Kiel

Working Papers

Kiel Institute
for the World Economy

Expectation-Driven Climate Treaties with Breakthrough Technologies

by Daiju Narita, Ulrich J. Wagner

No. 1732 | September 2011

Web: www.ifw-kiel.de

Kiel Working Paper No. 1732 | September 2011

Expectation-Driven Climate Treaties with Breakthrough Technologies*

Daiju Narita, Ulrich J. Wagner

Abstract: Recent research has shown that agreements centered on the adoption of breakthrough technologies can break the deadlock in international climate negotiations if the mitigation technology exhibits a network externality that transforms full cooperation into a self-enforcing outcome. This paper shows that the same externality also increases strategic uncertainty about future technology adoption, which makes coordination on the cooperative outcome more demanding. We analyze this coordination problem in a dynamic game of technology adoption with convex switching costs. We find that the adoption dynamics for some technologies depend exclusively on countries' expectations about future adoption. This possibility calls for a more prominent role of expectation management in climate policy. Another implication is that the choice of breakthrough technologies should be guided not only by economic efficiency and strategic adoption incentives but also by the amount of strategic uncertainty they engender.

Keywords: International environmental agreements (IEAs), climate policy, technology choice, expectations, multiple equilibria

JEL classification: Q54, O33, H87

Daiju Narita
Kiel Institute for the World Economy
24100 Kiel, Germany
E-mail: daiju.narita@ifw-kiel.de

Ulrich J. Wagner
Universidad Carlos III de Madrid
Spain
E-mail: uwagner@eco.uc3m.es

*Participants at the International Sustainability Forum at the University of Exeter and at the EAERE 2011 conference provided helpful comments. All errors are the sole responsibility of the authors. The authors gratefully acknowledge financial support from the German Research Foundation through the "Future Ocean" Cluster of Excellence program (Narita) and from the Spanish Ministry for Science and Innovation, reference number SEJ2007-62908 (Wagner).

---

1 Introduction

In recent years, a global consensus has emerged that stringent long-term goals for the emissions of greenhouse gases (GHG) are needed to mitigate climate change. In spite of this, international negotiations on a comprehensive climate treaty with targets and timetables are caught in a deadlock. Policy-makers are thus beginning to give serious consideration to alternative treaty designs, including the large-scale implementation of so-called breakthrough technologies. Since technology-oriented treaties of this kind could be tailored to individual sectors or technologies, they might be easier to negotiate than a single treaty with comprehensive emission targets (de Coninck et al., 2008). A fair number of historical examples of such agreements exist which effectively addressed various international environmental problems, albeit not in the realm of climate change (Barrett, 2003; de Coninck et al., 2008).

In the context of international climate treaties, recent theoretical research has shown that breakthrough technologies have the potential to improve participation by self-interested governments provided that they exhibit a network externality which makes the benefit to adoption proportional to the number of adopters. If the technology externality is strong, this effect dominates the free-riding incentive and full participation becomes a self-enforcing outcome (Barrett, 2006; Heal and Tarui, 2010; Hoel and Zeeuw, 2010). Appealing to the theory of the second-best, Barrett (2006) argues that this strategic advantage renders technology treaties superior to alternative treaty designs even if the technology comes at a higher cost than other forms of abatement.

Framing climate change mitigation as a technology adoption problem, however, also adds a layer of complexity to the issue. The nature of technology diffusion mechanisms is not fully understood to date. Many new technologies fail at penetrating the market despite proven technological feasibility and positive economic net benefits. The so-called McKinsey curve (see McKinsey & Company, 2010) purports that existing technologies for GHG abatement are not implemented even though they come at strictly negative cost. One reason behind this paradox is the very scale effect of technology deployment, as it renders adoption beneficial only if a sufficient number of adopters is reached in the not-too-distant future.

While the insight that technology treaties transform the cooperation problem into one of coordination is important, policy-makers also need to know how difficult a coordination problem arises with a given breakthrough technology in order to design optimal (or second-best) climate treaties. Previous research, however, has given little consideration to the coordination issue. Treaty formation has been modeled as a one-shot game, which is at odds with the reality that the diffusion of technology is not immediate but takes time. In combination with the technology externality, this delay engenders strategic uncertainty about future adoption decisions, transforming coordination on the good outcome into a non-trivial problem.[1]

This paper sheds new light on this issue. We model technology-oriented climate treaties in a richer framework to analyze the dynamics of treaty formation. We show that the possibility of multiple equilibria has important implications when countries cannot commit to future actions. Specifically, current treaty participation affects the country's future payoffs but strategic uncertainty exists about countries' future willingness to participate in the treaty. In this setting, a country takes its participation decision based on its beliefs about the others' intentions to participate. A salient implication is that participation dynamics could become self-fulfilling under certain conditions. Hence, the size of a self-enforcing treaty would be driven to some extent by subjective beliefs that

[1]Strategic uncertainty arises not due stochastic elements in the payoff functions but due to the uncertainty concerning the actions and beliefs of other players.

countries hold about their ability to coordinate policies. Such expectation-driven dynamics of technology adoption have been discussed in the contexts of development economics (Krugman, 1991) and climate policy (Narita, 2010), but so far no attempt has been made to analyze their effects in the context of self-enforcing treaties.

In this paper we study a dynamic game of treaty formation where country behavior is conditioned by a technology externality and by switching costs that are convex in the number of countries who switch technologies. The model gives rise to two types of equilibrium dynamics of technology adoption. In the first case, adoption follows a determinate path which leads to either full cooperation or no cooperation, depending on the initial state. In the second case, the dynamics are indeterminate, with stable paths leading to both full cooperation and no cooperation. The path chosen depends on the countries' expectations. In this setting, we show that in order to solve the coordination problem, a technology treaty must either focus on technologies that reduce the scope for multiple equilibria or include provisions that manage countries' expectations in a mutually beneficial way. The upshot is that these aspects need to be considered in addition to conventional features of the chosen technology such as its adoption and operating costs.

The paper is organized as follows. In Section 2, we extend the model by Barrett (2006) to sketch the basic mechanism of expectation-driven dynamics with irreversible actions. Section 3 analyzes a model with costly technology switching. We analyze both a two-period version of the model (in Section 3.1) and an infinite-horizon game (in Section 3.2). Section 3.3 explains the equilibrium dynamics and Section 3.4 discusses the implications for the design of climate treaties. Section 4 concludes.

2 Expectation-driven technology adoption

We consider the formation of a treaty that determines the adoption of a technology to reduce GHG emissions, as in Barrett (2006). Countries have a choice between conventional abatement q and the adoption of a breakthrough technology (technology X) that exhibits increasing returns to adoption. An example in the climate change mitigation context is a hydrogen-based transportation system (the combination of a new automobile technology and an infrastructure of fuel supply) or a set of CCS (carbon capture and storage) operations whose capture and storage sites are linked to each other by a carbon dioxide pipeline network.[2]

In this section we discuss a slightly modified version of Barrett's (2006) model of a technology adoption treaty. Barrett considers a one-shot game played among N countries where

$$\begin{aligned} \pi_i &= b_x \left(x_i + \sum_{j \neq i}^{N} x_j \right) - \frac{c_x}{N} \left(N - \sum_{j \neq i}^{N} x_j \right) x_i + \\ &\quad + b \left[(1 - x_i)\, q_i + \sum_{j \neq i}^{N} (1 - x_j) q_j \right] - c_0 (1 - x_i) \frac{q_i^2}{2} \end{aligned} \tag{1}$$

is the payoff function for country i, x_i is the indicator of adoption of the new technology by country i ($x_i \in \{0, 1\}$), q_i is country i's abatement rate by using the conventional technology, and b_x, c_x, b, c_0 are strictly positive coefficients representing the marginal benefit from technology X, the total cost of technology X, the marginal benefit from conventional abatement and the marginal cost of conventional abatement, respectively.

It is straightforward to show that without adopting technology X, the countries choose non-cooperative abatement amounts equal to $q_i = \frac{b}{c_0}$. Furthermore, Barrett shows that if the condition $c_x + \frac{b^2}{2c_0} > b_x > \frac{c_x}{N^2} + b^2 \frac{2N-1}{2Nc_0}$ holds, the game

[2] Regarding environmental issues other than climate change, Barrett (2003, Chapter 9) provides an extensive discussion of historical examples of environmental technologies that gave rise to network externalities.

has the quality of a "tipping treaty". That is, both universal adoption and non-adoption of technology X are Nash equilibria of the game, and all countries would be better off if X were adopted universally. The tipping point for X's attractiveness is the smallest integer greater than or equal to $N\left(\frac{b^2}{2c_0c_x}+1-\frac{b_x}{c_x}\right)$. Barrett concludes that coordination facilitates cooperation under a tipping treaty.

However, the prospects for successful coordination are likely to change if some countries are unable to adopt the technology at a given moment in time. Using a simple extension of Barrett's model, we can show that the addition of a temporal dimension limits the possibility of coordinated actions by countries and is conducive to a pattern of expectation-driven dynamics in the presence of the technology externality. We consider a game in two stages where countries are subdivided into two groups which could be labeled as technological frontrunners and followers. Decisions of technology choice by countries are made sequentially. In period 1, group 1 ($i=1,\ldots,M$ where $M<N-1$) countries – technological frontrunners with the technical capacity to use technology X from the beginning – make a decision about whether they introduce technology X. Group 2 countries ($i=M+1,...,N$) – followers – acquire the ability to introduce X only in period 2. The assumption of sequential decisions is chosen for the sake of simplicity and to reflect two types of costs, namely (i) a prohibitive cost of reverting from a new abatement technology to the conventional one and (ii) a prohibitive cost facing some countries associated with the early adoption of a new technology. We shall relax these assumptions in Section 3 below where we consider a simultaneous-moves game and reversible technology choices.

Group 1 countries adopt the new technology in period 1 if the present value payoff justifies adoption. The present value payoff for countries $1,\ldots,M$ evaluated in period 1 is

$$
\begin{aligned}
\Pi_i &= \pi_i^{t=1}(x_{11},\ldots,x_{i1},\ldots,x_{M1}) + \beta\pi_i^{t=2}(x_{12,},\ldots,x_{i2},\ldots,x_{N2}) \\
&= b_x\left(x_{i1}+\sum_{j\neq i}^{M}x_{j1}\right) - \frac{c_x}{N}\left(N-\sum_{j\neq i}^{M}x_{j1}\right)x_{i1} + \\
&\quad +b\left[(1-x_{i1})\,q_{i1}+\sum_{j\neq i}^{M}(1-x_{j1})q_{j1}\right] - \qquad\qquad (2)\\
&\quad -c_0(1-x_{i1})\frac{q_{i1}^2}{2}+\beta\left[b_x\left(x_{i2}+\sum_{j\neq i}^{N}x_{j2}\right)-\frac{c_x}{N}\left(N-\sum_{j\neq i}^{N}x_{j2}\right)x_{i2}\right] \\
&\quad +\beta\left\{b\left[(1-x_{i2})\,q_{i2}+\sum_{j\neq i}^{N}(1-x_{j2})q_{j2}\right]-c(1-x_{i2})\frac{q_{i2}^2}{2}\right\}
\end{aligned}
$$

where $x_{jt} \in \{0,1\}$ is the indicator for country j's technology X adoption in period $t \in \{1,2\}$, and $\beta \in (0,1)$ is the discount factor.

In period 2, group 2 countries (countries $M+1,\ldots,N$) may or may not adopt technology X depending on the rate of adoption by group 1 in period 1 and also on successful coordination among themselves in period 2. Yet the decision of adoption or non-adoption by group 1 countries in period 1 also depends on the success or failure of coordination by group 2 in period 2, and this is unknown to group 1 countries in period 1.

It is easy to see that there are certain cases in which expectations play a significant role in determining the outcome. Suppose that the adoption of technology X in period 1 makes frontrunners better off only if technology X is also adopted by followers in period 2. This is true if the following conditions are satisfied:

$$
\left(b_x-\frac{b^2}{2c_0}\right)M-\frac{c_x}{N}(N-M+1)+\frac{b^2}{2c_0}+\beta\left[\left(b_x-\frac{b^2}{2c_0}\right)N-\frac{c_x}{N}+\frac{b^2}{2c_0}\right]\geq 0 \quad (3)
$$

$$\left(b_x - \frac{b^2}{2c_0}\right) M - \frac{c_x}{N}(N - M + 1) + \frac{b^2}{2c_0} + \beta\left[\left(b_x - \frac{b^2}{2c_0}\right) M - \frac{c_x}{N}(N - M + 1) + \frac{b^2}{2c_0}\right] < 0 \tag{4}$$

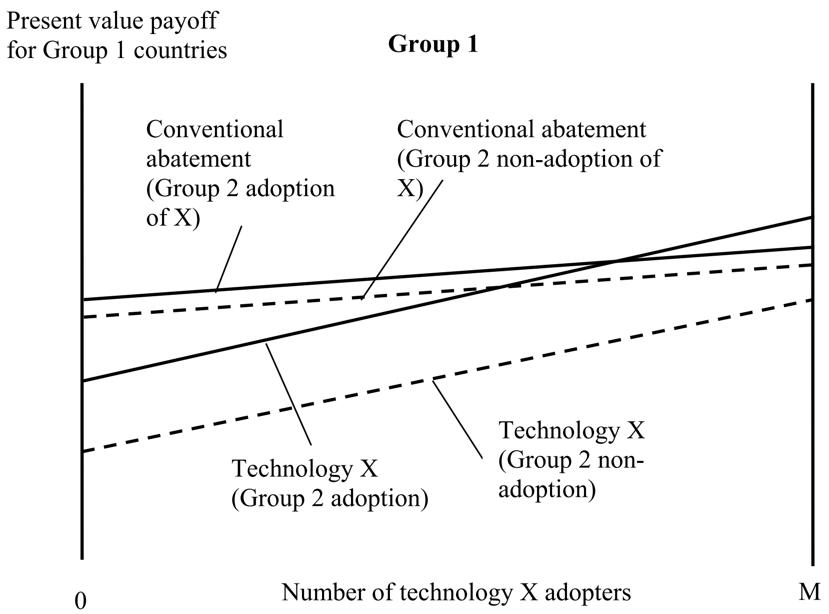
Figure 1 depicts the payoff schedules corresponding to this case. Non-adoption by all countries in both periods is one Nash equilibrium of the game, yet it is dominated by another one in which all countries adopt by the end of period 2.[3] This is similar to Barrett's model. However, since the decision on technology adoption is taken sequentially, a coordination problem arises in both periods. For instance, even if frontrunners manage to coordinate on adoption in period 1, coordination might fail in period as followers might still choose non-adoption of X. Conversely, conditions (3) and (4) imply that adoption of X by frontrunners in period 1 is not optimal if followers do not follow suit. And followers have no incentives to adopt X in period 2 if frontrunners have not adopted X beforehand.

As the incentives of the two groups are interrelated in a circular fashion there is more than one possible outcome. The outcome could in fact be determined by frontrunners' expectations about others' future intentions. Success or failure of treaty coordination in period 2 – and hence the rate of technology adoption in the future – is subject to strategic uncertainty in period 1. Thus frontrunners decide on technology adoption based on their beliefs about the future outcome. Despite the subjective nature of these beliefs, frontrunners determine the eventual adoption rate of the technology X by directly shaping the followers' incentive for technology adoption in period 2. As a consequence, the penetration of technology X could be driven by a subjective factor, and followers have little influence in shaping such subjective beliefs. Notice that this aspect does not arise in a one-shot game of technology adoption where the problem of coordination boils down

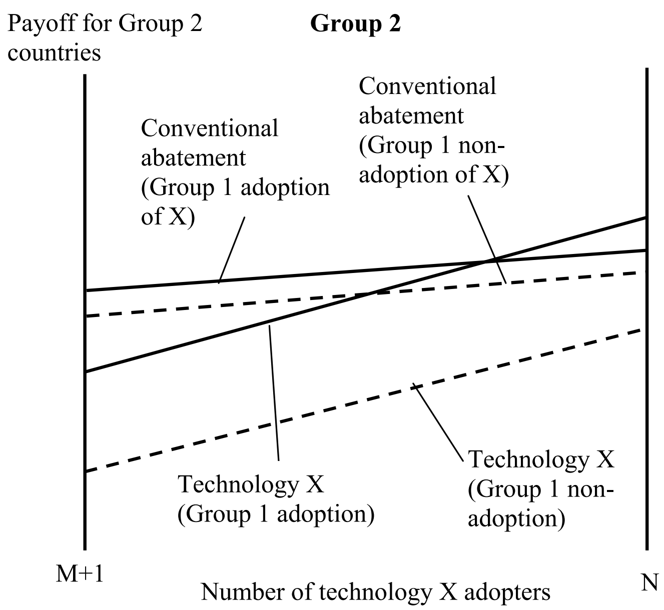
[3]Note that the tipping feature exists for followers as well since the above conditions imply that $c_x + \frac{b^2}{2c_0} > b_x > \frac{c_x}{N} + \frac{b^2(2N-1)}{2c_0}$.

Figure 1: Adoption incentives under irreversibility

(a) Payoff to frontrunners



(b) Payoff to followers



to a matter of successful political negotiations at one point in time. Previous analyses have thus given little attention to this issue.

To be sure, the problem of indeterminacy arising in the two-period model we consider here could be avoided when countries have a way of committing themselves to technology adoption in the long run. In practice, however, countries may be hard-pressed to find such a commitment device given that technology implementation covers a long time span and that political decision-makers face uncertainty about future election outcomes, economic growth, and the pace of technological progress. Perhaps as a reflection of this fact, the Kyoto Protocol has a commitment period of only five years.

The simple model discussed in this section clearly shows that the outcome of a technology X type treaty is partly determined by members' optimism or pessimism about future technology adoption, or, more precisely, about the collective capacity to coordinate technology adoption in the future. This begs the question of whether technology adoption could be promoted by manipulating countries' perceptions, even without any genuine change in the incentive structure. Countries' perceptions might be influenced by visible commitments to climate change mitigation or other international environmental problems besides the implementation of the breakthrough technology (for example, self-proclamation of long-term emission targets).[4] In contrast, a general lack of trust in stability and effectiveness of international agreements might negatively affect countries' belief about other countries' willingness for technology adoption. We shall return to this issue in Section 3.4 below.

[4]Note that in the context of the above model, it is the followers that need to display commitments to influence the frontrunners' choice.

3 Expectations and diffusion with two-way technology switching

This section extends the baseline model by dropping the assumptions of irreversibility and of a fundamental asymmetry between frontrunners and followers. The model allows for costly technology switching in both ways, i.e. both adoption and abandonment of technology X are possible.

3.1 Two-period case

We stick with the assumption of two periods but now assume that M_0 countries have adopted technology X already before period 1, and that ΔM_1 countries seek the introduction of technology X in period 1. By introducing an assumption of switching cost, which is to be described below, we lift the constraint of frontrunner-follower differentiation used in Section 2: the cost structure and technological capacity are identical for all countries, and all countries can adopt X from period 1.

In addition to the model primitives used in the previous section, we introduce an additional term SC representing the marginal initial costs of switching the technology. The initial costs may consist of various components. For simplicity, we adopt a notion of switching cost of production technology due to Mussa (1978). According to him, the costs of technology switching exclusively fall on the "moving industry" which takes care of all necessities associated with a shift of production technologies ("moving"). The moving industry is competitive, and the production of moving firms is determined by a fixed stock of resources specific to the industry and a variable factor (labor). This leads to the feature that the technology for the moving industry shows a diminishing return to the variable input. Based on this logic, both Mussa (1978) and Krugman (1991) assume that the aggregate switching costs across agents take a quadratic function of the number of agents switching their mode of production in a given period. This means that the marginal moving cost – the increase in the aggregate moving cost by entry of a new agent – is a linear function of the number of agents who switch technologies in this period. We apply this formulation to our case and assume that a switching country incurs initial costs up to $SC = f\Delta M_1$, where f is a positive constant and ΔM_1 is the number of countries adopting X.[5] We assume that $f > \frac{c_x}{N}(1+\beta)$, i.e. the switching cost outweighs the (present-value) externality effect on running costs for X. For the sake of simplicity, we assume that a symmetrical moving cost is incurred when $-\Delta M_1$ countries abandon technology X and switch to the conventional abatement.

Countries decide upon adoption taking into account the one-time switching cost as well as the present-value gain associated with using X instead of the conventional abatement option. Let λ denote this gain. λ is a function of M_0, ΔM_1, and ΔM_2 (the number of countries that switch technologies in period 2) given by

$$\lambda(M_0, \Delta M_1, \Delta M_2) \equiv \\ (1+\beta)\left[b_x - \frac{c_x}{N}(N - M_0 + 1) - \frac{b^2}{2c_0}\right] + (1+\beta)\Delta M_1 \frac{c_x}{N} + \beta \Delta M_2 \frac{c_x}{N} \quad (5)$$

[5]Convex costs are likely to arise when countries switch from nuclear energy to other forms of carbon-neutral electricity generation, as recently decided by the German government ("Energiewende"). Since Germany is the only country to take this step so far, it will be able to import cheap nuclear power from other European countries during the transition. However, if other European countries adopt similar decisions, this is bound to drive up the initial cost of technology switching as countries would bid up the price of (nuclear) power and the increased volume of transnational electricity trade could lead to congestion on the European transmission grid.

Define λ^1 and λ^2 as:

$$\lambda^1(M_0) = b_x - \frac{c_x}{N}(N - M_0 + 1) - \frac{b^2}{2c_0} \tag{6}$$

$$\lambda^2(\Delta M_1, \Delta M_2) = (1+\beta)\Delta M_1 \frac{c_x}{N} + \beta \Delta M_2 \frac{c_x}{N} \tag{7}$$

Then λ can be written as

$$\lambda(M_0, \Delta M_1, \Delta M_2) = (1+\beta)\lambda^1(M_0) + \lambda^2(\Delta M_1, \Delta M_2). \tag{8}$$

The balance of λ and the switching cost determines the number of adoption or abandonment of X in period 1. Let us first consider the case of progressive technology adoption, i.e. the number X of adopters increases over time. Then there is a maximum value of ΔM_1 (less than $N - M_0$) such that the payoff gain from switching from conventional abatement to X is positive (recall that a larger ΔM_1 reduces the expected payoff of adoption because of the switching cost). In equilibrium, the number of countries ΔM_1 switching technologies in period 1 is given by the largest integer ΔM_1 to satisfy

$$\lambda(M_0, \Delta M_1, \Delta M_2) \geq f \Delta M_1. \tag{9}$$

By contrast, countries might expect that others will abandon technology X in period 2. As technology X is attractive only with a large number of adopters, the fear of collective abandonment gives those that have adopted technology X an incentive to abandon it. As above, the switching costs limit the magnitude of abandonment in this period. If there is a set of negative ΔM_1 that satisfy

$$-\lambda(M_0, \Delta M_1, \Delta M_2) \geq -f \Delta M_1 \tag{10}$$

then the number of countries that abandon technology X in period 1 is given by the integer with the largest absolute value in this set.

It is straightforward to show possible cases in which expectations play a crucial role in determining the outcome. Figure 2a depicts the case where M_0 is located to the left of the tipping point A – mathematically, $\lambda^1(M_0) = 0$. In such a case, there is always a feasible combination of (ΔM_1,ΔM_2) where $\Delta M_1, \Delta M_2 \leq 0$ (a proof is given in the Appendix). However, under certain conditions, there may also be a feasible combination of (ΔM_1,ΔM_2) where $\Delta M_1, \Delta M_2 \geq 0$. To see this, notice that – by a logic similar to the one used to derive condition (9) – the number of countries that adopt technology X in period 2 should be the largest integer ΔM_2 to satisfy the inequality

$$\lambda^1(M_0 + \Delta M_1 + \Delta M_2) \geq f\Delta M_2 \tag{11}$$

The above two conditions are identical to:

$$\frac{N}{c_x}\lambda^1(M_0) + \left(1 - \frac{fN}{c_x(1+\beta)}\right)\Delta M_1 + \frac{\beta}{1-\beta}\Delta M_2 \geq 0 \tag{12}$$

$$\frac{N}{c_x}\lambda^1(M_0) + = \Delta M_1 + \left(1 - \frac{fN}{c_x}\right)\Delta M_2 \geq 0 \tag{13}$$

and can be satisfied by a set of weakly positive $(\Delta M_1, \Delta M_2)$. For example, positive ΔM_1 and ΔM_2 exist if $\frac{fN}{c_x(1+\beta)} - 1$ is very small (recall that, by assumption, $\frac{fN}{c_x(1+\beta)} - 1 > 0$) and there is a feasible ΔM_2 that satisfies

$$\frac{\beta}{1-\beta}\Delta M_2 > -\frac{N}{c_x\lambda^1(M_0)}$$

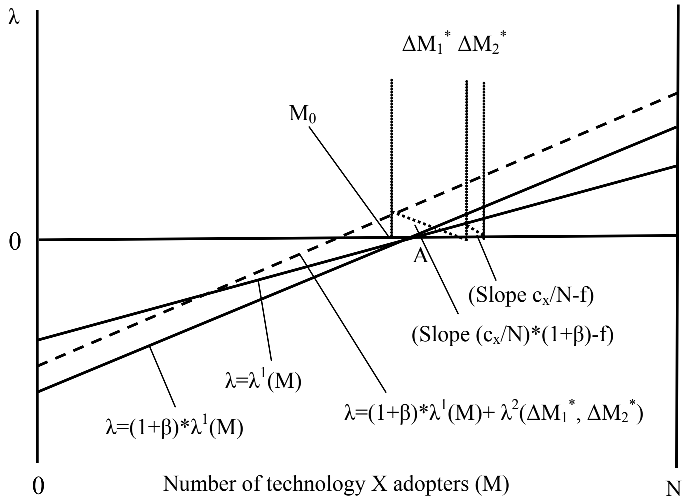
(note that $\lambda^1(M_0) < 0$).

An analogous reasoning can be developed for the case in which M_0 is located to the right of the tipping point, as is depicted in Figure 2b.

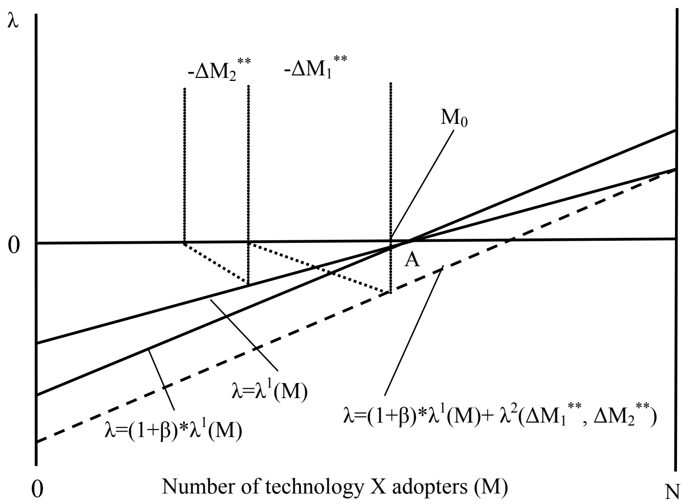
In summary, we have shown that the dynamics of technology switching may

Figure 2: Adoption incentives with costly technology switching

(a) Technology adoption



(b) Technology abandonment



be uniquely determined in the direction of either increasing adoption or abandonment, depending on the initial state of technology adoption. However, the system may also have feasible solutions for both directions of technology adoption and abandonment, in which case the outcome is determined entirely by countries' expectations.

3.2 Infinite-horizon game

Here we show that similar patterns to the ones described in the previous section could emerge in the case of an infinite-horizon game of technology adoption. In this setting, the effect of actions at any given stage is cumulative so that final outcomes differ drastically, depending on both the model primitives and players' expectations. There are two fundamentally different scenarios. In the first, expectations about future payoffs are perfectly aligned and lead to a unique outcome, either a universal adoption or zero adoption. In the second scenario, expectations may differ and the outcome is expectation-driven, akin to a self-fulfilling prophecy.

We analyze an infinite-horizon version of the game developed in the previous section. Play starts in period 0 with an initial number M_0 of adopters. Countries take into account the cumulative present-value payoffs associated with their chosen technology from the present to infinity. We focus on subgame perfect equilibria with the feature that countries immediately begin an optimal transition to either full adoption or no adoption – which one depends on M_0, payoff parameters, and expectations. Once this stage-game Nash equilibrium is reached, it will be repeated indefinitely as players have no incentives to further deviate. Since indefinite Nash play is a subgame perfect equilibrium of the continuation game, we can use backward induction to determine the individually rational transition towards this state. In the following exposition, we focus on the case of M_0 large enough so that countries have an incentive to adopt technology X. However, due to switching cost, adoption by the remaining countries will take place in L batches $\Delta M_1, \Delta M_2, \ldots, \Delta M_{L-1}, \Delta M_L$ where $\sum_{l=1}^{L} \Delta M_l = N - M_0$.

Consider the last batch of $\Delta M_L = N - M_{L-1}$ of adopters. The relative payoff to adoption for these countries is given by

$$\lambda^L = \frac{1}{1-\beta}\left[b_x - \frac{c_x}{N}(N - M_0 + 1) - \frac{b^2}{2c_0} + \frac{c_x}{N}\left(\sum_{l=1}^{L} \Delta M_l\right)\right] \tag{14}$$

A Nash equilibrium in this subgame requires that $\lambda^L \geq f\Delta M_L$ where $\Delta M_L = N - M_{L-1}$. Working backwards, in period $L-1$ a group of ΔM_{L-1} countries adopting technology X earns relative payoffs

$$\lambda^{L-1} = \frac{1}{1-\beta}\left[b_x - \frac{c_x}{N}(N - M_0 + 1) - \frac{b^2}{2c_0} + \frac{c_x}{N}\left(\sum_{l=1}^{L-2} \Delta M_l + \Delta M_{L-1} + \beta \Delta M_L\right)\right] \tag{15}$$

For there to be exactly ΔM_{L-1} adopters in Nash equilibrium, adoption must make all of them weakly better off, i.e.

$$\lambda^{L-1} \geq f\Delta M_{L-1}. \tag{16}$$

However, any additional adopter of technology X must be strictly worse off:

$$\lambda^{L-1} + \frac{c_x}{N} < f(\Delta M_{L-1} + 1). \tag{17}$$

Iterating backwards, we obtain the relative payoff to adoption on the equilibrium path for the kth batch of adopters, $k \in \{1, L-1\}$

$$\lambda^k = \frac{1}{1-\beta}\left[b_x - \frac{c_x}{N}(N - M_0 + 1) - \frac{b^2}{2c_0} + \frac{c_x}{N}\left(\sum_{l=1}^{k-1} \Delta M_l + \sum_{l=k}^{L} \beta^{l-k} \Delta M_l\right)\right] \tag{18}$$

and the equilibrium conditions

$$\lambda^k \geq f\Delta M_k \tag{19}$$

$$\wedge \ \lambda^k < f\Delta M_k + f - \frac{c_x}{N} \tag{20}$$

For given λ^k, conditions (19) and (20) pin down the number of adopters in a Nash equilibrium at stage k. The equilibrium path of λ^k can be constructed using backward induction where $\lambda^{L+s} = \frac{1}{1-\beta}\left[b_x - \frac{c_x}{N} - \frac{b^2}{2c_0}\right]$ for $s = 0, 1, \ldots$ serves as the end point.

To characterize the evolution of the relative payoff to adoption, we rewrite the relative payoff to adoption for adopters in the kth batch of adopters as follows

$$\lambda^k = \sum_{s=k}^{L} \beta^{s-k}\left[b_x - \frac{c_x}{N}(N - M_s + 1) - \frac{b^2}{2c_0}\right] + \frac{\beta^{L-k+1}}{1-\beta}\left[b_x - \frac{c_x}{N} - \frac{b^2}{2c_0}\right] \tag{21}$$

The difference in the relative payoffs to adoption for two subsequent batches of adopters k and $k+1$ can be written as

$$\lambda^{k+1} - \lambda^k = \delta\lambda^k - (1+\delta)\left[b_x - \frac{c_x}{N}(N - M_k + 1) - \frac{b^2}{2c_0}\right] \tag{22}$$

where $\delta \equiv \frac{1-\beta}{\beta}$. Along with the inequalities (19) and (20), equation (22) characterizes the dynamics of technology adoption in subgame perfect equilibrium. As in Section 3.1 above, the conditions for an equilibrium in which all countries switch back to the conventional technology can be derived in an analogous fashion.

3.3 Equilibrium dynamics

To analyze the dynamics of technology adoption along the equilibrium path, it is convenient to consider the limiting case where N goes to infinity and hence the rate of technology adoption γ can be modeled as a continuous variable γ ($0 \leq \gamma \leq 1$). The number of countries adopting technology X is thus given by $\lceil \gamma N \rceil$. Similar to the case discussed in the previous section, the equilibrium level of countries switching at each time period is one that balances the net present value of switching and the marginal switching costs for all countries. Along the equilibrium path, the net present value of switching from the conventional abatement to X at period t is given by

$$\lambda_t = \sum_{s=t}^{\infty} \beta^{s-t} \left[b_x - c_x(1-\gamma_s) - \frac{b^2}{2c_0} \right]. \tag{23}$$

The marginal switching cost for countries switching technologies between t and $t+1$ becomes $fN(\gamma_{t+1} - \gamma_t)$. Since N is fixed, we rewrite the marginal switching cost as $F(\gamma_{t+1} - \gamma_t)$ where $F \equiv fN$ is a positive constant. For λ_t continuous in γ, conditions (19) and (20) boil down to the difference equation

$$F(\gamma_{t+1} - \gamma_t) = \frac{1}{1+\delta} \lambda_{t+1}. \tag{24}$$

A second difference equation governs the evolution of λ_t

$$\lambda_{t+1} - \lambda_t = \delta \lambda_t + (1+\delta) \left[b_x - c_x(1-\gamma_t) - \frac{b^2}{2c_0} \right] \tag{25}$$

As the length of time periods goes to zero, the system of difference equations (24) and (25) can be approximated by the differential equations

$$\begin{aligned} F\dot{\gamma} &= \frac{1}{1+\delta} \lambda && (26) \\ \dot{\lambda} &= \delta \lambda + (1+\delta) \left[b_x - c_x(1-\gamma) - \frac{b^2}{2c_0} \right] && (27) \end{aligned}$$

This representation allows for a more tractable analysis of the dynamics along the equilibrium path. Eqs. (26) and (27) define a system of linear differential equations whose solution is given by a combination of exponential functions. If

$$0 < -b_x + c_x + \frac{b^2}{2c_0} < c_x \tag{28}$$

the system has a tipping pattern, i.e. both universal adoption and zero adoption of X are long-run (continuation) equilibria. In this case, the paths of λ and γ are obtained by tracing them backwards from two long-run equilibria where $\gamma = 0$ or $\gamma = 1$. The roots of the exponential functions determining the system are given by

$$\rho = \frac{1}{2}\left[\delta \pm \sqrt{\delta^2 - \frac{4c_x}{F}}\right] \tag{29}$$

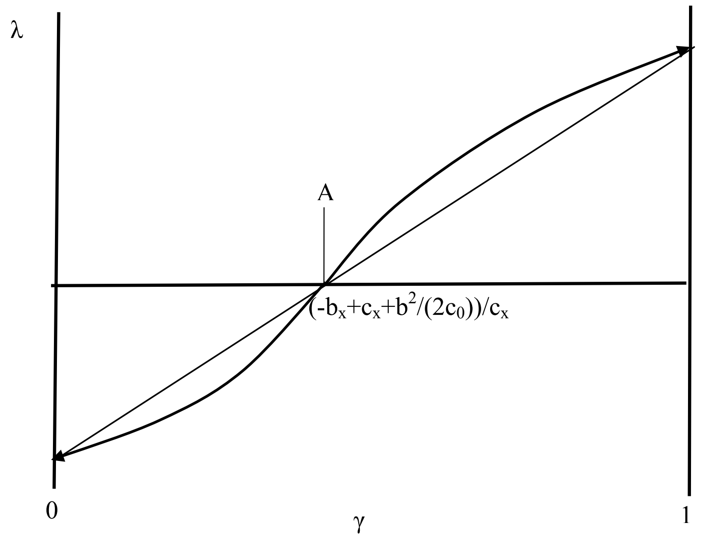
Note that the roots can be both real and complex depending on the parameter values, as the term $\delta^2 - 4c_x/F$ can be either positive or negative. The system dynamics exhibit remarkable differences depending on which type of roots prevail in eq. (29).

With a real root, the system is determinate and hence equilibrium play does not exhibit expectation-driven dynamics. This case is depicted in Figure 3a. Starting at the tipping point A, either of the two long-run equilibria can be attained as the dynamics evolve through a sequence of decisions by countries governed by the equilibrium conditions (26) and (27). The graph shows that each value of γ other than A corresponds to at most one point on one of the two trajectories. In other words, the initial state of adoption γ_0 uniquely determines the long-run penetration rate of technology X. Depending on whether $\gamma_0 \gtrless A$ either universal adoption ($\gamma = 1$) or zero penetration ($\gamma = 0$) result in long-run equilibrium.

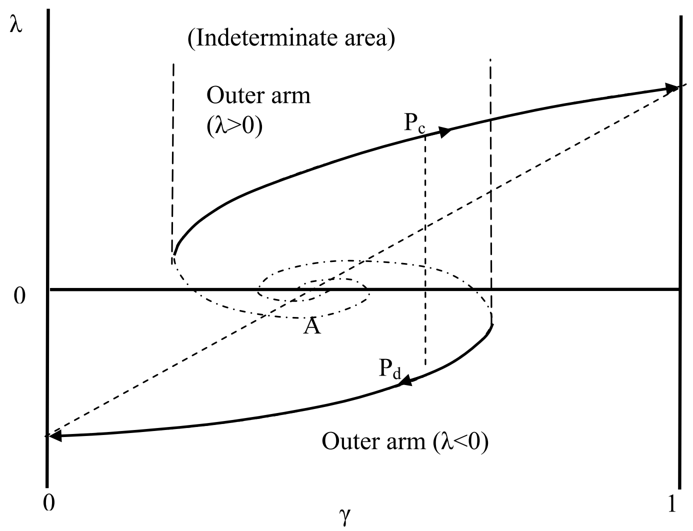
In contrast, expectations play a prominent role in the case with a complex root. In this case, the trajectories show oscillatory patterns, and their arms

Figure 3: Equilibrium dynamics in the infinite horizon game

(a) Determinate case (real roots)



(b) Indeterminate case (complex roots)



could cover a wide range of possible values for γ. When the two arms overlap over an interval of γ – as is depicted in Figure 3b – the initial state does not determine the direction of the path. In fact, there is an infinite number of feasible trajectories that the system can take. In such a circumstance, the model primitives do not condition countries to follow a unique equilibrium path. Rather, it is countries' *expectations* about future adoption of technology X that pace the growth (or decline) of technology penetration. This means that, even if there is a feasible equilibrium path leading to the universal adoption of X (for example, the path through point P_c in Figure 3b), sheer pessimism about future adoption by non-adopters could prevent the initial group of adopters from taking this path. Instead, they might choose to follow the trajectory to the zero adoption (for example, the path through point P_d in Figure 3b).

3.4 Discussion

The dynamic model highlights two distinct patterns of technology choice under a technology treaty which deserve further discussion from a policy point-of-view.

In the determinate case, there is a unique equilibrium path leading to a unique long-run outcome. This outcome can be either universal or zero adoption and is determined by the initial state of technology adoption γ_0 and by the tipping point

$$\gamma^* = 1 - \frac{b_x - \frac{b^2}{2c_0}}{c_x} \tag{30}$$

Only if the initial proportion of adopters is sufficiently large will the technology be adopted by everyone in the long run. Otherwise, all countries will switch back to the conventional technology. The policy recommendation growing out of this is to pick the most efficient technology to satisfy the constraint that $\gamma^* < \gamma_0$. The tipping point is likely to be lower the more affordable the breakthrough technology, the more expensive the conventional technology and the larger the relative benefits of technology X compared to those associated with the conventional technology. In this scenario, the earlier results by Barrett (2006) and Hoel and Zeeuw (2010) go through and the coordination problem is negligible.

However, this is not the case if the technology is such that the long-run outcome is expectation-driven, akin to a self-fulfilling prophecy. Eq. (29) implies that this case arises if $\delta^2 < 4c_x/F$ and hence the system of differential equations has complex roots. It is easily seen from this inequality that a higher discount rate δ and a higher switching cost F parameter both promote determinacy of the system. This is because both myopia and high costs of technology switching enhance the relative importance of current over future payoffs, which are subject to strategic uncertainty. Conversely, a large scale effect of technology X c_x is conducive to indeterminacy of the system, as it makes countries' present-value payoff more susceptible to others' technology choices in the future.

Climate policy can address this strategic uncertainty in two ways. The first approach is to reduce strategic uncertainty by managing expectations. This could be implemented, for example, by setting long-term targets for carbon emissions or policies which coordinate expectations across countries on the path leading to full adoption. In fact, the recognition that the increase in global temperature should be below 2 degrees Celsius under the Copenhagen Accord can be interpreted in this way. This target does not require any country to reduce its emissions at present, but it may align countries' expectations and thereby make a future technology treaty tip towards adoption.

The second approach to this is to reduce strategic uncertainty by choosing technologies that minimize the potential for indeterminacy of the dynamic system. Instead of choosing the most efficient breakthrough technology, policymakers might favor a technology with high switching cost as this locks the frontrunners into their decisions while also reducing strategic uncertainty of followers. This aspect of technology adoption arises only in our explicitly dynamic framework and is an important extension to the second-best argument by which technologies with scale effects are superior to alternative treaty designs even if they come at a higher cost because they reduce the incentive to free ride (Barrett, 2006).

Finally, our model highlights an important tradeoff that arises in the design of climate treaties with breakthrough technologies for fixed switching costs. If countries are very impatient, policymakers should give priority to technologies with a low tipping point as impatience promotes determinacy of the system. On the contrary, if countries are very patient, policymakers should focus on complementing technology agreements with expectation management as the system is more likely to be indeterminate.

In sum, when technology-oriented climate treaties are analyzed under more realistic assumptions about the timing and costs of technology switching than in previous work, a variety of new results emerge which are of crucial importance for the success of such treaties.

4 Conclusion

Technology-oriented treaties are beginning to receive wide attention as a potentially useful approach to the conundrum of international climate policy. The scale effect – or network externality – of some technologies may work in favor of countries' participation in such technology treaties, but it also engenders a problem of coordination among multiple equilibria which adds to other obstacles to the diffusion of new technologies. The literature on self-enforcing international environmental treaties so far has emphasized the former point but given little consideration to the coordination problem. This omission leaves a number of open questions, for example, why coordination about technology adoption is often missing or ineffective in reality, despite the fact that all countries are expected to benefit under universal adoption.

In this study, we have investigated this proposition within a multistage model of treaty formation. In the presence of a network externality, the benefit of technology adoption for countries depends on the total number of technology adopters. Since countries evaluate their choices based on the present value of technology options, their current decisions of technology choice are influenced by the future adoption rates of technologies, which they do not know at present and can only expect. We have shown that under certain conditions, the adoption of breakthrough technologies becomes expectation-driven: technologies may not diffuse when countries are struck with pessimism, but this also means that adoption of mitigation technology may be enhanced substantially if countries can display some ability to commit.

This paper stresses the vital role of sustained coordination – e.g. through managing expectations regarding future technology use – in the diffusion of technologies through treaties. In so doing, our study offers a new perspective on the debate about the economics of climate change and uncertainty, emphasizing that heterogeneity of expectations can be influential in climate policy. This emphasis on strategic uncertainty is markedly different from uncertainty about the economic primitives of climate policy which has been the focus of economic analyses so far (e.g. Heal and Kriström, 2002; Weitzman, 2009).

References

Barrett, S. (2003). *Environment and Statecraft.* Oxford University Press, New York.

Barrett, S. (2006). Climate treaties and "breakthrough" technologies. *The Amer-*

ican Economic Review, 96(2):pp. 22–25.

de Coninck, H., Fischer, C., Newell, R. G., and Ueno, T. (2008). International technology-oriented agreements to address climate change. *Energy Policy*, 36(1):335–356.

Heal, G. and Kriström, B. (2002). Uncertainty and climate change. *Environmental & Resource Economics*, 22(1):3–39.

Heal, G. and Tarui, N. (2010). Investment and emission control under technology and pollution externalities. *Resource and Energy Economics*, 32(1):1 – 14.

Hoel, M. and Zeeuw, A. (2010). Can a focus on breakthrough technologies improve the performance of international environmental agreements? *Environmental & Resource Economics*, 47(3):395–406.

Krugman, P. (1991). History versus expectations. *The Quarterly Journal of Economics*, 106(2):651–67.

McKinsey & Company (2010). Impact of the financial crisis on carbon economics: Version 2.1 of the global greenhouse gas abatement cost curve. Technical report, McKinsey & Company.

Mussa, M. (1978). Dynamic adjustment in the Heckscher-Ohlin-Samuelson model. *Journal of Political Economy*, 86(5):pp. 775–791.

Narita, D. (2010). Climate policy as expectation management? *Kiel Working Paper*, 1624. Kiel Institute for the World Economy.

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

A Appendix - Not for publication

A.1 Proof.

The given conditions imply $-\lambda(M_0, 0, 0) > 0$ and $.-\lambda^1(M_0) > 0$. Since $f > \frac{c_x}{N}(1+\beta)$, the function

$$-\lambda(M_0, \Delta M_1, \Delta M_2) - \lambda^1(-\Delta M_1)$$

is increasing in ΔM_1 (decreasing in $-\Delta M_1$) and decreasing in ΔM_2 (increasing in $-\Delta M_2$). As , this means that there is at least one feasible $\Delta M_1 \leq 0$ for all ΔM_2 satisfying $\Delta M_2 \leq 0$.

Meanwhile, a negative ΔM_2 satisfies the following inequality

$$-\lambda^1(M_0 + \Delta M_1 + \Delta M_2) \geq f(-\Delta M_2)$$

Since $f > \frac{c_x}{N}(1 + \beta)$, the function

$$-\lambda^1(M_0 + \Delta M_1 + \Delta M_2) - f(-\Delta M_2)$$

is increasing in ΔM_2 (decreasing in $-\Delta M_2$). As $-\lambda^1(M_1) > 0$, this means that there is at least one $\Delta M_2 \leq 0$ that satisfies the above inequality for all ΔM_1 such that $\Delta M_1 \leq 0$.

The above means that if M_0 is located on the left of the tipping point, there is always a feasible combination of $(\Delta M_1, \Delta M_2)$ such that $\Delta M_1, \Delta M_2 \leq 0$.

A.2 Difference equation for λ in the discrete game

To characterize the evolution of the relative payoff to adoption, we rewrite the relative payoff to adoption for adopters in the kth batch of adopters as follows

$$\lambda^k = \sum_{s=k}^{L} \beta^{s-k} \left[b_x - \frac{c_x}{N}(N - M_s + 1) - \frac{b^2}{2c_0} \right] + \frac{\beta^{L-k+1}}{1-\beta} \left[b_x - \frac{c_x}{N} - \frac{b^2}{2c_0} \right]$$

The relative payoff to adoption for the subsequent batch of adopters *k+1* is given by

$$\lambda^{k+1} = \sum_{s=k+1}^{L} \beta^{s-(k+1)} \left[b_x - \frac{c_x}{N}(N - M_s + 1) - \frac{b^2}{2c_0} \right] + \frac{\beta^{L-k}}{1-\beta} \left[b_x - \frac{c_x}{N} - \frac{b^2}{2c_0} \right]$$

Let $\delta \equiv \frac{1-\beta}{\beta}$ and calculate

$$\begin{aligned} \lambda^{k+1} - (1+\delta)\lambda^k &= \lambda^{k+1} - \frac{\lambda^k}{\beta} \\ &= \sum_{s=k+1}^{L} \beta^{s-(k+1)} \left[b_x - \frac{c_x}{N}(N - M_s + 1) - \frac{b^2}{2c_0} \right] - \\ &\quad - \sum_{s=k}^{L} \beta^{s-k-1} \left[b_x - \frac{c_x}{N}(N - M_s + 1) - \frac{b^2}{2c_0} \right] \\ &= -\frac{1}{\beta} \left[b_x - \frac{c_x}{N}(N - M_k + 1) - \frac{b^2}{2c_0} \right] \end{aligned}$$

Simple manipulation of this expression yields

$$\lambda^{k+1} - \lambda^k = \delta\lambda^k - (1+\delta) \left[b_x - \frac{c_x}{N}(N - M_k + 1) - \frac{b^2}{2c_0} \right] \tag{31}$$

A.3 Difference equation for λ with a continuum of countries

$$\lambda_t = \sum_{s=t}^{\infty} \beta^{s-t} \left[b_x - c_x(1-\gamma_s) - \frac{b^2}{2c_0} \right]$$

and

$$\lambda_{t+1} = \sum_{s=t+1}^{\infty} \beta^{s-t-1} \left[b_x - c_x(1-\gamma_s) - \frac{b^2}{2c_0} \right]$$

Calculate

$$\begin{aligned}
\Delta\lambda_{t+1} - \delta\lambda_t &= \sum_{s=t+1}^{\infty} \beta^{s-(t+1)} \left[b_x - c_x(1-\gamma_s) - \frac{b^2}{2c_0} \right] - \\
&\quad - \underbrace{(1+\delta)}_{=1/\beta} \sum_{s=t}^{\infty} \beta^{s-t} \left[b_x - c_x(1-\gamma_s) - \frac{b^2}{2c_0} \right] \\
&= \sum_{s=t+1}^{\infty} \beta^{s-(t+1)} \left[b_x - c_x(1-\gamma_s) - \frac{b^2}{2c_0} \right] - \\
&\quad - \sum_{s=t}^{\infty} \beta^{s-t-1} \left[b_x - c_x(1-\gamma_s) - \frac{b^2}{2c_0} \right] \\
&= -\frac{1}{\beta} \left[b_x - c_x(1-\gamma_t) - \frac{b^2}{2c_0} \right]
\end{aligned}$$

Hence

$$\begin{aligned}
\lambda_{t+1} - \lambda_t &= \frac{1-\beta}{\beta}\lambda_t - \frac{1}{\beta} \left[b_x - c_x(1-\gamma_t) - \frac{b^2}{2c_0} \right] \\
&= \delta\lambda_t - (1+\delta) \left[b_x - c_x(1-\gamma_t) - \frac{b^2}{2c_0} \right]
\end{aligned}$$