The Economics of the Thermohaline Circulation – A Problem with Multiple Thresholds Of Unknown Location*

Eric Nævdal

Department of Economics and Resource Management

Norwegian University of Life Sciences

and

Michael Oppenheimer

Woodrow Wilson School, Princeton University.

Running title: The Economics of the Thermohaline Circulation

---

* Correspondence to: Eric Nævdal, Department of Economics and Resource Management, Norwegian University of Life Sciences, PO Box 5003, N-1432 Aas, Norway. E-mail: eric.navdal@umb.no. Telephone: (+47) 64 96 57 00. Fax: (+47) 64 94 30 12. This paper was written while Eric Nævdal was a post-doctoral fellow at Woodrow Wilson School, Princeton University. Funding from the Princeton Carbon Mitigation Initiative is gratefully acknowledged.

The Economics of the Thermohaline Circulation – A Problem with Multiple Thresholds Of Unknown Location

Abstract

A potentially serious environmental threat facing humanity is the possibility of a collapsing thermohaline circulation. The collapse of the thermohaline circulation may be triggered by temperature and/or rate of temperature change exceeding certain thresholds. The locations of these thresholds are unknown. Economic regulation of this problem requires solution methods for a class of dynamic optimisation problems with multiple thresholds located in n-dimensional space. This class of problems has hitherto not been discussed in the literature. We present a model for the economic regulation of CO_2 emissions in the presence of threshold-triggered risk of a collapsing thermohaline circulation.

Keywords: Optimal Control, Catastrophic Risk, Threshold effects, Climate Change.

JEL classification: C61, Q25.

Introduction

One of the many risks associated with global warming is the possibility of a collapse in the Thermohaline Circulation, (THC), which is one component of the Meriodonal Overturning Circulation (MOC). The consequences of such a collapse are at present hard to estimate. However, dramatic effects may occur on a global scale, although with considerable regional variations, [1] and [21] . Rapid (decadal scale) changes in Greenland temperatures on the order of several degrees are associated with shutdown of the THC, as indicated in ice core records. Association of THC shutdown with global scale climate changes is less clear in other paleo-climate records. The effect of THC changes on temperature arises from changes in the hydrological cycle at high latitudes, particularly higher precipitation and melting of land and sea ice, which reduce surface water density. Warming of surface waters also plays a role [7] .

The effect that future increases in CO_2 will have on the THC is uncertain. Whether the past provides a useful analogous is questionable due to the current absence of the large ice sheets of the past glacial epoch. Experiments with simplified forcing, such as arbitrary increases in freshwater flux ("hosing"), can induce complete shutdown [11] . In simplified ocean models, increases in radiative forcing corresponding to levels anticipated for late in this century or early in the next can also cause shutdown of the circulation [16] . More realistic general circulation models forced by gradually increasing radiative forcing generally lead to a slowdown of the circulation within this century, although one respected model produces no such effect [8] . GCM simulations beyond the 21st century with one model show either slowdown or recovery on a multi-century timescale or total, permanent shutdown of the THC beyond forcing equivalent to a doubling of pre-industrial carbon dioxide [10] . In any event, impacts of THC shutdown could include a dramatic but absolute cooling (versus today) in a restricted area of the North Atlantic basin and northwest Europe and hydrological changes elsewhere but little effect on global mean temperature [22] .

In summary, for the purposes of this paper we hypothesize that either the THC will "flicker," and recover fully or partially in the course of centuries if for sufficient forcing, or the THC will shut down and remains shut down for millennia, [10] , [16] , [24] and [15] . Here we will focus on the latter prospect. Simple models indicate that the THC collapse depends not only on the level of temperature caused by greenhouse gases, but also on how fast temperatures rise. If temperatures increases on average by, say 5° C over 500 years, the THC may be sustained. However if the same increase occurred over a century this could trigger a complete collapse. The reason is that both the temperature level and the rate of temperature change affect the budget of freshwater discharge in the North Atlantic, which is a critical component in how global warming affects the THC.[1] The analysis in the cited papers strongly suggests that there are multiple threshold effects present in the dynamics of the THC. If temperature, temperature change or both become too high, a THC shutdown may result.

The term 'threshold' as used in resource economic analysis has at least three interpretations. The first is related to the existence of Skiba points (or surfaces) in the Hamiltonian system that defines optimality conditions, [17] . A second type is concerned with regulatory thresholds where regulatory action is triggered by certain variables such as return on some environmental asset, [3] . The present paper examines a third type of threshold problem, which is concerned with physical shocks that occur when surfaces in state space are crossed. Such shocks have been analysed in both deterministic and stochastic settings. Deterministic analysis of one-dimensional thresholds with known location may be found in [6] , [12] and [13] . Here we consider the case where we do not know the location of the thresholds. Thresholds with unknown locations have been analysed by, among others, [4] , [18] , [19] and [13] . To our knowledge, this literature on

[1] It is unlikely that temperature growth rates above the threshold for a short duration will trigger a THC collapse. The underlying rationale for rather sharp formulation in the present paper is that there is inertia to the way emissions translate into temperature change. If temperature change crosses the threshold, the threshold is likely to be crossed for a significant amount of time.

threshold effects is only concerned with the case where the threshold is a single point, distributed along some subset of the real line.[2] The only exception, as far as we know, is [14] which discuss a problem with a single multidimensional threshold. However, problems of the type studied in the present paper where there are more than one threshold distributed in several dimensions has as yet not been studied. The present paper adds to the economic toolkit by studying a problem where there are two thresholds distributed in two dimensions. The steps needed to generalise to a problem with k thresholds distributed in n dimensions are straightforward.

The type of risk management problem examined in this article may be categorized as catastrophic risk. There are two classes of catastrophic risk. Most of the literature analysing this type of risk is concerned with catastrophic events distributed over time, see e.g. [3] and [5] . In the present paper, as well as in the other articles that discuss the present type of thresholds, the events are distributed in state space. The distinction between the two types is discussed in [20] where the two types are referred to as endogenous and exogenous risk and in [13] where the thresholds are referred to as distributed in state-space as opposed to distributed in time. The key to solving the threshold problems is to recognize that state variables are functions of time. The distributions of the thresholds can then be translated into a single distribution over time. However, this transformation is not trivial when thresholds are multi-dimensional and there is more than one threshold. After transforming the risk structure, the threshold risk can be analysed with the same tools as catastrophic events distributed over time.

[2] The cited references are concerned with natural resource economics. In other strands of the economic literature, other applications are discussed, in particular in R&D. This literature is also concerned with single, one-dimensional thresholds. See e.g. [9] and [23] .

There is a technical detail that must be mentioned. With the exception of [13] and [14] , the literature concerned with catastrophic risk examines shocks to utility. The alternative approach used here is to examine shocks to a subset of state variables. The method, developed by Atle Seierstad, University of Oslo, is described in the appendix. The present set-up has three advantages to more common methods. First, it is slightly more general. Second, the method often leads to autonomous differential equation, which simplifies the calculations. Third, the economic interpretation of co-state variables is similar to co-state variables in non-stochastic models.

A Climate Model with risk of THC collapse

The problem is explored within the following simple climate model. Let u be CO_2 emissions. Let c be the stock of atmospheric CO_2 measured by increase above pre-industrial levels. Let F be the global average increase in temperature measured by increase above pre-industrial levels. c and F are assumed to be driven by the following differential equations:

$$\dot{c} = u - \delta c \tag{1}$$

$$\dot{F} = ac - \gamma F \tag{2}$$

In Equation (1) atmospheric CO_2 increases with emissions. A fraction of atmospheric CO_2 is sequestered in carbon sinks. In Equation (2) the stock of carbon is translated into radiative forcing which is the flux that drives temperature increases. In accordance with thermodynamic laws, a fraction of the energy stored as heat in the atmosphere dissipates. Based on the above discussion, the existence of two thresholds is postulated. If F ever becomes larger than or equal to y, then the THC is assumed to collapse. However, if $\dot{F}$ ever becomes larger than or equal to x, then the THC will also collapse. The numbers x and y are assumed to be realisations of random variables X and Y respectively

with known distributions $f(X)$ and $g(Y)$. The domain of $f(X)$ is assumed to be given by $[x_l, x_h] \subseteq \mathbb{R}$ and the domain of $g(Y)$ is assumed to be given by $[y_l, y_h] \subseteq \mathbb{R}$. To illustrate how this system may trigger a THC breakdown, a path starting at $c(0) = F(0) = 0$ is illustrated in Figure 1. In Figure 1, it is assumed that u is constant. The steady state values for F and c are $\frac{au}{\gamma\delta}$ and $\frac{u}{\delta}$ respectively are shown and the path will converge towards the steady state as indicated by the curve starting at the origin and ending at $(\frac{u}{\delta}, \frac{au}{\gamma\delta})$. The line $\dot{F} = 0$ is drawn to show the slope of the threshold defined by $\dot{F} = x$ as well as indicate a lower bound for that threshold. There are three possible values indicated for each threshold. If the true values of the thresholds are $F = y_3$ and $\dot{F} = x_3$, then the THC is safe as the path does not cross any thresholds.

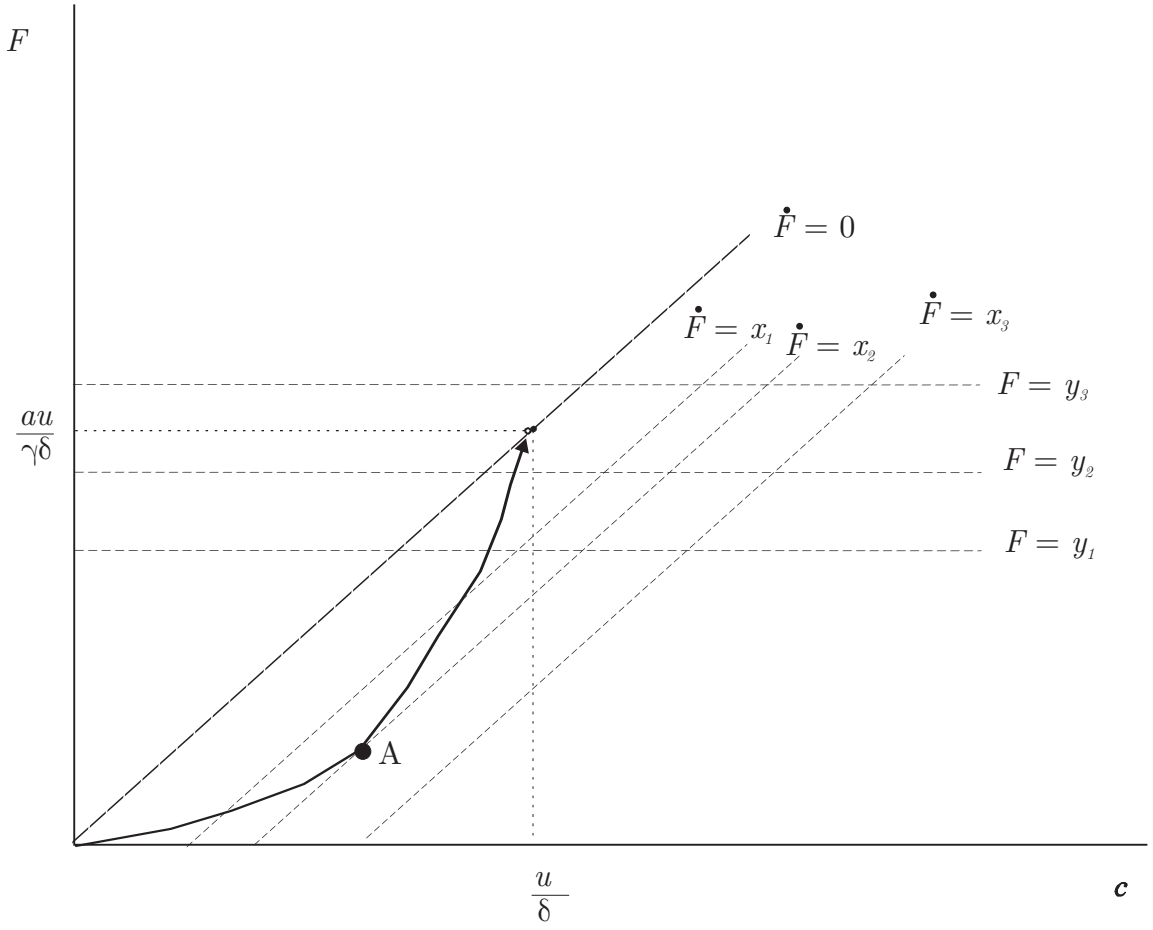


Figure 1, Possible threshold locations in (c, F) space.

If the true values of the thresholds are $F = y_2$ and $\dot{F} = x_2$, then only the temperature threshold will be crossed. If the $F = y_3$ and $\dot{F} = x_1$, then only the temperature growth rate threshold will be crossed. In both of these cases, the result will be the same. The THC collapses.

One important aspect of facing multiple thresholds is that they may not all be "active" in the sense that they may not all contribute to the "riskiness" of any particular path. In the path illustrated, one can see that as the state variables approach the point A, F and $\dot{F}$ are both increasing. There is therefore a risk that both thresholds may be crossed. However, after the point A has been reached $\dot{F}$ is decreasing. In a sense one is therefore going through possible values of $\dot{F}$ that are already known to be safe. The only threshold that is active along the path is the F threshold as F is everywhere increasing.

The question is now how to optimise a system that contains such a risk structure. Before we present conditions for optimality, this risk structure must be suitably transformed. The events $ac - \gamma F = x$ and $F = y$ are distributed in state space. However, because c and F are functions of time, these events must also be distributed over time. Let τ_1 be the point in time such that $ac(\tau_1) - \gamma F(\tau_1) = x$ and let τ_2 be the point in time defined by $F(\tau_2) = y$. It is assumed that these variables are statistically independent. Using standard properties of the integral, one can prove that:[3]

$$\tau_1 \sim f\left(ac(t) - \gamma F(t)\right) \upsilon_1(t) \tag{3}$$

$$\tau_2 \sim g(F)\, \upsilon_2(t) \tag{4}$$

The functions υ_1 and υ_2 are defined by:

[3] This transformation is discussed in detail in Nævdal (2003).

$$\upsilon_1\left(c(t),F(t)\right)=\begin{cases}a\dot{c}-\gamma\dot{F}\text{ for }a\dot{c}-\gamma\dot{F}\geq 0\text{ and }\dot{F}(t)\geq\sup\limits_{s\in[0,t)}\dot{F}(s)\\ 0\text{ elsewhere}\end{cases}\tag{5}$$

$$\upsilon_2\left(c(t),F(t)\right)=\begin{cases}\dot{F}\text{ for }\dot{F}\geq 0\text{ and }F(t)\geq\sup\limits_{s\in[0,t)}F(s)\\ 0\text{ elsewhere}\end{cases}\tag{6}$$

These definitions are required to ensure that no probability is assigned to the event $t=\tau_1$ or $t=\tau_2$ if F and $\dot{F}$ takes values that have already occurred. Note that υ_1 and υ_2 are both forward differentiable as long as they satisfy the upper part of their definitions. To simplify the discussion the following nomenclature is employed.

Definition

A threshold i is stochastically active if and only if $\upsilon_i>0$.

Whether a threshold is stochastically active or not will obviously impact on optimal policies. Indeed, as we shall see below, whether a threshold is active is in part endogenous. However, from some starting points, in (c, F) space one or both thresholds will be stochastically active regardless of the choice of u. This is because $\dot{F}$ depends on state variables that move only slowly. Thus certain values of c and F imply a commitment to accepting risk for some period of time. Strictly speaking, the distributions in Equations (3) and (4) are only proper probability functions if $\upsilon_1(t)$ spans the domain of $f(\cdot)$ and $\upsilon_2(t)$ spans the domain of $g(\cdot)$. Otherwise, probability mass must be assigned to the events that $\tau_1\notin[0,\infty)$ and $\tau_2\notin[0,\infty)$. Regardless of this possibility, the hazard rates associated with the two distributions are given by:

$$\mu_1\big(c(t),F(t)\big)=\lim_{dt\to 0}\Pr\big(\tau_1\in[t,t+dt]\,|\,\tau_1>t\big)dt\approx\frac{f\big(ac(t)-\gamma F(t)\big)\upsilon_1(t)}{1-\int_{x_l}^{ac(t)-\gamma F(t)}f(x)\,dx}\quad(7)$$

$$\mu_2\big(F(t)\big)=\lim_{dt\to 0}\Pr\big(\tau_2\in[t,t+dt]\,|\,\tau_2>t\big)dt\approx\frac{f\big(F(t)\big)\upsilon_2(t)}{1-\int_{y_l}^{F(t)}g(x)\,dx}\quad(8)$$

In the process of transforming an event distributed in state-space to an event distributed over time, we have transformed the stochastic process to two Poisson processes with intensity/hazard rate μ_1 and μ_2 respectively. One important aspect of (3) and (4) is that μ_1 differentiable with respect to c and F as long as $a\dot{c}-\gamma\dot{F}\geq 0$ and μ_2 is differentiable as long as $\dot{F}\geq 0$. Because it is sufficient if only one of these events to occur for the THC to break down, we are interested in the distribution of the event $\tau = \text{Min}[\tau_1, \tau_2]$. Or to be precise, we are interested in the hazard rate associated with the distribution of τ. This hazard rate is given by:[4]

$$\lambda\big(c(t),F(t)\big)=\lim_{dt\to 0}\Pr\big(\tau\in[t,t+dt]\,|\,\tau>t\big)dt=\mu_1\big(c(t),F(t)\big)+\mu_2\big(F(t)\big)\quad(9)$$

In order to simplify the exposition, it is assumed that $f(X) = \lambda_1\exp(-\lambda_1 X)$ and $g(Y) = \lambda_2\exp(-\lambda_2 Y)$ both defined over $[0, \infty)$. One can then write $\lambda(c(t), F(t))$ as:

$$\begin{aligned}\lambda(u,c,F)&=\lambda_1\big(a\dot{c}-\delta\dot{F}\big)+\lambda_2\dot{F}\\&=a\lambda_1(u-\delta c)+\big(\lambda_2-\delta\lambda_1\big)(ac-\gamma F)\end{aligned}\quad(10)$$

For this expression to be correct, it must hold that $a\dot{c}-\delta\dot{F}\geq 0$ and $\dot{F}\geq 0$ in accordance with the definition of the hazard rates in Equations (7) and (8). For

---

[4] This follows from the fact that for statistically independent variables $\Pr(\tau\in[t,\ t+\ dt]|\ \tau\geq t) = \Pr(\tau_1\in[t,\ t+\ dt]|\ \tau_1\geq t,\ \tau_2\geq\ t+\ dt) + \Pr(\tau_2\in[t,\ t+\ dt]\ |\ \tau_2\geq t,\ \tau_1\geq\ t+\ dt) + \Pr(\tau_1\in[t,\ t+\ dt],\ \tau_2\in[t,\ t+\ dt])$.

future reference note that $\ddot{F} = a\dot{c} - \gamma\dot{F} < 0 \Rightarrow \upsilon_1 = 0$. If $\upsilon_2 > 0$ and $\upsilon_1 = 0$, then $\lambda(u,c,F) = \lambda_2\dot{F}$. $\upsilon_2 = 0$ and $\upsilon_1 > 0$, is not economically interesting, so this case will not be further discussed.[5] Having described a simple probabilistic model of THC breakdown, a model of the economic costs of CO_2 emission reduction is required. In order to focus on the problem with multiple thresholds, this is done quite simply. The instantaneous cost of emission reduction is assumed given by:

$$\frac{K}{2}\left(u^0 - u\right)^2 \tag{11}$$

Here K is a positive parameter and u^0 is the unregulated level of carbon emissions. Further, it is assumed that the only negative impact from climate change is the possibility of THC breakdown. This negative impact incorporated into the model as a state variable, ψ, that takes the value 0 before the thresholds is crossed. If the event $t = \tau$ occurs, then $\psi = G$. For all $t > \tau$, $\dot{\psi} = 0$. We can then write the objective function as

$$J\left(c(0), F(0)\right) = \max_u E\left(\int_0^\infty \left(-\psi - \frac{K}{2}\left(u^0 - u\right)^2\right) e^{-rt} dt\right) \tag{12}$$

The maximisation is performed subject to the differential equations in Equations (1) and (2), the stochastic process generating τ given in Equation (9) and the jump in ψ which may formally be written:

$$\psi\left(\tau^+\right) - \psi\left(\tau^-\right) = -\psi\left(\tau^-\right) - G \tag{13}$$

[5] In this case, the growth in temperature is negative but increasing. This will not occur in the present model except when initial values of c and F are unrealistic.

Optimal regulation after a THC breakdown.

In order to derive optimal policies we must specify optimal policies after a THC breakdown. Formally, this is done by solving the problem:

$$J\big(c(\tau),F(\tau)\mid\tau\big)=\max_{u}\int_{\tau}^{\infty}\left(-\psi-\frac{K}{2}\big(u^{0}-u\big)^{2}\right)e^{-r(t-\tau)}dt \tag{14}$$

The maximisation is performed subject to the differential equations in Equations (1) and (2). This is a deterministic problem. Because the only climate problem in the model is the breakdown of the THC and this has already happened, Equation (14) has the trivial solution $u = u^0$ and $J\big(c(\tau),F(\tau)\mid\tau\big)= -G/r$. The problem also yields co-state variables, associated with both F and c. These are both zero for all $t > \tau$.

Optimal regulation prior to a THC breakdown.

With the results from the previous section, one can formulate necessary conditions for optimality. The necessary conditions are derived using a suitable version of the Maximum Principle presented and briefly discussed in the Appendix. The solution is framed in terms of a risk-augmented Hamiltonian. We shall start by discussing the case where $F(0) = c(0) = 0$. This assumption is obviously not realistic, as it now seems certain that both atmospheric CO_2 and temperature are above pre-industrial levels. However, this simple condition enables us to start by examining the case where both thresholds will be stochastically active if F is to move at all. The appendix gives the general maximum principle that is applicable. In the present case this Hamiltonian is given by:

$$\begin{aligned}H(u,c,F,\chi,\phi)=&-\psi-\frac{K}{2}\big(u^{0}-u\big)^{2}+\chi(u-\delta c)+\phi(ac-\gamma F)\\&+\lambda(u,c,F)\Big(J\big(c(t),F(t)\mid\tau\big)-z\Big)\end{aligned} \tag{15}$$

$\lambda(u,\ c,\ F)$ is given by (10). The term $J(c(t),\ F(t)\mid\tau) - z$ requires an explanation. $J(c(t),\ F(t)\mid\tau)$ is defined above. The variable z has an interesting interpretation. Along the optimal path $z(t)$ is the value of the objective function net of the part that will is consumed in $[0,\ t)$. Formally, z is given by:

$$z(t) = \max_{u} E\left(\int_{t}^{\infty}\left(-\psi(s) - \frac{K}{2}\left(u^{0} - u(s)\right)^{2}\right)e^{-r(s-t)}ds\right) \tag{16}$$

The expression $J(c(t),\ F(t)\mid\tau) - z$ is thus the economic value gained if the threshold should happen to be crossed at time t. If $J(c(t),\ F(t)\mid\tau) - z > 0$, then crossing the threshold is desirable. If $J(c(t),\ F(t)\mid\tau) - z < 0$, it is not. In the present model, it is straightforward to see that $z \leq 0$. The easiest way to see this is to note that the objective function in (12) consists solely of terms less than or equal to zero. Somewhat harder to prove, but intuitively obvious is that $z(t) > -G/r = J(c(t),\ F(t)\mid\tau)$. This follows from the fact that along the optimum path, one is in the present model always better off if the THC has not yet collapsed.

The analysis below is built on the assumption that optimal paths converge towards a steady state. This is an assumption rather than a result because even showing local stability in the saddle path sense would require the finding the eigenvalues of a 5-dimensional matrix, which implies finding roots for a polynomial of degree 5. It was proven by Niels Henrik Abel that no general formulas exist that enables an analytical solution for such polynomials and therefore a formal stability analysis of the analytical model is awkward. However, by examining the structure of the model one may be reasonably certain that stability holds. CO_2 emissions are bounded between zero and u^0. This implies that all paths of state variables are bounded and therefore also the economically optimal paths. This precludes the possibility of solutions where state variables approach $\pm\infty$. Further, paths that do not converge to steady state would imply that $(u(t),\ F(t))$ would exhibit cyclic behaviour, which seems improbable as this would imply that the system would switch indefinitely between deterministic and stochastic deterministic segments. The stability properties of this type of model

have been discussed in a model with a single two-dimensional threshold by [14] where state variables are shown to converge to the steady state.

Now assume that both thresholds are stochastically active. Applying the Maximum principle to (15) gives the following necessary conditions:[6]

$$\begin{aligned} u &= \arg\max_{y} H(u, c, F, \chi, \phi) \\ &= u^0 + \frac{\chi}{K} - \frac{a\lambda_1}{K}\left(z + \frac{G}{r}\right) \end{aligned} \tag{17}$$

$$\begin{aligned} \dot{\chi} &= r\chi - \frac{\partial H}{\partial c} \\ &= r\chi + \delta\chi - \phi a + \lambda\chi + a\left(\lambda_2 - \lambda_1(\delta + \gamma)\right)\left(z + \frac{G}{r}\right) \end{aligned} \tag{18}$$

$$\dot{\phi} = r\phi - \frac{\partial H}{\partial F} = r\phi + \gamma\phi + \lambda\phi + \gamma\left(\gamma\lambda_1 - \lambda_2\right)\left(z + \frac{G}{r}\right) \tag{19}$$

$$\dot{z} = rz + \psi + \frac{K}{2}\left(u^0 - u\right)^2 + \lambda\left(z + \frac{G}{r}\right) \tag{20}$$

These conditions, coupled with Equations (1) and (2), initial conditions and appropriate transversality conditions determine optimal paths.

Unfortunately, it is hard to find explicit solutions to this system. Steady state solutions are however easy to compute. Inserting u from (17) into the relevant differential equations, setting all time derivatives equal to zero and solving yields the following solution.

$$z^* = -\frac{K - \sqrt{K^2 + 2K\theta^2 G}}{r\theta^2} - \frac{G}{r} \tag{21}$$

$$u^* = u^0 + \frac{1}{\theta K}\left(K - \sqrt{K^2 + 2K\theta^2 G}\right) \tag{22}$$

[6] Interior solutions are assumed.

$$c^* = \frac{u^0}{\delta} + \frac{1}{\delta\theta K}\left(K - \sqrt{K^2 + 2K\theta^2 G}\right) \tag{23}$$

$$F^* = \frac{au^0}{\gamma\delta} + \frac{a}{\theta\delta\gamma K}\left(K - \sqrt{K^2 + 2K\theta^2 G}\right) \tag{24}$$

Here θ is given by:

$$\theta = a\frac{r\lambda_1 + \lambda_2}{(r+\delta)(r+\gamma)} \tag{25}$$

This steady state solution has some nice properties. It is easy to check that $-G/r \leq z \leq 0$. Thus in steady state, the current value of future expected utility lies somewhere between expected utility after a THC breakdown and expected utility without risk of THC breakdown. Further, setting $G = 0$ implies that all variables revert to their unregulated levels $u = u^0$, $c = u^0/\delta$ and $F = au^0/\gamma\delta$. Unfortunately, this solution is of limited value as it is only valid for parameter values such that $F^* \leq 0$.

Proposition 1.

Assume that $F(0) < F^*$. Then an optimal path cannot be driven by the differential equations in Equations (18) – (20) for all t. It follows that F will not converge to F^*

Proof: Note that $\dot{F}$ depends only on state variables, so that $\dot{F}$ must be continuous. If an optimal path converges to $F = F^*$, then by the vary nature of a steady state $\dot{F}$ must decrease. But this implies that the threshold $\dot{F} = x$ cannot be stochastically active. Therefore the differential equations in Equations (17) – (20) are invalid over the period of time where $\dot{F}$ is decreasing while F is increasing.

■

The situation is the same as in Figure 1. There is a point in state space such as A. Over the segment between A and SS, the probability of $\dot{F} = y$ over this segment is zero because $\dot{F}$ in this segment $\dot{F}$ only takes values that has been experienced earlier in the segment between the origin and A. Therefore, along an optimal path, the threshold $\dot{F} = x$ cannot be stochastically active for all t. The threshold $F = y$ remains stochastically active for all t as $\dot{F}$ decreases and converges towards $\dot{F} = 0$. The proposition 2 does not cover the case where $F^* \leq 0$. If this is the case and $F(0) = 0$, then F^* is a valid steady state as it never optimal to allow F to increase. The steady state values in Equation (21) – (24) may be thought of as conditions for when the precautionary principle should apply and no risk is accepted.

If $F(0) < F^*$, the above discussion indicates that there is some t^* such that $\upsilon_1 = 0$ for all $t > t^*$. One can then try to examine under what conditions it is optimal that both constraints are active in the interval $[0, t^*)$ and only the temperature level constraint is active in (t^*, ∞). Having established this one can examine optimality conditions over the interval $[t^*, \infty)$. The problem to be solved is

$$J\left(c\left(t^*\right), F\left(t^*\right)\right) = \max_{u} E\left(\int_{t^*}^{\infty}\left(-\psi - \frac{K}{2}\left(u^0 - u\right)^2\right)e^{-rt}dt\right) \tag{26}$$

The maximisation is performed subject to the differential equations in Equations (1) and (2), the stochastic process generating τ given in Equation (7) and the jump in (13). The Hamiltonian is now given by:

$$\begin{aligned} H\left(u, c, F, \chi, \phi\right) = -\psi - \frac{K}{2}\left(u^0 - u\right)^2 + \chi\left(u - \delta c\right) + \phi\left(ac - \gamma F\right) \\ +\lambda_2\left(ac - \gamma F\right)\left(J\left(c\left(t\right), F\left(t\right) \mid \tau\right) - z\right) \end{aligned} \tag{27}$$

Using the Maximum principle in the appendix now gives the following conditions.

$$u = u^0 + \frac{\chi}{K} \tag{28}$$

$$\dot{\chi} = r\chi + \delta\chi - \phi a + \lambda_2 (ac - \gamma F)\chi + a\lambda_2 \left(z + \frac{G}{r} \right) \tag{29}$$

$$\dot{\phi} = r\phi - \frac{\partial H}{\partial F} = r\phi + \gamma\phi + \lambda_2 (ac - \gamma F)\phi + \gamma\lambda_2 \left(z + \frac{G}{r} \right) \tag{30}$$

$$\dot{z} = rz + \psi + \frac{K}{2}\left(u^0 - u\right)^2 + \lambda_2 (ac - \gamma F)\left(z + \frac{G}{r} \right) \tag{31}$$

These conditions, coupled with Equations (1) and (2), are necessary for optimality, conditional on the stochastic process in (7) not being active. Again, the full solution is hard to derive analytically. However, the steady state values are easy to calculate.

$$u = u^0 - \frac{-K + \sqrt{K^2 + 2\alpha^2 KG}}{\alpha K} \tag{32}$$

$$c = \frac{u^0}{\delta} - \frac{-K + \sqrt{K^2 + 2\alpha^2 KG}}{\alpha\delta K} \tag{33}$$

$$F = \frac{au^0}{\gamma\delta} - a\frac{-K + \sqrt{K^2 + 2\alpha^2 KG}}{\alpha\gamma\delta K} \tag{34}$$

$$z = \frac{-K + \sqrt{K^2 + 2\alpha^2 KG}}{\alpha^2 r} - \frac{G}{r} \tag{35}$$

Here α is given by:

$$\alpha = a\frac{\lambda_2}{(r + \delta)(r + \gamma)} \tag{36}$$

The discussion so far suggests that along the optimal path, the system is driven by Equations (17) – (20) until t^* is reached. At t^*, $\dot{F}$ reaches a maximum. At this point along the optimal path $\dot{F}$ may decrease or it may remain constant for some period of time. Proposition 2 shows that it is the latter that is the case.

Proposition 2.

Assume that along an optimal path there is a time interval $[0,\ s_1)$ such that both thresholds are stochastically active and an interval $(s_2,\ \infty)$ such that only the threshold $F = y$ is active. Then $s_1 \neq s_2$.

Proof: Assume $s_1 = s_2 = t^*$. At time t^*, it must hold that $\dot{F}$ is maximised as a function of time and that this maximum is unique. The optimal choice of u need not be continuous at t^*, so the necessary condition $\ddot{F} = 0$ evaluated at t^* can not be used. However, with piecewise continuity it must hold that:

$$\lim_{t \to t^*_-} \ddot{F} \geq 0 \text{ and } \lim_{t \to t^*_+} \ddot{F} \leq 0 \tag{37}$$

Equation (37) is merely a formal way of stating that the second derivative of F with respect to time must change sign from non-negative positive to non-positive if t^* is to maximise $\dot{F}$. By using the fact that $\ddot{F} = a\dot{c} - \gamma\dot{F}$ one can show that Equations (17), (28) and (37) imply that:

$$-\frac{a\lambda_1}{K}\left(z + \frac{G}{r}\right) \geq 0 \tag{38}$$

But $z > -G/r$, so (38) implies a contradiction.

■

From Proposition 1 it follows that there is an upper bound on $\dot{F}$ that is reached in finite time. Therefore υ_1 must be zero at some point. Also, from Proposition 2 it follows that $\dot{F}$ will not increase up to some point in time and then immediately start to decrease. The only possible optimal path is then some path such that $\dot{F}$ reaches some maximum value and remains at this value for some period of time. Formally, this implies that there is an interval $(t^*,\ t') \subseteq (t^*,\ \infty)$ such that $\ddot{F} = 0$ for all $t \in (t^*,\ t')$. Therefore, the regulation problem may be written:

$$\max_{u,t^*} E\left(\int_0^{t^*}\left(-\psi-\frac{K}{2}\left(u^0-u\right)^2\right)e^{-rt}dt+J\left(c\left(t^*\right),F\left(t^*\right)\right)\right) \tag{39}$$

subject to the differential equations in Equations (1), (2) and the additional constraint that $\ddot{F}=a(u-\delta c)-\gamma(ac-\gamma F)\leq 0$ for all $t > t^*$. Here $J\left(c\left(t^*\right),F\left(t^*\right)\right)$ is defined by the problem in Equation (26). It is obvious that t' must be less than ∞. Otherwise, if $\ddot{F}=0$ for all $t\in(t^*,\infty)$, temperature would increase linearly to infinity. This could only be achieved by letting u go to infinity, which is obviously not optimal or even feasible.

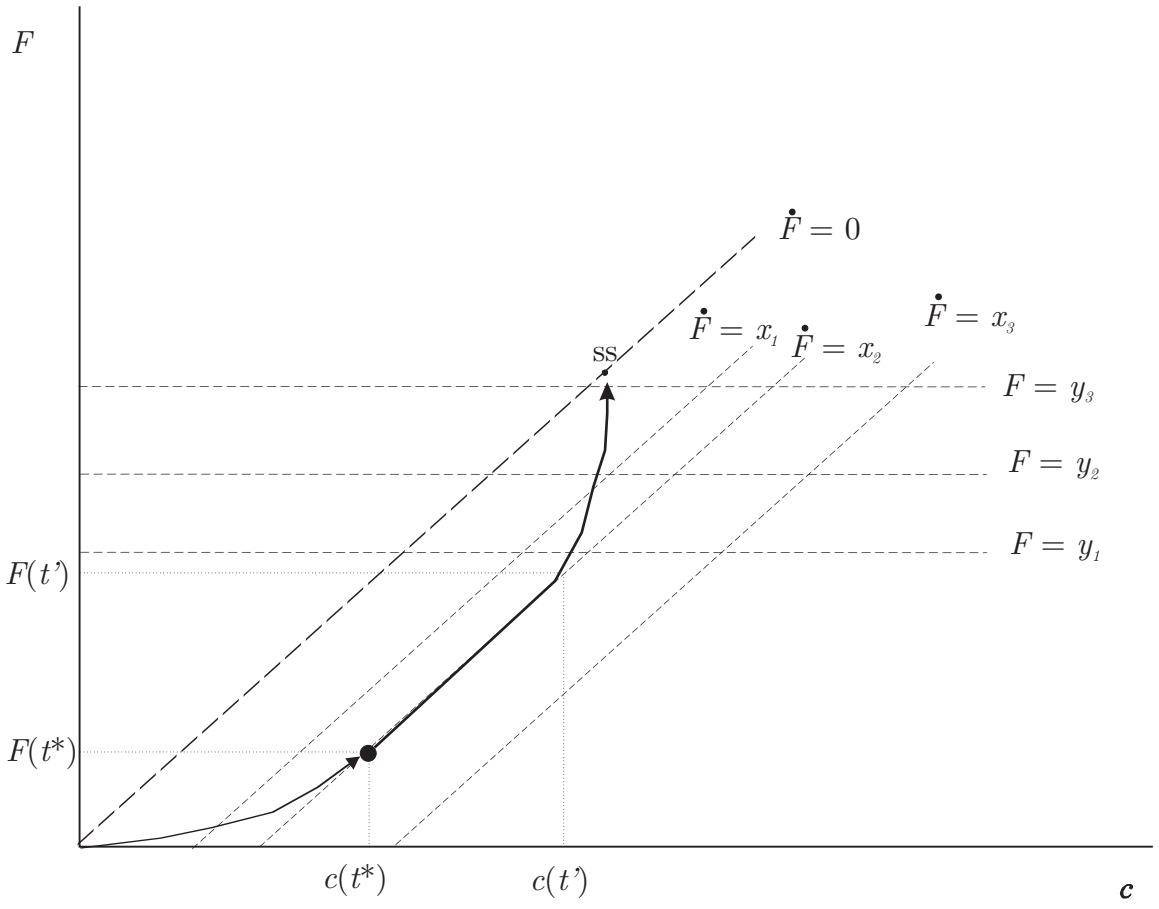


Figure 2. Optimal state variable paths in (*c*, *F*) space.

Using this, the optimal solution has been sketched in (c, F) space in Figure 2 and in (t, F) space in Figure 3. Initially, both thresholds are stochastically active as there is a positive probability that both τ_1 and τ_2 may occur. At the point ($c(t^*)$, $F(t^*)$), $\ddot{F}$ takes the value zero. The state-variables then proceeds along the line $\dot{F} = x_2$ until the point ($c(t')$, $F(t')$) is reached. Over this interval, the threshold $\dot{F} = x$ is not active. Rather, it may be thought of as *binding* in the sense that the regulator is not willing to accept positions in state-space associated with higher values of $\dot{F}$.

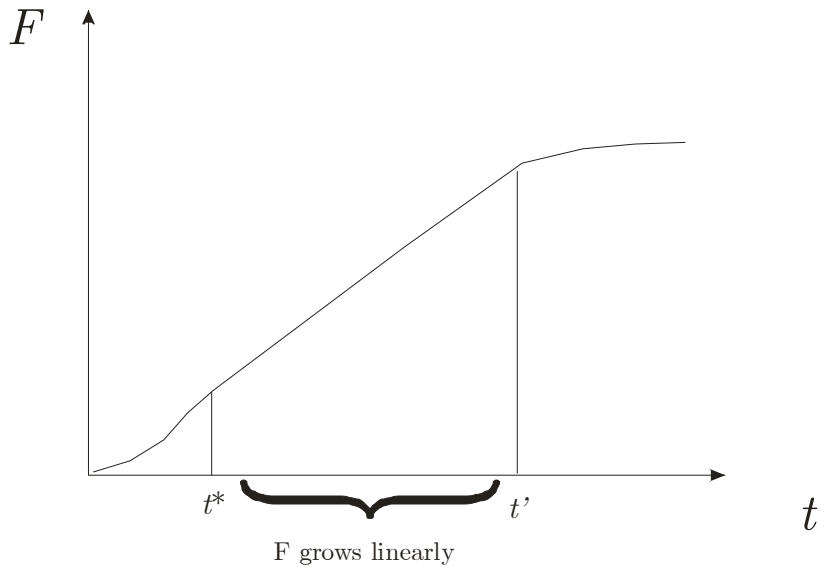


Figure 3, Optimal temperature path in (t, F) space.

At the point ($c(t')$, $F(t')$), the threshold associated with $\dot{F}$ ceases to be binding. In fact it ceases to be relevant at all because the regulator would not want to increase $\dot{F}$ even if this was risk free. In fact for all $t > t'$ such a small increase in $\dot{F}$ would in fact be partially risk free, since all along this part of the path $\dot{F}$ will only take values that has previously been experienced and are known to be safe.[7]

[7] Obviously an increase in $\dot{F}$ would imply that one would reach higher values of F sooner and this would imply increased risk.

Optimal Regulation with Alternative Initial Conditions

So far it has been assumed that $c(0)$ and $F(0)$ has been zero. Unfortunately there is every reason to believe that this state of affairs has passed and will not return in the foreseeable future. As it turns out, the qualitative nature of optimal paths is quite sensitive to initial conditions. The discussion will therefore start by identifying a set of feasible/interesting initial conditions. This set will depend on possible of CO_2 emission histories. Figure 4 divides the state space into different regions. The lines dividing the different regions are $\dot{F} = 0$, the path through state space given that $u(t) = u^0$ and the optimal path given $c(0) = F(0) = 0$. The regions are contained within the set $[0, c_{max}]\times[0,F_{max}]$. This set is defined by the points that are feasible for paths where $u(t) \leq u^0$ so discussion of these starting points are omitted. For the same reason is discussion of starting points in the region labelled B omitted. The set C contains starting points where emissions have been regulated, but sub-optimally and less stringently than optimal. The set B contains starting points where regulation of u has been too stringent compared to the optimal solution.[8] Although this does not seem very probable today, one cannot completely disregard this possibility. Finally, the set E contains starting points where the growth rate of F is negative, which implies by definition that none of the thresholds are stochastically active at the outset. Since starting points with the property that $\dot{F} < 0$ seems to be highly unlikely in the foreseeable future, starting points in this region is not further discussed. Thus the discussion will assume that starting points are contained in the closure of the set $C \cup D$. The point SS is the steady state point when the threshold $F = y$ is active while the threshold $\dot{F} = y$ is not. Finally there is a shaded area. In this area are starting points such that $F(0)$ is larger than the F value associated with SS. This

[8] There is in principle nothing that precludes the possibility that for a given emission history, (c, F) may have switched from one set to another.

area raises a set of issues that are peculiar and will be discussed in more detail below.

It will be shown that any optimal policy starting in $C \cup D$, will terminate in *SS*. This may seem surprising. First, it is straightforward to see that within the current set of differential equations there is an inherent inertia in the development of F. Because $\dot{F}$ only depends on state variables, it is not possible to abruptly change the sign $\dot{F}$. Thus, for a given starting point such that $\dot{F} > 0$, there are clear restrictions on which values that F and $\dot{F}$ can take in the immediate future. In the longer run it turns out that these restrictions are not important. Proposition 3 gives a relevant result for the feasibility of solutions in the long run.

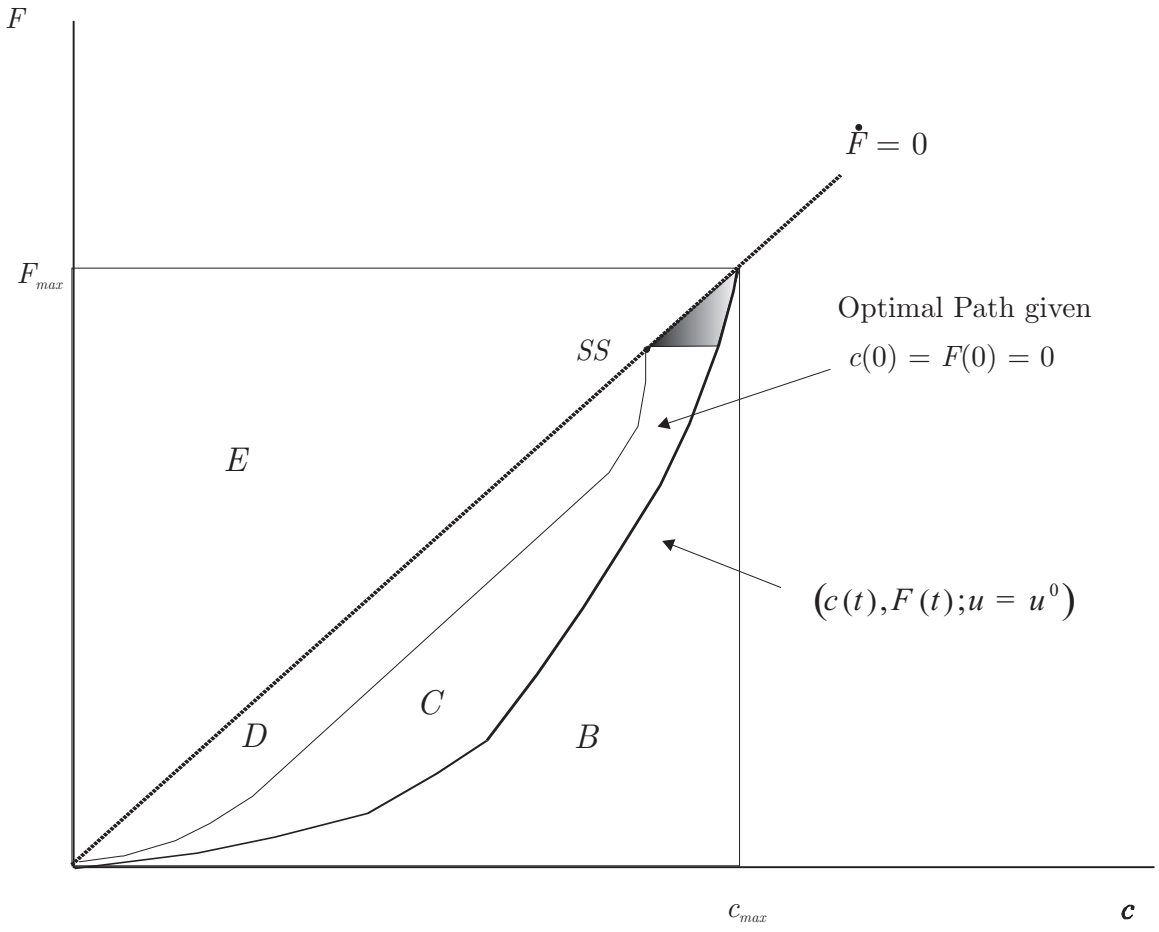


Figure 4. Possible Starting Points in (c, F) space.

Proposition 3.

For any $(c(0), F(0)) \in C \cup D$, $\lim_{t\to\infty}(c(t), F(t)) = SS$ is feasible through a path such that $(c(t), F(t)) \in C \cup D$ for all $t < \infty$.

Proof (sketch)

If it can be proven that there exists a constant $u \in [0, u^0]$ such that $\lim_{t\to\infty}(c(t), F(t)) = SS$, the proposition is proven. If $u = 0$, then $\lim_{t\to\infty}(c(t), F(t)) = (0, 0)$. If $u = u^0$, then $\lim_{t\to\infty}(c(t), F(t)) = (0, 0)$. Now note that SS can be written as a linear combination of $(0, 0)$ and (c_{max}, F_{max}). The existence of u then follows from the completeness of solutions to linear differential equations.

It follows from Proposition 3 that for any path originating in $C \cup D$, SS is a possible steady state. In fact, if the $F = y$ is the only stochastically active threshold; it is the only steady state. Given the assumption of convergence towards a steady state, this implies that all paths originating in $C \cup D$ will converge towards SS if there is a time interval $\left[\tilde{t}, \infty\right)$ such that only the threshold $F = x$ is stochastically active in that interval. This does not imply that optimal paths will necessarily look like the optimal path when $c(0) = F(0) = 0$. In fact, one can prove that for some starting points in $C \cup D$ the threshold $F = x$, and only this threshold, is stochastically active over the interval. This can be proven in a number of ways. The easiest is to take the path that is optimal when $c(0) = F(0) = 0$. On this path one can pick a point where only $F = x$ is active and then refer to the principle of optimality. An alternative way of proving this point is to pick a point on the curve $\left(c(t), F(t); u = u^0\right)$ for a value of s such that $F'(s) < \sup F'\left(t \mid u = u^0\right)$. With such a starting point, it will never pay to

choose u such that $\dot{F} > \sup F'\left(t \mid u = u^0\right)$ and therefore the threshold $\dot{F} = y$ will never be stochastically active. The main point of this discussion is that for paths starting in $C \cup D$ there will be qualitatively different paths with some paths, satisfying necessary conditions, exhibiting a constant value of $\dot{F}$ over some period of time. There will be other paths, also satisfying necessary conditions but with different starting points where $\dot{F}$ is decreasing for all t.

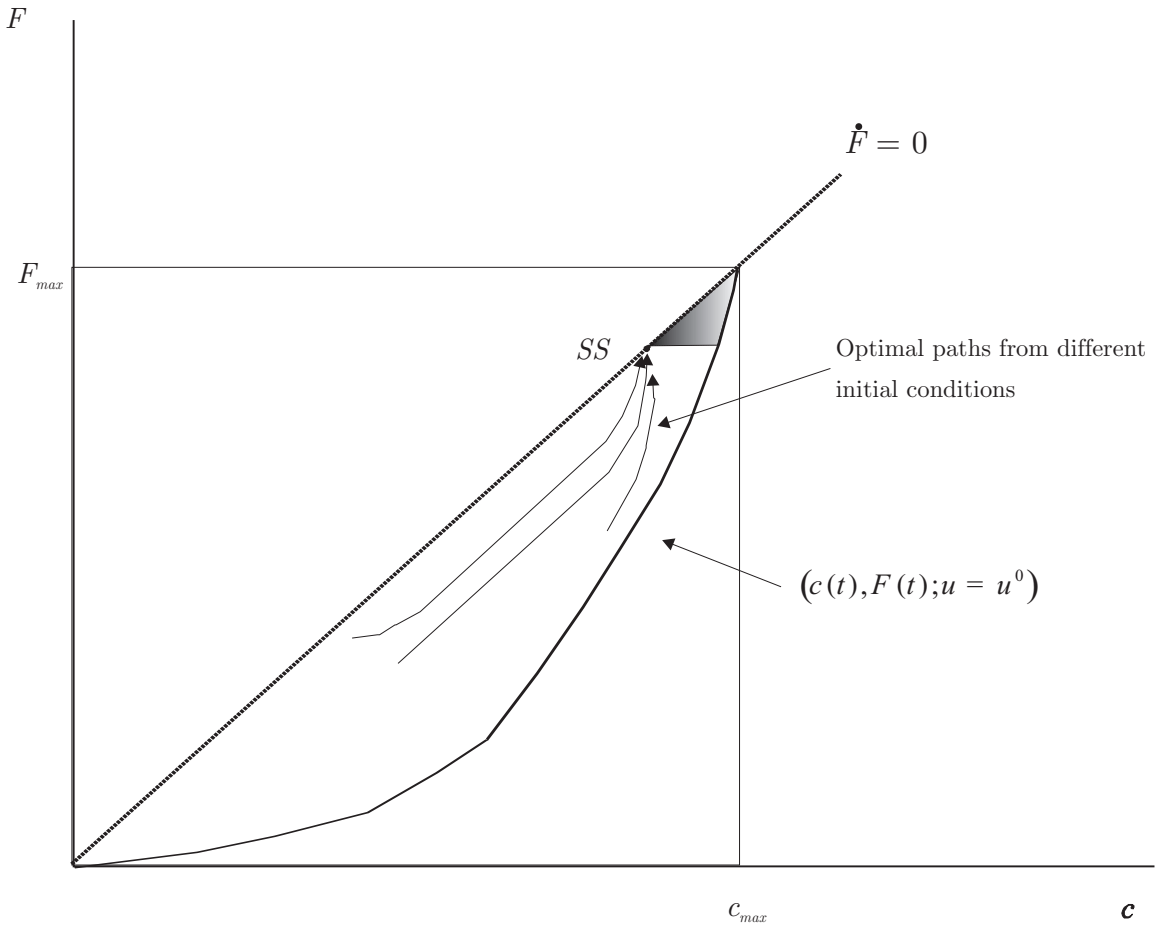


Figure 5. Possible optimal paths from different starting conditions

Regulation from Starting Points in the Shaded Triangle

Optimal paths originating in this triangle will only be briefly discussed. First, paths originating in this area are technically tricky. Second, it is hoped that

proper regulation of CO_2 emissions is implemented so that starting points in this area never become relevant. Third, if these points ever become relevant without the THC actually collapsing first, this will occur a long time into the future.

The first observation to make is that $F(0) > F_{ss}$. This means that one starts out with a larger value of F than the maximum value one would allow if the path was optimally regulated from a $F(0) < F_{ss}$. This implies that one does not want F to increase above $F(0)$. Unfortunately, in the interior of the triangle it is not possible to immediately freeze F, so the threshold $F = x$ is stochastically active. In fact, depending on the history leading up to the realisation of an initial condition in this triangle, the threshold $\dot{F} = y$ may also be active. This may lead to complicated optimisation programs. If the optimal solution converges towards some steady state, one can have that both, either one or none of the thresholds are stochastically at various points in time. The only thing that will always hold is that the threshold $F = x$ will, at least initially, be stochastically active.

Conclusions

The present paper has analysed optimal regulation of CO_2 emissions if there is a risk that the thermohaline circulation could collapse if one of two possible thresholds are crossed. The problem is analysed using a novel extension of the theory of optimal control in the presence of catastrophic risk. Results indicate that as time goes by, the probability that one of the thresholds is crossed becomes zero and this threshold becomes irrelevant. Results also indicate that there are upper bounds on the temperatures and the rate of temperature growth that a regulator is willing to accept. The upper bound for the rate of temperature growth is reached in finite time whereas the upper bound on temperature is reached only as time goes to infinity.

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 (2003). Although alternative, but equivalent, formulations exist in the literature this method is to my 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:

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

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

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

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

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 $x(\tau^+) - x(\tau^-) = q(x(\tau^-))$. Nævdal (2000) has shown that the conditions derived by Seierstad can be stated in terms of a *Risk-Augmented Hamiltonian*, defined by:[9]

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

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

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

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

$$\sigma \sim \lambda(x(s)) e^{-\int_0^s \lambda(x(\zeta))d\zeta} \text{ over } [t, \infty) \tag{A.8}$$

---

[9]Seierstad (2003) and Nævdal (2000) both work with present value formulations. The conditions presented here have a current value interpretation. Transforming the present value necessary conditions to current value conditions is a trivial and well-known process.

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

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:

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

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

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 $J(t,\ x\ |\tau=t)\ =\ \int_{t}^{\infty} f\left(y\left(s \mid t, x\right), u\left(s \mid t, x\right)\right) e^{-rs} ds$ is the value of criterion after a shock has driven the system to some arbitrary state x at time t. $J(t,\ x\ |\tau=t)$ is thus the criterion conditional on the event τ occurring at time t. The interpretation of $\left(J\left(t, x+q(x) \mid \tau\right)-J\left(t, x\right)\right)$ 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 a the state variable taking the value x. Now apply the maximum principle to the Hamiltonian in (A.5). Doing so yields the following conditions:

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

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

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:

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

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 $\frac{\partial}{\partial x}\left(J(t,x)\right)$ are needed in order to solve (A.13).

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

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

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:

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

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

$$u = \arg\max_{y} \Big(f(x,u) + \mu g(x,u) + \lambda(x) \Big(J\big(t, x + q(t,x) \mid \tau\big) - z \Big) \Big) \tag{A.17}$$

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

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:

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

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

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 $\dot{y} = g(y, u)$.

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

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

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.[10] 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.

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

References

[1] Broecker, W.S. (1997), Thermohaline Circulation, the Achilles Heel of Our Climate System: Will Man-Made CO_2 Upset the Current Balance. *Science*, **278**, 28 November, pp: 1582-1588.

[2] J. M. Conrad, (2000), Wilderness: options to preserve, extract, or develop, *Resource and Energy Economics*, **22**, 3: 205-219.

[3] Clark, H. R. and W. J. Reed (1994). Consumption/pollution tradeoffs in an environment vulnerable to pollution-related catastrophic collapse, *Journal of Economic Dynamics and Control*, **18**, 991 – 1010.

[4] Cropper, M. L. (1976), Regulating Activities with Catastrophic Environmental Effects. *Journal of Environmental Economics and Management*. **3**, 1 – 15.

[5] Dasgupta, P. and G. Heal (1974), The Optimal Depletion of Exhaustible Resources, 41 (*Review of Economic Studies*, Symposium on the Economics of Exhaustible Resources), **3** -28.

[6] Farzin, Y. H., (1996), Optimal Pricing of Environmental and Natural Resource Use with Stock Externalities, *Journal of Public Economics*, **62**, 31-57.

[7] Hemming, S. R. (2004), Heinrich events: Massive late Pleistocene detritus layers of the North Atlantic and their global climate imprint, *Reviews of Geophysics*, **42**, RG1005.

[8] Cubasch, U., G.A. Meehl, G.J. Boer, R.J. Stouffer, M. Dix, A. Noda, C.A. Senior, S. Raper, K.S. Yap. IPPC (2001), Projections of Future Climate Change In: *Climate Change 2001, The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, [Houghton, J.T.,Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, X. Dai, K. Maskell, and C.A. Johnson (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

[9] Kamien, M. and N. Schwartz (1971), Expenditure Patterns for Risky R&D Projects. *Journal of Applied Probability*. **8**, 60-73.

[10] Manabe, S and R.J. Stouffer (1993), Century-scale effects of increased atmospheric $C0_2$ on the ocean–atmosphere system, *Nature*, **364**, 215.

[11] 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.

[12] Nævdal, E., (2001), Optimal Regulation of Eutrophying Lakes, Fjords and Rivers, *American Journal of Agricultural Economics*, **83**, 4:972-984.

[13] Nævdal, E. (2003), Optimal Regulation of Natural Resources in the Presence of Irreversible Threshold Effects, *Natural Resource Modelling*. **16**, 3:305–333.

[14] Nævdal, E. (2003), 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, Forthcoming in *Journal of Economic Dynamics and Control*.

[15] Rahmstorf, S. (1996), *On the freshwater forcing and transport of the Atlantic thermohaline circulation*, Climate Dynamics, **12**:799-811.

[16] Stocker, T.F. & A. Scmittner, (1997), Influence of CO_2 emission rates on the stability of the thermohaline circulation. *Nature*, **388**, pp:862-865.

[17] Skiba, A. K. (1978). Optimal growth with a convex-concave production function. *Econometrica*, **46**:527-540.

[18] Tsur, Y and A. Zemel (1995), Uncertainty and Irreversibility in Groundwater Management, *Journal of Environmental Economics and Management*, **29**, 149 – 161.

[19] 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.

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

[21] Vellinga, M. & R. A. Wood (2002), Global Climatic Impacts of a Collapse of the Atlantic Thermohaline Circulation, *Climatic Change*, **54**, 3:251–267.

[22] Vellinga, M., R. A. Wood and J. M. Gregory (2002), Processes Governing the Recovery of a Perturbed Thermohaline Circulation in HadCM3, Journal of Climate: Vol. **15**, 7:764–780.

[23] Zemel, A., I. David and A. Mehrez (2001), On Conducting Simultaneous versus Sequential Engineering Activities in Risky R&D. *International Transactions in Operational Research*, **8**, 585-601.

[24] Zickfeld, K., T. Slawig & Stefan Rahmstorf (2004), A low-order model for the response of the Atlantic thermohaline circulation to climate change, *Ocean Dynamics*, **54**: 8-26.