Climate Change, Catastrophic Risk and the Relative Unimportance of the Pure Rate of Time Preference

Eric Nævdal[1] & Jon Vislie[2]

[1]Ragnar Frisch Centre for Economic Research
Gaustadalléen 21
N-0349 Oslo
Norway
e-mail: eric.navdal@frisch.uio.no

[2]Department of Economics
University of Oslo
Moltke Moes vei 31
NO-0851 Oslo.

Acknowledgements: We thank Geir Asheim, Michael Hoel, Michael Oppenheimer, Reyer Gerlagh and Atle Seierstad for useful comments.

Climate Change, Catastrophic Risk and the Relative Unimportance of the Pure Rate of Time Preference

Abstract

The role of discounting in the management of climate change is a hotly debated issue. Many scientists and laymen concerned with potentially catastrophic impacts feel that if an increase in the discount rate drastically increases the likelihood of catastrophic outcomes, this discredits economic cost-benefit calculations. This paper argues that this intuition is sound. If cost-benefit calculations are done within a model that encompasses the type of catastrophic threshold effects that these scientists worry about, the resulting stabilization target will only be slightly influenced by the discount rate.

Key words: climate change, discounting, catastrophic risk, optimal control.

1. Introduction

Discourse on policy responses to climate change has a tendency to become a debate on the appropriate method of discounting. The Stern Review was severely criticized by influential economists such as William Nordhaus and Martin Weitzman who claimed that much of the results where artificially driven by low discount rates [15], [9], [21]. Indeed, much of the discussion about these models boils down to the appropriate choice of numerical value for the pure rate of time preference and measures of income inequality aversion, [3]. To the extent that the pure rate of time preference actually matters for climate policy this is a fruitful debate, but it is not clear that the pure rate of time preference is of paramount importance if climate change induces catastrophic risk which must be managed. In the literature there are numerous attempts to take account of an uncertain future, like irreversibilities. As demonstrated, for instance by Gollier [5],[6], and by Weitzman [19],[20], a time-dependent and declining discount rate can be rationalised so that future uncertainty or risk will be properly accounted for. It has for some time been recognized that climate change carries with it the risk of catastrophes when certain boundaries, termed thresholds or tipping points are crossed. Examples of possible catastrophic scenarios include coral bleaching, marine ice sheet instability, methane hydrate destabilization and disruption of the thermohaline circulation (Gulf Stream),[12], [7], [1], [8]. There is unfortunately a disconnect between scientists concerned with potentially catastrophic threshold effects and economists who do not include them in their models, or simply ignore them because the catastrophic event is expected to occur in the far-distant future. This has led to an unfortunate breakdown of communication between the scientists who feel that the intelligent management of catastrophic risk should not be very sensitive to discounting while economists armed with results from integrated assessment models claim that the pure rate of time preference is a crucial parameter in climate policy.

The economic analysis of problems with threshold risk is obviously confounded by the lack of precise knowledge about the location of these thresholds. Partha Dasgupta has even suggested that the existence of such tipping points may severely restrict the usefulness of cost-benefit analysis, [4]. It is therefore all the more worrisome that threshold risk is not an integral part of current economic models of climate change. Further, if threshold effects are an important part of the possible damages induced by climate change, one may argue that economic discourse on the role of discounting is premature until its effect in dynamic models with threshold risk is properly understood.

There is very little formal economic analysis of threshold effects with unknown threshold location, and what there is pays very little attention to the role of the pure rate of time preference, [10], [11], [16], [17]. The exception is [18] where it is shown how threshold risk turns the socially optimal discount rate endogenous to climate change policy. However, the importance of the pure rate of time preference is not addressed here either. Here we present a stylized model showing that if catastrophic risk of crossing a crucial climate threshold is incorporated into an economic decision model, the pure rate of time preference is of little importance for the question of what level to stabilize atmospheric CO_2. The model is solved analytically and contains a number of simplifying assumptions in order to clarify the role of the pure rate of time preference rate when attempting to control catastrophic climate risk. We assume risk-neutrality and standard exponential discounting. The consequence of a catastrophe is modelled as a fixed cost which does not entail the possibility of the marginal utility of consumption becoming infinite. Thus our model is different from [22], where the discount rate does not matter because the fat tails associated with statistical estimation of parameters implies a positive probability of an outcome with infinite marginal utility and therefore an infinite willingness to pay for avoiding this outcome. The model is aimed to capture the rational deliberations of a standard economic decision maker who faces the possibility of a severe catastrophe, which does not however entail an outcome where the human race is pushed to or below a minimum subsistence level. Our results indicate that when the threshold nature of catastrophic climate change is properly incorporated into an economic decision model, the numerical value of the discount rate is of marginal importance for the long-term choice of CO_2 stabilization level. Although it remains to be seen whether our results carry over to more realistic numerical models of climate management, it may be that much of the discussion about discounting and climate change is not as relevant as one could believe when examining results from the current crop of largely deterministic numerical models.

2. A Simple Model of Carbon Emissions and Catastrophic Risk

Here we present a stylized model of catastrophic climate change. Threshold effects require somewhat specialized optimal control techniques. Let the stock of atmospheric CO_2 above pre-industrial levels be determined by the following differential equation:

$$\frac{dx(t)}{dt} = u\left(t\right) - \delta x(t), \;\; x\left(0\right) \text{ given.} \tag{1}$$

Here x is the stock of atmospheric carbon above pre-industrial levels, u is the flow of CO_2 emissions and δ is the inverse of the mean atmospheric lifetime of CO_2. Assume further that there is a threshold $\bar{x}$ such that if $x = \bar{x}$ then an irreversible catastrophic event is triggered. The threshold location $\bar{x}$ is a random variable with a positive density function $f(x)$ on $[x_L, \infty)$. We have defined x_L to be the highest value of x known to be below the true threshold. As x is a function of t, then for any given function $u(t)$ the point in time τ such that $x(\tau) = \bar{x}$ is a random variable. Thus for any path $x(t)$ one can translate the distribution of $\bar{x}$ over x into a distribution over time. This translation is a bit involved and is therefore developed in a heuristic manner. A more technical treatment is given in the Appendix. This is illustrated in Figure 1 which shows an arbitrary sample path $x(t)$, which should not be taken to be optimal. Every point on the $x(t)$-axis is a possible threshold location. The path oscillates until t = D for then to converge to $x(\infty)$. The key to understanding the stochastic process generated by the threshold is that there is only a risk of crossing the threshold if $x(t)$ is taking values that have not previously been attained. Thus in the interval [O, A], $x'(t)$ is positive and $x(t) > x(s)$ for all $t > s$. There is therefore some probability that the threshold will be crossed in the time interval [O, A]. At A, $x'(t)$ changes sign and over the interval [A, B], $x(t) \leq x_A$ which implies that x is running through values known to be safe. At B, $x(t)$ again enters uncharted territory with some risk of crossing the threshold until time C when $x(t) = x_C$. At time C, $x(t)$ takes another dip and there is again no probability of crossing the threshold until time D when x(D) again equals x_C .

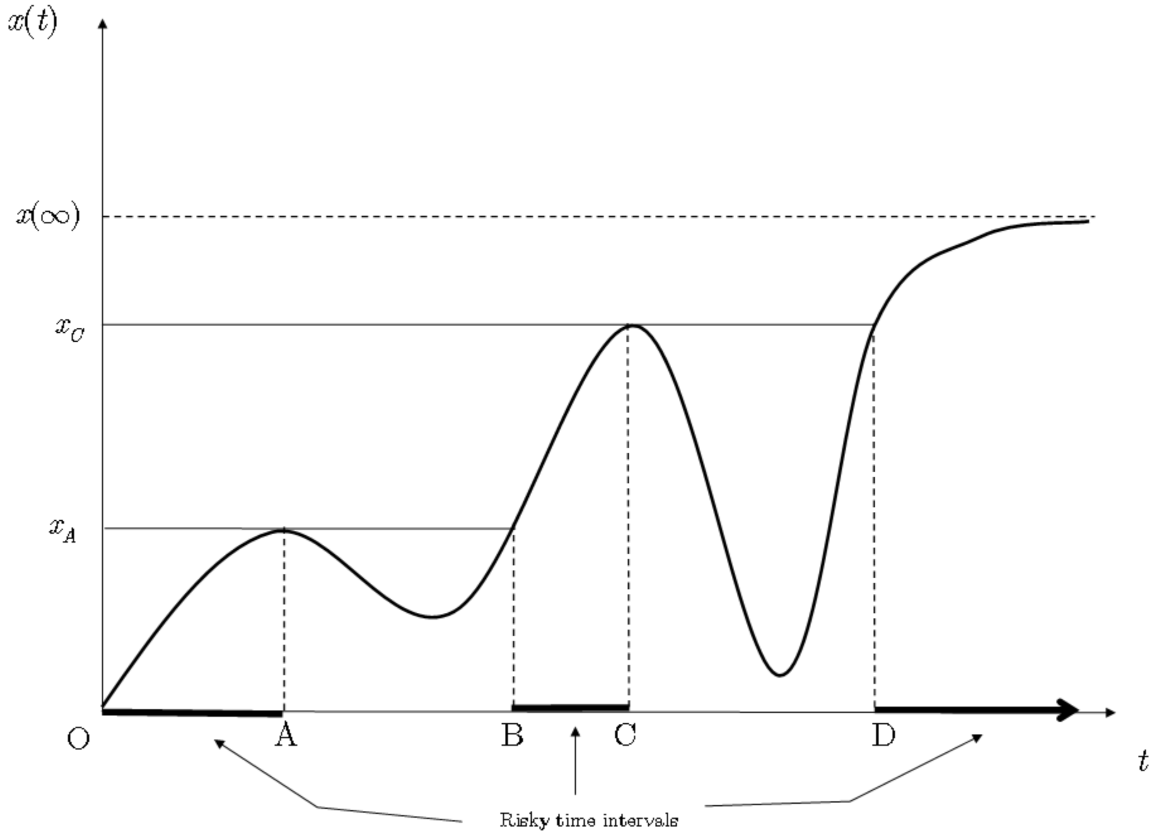

Figure 1, Threshold effects under uncertainty

In Figure 1, $x(t)$ converges towards $x(\infty)$ as time increases. $x(t)$ increases monotonously from time D, so there is always some probability that the threshold will be crossed in any given time interval. However, as the rate of increase in $x(t)$ becomes smaller and smaller, the probability per unit of time that the threshold will be crossed becomes smaller and smaller and goes to zero as time goes to infinity. The probability that the threshold will be crossed at some point in time is then $\int_{x_L}^{x(\infty)} f\left(x\right)dx$.

When optimizing processes with catastrophic risk it is often convenient to work with the hazard rate. The hazard rate of $f(x)$ is given by $\lambda_{\bar{x}}\left(x\right)$ and is defined by:

$$\lambda_{\bar{x}}\left(x\right) = \lim_{dx \to 0} \frac{\Pr\left(\bar{x} \in \left[x, x+dx\right) \mid \bar{x} > x\right)}{dx} = \frac{f\left(x\right)}{1 - \int_{x_L}^{x} f\left(y\right)dy} \tag{2}$$

For the purpose of optimization we need to transform this hazard rate to the time domain. It is shown in the appendix that the hazard rate in the time domain is given by:

$$\lambda_\tau\left(t\right)=\begin{cases}\lambda_{\bar{x}}\left(x(t)\right)\dot{x}(t) \text{ for } \dot{x}(t)\geq 0 \text{ and } x\left(t\right)\geq \sup_{s\in[0,t)} x\left(s\right)\\ 0 \text{ elsewhere}\end{cases} \tag{3}$$

This rather awkward definition holds for an arbitrary $x(t)$. It is shown below that along an optimal path, the hazard rate will simplify to $\lambda_{\bar{x}}\left(x\right)\times\max\left(0,\dot{x}\right)$. For ease of exposition we assume an exponential distribution for the threshold location, with a constant intensity λ, distributed over $[x(0), \infty)$. Thus the hazard rate is:

$$\lambda_\tau\left(t\right)=\lambda\times\max\left(0,u\left(t\right)-\delta x\left(t\right)\right) \tag{4}$$

The catastrophic event that occurs when $x(\tau) = \bar{x}$ is that society incurs a constant loss of utility flow given by G per unit of time.[1] This loss is assumed to be irreversible.[2] Formally we define a state-variable $\gamma(\tau)$, with $\gamma(0) = 0$, so that $\forall t \neq \tau$, $\frac{d\gamma(t)}{dt}=0$ and having a jump at the unknown τ, as given by $\gamma(\tau^+)-\gamma(\tau^-)=-G$. Finally, assume that the cost of emission reduction is given by:

$$C\left(u\right)=\frac{c}{2}\left(u^0-u\right)^2 \tag{5}$$

Here u^0 denotes the "business as usual" emission levels and represents the optimal emissions in the absence of environmental consequences. Setting u to u^0 implies that emission reduction costs are minimized and so u^0 may be thought of as the emissions in the absence of regulation and therefore an upper bound on emissions. In order to focus on the role of catastrophic risk, no other damages from CO_2 emissions are included in the model. In addition to the threshold effect, CO_2 is also assumed to have a stock pollutant effect with the marginal damage of the stock of CO_2 for simplicity assumed to be a. The principles of conventional economic analysis then lead to the following planning problem, where E is the expectation operator:

$$\max_{u(t)} E\left(\int_0^\infty\left(\gamma\left(t\right)-ax-\frac{c}{2}\left(u^0-u\right)^2\right)e^{-rt}dt\right) \tag{6}$$

[1] The assumption that the disaster gives rise to a constant flow of disutility is not crucial as it is always possible to replace the integral for net present value of actual damages with an equivalent annuity of damages. Furthermore, the chosen hazard rate λ_τ implies a rather "optimistic" view as to the occurrence of the catastrophe. A more realistic approach would require this hazard rate to increase with x.

[2] We have irreversible consequences of climate change as opposed to e.g. [13] where a regulator chooses an irreversible action.

subject to (1), (4), and (7) below, with $x(0)$ given, with $u \in \left[0, u^0\right]$ for all t, and r as the much maligned rate of time preference which we from hereon will simply term the discount rate. Note that the model employs traditional exponential discounting with a constant utility discount rate and no risk aversion. The state variable γ satisfies:

$$\dot{\gamma}(t) = 0 \ \forall t \neq \tau, \gamma\left(0\right) = 0, \ \gamma\left(\tau^+\right) - \gamma\left(\tau^-\right) = -G \tag{7}$$

2.1. Optimal Stabilization Targets

The solution to the optimization problem in (6) is an optimal path of emissions and a corresponding time path of CO_2, contingent on the threshold effect not occurring. These paths are to be chosen as long as the threshold is not crossed. In order to calculate these paths we also need to calculate optimal paths contingent on the occurrence of the catastrophe. A general algorithm for solving threshold problems may be found in [10], based on a general algorithm for piecewise deterministic control problems derived in [14]. The solution is found recursively. First one solves the problem conditional on the threshold effect having occurred at some point in time τ. This problem is given by:

$$J\left(\tau, x\left(\tau\right)\right) = e^{r\tau} \max_{u(s)} \left(\int_{\tau}^{\infty} \left(-G - ax - \frac{c}{2}\left(u^0 - u\right)^2 \right) e^{-rs} ds \right) \tag{8}$$

Note that we here have scaled the objective in order to get the maximum expressed in current value terms. This expression is maximised subject to $\dot{x} = u - \delta x$ and that $x(\tau)$ has some arbitrary value. Note that the problem is now deterministic and that the magnitude of constant G will not affect the solution. The solution to (8) is straightforward to solve with standard control techniques.

$$\begin{aligned} u\left(s \mid \tau, x\left(\tau\right)\right) &= u^0 - \frac{a}{c\left(r + \delta\right)} \\ \mu\left(s \mid \tau, x\left(\tau\right)\right) &= -\frac{a}{r + \delta} \\ x\left(s \mid \tau, x\left(\tau\right)\right) &= \frac{u^0 c\left(r + \delta\right) - a}{c\delta\left(r + \delta\right)} \end{aligned} \tag{9}$$

Here $\mu(s|\tau, \ x(\tau))$ is the standard current value co-state variable. We will also need the expression for $J\left(\tau, x\left(\tau\right)\right)$, which by integrating (8) after inserting from (9) is found to be:

$$J\left(\tau, x\left(\tau\right)\right) = -\frac{a}{r+\delta}x\left(\tau\right) + \frac{a^2 - 2acu^0(r+\delta)}{2cr(r+\delta)^2} - \frac{G}{r} \tag{10}$$

Having characterized the optimal solution after the threshold has been crossed, we may proceed to solve for the optimal emissions path prior to crossing the threshold. Here we will need the expression for $\mu\left(s \mid \tau, x\left(\tau\right)\right)$ in (9) and $J\left(\tau, x\left(\tau\right)\right)$ (10). In particular we will use $\mu\left(t \mid t, x\left(t\right)\right)$ and $J\left(t, x\left(t\right)\right)$ which is interpreted as the shadow price on x and the value function respectively, conditional on crossing the threshold at t. The solution is expressed in terms of a risk-augmented Hamiltonian given by:

$$H = \gamma - ax - \frac{c}{2}\left(u^0 - u\right)^2 + \mu(u - \delta x) + \lambda_\tau\left(t\right)\left(J\left(t, x\left(t\right)\right) - z\left(t\right)\right) \tag{11}$$

Here $\lambda_\tau(t)$ is the hazard rate defined in (3). Note that t in $J\left(t, x\left(t\right)\right)$ now denotes running time. $z(t)$ is an auxiliary variable which has the interpretation of being the value of the objective function evaluated from time t, conditional on the threshold not being crossed at that any time less than or equal to t. The term $J\left(t, x\left(t\right)\right) - z\left(t\right)$ is thus the net cost of the threshold being crossed at time t.

$$u = u^0 + \frac{\mu}{c} + \frac{\lambda}{c}\left(J\left(t, x\left(t\right)\right) - z\right) \tag{12}$$

$$\dot{\mu} = r\mu - \frac{\partial H}{\partial x} = a + \left(r+\delta\right)\mu + \lambda(u - \delta x)\left(\mu - \mu\left(t \mid t, x\left(t\right)\right)\right) + \lambda\delta\left(J\left(t, x\left(t\right)\right) - z\right) \tag{13}$$

$$\dot{z} = rz + ax + \frac{c}{2}\left(u^0 - u\right)^2 - \lambda\left(u - \delta x\right)\left(J\left(t, x\left(t\right)\right) - z\left(t\right)\right) \tag{14}$$

After inserting for $\mu\left(t \mid t, x\left(t\right)\right)$ from (9) and $J\left(t, x\left(t\right)\right)$ from (10), equations (1), (12) – (14), coupled with appropriate transversality conditions define the optimal paths, possibly only prior to possibly crossing the threshold. Setting time derivatives equal to and solving these equations along with (1) for u, x, μ and z gives the steady state solution. The solution for x is the may be interpreted as an Optimal Stabilization Target (OST) above which CO_2 should not be allowed to increase. This level is given by:

$$x^{ss} = \lim_{t\to\infty} x\left(t\right) = \frac{u^0}{\delta} - \frac{a}{c\delta\left(r+\delta\right)} + \frac{1}{\delta\lambda}\left(\left(r+\delta\right) - \sqrt{\left(r+\delta\right)^2 + 2G\frac{\lambda^2}{c}}\right) \tag{15}$$

Emissions will converge to:

$$u^{ss} = \lim_{t\to\infty} u\left(t\right) = u^0 - \frac{a}{c\left(r+\delta\right)} + \frac{1}{\lambda}\left(\left(r+\delta\right) - \sqrt{\left(r+\delta\right)^2 + 2G\frac{\lambda^2}{c}}\right) \tag{16}$$

The steady state stock of CO_2 may be decomposed in the following manner:

$$x^{ss} = \lim_{t\to\infty} x(t) = \frac{u^0}{\delta} + \Delta x_a^{ss} + \Delta x_G^{ss} \text{ where}$$
$$\Delta x_a^{ss} = -\frac{a}{c\delta(r+\delta)},\ \Delta x_G^{ss} = \frac{1}{\delta\lambda}\left((r+\delta) - \sqrt{(r+\delta)^2 + 2G\frac{\lambda^2}{c}}\right) \tag{17}$$

Δx_a^{ss} and Δx_G^{ss} are the respective steady state changes in CO_2 stock due to the stock pollutant effect and the threshold effect. Note that both these terms are strictly negative. Δx_G^{ss} has some intuitive properties. E.g.:

$$\lim_{\lambda\to 0} \Delta x_G^{ss} = \lim_{G\to 0} \Delta x_G^{ss} = 0 \tag{18}$$

If there is almost no risk or the cost of crossing the tipping point is close to zero, then the reduction in steady state stock of atmospheric CO_2 due to the threshold effect goes to zero. These steady states values may be interpreted as stabilization targets. However, some care must be taken when interpreting these steady state values. First, it is only optimal to let $x(t)$ and $u(t)$ converge to x_{ss} and u_{ss} if $x(0) \leq x_{ss}$. Also note that for some parameter values, e.g. sufficiently high values of G, the steady state levels will become negative. Obviously this is not realistic. Indeed, according to the following proposition it is never optimal to let $x(t)$ be decreasing over any time interval. These assertions are formally proven in Propositions 1 and 2.

Proposition 1.
Suppose that $x(0) \leq u^0/\delta + \Delta x_a^{ss}$. The optimal solution will then exhibit a non-decreasing path for the stock variable $x(t)$.

The proof of this proposition is given in the appendix. Intuitively, the result follows from the existence of a threshold effect. In the present model, the environmental damage occurs only if the threshold is crossed. If x^* is the highest level of x that has previously occurred, then it is known that all values of $x < x^*$ are below the threshold and therefore safe. There is therefore no incentive to reduce x below x^*. A corollary to Proposition 1 is given in Proposition 2.

Proposition 2.

Suppose we have $u^0/\delta + \Delta x_a^{ss} > x(0) \geq x_{ss}$. Then the optimal path requires that $x(t) = x(0)$ for all t and that the optimal control should take the value $u(t) = \delta x(0)$ for all t.

Proof: The proof is quite simple. Proposition 1 rules out the possibility of $x(t)$ oscillating or decreasing, so if Proposition 2 is false, $x(t)$ must strictly increasing and non-convergent for all t or converge to some steady state in the interval $(x_{ss}, u^0/\delta + \Delta x_a^{ss})$. $x(t)$ cannot be strictly increasing and non-convergent as this would imply that $u(t)$ at some point increases to levels above u^0, which is not optimal. Nor can $x(t)$ converge to a steady state in $(x_{ss}, u^0/\delta + \Delta x_a^{ss})$ as no such steady state exists. ■

Intuitively, Proposition 2 says that if the system is not regulated until after $x(t)$ has increased above the desired stabilization level implied by (15), then this stabilization level loses its relevance. By luck one has been able to reach a stock level of $x(t)$ that is too high from an optimality perspective and can therefore enjoy the decreased costs from emission reductions that is induced by this luck. Having had this luck however, it does not pay to stretch it further by allowing even larger increases in $x(t)$ relative to x_{ss}.

2.2. Discounting and the Effect on Stabilization Targets

Evidently, the steady-state solutions in (15) and (16) depend on the discount rate. However, a closer examination shows that the discount rate affects Δx_a^{ss} and Δx_G^{ss} in very different ways. In Δx_a^{ss}, r enters the *denominator* multiplicatively as a very small number it is therefore not surprising that small changes in r may have a large impact on stabilization targets. In Δx_G^{ss}, however r enters the expressions additively in the numerator. Adding small numbers to a numerator will, roughly speaking, have a very small effect on a number. To see this, examine the terms within the parenthesis:

$$\left((r+\delta) - \sqrt{(r+\delta)^2 + 2G\frac{\lambda^2}{c}} \right) \tag{19}$$

Defining $r + \delta$ to be A and $2G\lambda^2 c^{-1}$ to be B, the non-positive expression in (13), may be written:

$$A - \sqrt{A^2 + B} \tag{20}$$

Let the unit of time be "one year". The annual discount rate is then typically lower than 0.07. $1/\delta$ is the average lifetime of a CO_2 molecule in the atmosphere. This number was popularly believed to be of the order of a few hundred years, but recent work indicates that it may be considerably higher, which implies that δ is at the very highest 1/200, but may be considerably smaller, see [1]. In any case, the number A is of the order of magnitude 10^{-1}. The number B depends on the ratio of the cost of catastrophe G and, roughly speaking, the cost of emission reduction c. If the catastrophe has consequences that are truly serious so that the number B is of an order of magnitude, say 10^6 or more, then B will clearly dominate the expression in (13). Indeed, the expression has A minus the root of the square of A plus something and will tend to disappear. We can formalize this by examining the respective elasticities of Δx_G^{ss} and Δx_a^{ss}.

$$\mathrm{El}_r \Delta x_a^{ss} = \frac{\frac{\partial}{\partial r}\left(\Delta x_a^{ss}\right)}{\Delta x_a^{ss}} r = -\frac{r}{r+\delta}, \qquad \mathrm{El}_r \Delta x_G^{ss} = \frac{\frac{\partial}{\partial r}\left(\Delta x_G^{ss}\right)}{\Delta x_G^{ss}} r = -\frac{r}{\sqrt{(r+\delta)^2 + \dfrac{2G\lambda^2}{c}}} \tag{21}$$

Remember that δ is at the most 1/200. To simplify, let us examine these elasticities when $\delta = 0$.

$$\mathrm{El}_r \Delta x_a^{ss} = -1, \qquad \mathrm{El}_r \Delta x_G^{ss} = -\frac{r}{\sqrt{r^2 + \dfrac{2G\lambda^2}{c}}} \tag{22}$$

The difference is quite striking. If we only concern ourselves with the deterministic stock pollutant effect, a 1% increase in r, say from 5% to 5.05% would imply that steady state CO_2 stocks should be allowed to increase by 1%. In the present model, this implies that an increase in r by one percentage point implies a decrease in reductions of 20%! On the other hand, if we are concerned only about the threshold effect, the elasticity $\mathrm{El}_r \Delta x_G^{ss}$ is a small negative number. Indeed, if B is a number of some magnitude, $\mathrm{El}_r \Delta x_G^{ss}$ is for practical purposes indistinguishable from 0.

It should be clear from this discussion that the discount rate does not matter much for what level one should stabilize atmospheric CO_2 if one is primarily concerned with tipping points or threshold effects. As the probability of crossing the threshold is given by the integral $\int_{x(0)}^{x_{ss}} f(x)dx$, this probability is not very dependent on the discount rate either. Any fruitful scientific and economic discussion about this topic should therefore focus on the magnitude of the parameters G, c and λ. This does not imply that the interest rate is completely insignificant. The path of emissions and atmospheric CO_2 leading up to the stabilized levels in (15) and (16) will in general be sensitive to changes in interest rates, but for the determination of the actual stabilization targets, the discount rate plays a minor role.

3. Summary

The debate between proponents of conventional discounting and sceptics concerned about catastrophic risk is somewhat misplaced as the role of discounting in catastrophic risk is minor if the threshold nature of the risk structure is accounted for. To the extent that threshold effects are important in climate change, this should be incorporated into integrated assessment models and thereby conciliate the results of these models with the concerns of climate scientists.

Appendix

Derivation of the hazard rate, λ_τ.

The threshold location is distributed over the interval (x_L, x_H) where $x_H \leq \infty$ with a pdf given by $f(x)$ and a cdf given by $F(x)$. By definition the hazard rate associated with $f(x)$ is given by:

$$\lambda_{\bar{x}}\left(x\right) = \lim_{dx \to 0} \frac{\Pr\left(\bar{x} \in \left[x, x + dx\right] \mid \bar{x} > x\right)}{dx} = \frac{f\left(x\right)}{1 - F\left(x\right)}$$

Now let $x(t)$ be an arbitrary continuous and piecewise differentiable function such that $x(0) = x_L$, $x'\left(t\right) = h\left(t, x\left(t\right)\right)$ in points of differentiability and let τ solve the equation $x(\tau) = \bar{x}$. If $h(t, x(t))$ is everywhere non-negative, if follows from a standard property of the integral operator that:

$$F\left(x\left(t\right)\right) = \int_{x_L}^{x(t)} f\left(y\right) dy = \int_0^t f\left(x\left(s\right)\right) x'\left(s\right) ds = \int_0^t f\left(x\left(s\right)\right) h\left(s, x\left(s\right)\right) ds$$

If $h(t, x(t))$ is not everywhere non-negative, we must avoid assigning positive probability to time intervals where x take values known to be safe. This is done by defining a function $\psi(t)$ with the property that:

$$\psi\left(t\right) = \begin{cases} h\left(t, x\left(t\right)\right) & \text{for } h\left(t, x\left(t\right)\right) \geq 0 \text{ and } x\left(t\right) \geq \sup\limits_{s \in [-\infty, t)} x\left(s\right) \\ 0 \text{ otherwise} \end{cases}$$

The cdf for the distribution of the event $t = \tau$ is then given by:

$$F\left(x\left(t\right)\right) = \int_0^t f\left(x\left(s\right)\right) \psi\left(s\right) ds$$

The corresponding pdf is then given by:

$$f_\tau\left(t\right) = f\left(x\left(t\right)\right) \psi\left(t\right)$$

It follows from the definition of the hazard rate that the hazard rate for the point in time of event occurrence is given by:

$$\lambda_\tau = \frac{f\left(x\left(t\right)\right) \psi\left(t\right)}{1 - \int_0^t f\left(x\left(s\right)\right) \psi\left(s\right) ds} = \frac{f\left(x\left(t\right)\right) \psi\left(t\right)}{1 - F\left(x\left(t\right)\right)}$$

Proof of Proposition 1

To avoid cluttered notation we show the proposition under the assumption that $a = 0$. If the proposition is false, then one of the following conditions must hold:

Condition 1: There must exist a t^* such that $x(t) < x(t^*)$ for all $t > t^*$.

Condition 2: There must exist a t^* and $t^{**} > t^*$ such that $x(t^*) = x(t^{**})$ and $x(t) < x(t^*)$ for all $t \in (t^*, t^{**})$.

Bear in mind that emissions will never exceed u^0 implying that x will never exceed u^0/δ. If Condition 1 holds, then $\psi(t) = 0$ for all $t > t^*$. If this is the case, then the optimal path must solve the deterministic control problem

$$\max_u \int_{t^*}^{\infty} [-\tfrac{c}{2}\left(u^0 - u\right)^2]e^{-rt}dt \quad s.t. \quad \dot{x} = u - \delta x, \;\; x\left(t^*\right) \text{ given}$$

It is straightforward to see that this problem has the unique solution $u(t) = u^0$. For all $x(t^*) \leq u^0/\delta$, $x(t)$ will therefore be increasing; hence we have a contradiction.

If Condition 2 holds, optimality implies that the optimal path over $[t^*, t^{**}]$ solves the following optimization problem:

$$\max_{u(t)} \int_{t^*}^{t^*+\Delta} [-\frac{c}{2}\left(u^0 - u\right)^2]e^{-rt}dt, \;\; s.t. \; \dot{x} = u - \delta x, \; x\left(t^*\right) = x\left(t^{**}\right) \text{given}$$

Here $\Delta = t^{**} - t^*$. This is again a straightforward deterministic optimal control problem. Solving this problem yields that, for any Δ, $x(t) = x(t^*)$ for all $t \in \left[t^*, t^{**}\right]$, implied by a constant emission rate $u = \delta x(t^*)$ for any $t \in \left[t^*, t^{**}\right]$; which contradicts our assumption.

■

References

[1] Archer, D., (2007), Methane hydrate stability and anthropogenic climate change, *Biogeosciences*, **4**, pp:521–544.

[2] Archer, D. (2005), Fate of fossil fuel CO2 in geologic time, *Journal Of Geophysical Research*, **110**, C09S05.

[3] Dasgupta, P. (2008), Discounting climate change, *Journal of Risk and Uncertainty*, In Press, 10.1007/s11166-008-9049-6.

[4] Dasgupta, P. (2007), A challenge to Kyoto, *Nature*, **449**, 143 - 144 (12 Sep 2007).

[5] Gollier, C., (2002), Discounting an uncertain future, *Journal of Public Economics* **85**, 149-166.

[6] Gollier, C., (2009), Should we discount the far-distant future at its lowest possible rate?, Economics, The Open-Access, Open-Assessment E-Journal, Discussion paper 7, January 9th.

[7] Hoegh-Guldberg, O., (1999), Climate change, coral bleaching and the future of the world's coral reefs, *Marine and Freshwater Research* **50**, pp. 839–866.

[8] Manabe, S., and R.J. Stouffer, (1995), Simulation of Abrupt Climate Change Induced by Freshwater Input to the North Atlantic Ocean, *Nature*, 378:165-167.

[9] Nordhaus, W. (2007), The Stern Review on the Economics of Climate Change, unpublished manuscript available at:http://nordhaus.econ.yale.edu/stern_050307.pdf

[10] Naevdal, Eric, (2006), Dynamic optimisation in the presence of threshold effects when the location of the threshold is uncertain - with an application to a possible disintegration of the Western Antarctic Ice Sheet, *Journal of Economic Dynamics and Control, Elsevier*, **30**, 7:1131-1158.

[11] Naevdal, E. & Michael Oppenheimer (2007), The economics of the thermohaline circulation - A problem with multiple thresholds of unknown locations, *Resource and Energy Economics*, **29**,4:262-283.

[12] Oppenheimer, M. (1998), Global warming and the stability of the West Antarctic ice sheet, *Nature* **393**:pp. 325–332.

[13] Pindyck R.S. (2000), Irreversibilities and the timing of environmental policy, *Resource and Energy Economics*, **22**, 233–259.

[14] Seierstad, A. (2008), Stochastic Control in Discrete and Continuous Time, Springer-Verlag New York.

[15] Stern, N. (2007). The Economics of Climate Change: The Stern Review. Cambridge University Press.

[16] Tsur, Y and A. Zemel (1996), Accounting for Global Warming Risks: Resource Management under Event Uncertainty, *Journal of Economic Dynamics and Control*, **20**, 1289 – 1305.

[17] Tsur, Y & Zemel A. (1998), Pollution Control in an Uncertain Environment, *Journal of Economic Dynamics and Control*, **22**, 967-975.

[18] Tsur, Y & Zemel A. (2009), Endogenous Discounting and Climate Policy, *Environmental and Resource Economics*, **44**, 4: 507-520.

[19] Weitzman, M.L., (1994), On the "Environmental" Discount Rate, *Journal of Environmental Economics and Management,* **26**, 200-209.

[20] Weitzman, M.L., (1998), Why the Far-Distant Future Should Be Disconnected at Its Lowest Possible Rate, *Journal of Environmental Economics and Management,* **36**, 201-208.

[21] Weitzman, M. L. (2007), A Review of the Stern Review on the Economics of Climate Change, *Journal of Economic Literature*, **45**, 3:703-724.

[22] Weitzman, M. L. (2009), On Modeling and Interpreting the Economics of Catastrophic Climate Change, Forthcoming in: *The Review of Economics and Statistics* 91:1.

Appendix – Piecewise Deterministic Optimal Control of Poisson Processes.

This appendix presents necessary conditions for Piecewise Deterministic Optimal Control problems. The conditions presented here are due to Seierstad (2008). Similar expositions to this one may be found in [10] or [11] referenced in the main text. Although alternative, but equivalent, formulations exist in the literature this method is to our knowledge the most general. In addition, this formulation has two advantages that other formulations do not have.

1. The Hamiltonian and co-state variables have interpretations that are equivalent to the interpretation of these quantities in deterministic control theory.
2. The necessary conditions often take the form of *autonomous* differential equations. This facilitates steady state analysis.

The general problem to be studied is:

$$J\left(0,x\left(0\right)\right)=\max_{u\in U}E\left(\int_0^T f\left(x,u\right)e^{-rt}dt\right) \tag{A.1}$$

$$s.t: u\subseteq\mathbb{R}^m,\ x\left(0\right)\subseteq\mathbb{R}^n, \dot{x}=g\left(x,u\right) \tag{A.2}$$

$$\tau\sim\lambda\left(x\left(t\right)\right)e^{-\int_0^t\lambda\left(x\left(\sigma\right)\right)d\sigma}\text{ over }[0,\infty) \tag{A.3}$$

$$x\left(\tau^+\right)-x\left(\tau^-\right)=q\left(x\left(\tau^-\right)\right) \tag{A.4}$$

All functions are assumed to be twice differentiable. The interpretation of this problem is that of controlling a process that yields instantaneous utility $f(\cdot)$ over some time span. The state variable, x, is controlled by choosing a control u. There is a Poisson process going on in the background distributed over time. This process has a hazard rate given by $\lambda(x(t))$. If or when, the random event driven by the Poisson process occurs at a time τ there is a shock to the state variable given by $x\left(\tau^+\right)-x\left(\tau^-\right)=q\left(x\left(\tau^-\right)\right)$.

$$H=f\left(x,u\right)+\mu g\left(x,u\right)+\lambda\left(x\right)\left(J\left(t,x+q\left(x\right)\mid\tau=t\right)-J\left(t,x\right)\right) \tag{A.5}$$

This Hamiltonian differs from the Hamiltonian from deterministic control theory only by the term $\lambda(x)(J\left(t, x+q\left(x\right) \mid \tau=t\right)-J\left(t, x\right))$. $J(t,\ x)$ is defined by the solution to problem:

$$J\left(t, x\right)=\max_{u \in U} E\left(\int_{t}^{T} f\left(y, u\right) e^{-r\left(s-t\right)} ds\right) \tag{A.6}$$

$$s.t: u \subseteq \mathbb{R}^{m},\ y\left(t\right)=x \subseteq \mathbb{R}^{n}, \dot{y}=g\left(y, u\right) \tag{A.7}$$

$$\sigma \sim \lambda\left(x\left(s\right)\right) e^{-\int_{0}^{s} \lambda\left(x\left(\varsigma\right)\right) d\varsigma} \text{ over } [t, \infty) \tag{A.8}$$

$$x\left(\sigma^{+}\right)-x\left(\sigma^{-}\right)=q\left(x\left(\sigma^{-}\right)\right) \tag{A.9}$$

This problem is exactly the same as the problem posed in Equations (A.1) - (A.4) except that the problem starts from an arbitrary point $(t,\ x)$. $J(t,\ x)$ is thus the value to the objective function when the problem starts from some arbitrary point in $(t,\ x)$ space. The term $J\left(t, y \mid \tau=t\right)$ is defined by:

$$J\left(t, x \mid \tau=t\right)=\max_{u \in U} \int_{t}^{T} f\left(y, u\right) e^{-r\left(s-t\right)} ds \tag{A.10}$$

$$s.t: u \subseteq \mathbb{R}^{m},\ x \subseteq \mathbb{R}^{n}, y=g\left(y, u\right) \tag{A.11}$$

This problem differs from the one posed in Equations (A.1) - (A.4) in two respects. The problem is a deterministic problem and the starting point is an arbitrary point in $(t,\ x)$ space after the shock has happened. In order to solve the problem in equation (A.1) one must find a solution to (A.10). The solution to (A.10) will be a function $y\left(s \mid t, x\right)$, a control $u\left(s \mid t, x\right)$ and a co-state $\mu\left(s \mid t, x\right)$. It is clear that $J(t,\ x \mid \tau = t) = \int_{t}^{\infty} f\left(y\left(s \mid t, x\right), u\left(s \mid t, x\right)\right) e^{-rs} ds$ is the value of criterion after a shock has driven the system to some arbitrary state x at time t. $J(t,\ x \mid \tau = t)$ is thus the criterion conditional on the event τ occurring at time t. The interpretation of $\left(J\left(t, x+q\left(x\right) \mid \tau\right)-J\left(t, x\right)\right)$ should now be clear. It is the net loss (or gain) to the objective system if the shock occurs at an arbitrary point in time t and results in the state variable taking the value x. Now apply the maximum principle to the Hamiltonian in (A.5). Doing so yields the following conditions:

$$u=\arg\max_{y}\left(H\right) \tag{A.12}$$

$$
\begin{aligned}
&\dot{\mu} = r\mu - \frac{\partial H}{\partial x} = r\mu - f_x'\left(x,u\right) - \mu g_x'\left(x,u\right) + \\
&\lambda\left(x\right)\left(\frac{\partial}{\partial x}J\left(t,x\right) - \frac{\partial}{\partial x}J\left(t,x+q\left(x\right)\mid\tau\right)\right) + \lambda'\left(x\right)\left(J\left(t,x\right) - J\left(t,x+q\left(x\right)\mid\tau\right)\right)
\end{aligned}
\tag{A.13}
$$

Coupled with the appropriate transversality condition, the solution is determined by the equation for $\dot{x}$, (A.17) and (A.18). It follows from standard results in deterministic control theory that:

$$
\frac{\partial}{\partial x}J\left(t,x+q\left(t,x\right)\mid\tau\right) = \mu\left(t\mid t,x+q\left(x\right)\right)\left(I^n + q_x'\right) \tag{A.14}
$$

Here I^n is the n-dimensional identity matrix. The final piece of information required to solve the problem in (A.1) is an expression for $J(t, x)$, as this expression and an expression for $\frac{\partial}{\partial x}\left(J\left(t,x\right)\right)$ are needed in order to solve (A.13).

To find an expression for $J(t, x)$, define the following differential equation:

$$
\dot{z} = rz - f\left(x,u\right) + \lambda\left(x\right)\left(z - J\left(t,x+q\left(t,x\right)\right)\right) \tag{A.15}
$$

The solution to (A.15) is a function $z(t)$ that is equal to $J(t, x(t))$ along the optimal path. Seierstad (2003) has proven that:

$$
\frac{\partial}{\partial x}J\left(t,x\right) = \mu\left(t\right) \tag{A.16}
$$

Rewriting (A.17) and (A.18), using (A.14), (A.16) and exchanging $J(t, x)$ with z gives:

$$
u = \arg\max_{u'}\left(f\left(x,u'\right) + \mu g\left(x,u'\right) + \lambda\left(x\right)\left(J\left(t,x+q\left(t,x\right)\mid\tau\right) - z\right)\right) \tag{A.17}
$$

$$
\begin{aligned}
&\dot{\mu} = r\mu - \frac{\partial H}{\partial x} = r\mu - f_x'\left(x,u\right) - \mu g_x'\left(x,u\right) - \\
&\lambda\left(x\right)\left(\mu\left(t\mid t,x+q\left(x\right)\right)\left(I^n + q_x'\right) - \mu\right) - \lambda'\left(x\right)\left(J\left(t,x+q\left(x\right)\mid\tau=t\right) - z\right)
\end{aligned}
\tag{A.18}
$$

The differential equations in (A.15), (A.17), (A.18) and the differential equation $\dot{x} = f(x, u)$ gives the necessary conditions required to solve the problem at hand when coupled to the appropriate transversality conditions. For the case where $T < \infty$, the transversality conditions are given by:

$$\mu\left(T\right) = 0 \tag{A.19}$$

$$z\left(T\right) = 0 \tag{A.20}$$

Equation (A.19) is the transversality condition on the co-state. Paralleling the interpretation of the co-state variable in the deterministic problem, the interpretation is that at the end of the planning horizon, the marginal value of x is zero in the absence of any scrap value. The condition that $z(T) = 0$, is best understood by noting from the definition of $z(t)$ that $z(T) = J(T,\ x(T))$. Thus, $z(T)$ is the "remaining" utility to be consumed at the end of the planning horizon and equal to zero. If $T = \infty$, then as long as instantaneous utility is bounded, the following conditions will usually work and be consistent with Catching Up Optimality. If x is the optimal path, then for all admissible paths y satisfying $u \in U$ and $\dot{y} = g(y,\ u)$.

$$\lim_{t\to\infty} \mu e^{-rt}\left(y\left(t\right) - x\left(t\right)\right) \geq 0 \tag{A.21}$$

$$\lim_{t\to\infty} z\left(t\right)e^{-rt} = 0 \tag{A.22}$$

These conditions are required to take care of some special cases that turn up in infinite horizon models. These conditions may often be replaced by $\mu(\infty) = z(\infty) = 0$. In particular, this is the case if the steady state is unique.[3] If the limit in equation (A.21) does not exist, which will only be the case in very rare problems, the lim operator must be replaced by lim inf.

[3]The issues involved here are parallel to the problems encountered in deterministic control theory. See Seierstad and Sydsæter (1987), pp 229-250.

References for Appendix for Reviewers

Seierstad and Sydsæter, 1987 A. Seierstad and K. Sydsæter, Optimal Control Theory with Economic Applications, North-Holland, Elsevier Science, Amsterdam (1987).

Seierstad, A. (2008), Stochastic Control in Discrete and Continuous Time, Springer-Verlag New York.