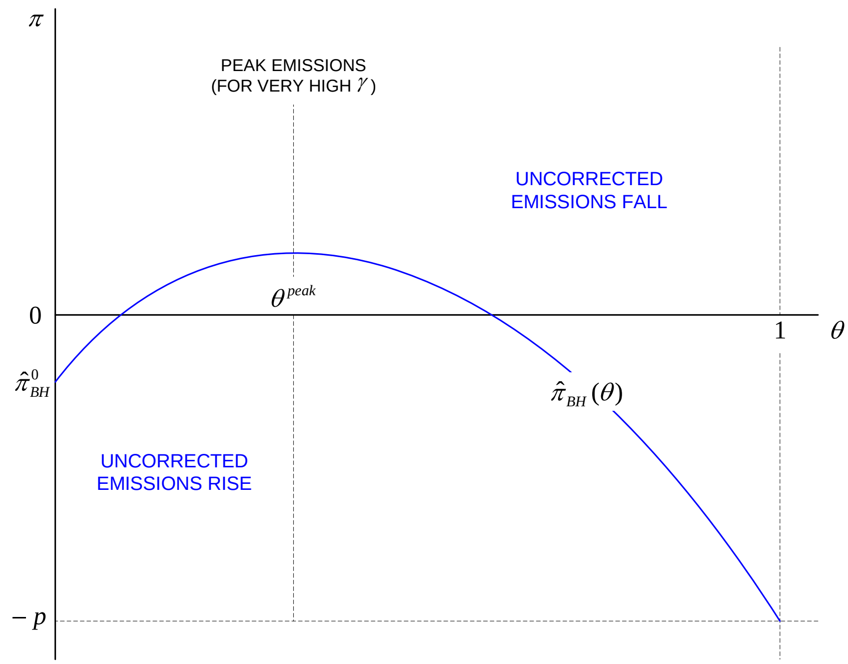


FIGURE 5

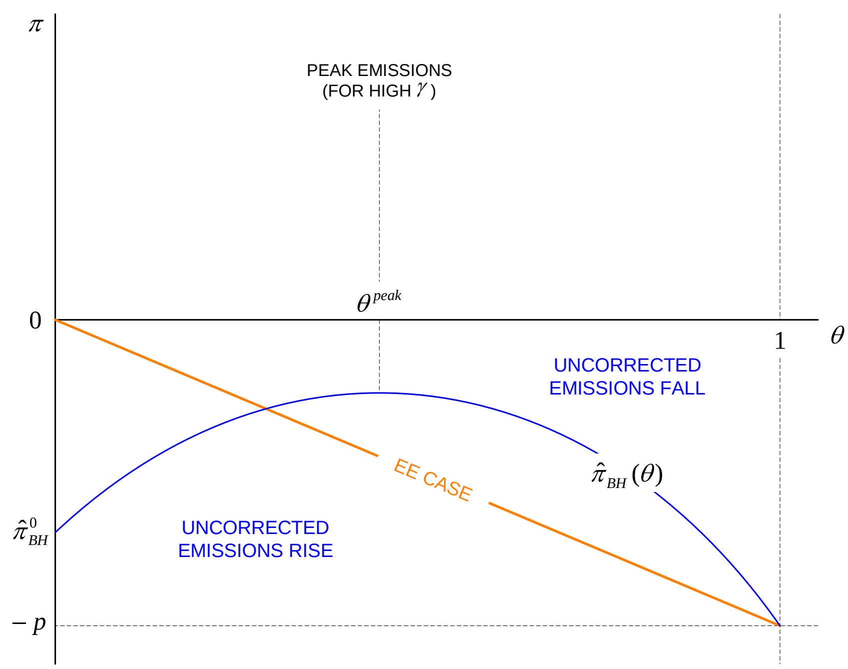


FIGURE 6

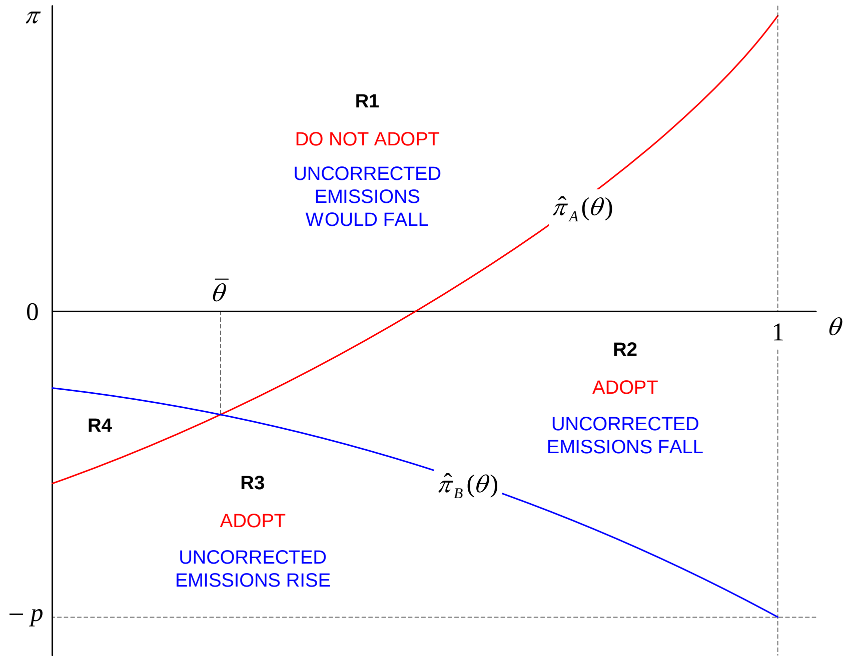


FIGURE 7

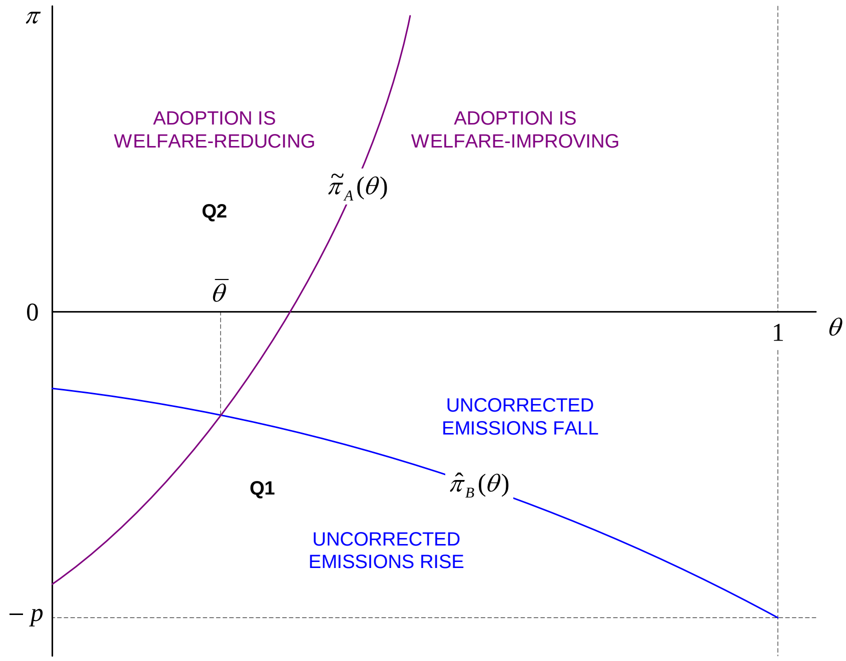


FIGURE 8A

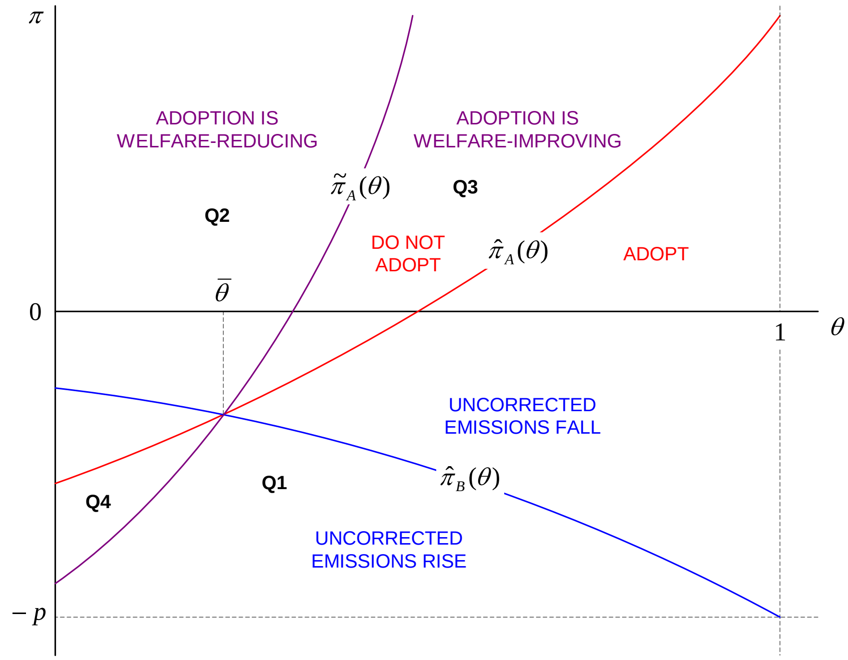


FIGURE 8B

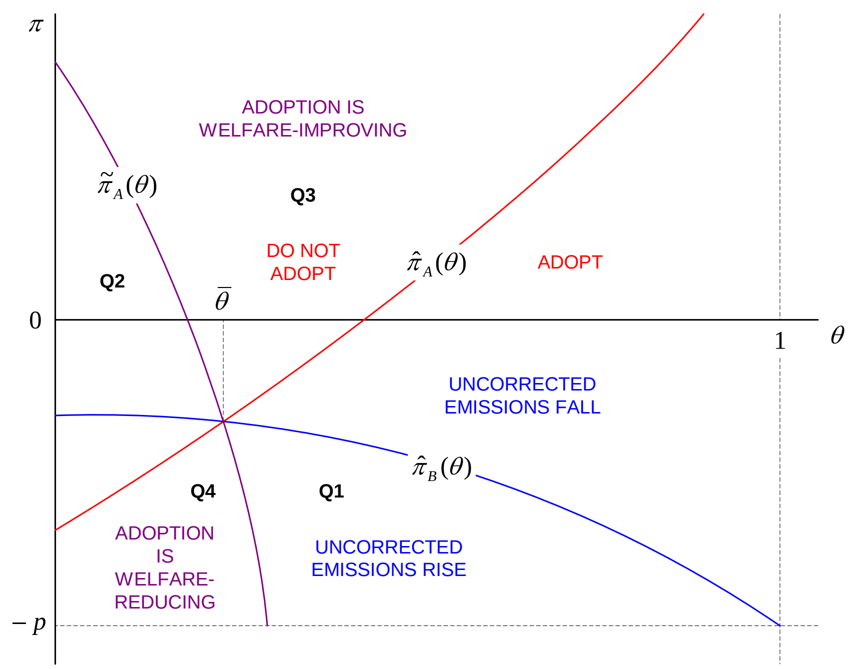


FIGURE 9

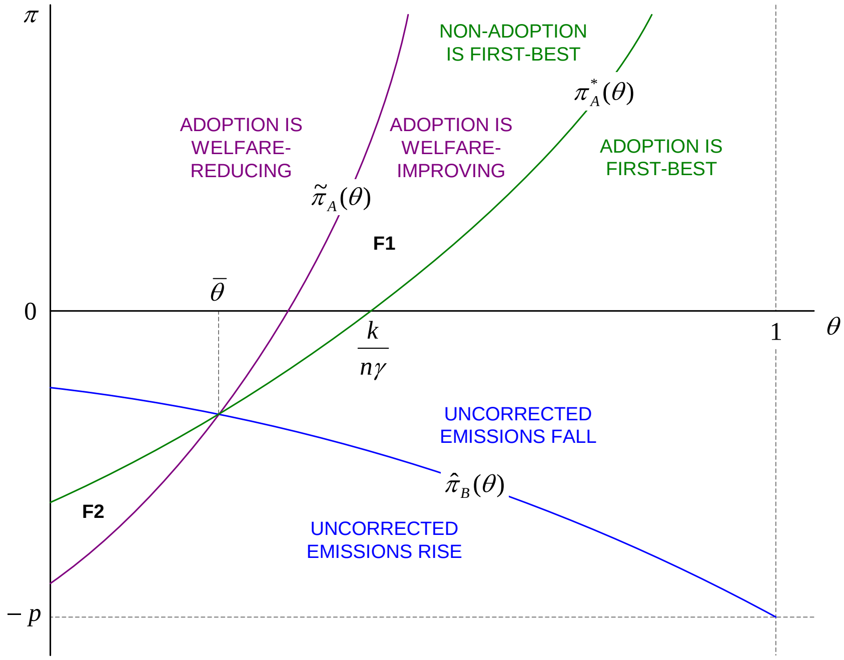


FIGURE 10

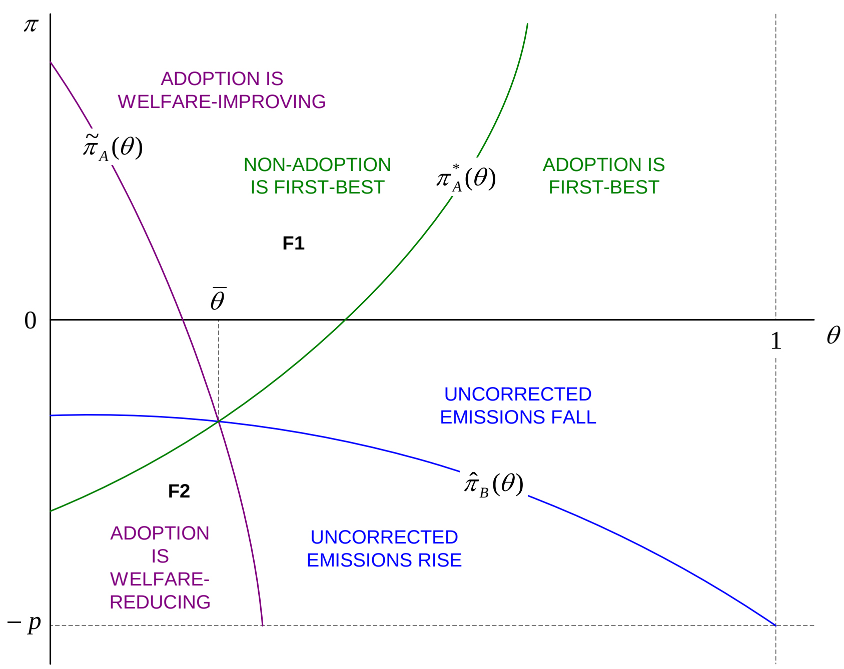


FIGURE 11

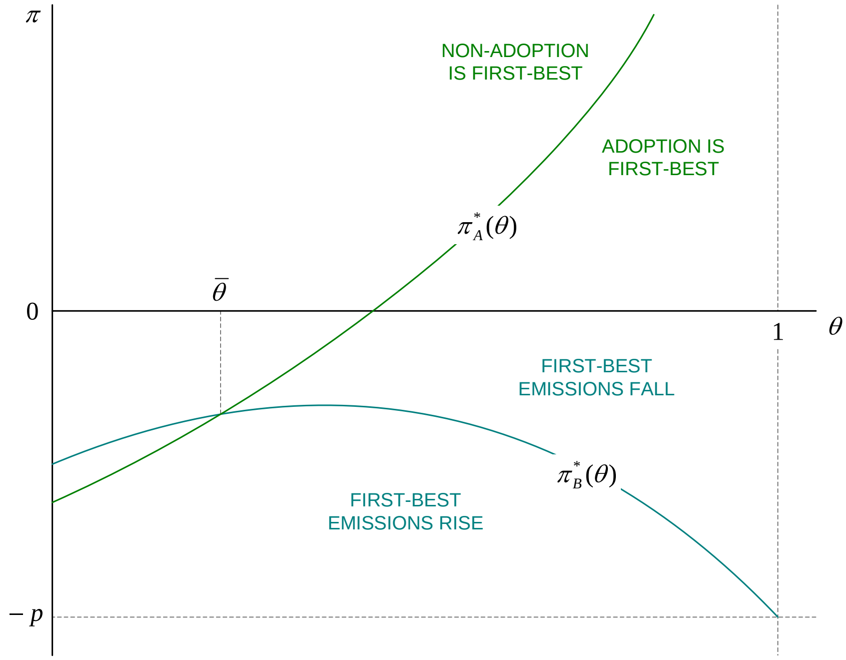


FIGURE 12

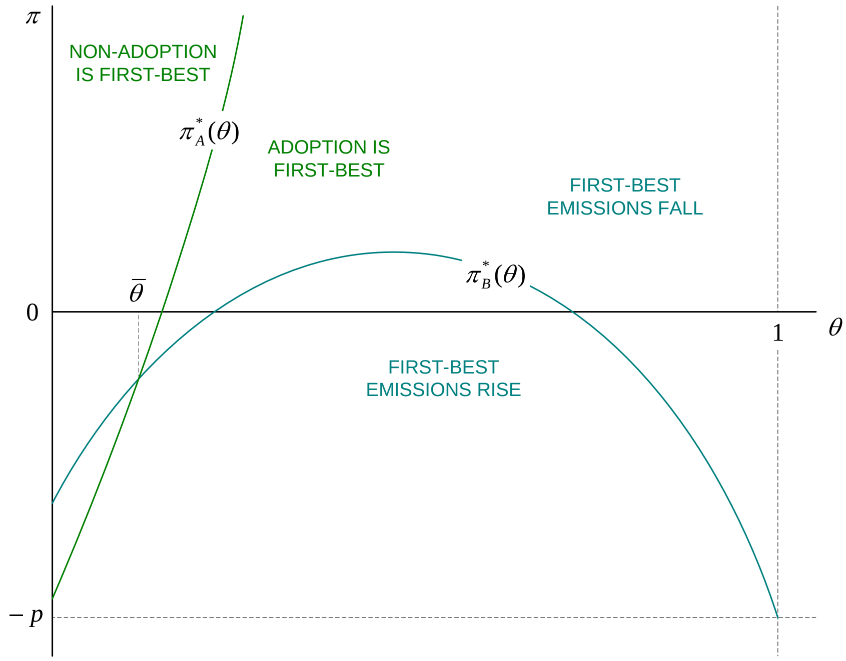


FIGURE 13

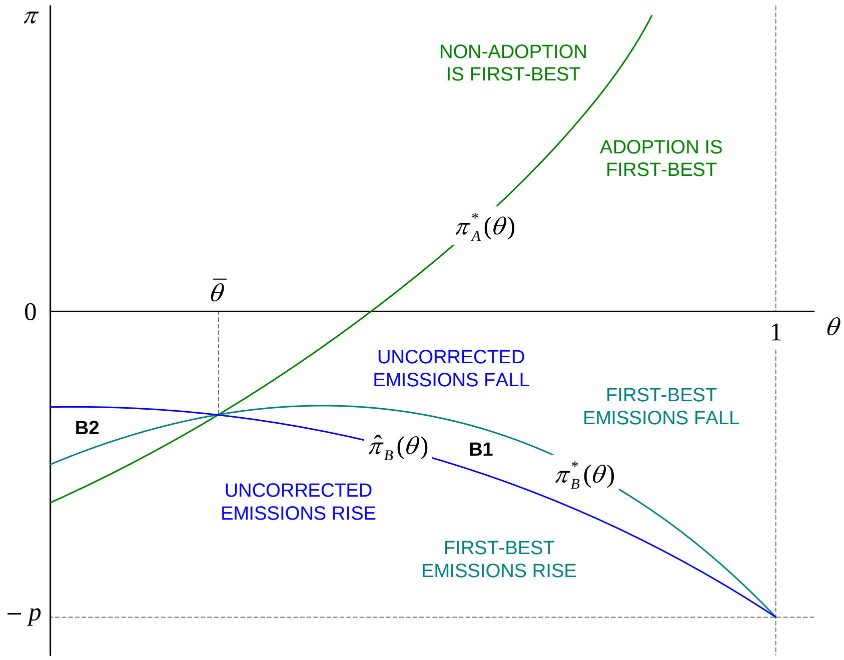


FIGURE 14

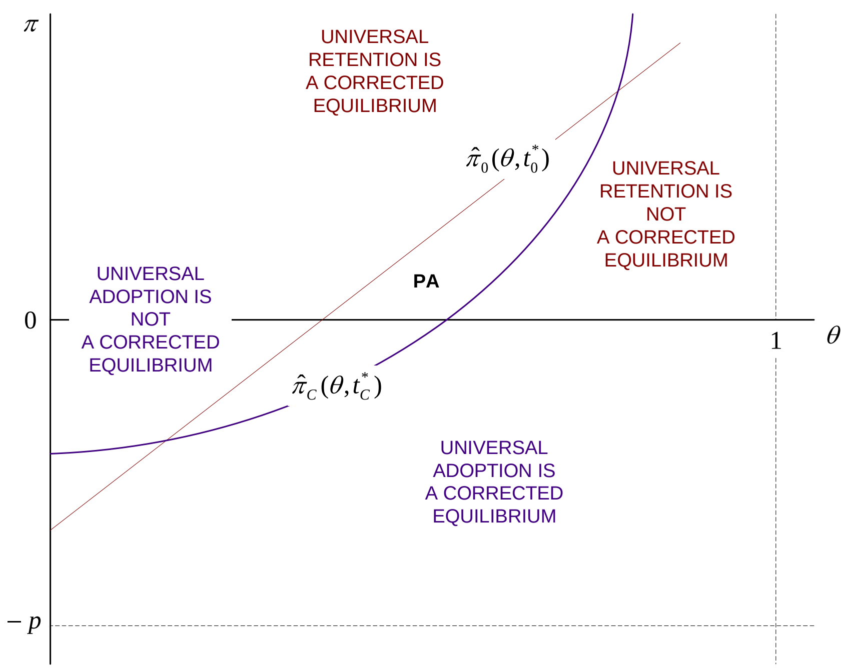


FIGURE 15

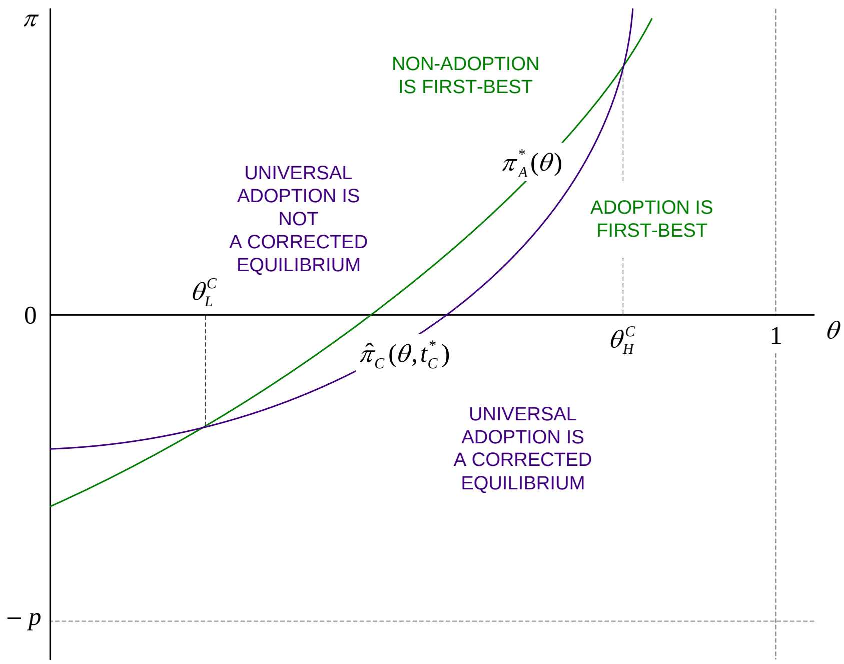


FIGURE 16

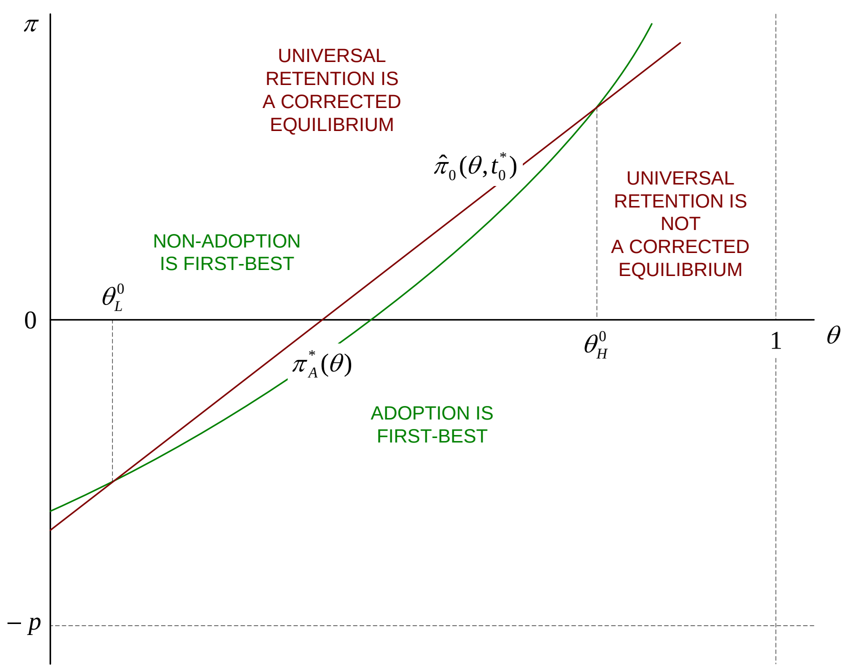
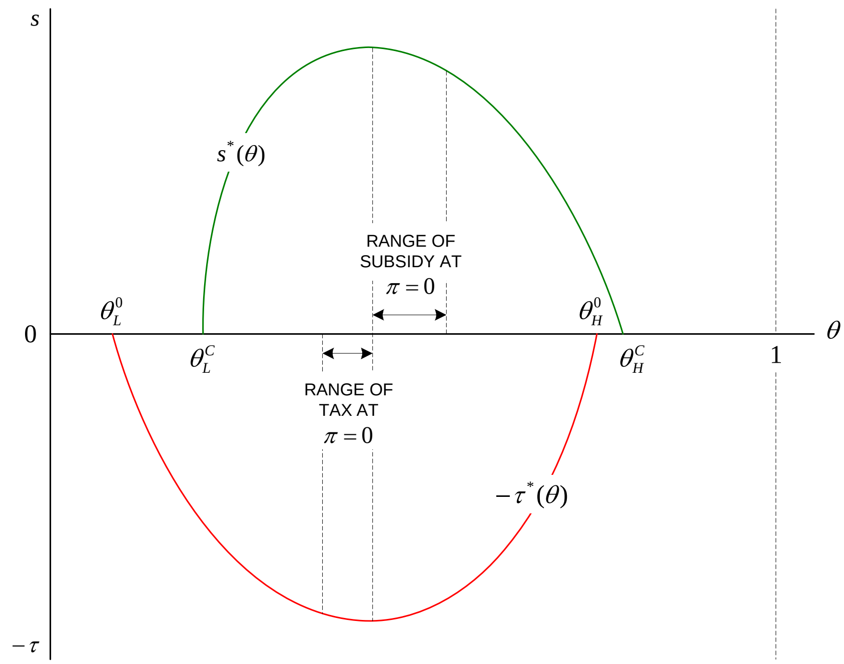


FIGURE 17



s
$s^*(\theta)$
RANGE OF
SUBSIDY AT
$\pi = 0$
θ_L^0
θ_H^0
0
θ
θ_L^C
θ_H^C
1
RANGE OF
TAX AT
$\pi = 0$
$-\tau^*(\theta)$
$-\tau$

FIGURE 18