Inertia Risk: Improving Economic Models of Catastrophes *

Anne-Sophie Crépin

The Beijer Institute, The Royal Swedish Academy of Sciences, SE-104 05 Stockholm,

Sweden, e-mail: asc@beijer.kva.se. (Corresponding author)

Eric Nævdal

The Frisch Centre, 0349 Oslo, Norway, e-mail: eric.navdal@frisch.uio.no.

Abstract

We model endogenous catastrophic risk in a new way we term *inertia risk*, which

accounts for delays between physical variables and the hazard rate—a characteristic

often observed in reality. The added realism significantly impacts optimal policies

relative to the standard model of catastrophic risk. The probability of a catastrophe

occurring at some point in time may span the entire interval [0, 1] and is not 0 or 1 as is

typical in standard models. Inertia risk may also generate path dependencies. We

illustrate the implications for policy in a simple model of climate change.

I. Introduction

Recent years have seen growing awareness that human activity increases the risk of regime shifts (substantial, persistent and abrupt system changes) (Biggs *et al.*, 2012a), some with potentially catastrophic impacts (Steffen *et al.*, 2004; Rockström *et al.*, 2009; Steffen *et al.*, 2015).[1] Given the tight social and ecological interactions in many managed systems, economic models focusing on regime shifts should incorporate processes likely to trigger such shifts (Levin *et al.*, 2012; Crépin and Folke, 2014). Economic analysis of regime shifts leads to dynamic resource management problems where stochastic processes may trigger rapid and dramatic changes. In continuous time models, economists usually model catastrophic risk by assuming the existence of a hazard rate that may depend on endogenous variables such as CO_2 concentrations in the atmosphere or fish stock size (Heal, 1984; Clarke and Reed, 1994; Tsur and Zemel, 1998; Gjerde *et al.*, 1999). The literature on the topic is largely theoretical, but with some empirical applications (Cai *et al.*, 2013; Lemoine and Traeger, 2014; Lontzek *et al.*, 2015).

The two most common approaches to modelling catastrophic risk in a dynamic setting consider time-distributed catastrophes (TDC, by far the most common) (*e.g.* Polasky *et al.*, 2011; Reed and Heras, 1992) or state-space distributed catastrophes (SDC) (*e.g.* Nævdal, 2006; Tsur and Zemel, 1995).[2] These risk structures imply that the catastrophe either occurs with probability one (TDC) or that the hazard rate is instantly responsive to changes in control variables (SDC). These properties are inadequate for many real-world situations. We present an alternative risk structure we call *inertia risk*, which is a hybrid of both standard approaches. Inertia risk represents real-life problems in a more realistic way, especially

---

[1]Impacts may also be beneficial. In order to simplify the discussion, we only discuss negative impacts.

[2]We introduce this terminology and are not aware of any alternative terminology for TDCs and SDCs.

because it accounts for dynamic lags between physical variables and the hazard rate and it allows for the possibility that substantial use of a resource at a certain level may eventually be deemed safe.[3] This approach is particularly appropriate for modelling renewable resource stocks like fisheries (Polasky *et al.*, 2011), lakes (Scheffer and Carpenter, 2003), coral reefs (Hughes, 1994; Norström *et al.*, 2009), grasslands/savannah (Hirota *et al.*, 2011) and global phenomena like planetary boundaries (Rockström *et al.*, 2009; Steffen *et al.*, 2015). Some planetary boundaries relate to global-scale thresholds linked to climate change, ocean acidification and stratospheric ozone depletion, while others like biogeochemical flows, land-system change, and biodiversity loss relate to regional thresholds (Rockström *et al.*, 2009).

We study the implications of using our new risk structure, compared with conventional models, in management models, using examples from planetary boundaries and fisheries. However, we argue that the inertia risk approach is applicable to other economic areas such as knowledge spillover in relation to investments in R&D activities. The literature considers different types of risk structures (Kamien and Schwartz, 1972; Doraszelski, 2003), but not inertia risk.

We show that inertia risk gives rise to optimal paths that differ from the standard TDC approach in the following important respects:

- Optimal paths derived from TDC models are Markovian. Paths derived from inertia risk models are possibly path-dependent and optimal policy may depend on history as well as current location in state space.

- Optimal paths derived from TDC models imply that a catastrophe will happen with probability 1 unless a corner solution is optimal. For optimal paths derived from

---

[3]We use the term 'inertia' synonymously with 'sluggishness' and not in the technical sense used in physics.

inertia risk models, the probability of catastrophe may be any number in the interval [0 1].

- Steady state solutions derived from TDC models exhibit qualitatively different comparative statics from steady state solutions for inertia risk models.

Section II presents and models the different risk structures involved and discusses their respective advantages and drawbacks. Section III applies the inertia risk structure to a simplified model that captures some basic features of many systems, including the climate, lakes, coral reefs and grasslands/savannah. Section IV discuss how to generalise and apply the results in management. Section V provides a summary and suggestions for future research.

II. Risk structures

Consider a system with uncertain dynamics that can undergo a regime shift whereby some event can trigger a persistent non-marginal perturbation, termed a catastrophe. Here we focus on stochastic processes where an event occurs at some random point in a continuous time domain. It is common to model such processes using a *hazard rate*. If Y is a random variable defined over some interval $(y_1, y_2) \subseteq (-\infty, \infty)$, with a pdf given by $f(y)$ and a cdf given by $F(y)$, then the hazard rate $\lambda(y)$ is given by:

$$\lambda(y) = \lim_{h \to 0} \frac{\Pr(Y \in [y, y+h] | Y \geq y)}{h} = \frac{f(y)}{1 - F(y)}. \tag{1}$$

Roughly speaking, the hazard rate gives the probability of a regime shift occurring in the next unit of time. A particular hazard rate uniquely determines its associated pdf, $f(y)$, and cdf,

$F(y)$, as long as $\lim\limits_{\phi \to y_2} \int_{y_1}^{\phi} \lambda(y)\,dy = \infty$ and $\lambda(y) \geq 0$ for all y, implying that we can choose

whether to work with $\lambda(y), f(y)$ or $F(y)$.[4]

Existing models of catastrophic risk

One way to categorise dynamic models of catastrophic risk is to distinguish between models where the hazard rate is exogenous and where it is endogenous (Polasky *et al.*, 2011). We focus only on endogenous risk, *i.e.* where decision makers' actions may affect the risk of a catastrophe occurring. Such risk structures are particularly relevant when studying problems of pollution release and resource exploitation, where human activities, besides producing welfare, may also affect the risk of the system making a critical transition to an alternate regime. In standard models with endogenous hazard rates, the hazard rate is a function of some state variable, denoted x, which in turn may be time–dependent (TDCs) and/or determined by the path of one or several control variables (SDCs, respectively).

Models of TDCs typically specify a hazard rate in the form of the probability of a particular event occurring at each point t in time in the following way:

$$\lambda_t\big(x(t)\big), \tag{2}$$

where $x(t)$ is an arbitrary continuous function possibly determined by a controlled differential equation (Reed and Heras, 1992). This risk structure has been applied to a wide range of topics, including regime shifts in fish stocks (Polasky *et al.*, 2011), climate policy (van der Ploeg and de Zeeuw, 2013; van der Ploeg, 2014), epidemic outbreaks (Berry *et al.*, 2015) and ecosystem services (Barbier, 2013).

---

[4]This can be proven by viewing (1) as a differential equation for the unknown function $F(y)$ with the initial condition $F(y_1) = 0$.

Models of SDCs are usually specified as pdfs in state space, *i.e.* some critical boundary in state space must not be crossed. Typically, this point is an unknown number representing a particular value of some indicator (*e.g.* fish stock, pollution stock). If the pollution stock exceeds or the fish stock falls below some critical level, a regime shift is triggered (Tsur and Zemel, 1995; Nævdal, 2003; Kama *et al.*, 2014). More general specifications where the critical boundary is a curve in state space or there is more than one critical boundary have also been analysed (Nævdal, 2006; Nævdal and Oppenheimer, 2007).

Formally, there is a random variable $X \sim h(x)$, with $h(x) = H'(x)$, where $H(x)$ is the cdf of X such that $x = X$ triggers a regime shift. $x(t)$ is also a solution to a differential equation. Using this approach in dynamic optimisation models requires transforming the event $x = X$ distributed in state space to the event $x(\tau) = X$ distributed over time. The details of this transformation depend to a certain extent on the properties of the system studied. The hazard rate, denoted $\lambda_x(x)$, is given as:

$$\lambda_x(x) = \max\left(\frac{h(x)}{1 - H(x)} \frac{dx}{dt}, 0 \right). \tag{3}$$

Strictly speaking, eq. (3) is only valid when the term $x'(t) = 0$ has at most one solution. However, as long as $x(t)$ is one-dimensional and an optimally controlled function, $x(t)$ is either strictly increasing or decreasing. This is usually not an issue for problems with a single state variable. We return to this topic in section III.

A more general risk structure – inertia risk

We present a more general risk structure that incorporates delayed effects. Many real-world problems involve an endogenous risk of system shift, while complex system dynamics with slow reinforcing processes mean that regime shift can occur long after the initial trigger (Carpenter and Turner, 2000). For example, in lake fisheries the fast variable 'angling' or the

slow 'shoreline development' can both trigger a regime shift (Biggs *et al.*, 2009). Outbursts of methane due to melting permafrost (IPCC, 2013), coral reef bleaching (Hughes, 1994; Norström *et al.*, 2009) and sudden collapse of the Greenland and Antarctica ice sheets (Joughin *et al.*, 2014) are all catastrophes that could occur if a critical temperature threshold is reached. Other examples include damage and healing of the ozone layer (Solomon *et al.*, 2016) and the economics of R&D. Investing in knowledge is a slow process, with a potential lag between the build-up of knowledge capital and the achievement of results, so inertia appears to be a pertinent assumption. Moreover, even if a certain stock of knowledge capital is maintained indefinitely, the probability of a breakthrough should diminish with time. Some breakthroughs simply do not happen regardless of how long a knowledge capital stock is maintained. In the SDC approach, if the knowledge stock does not increase, the probability of a breakthrough immediately becomes zero. However, it seems sensible to allow for a breakthrough occurring even if knowledge capital has started to decline. Introducing a more realistic inertia risk structure into R&D models may have policy-relevant implications.

Systems with slowly moving variables can be represented by models where some parameter value, rather than being constant, changes slowly over time. Examples include the build-up of corals (Crépin 2007) and the accumulation of phosphorus in the mud layer between water column and sediment in lakes (*e.g.* Srinivasu, 2004; Graß *et al.* 2017). The disadvantage of this approach is that proper analysis means adding new dynamic equations to represent each of these slow variables and their interactions with all other variables. The analysis then becomes tedious, despite techniques like perturbation theory (see Crépin, 2007) and detailed parameter sensitivity analysis (see Wagener, 2003). The representative set of variables and their dynamics are also likely to differ in different systems, making it difficult to obtain general results. Furthermore, the complex adaptive system dynamics (Levin *et al.*, 2012) in

some systems mean that a minimal change in model configuration can have substantial impacts on the outcome. Hence, more detail may not necessarily improve the solution.

We propose a different approach that builds on a social-ecological systems perspective (Berkes and Folke, 1998) and resilience theory (Walker and Salt, 2012). In most social-ecological systems, a limited set of variables and internal feedback processes interact to steer the system configuration (Walker and Salt, 2012; Biggs *et al.*, 2012b). We suggest that focusing on these essential dynamics can generate ways of overcoming some of the challenges posed by delayed dynamics. Rather than aiming for fine-tuned management including all known details of the system and still risking being wrong, our approach is more stylised and general and aims for approximately 'right' management promoting a regime that is desirable for society. Hence we introduce an additional variable stress, denoted *s*, which can represent all kinds of system pressures and loss of resilience that occur slowly, but are serious enough to trigger a regime shift. *s* does not represent one particular slow variable, but is rather an aggregate measure of key variables that could potentially influence system resilience and trigger a catastrophe. A catastrophe consists of two distinct dynamic phases: the process that generates a probability of regime shift, typically modelled with a hazard rate, and the regime shift itself. Our dynamic model contains an explicit lag between cause and effect. An alternative is to model a lag between effect and consequence, where it takes time from a disaster occurs until its consequences are felt (Liski and Salanié, 2018). An approach building on a pure TDC framework with added lags is also possible. However, neither of these approaches would fit the general type of problem we discuss here.

In our approach, human activities aiming to control the system through *e.g.* harvest or nutrient release are denoted *u* and have a direct cumulative effect on a state variable *x* through a differential equation $\dot{x} = f(x, u)$. The change in stress in a system depends on previous amounts of accumulated stress (*s*) and the level of the state variable (*x*):

$$\dot{s} = g(s, x). \tag{4}$$

Given standard regularity conditions, the solution $s(t)$ has a continuous derivative and therefore responds sluggishly to changes in u. We assume a random variable S with known pdf $h(s)$ and cdf $H(s)$. A large realisation of S implies a more resilient system, although the true value of S is only known if the regime shift is triggered. The event $s(t) = S$ has a distribution over time. The inertia hazard rate for the process is given by: $\lambda_s(s, x) = h(s)/(1 - H(s)) \times \max(\dot{s}, 0)$. The *inertia hazard process* consists of the two differential equations for x and s, and the hazard rate generated by the paths of these variables. Note that the definition of $\lambda_s(s, x)$ implies that the hazard rate only depends on slow variables determined by differential equations and is therefore itself a slow variable.

A *regime shift* is a discrete shock[5], which could take many forms depending on the problem at hand. In section III, we model a situation where the differential equation shifts from $\dot{x} = f(x, u, 0)$ to $\dot{x} = f(x, u, \beta)$. Alternatively, a regime shift could be a shock to utility or a jump in one of the state variables.

We give an example of an inertia hazard process without specifying the regime shift or performing any economic analysis. Assume that the stock of some pollutant is determined by:

$$\dot{x} = u(t) - \delta x, \tag{5}$$

where $u(t)$ denotes the rate of pollution emissions and δ the rate of depreciation (*i.e.* the rate at which the pollutant is absorbed by the environment and becomes harmless). Stress is determined by the following differential equation, where α and γ are parameters:

$$\dot{s} = \alpha x - \gamma s. \tag{6}$$

---

[5]A regime shift could also be modelled as an endogenous change in system dynamics due to *e.g.* system bifurcation (Wagener, 2003; Crépin, 2007; Polasky *et al.,* 2011). However, for our purposes a jump simplifies model analysis substantially, while remaining quite realistic.

Here α is a measure of inertia. The expected time before an additional unit of x contributes to an increase in stress is α^{-1}. Parameter γ is the system's capacity to absorb stress. If s increases above some threshold level S, this triggers a regime shift (unspecified for now). S is distributed over $[0, \infty)$ with mean μ and standard deviation σ. For simplicity, we assume that $S \sim h(s) = \mu s \exp\left(-\frac{1}{2}\mu s^2\right)$ and $H(s) = \int_0^s h(q)\,dq = 1 - \exp\left(-\frac{1}{2}\mu s^2\right)$. This functional form has a hazard rate given by μs. The time-distributed inertia hazard rate is then

$$\lambda_t = \mu s \times \max[\dot{s}, 0]. \tag{7}$$

We illustrate the risk structure with two numerical examples (panels A and B in Figure 1). In panel A, emission level, $u(t)$, is a step function with $u(t) = 0$ for $t < 0$ and $u(t) = k$ for $t \geq 0$, where k is some positive constant, *e.g.* an increase in pollution from new agricultural activities. The stock of pollutant increases and the hazard rate increases at first, representing an increasing risk of regime shift due to accumulation of the pollutant creating increased stress.
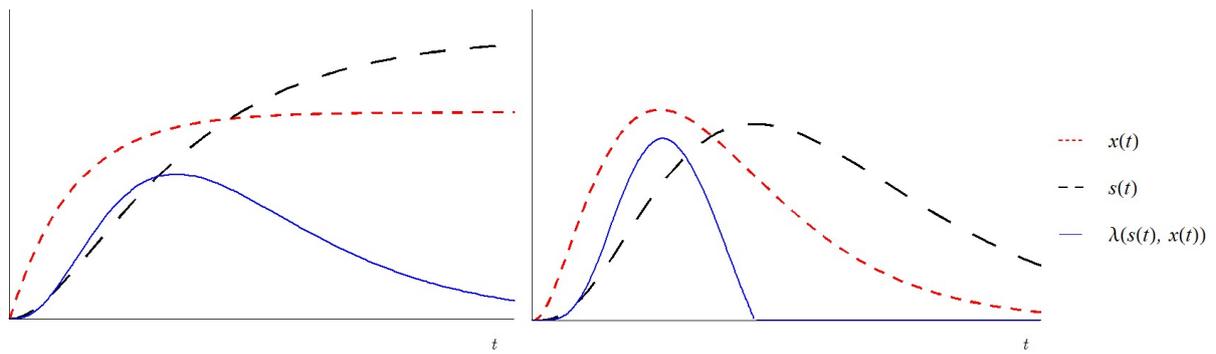

Figure 1: Time path of the inertia hazard rate (λ), stress (s) and state (x) when $u(t)$ is a constant k (panel A) or a pulse $t \times \mathrm{Exp}(-t)$ (panel B). Note that the relative magnitude of $x(t)$, $s(t)$ and $\lambda(s(t), x(t))$ is not important, as they are measured in different units.

However, if this constant flow of pollutant does not trigger a regime shift after some time, then the system can apparently absorb the pollutant without any particular disturbance. This means that the ecosystem is probably sufficiently resilient to stress and the hazard rate begins to decrease and goes asymptotically to zero. Two trends co-occur here. First, the hazard rate exhibits a lagged response to changes in emissions. Second, as *s* converges to steady state, it becomes increasingly likely that the ecosystem can handle the stress unless emissions increase further.

Panel B in Figure 1 shows the effect of an emission pulse, for example an oil spill at sea. As in panel A, the hazard rate increases initially and continues to increase after the stock of the pollutant has started to decline, as stress accumulates even then. Note also that after the stress $s(t)$ begins to decline, the hazard rate is zero. In both graphs a higher level of inertia (high α) would result in a higher level of the stress curve, while a higher capacity to absorb stress (high γ) would make the curve flatter.

When the hazard rate starts decreasing and how long it takes depends on the system. Slow variables can substantially influence the timing of the increase and the time it takes for the risk to come back to 0 after disturbance. This decrease is presumably relatively late if there is a slow accumulation process with potentially late release due to slow variables (see *e.g.* Crépin, 2007, for a continuous model with such properties).

Conceptual implications of TDC, SDC and inertia risk

Comparing particular properties of the different risk structures illustrates that each is differently suited for different applications, *e.g.* steady state properties and long-run dynamics (note that the subtle transition dynamics in hazard rate illustrated in Figure 1 only exist with our inertia risk structure).

One particular property of TDC risk structure is that if the state variable, $x(t)$, is constant, the hazard rate, $\lambda_t(x(t))$, is also constant. Hence, even if no more stress is added to the system and

it remains in a particular state for a very long time, the risk of a regime shift still remains unchanged. As time goes to infinity, the event will occur with probability 1. A resource exploitation model or pollution model that leads to a catastrophe may be less interesting for managers, and often less realistic, than a model that at least incorporates the *possibility* of sensible management being sustainable in the long run.[6] The steady state risk could be very small of course: an expected waiting time of 1 million years before disaster occurs would probably suffice for a process to be termed sustainable. However, even if a model with inertia risk yields a small total probability of catastrophe, using a TDC would not necessarily imply low hazard rates. Furthermore, if a model with a TDC yields a very small hazard rate in steady state, this is in general because the cost of a catastrophe and/or the *ex post* marginal utility is close to infinity. This probably only applies to catastrophes of a global nature.

Models with SDC do not have this shortcoming. If, for some reason, one is able to freeze state variables, the rate of change jumps from strictly positive to exactly zero. For example, if a fish stock is driven down to *e.g.* 20% of its carrying capacity and does not immediately collapse, then that level is known to be safe. However, this assumption is often too optimistic for models aiming to capture real-world dynamics. When a fish stock is down to 20%, there may be a period during which it is uncertain whether that fish stock can remain indefinitely at that level or will collapse. Models with SDCs fail to account for cumulative effects that may emerge in the long run as a consequence of a disturbance. Thus they are inappropriate for modelling *e.g.* knowledge spillover, which can occur long after some intervention has been made. Our definition of inertia risk $\lambda_s(s, x)$ means that the hazard rate only depends on

---

[6]One possible alternative is to specify that there is only risk above or below a certain threshold (Margolis and Nævdal, 2008). However, in such a model the result holds that if it is optimal to accept a strictly positive probability of regime shift for some time interval $[t^*, t^* + \varepsilon]$, it is also optimal to accept it for all t over $[t^*, \infty)$ and the system will experience a regime shift with probability 1.

continuous variables, and is therefore itself a continuous variable. In contrast, in the SDC approach the hazard rate may have discontinuous jumps or may respond rapidly to changes in a continuous manner if controls change quickly enough.

Another property worth comparing is the potential for learning. TDCs do not allow for transiently high hazard rates and optimally managed systems are Markovian, so for given levels of state variables, a system that has been stressed holds the same information as a system that has not been stressed. An optimal policy only depends on the current value of the state variable and not on the history, which is not necessarily the case with inertia risk. Hence assuming a TDC suggests that there is no learning and that long-run steady states are independent of initial conditions. In some cases this is an innocuous assumption, but in real life non-monotonic hazard rates are often perfectly sensible. For example, exposing a natural resource system to stress results in learning about the system's capacity to cope with that stress. Hence, optimally managed systems could be expected to exhibit path dependency, which does not happen with a TDC, at least unless the problem itself is convex-concave. In contrast, SDCs assume that learning is immediate because once a particular point in space is reached and no regime shift has occurred, it is known to be safe.

Hence, TDCs and SDCs have properties that are unfortunate as building blocks for models of resource management, climate change or knowledge spillover, but our inertia risk approach does not. If inertia risk is the most realistic structure, then imposing a TDC or a SDC structure with optimal paths that are not path-dependent will probably make regulation more costly than necessary. Below we analyse the implications of inertia risk in a stylised model and compare these with corresponding outcomes with TDC and SDC.

III. Application: A Model of Climate Change

We now examine the implications of the different risk structures in a simple climate model. Assume that x denotes global temperature measured as increase above pre-industrial levels. Building on evidence of an almost constant carbon-climate response (Matthews *et al.*, 2009) we assume that temperature dynamics are given by eq. (8), where u is the rate of carbon emissions and δ is the natural rate of cooling. Following Steffen *et al.* (2018), we include a climate feedback, represented by β, that plays out if average global temperature exceeds a certain level (they suggest around 2 degrees Celsius) where self-reinforcing mechanisms mean that when this threshold level is passed at time τ, climate dynamics undergo a regime shift into a "Hothouse Earth". Transgressing tipping elements (Lenton *et al.*, 2008), like substantial release of methane from melting permafrost (Walter *et al.*, 2006; Anthony *et al.*, 2012) and collapse of Greenlandic and Antarctic ice sheets (Joughin *et al.*, 2014) may activate feedback, β.

$$\dot{x} = \begin{cases} u - \delta x \text{ for } t \leq \tau \\ u - \delta x + \beta \text{ for } t > \tau \end{cases}. \qquad (8)$$

Let a be the amount of stress that triggers a shift in carbon cycle dynamics. Now define τ as the earliest point in time t at which stress reaches the level a: $\tau = \inf_t \left(t : s(t) = a \right)$. For simplicity, assume that a is the realisation of a random variable S, which is exponentially distributed with intensity λ. This allows steady state expressions to be calculated. More realistic assumptions would not alter the fundamental results, but would require numerical analysis. We assume that the amount of stress s in the system is governed by eq. (6).

Assume that the damage done by temperature, x, is Ax and the cost of reducing greenhouse gas emissions is $\frac{1}{2}c(u^0 - u)^2$ where $u \in [0, u^0]$. We model the feedback effect as a state variable B, where $B = \beta$ with feedback and $B = 0$ without. This leads to the following optimisation problem:

$$\max_{u(t)\in\left[0,u^0\right]} \mathbb{E}_\tau\left(\int_0^\infty \left(-Ax-\tfrac{c}{2}\left(u^0-u\right)^2\right)e^{-rt}dt\right), \tag{9}$$

subject to:

$$\dot{x}=u-\delta x+B, \ x(0) \text{ given,}$$
$$\dot{s}=\alpha x-\gamma s, \ s(0) \text{ given, and} \tag{10}$$
$$\dot{B}=0, \ B(0)=0.$$

The stochastic process is:

$$\Lambda(s,x)=\begin{cases}\displaystyle\lim_{dt\to 0}\frac{\Pr\left(\tau\in[t,t+dt]\,|\,\tau\ge t\right)}{dt}=\lambda\max(\dot{s},0) \text{ if } s(t)\ge\sup_\xi s(\xi)\\[2mm]0 \text{ if } s(t)<\sup_\xi s(\xi)\end{cases}. \tag{11}$$

$$\lim_{t\to\tau^+}B(t)-\lim_{t\to\tau^-}B(t)=\beta.$$

We solve the problem recursively using optimality conditions found in Nævdal (2006) based on more general conditions found in Seierstad (2009) and reproduced in online appendix A. First, we calculate the general optimal solution after the catastrophe has occurred from an arbitrary post-catastrophe state $(\tau, x(\tau))$. We then calculate the optimal solution conditional on the catastrophe having not yet occurred.

Optimal solution post catastrophe

Let $J\left(x(\tau),\beta\right)$ denote the current value function post catastrophe and $\mu(t|\ \tau)$ the ex-post shadow price of *x*. Note that *s*(*t*) no longer affects the management problem, as the damage inflicted by stress has already been done and is irreversible. The notation $(\cdot\ |\ \tau)$ indicates that the expressions are conditional on the event τ having occurred. When the catastrophe has occurred, *B* is equal to β and the post-catastrophe problem is:

$$J(x,\beta\,|\,\tau)=\exp(r\tau)\max_{u(t)\in\left[0,u^0\right]}\int_\tau^\infty\left(-Ax-\tfrac{c}{2}\left(u^0-u\right)^2\right)e^{-rt}dt \ . \tag{12}$$
$$s.t:\dot{x}=u-\delta x+\beta, \ x(\tau) \text{ given.}$$

The solution to this problem is:

$$u(t \mid \tau) = u^0 - \frac{A}{c(r+\delta)}, \quad \mu(t \mid \tau) = \frac{-A}{r+\delta}.$$

$$x(t \mid \tau) = \left(\frac{u^0 + \beta}{\delta} - \frac{A}{c\delta(r+\delta)} \right)\left(1 - e^{-\delta(t-\tau)}\right) + x(\tau)e^{-\delta(t-\tau)}.$$

(13)

Evaluating (13) as times goes to infinity gives the steady state level of $x(t)$ conditional on the event having occurred:

$$x(\infty)_{\beta>0} = \frac{u^0 + \beta}{\delta} - \frac{A}{c\delta(r+\delta)}, \quad s(\infty)_{\beta>0} = \frac{\alpha}{\gamma}\left(\frac{u^0 + \beta}{\delta} - \frac{A}{c\delta(r+\delta)} \right).$$

(14)

Inserting from (13) into (12) and calculating the integral gives the value function after the regime shift for an arbitrary value of x:

$$J(x, \beta \mid \tau) = -\frac{Ax}{r+\delta} + \frac{A(A - 2c(u^0 + \beta)(r+\delta))}{2cr(r+\delta)^2} < 0.$$

(15)

Note that setting $\beta = 0$ and $\tau = 0$ in equations (12)-(15) provides the steady state solutions of x and s when there is no risk.

$$x(\infty)_{\beta=0} = \frac{u^0}{\delta} - \frac{A}{c\delta(r+\delta)}, \quad s(\infty)_{\beta=0} = \frac{\alpha}{\gamma}\left(\frac{u^0}{\delta} - \frac{A}{c\delta(r+\delta)} \right).$$

(16)

Pre-catastrophe solution

Let z be the *ex ante* value function. Then, using (15), the cost of the regime shift, should it occur at time τ, is given by: $J(x, \beta \mid \tau) - z$. We form the Hamiltonian to solve for optimal policies prior to hitting the threshold, where μ_x denotes the shadow price of temperature x and μ_s the shadow price of stress s. $\Lambda(s, x)$ is the hazard rate defined in eq. (11). Note that $\Lambda'_s = -\lambda\gamma$ if $\dot{s} > 0$ and $\Lambda'_s = 0$ if $\dot{s} \leq 0$. Similarly, $\Lambda'_x = \lambda\alpha$ if $\dot{s} > 0$ and $\Lambda'_x = 0$ if $\dot{s} \leq 0$.

$$H(u, x, s, \mu_x, m_s) = -Ax - \frac{c}{2}(u^0 - u)^2 + \mu_x(u - \delta x + B) + \mu_s(\alpha x - \gamma s)$$
$$+ \mu_B \times 0 + \Lambda(s, x)(J(x, \beta \mid \tau) - z).$$

(17)

This Hamiltonian consists of that used in deterministic control theory with an additional term, $\Lambda(s,\ x)(J(x,\ \beta|\tau) - z)$, representing the expected cost of the catastrophe. Applying the maximum principle to (17) yields:

$$u = \underset{v \in \left[0, u^0\right]}{\arg\max} H\left(v, x, s, \mu_x, m_s\right) = \min\left(\max\left(0, u^0 + \frac{\mu_x}{c}\right), u^0\right), \tag{18}$$

$$\begin{aligned}\dot{\mu}_x &= r\mu_x - \frac{\partial H}{\partial x} \\ &= (r+\delta)\mu_x + A - \alpha\mu_s + \Lambda\left(s, x\right)\left(\mu_x - \mu\left(t\,|\,\tau\right)\right) - \left(J\left(x, \beta\,|\,\tau\right) - z\right)\Lambda_x',\end{aligned} \tag{19}$$

$$\dot{\mu}_s = r\mu_s - \frac{\partial H}{\partial s} = (r+\gamma)\mu_s - \left(J\left(x, \beta\,|\,\tau\right) - z\right)\Lambda_s', \tag{20}$$

$$\dot{\mu}_B = r\mu_B - \frac{\partial H}{\partial B} = r\mu_B - \mu_x, \tag{21}$$

$$\dot{z} = rz + \left(Ax + \tfrac{c}{2}\left(u^0 - u\right)^2\right) - \Lambda\left(s, x\right)\left(J\left(x, \beta\,|\,\tau\right) - z\right). \tag{22}$$

Together with appropriate transversality conditions and the equations of motion, equations (18)-(22) characterise the optimal solution assuming that $s(t)$ is non-decreasing. Note that the expression for μ_B does not affect the optimal solution and does not need to be computed. Steady state solutions can be found by setting all time derivatives in (19)-(22) equal to zero and solving the resulting equations together with (18).[7]

$$x_{ss} = \frac{u^0}{\delta} - \underbrace{\frac{A}{c\delta(r+\delta)}}_{\text{Abatement without risk}} + \underbrace{\frac{(r+\gamma)(r+\delta) - \sqrt{\left((r+\gamma)^2(r+\delta)^2 + \dfrac{2A\alpha^2\beta\lambda^2}{c(r+\delta)}\right)}}{\alpha\delta\lambda}}_{\text{Abated because of threshold risk}}. \tag{23}$$

---

[7]These steady state conditions must be interpreted with some care because the derivative of the hazard rate is discontinuous when $\dot{s} = 0$. x_{ss} and s_{ss} are the limits of $s(t)$ and $x(t)$ if $\dot{s} > 0$ along the optimal path. As a steady state is never actually reached in finite time $\Lambda_s' = -\lambda\gamma$ and $\Lambda_x' = \lambda\alpha$ for all t. Note also that there exists a steady state where $x_{ss} > u^0/\delta$. This steady state has no economic interest and is a consequence of the quadratic instantaneous utility function.

Steady state emission and stress levels are given by:

$$u_{ss} = \delta x_{ss}, \quad s_{ss} = \alpha \gamma^{-1} x_{ss}. \tag{24}$$

Note that if $A = 0$ then $x_{ss} = u^0/\delta$, which is the steady state if u is unregulated. Note also that if $\beta = 0$ or $\lambda \to 0$, then $x_{ss} = u^0/\delta - A/(c\delta(r + \delta)) = x(\infty)_{\beta=0}$, which is the steady state of the deterministic problem with no risk of B jumping from 0 to β from eq. (16). The solution has several interesting properties. Note that the magnitude of abatement does not depend on unregulated emissions u^0. Further, if the system is relatively slow (*i.e.* δ and γ are small), then $(r + \delta)(r + \gamma)$ is also a small number and the terms containing A dominate the steady state abatement level. Obviously, if the term $A\alpha\lambda/ c(r + \delta)$ becomes sufficiently large, then $x_{ss} \leq 0$. When this happens, it is optimal to let $x(t)$ go to 0. This happens if *e.g.*:

$$A \geq \left(\phi + \left(u^0 + \beta \right)\alpha\lambda + \sqrt{\phi^2 + 2\phi\alpha\lambda\beta + \left(2u^0 + \beta \right)\alpha^2\lambda^2} \right)\frac{c\left(r + \delta \right)}{\alpha\lambda}. \tag{25}$$

where $\phi = (r + \delta)(r + \gamma)$. Hereafter, unless explicitly stated, we assume that A does not satisfy (25) and positive emissions are optimal. It is also interesting to note that threshold risk means that discounting works differently. In standard deterministic and TDC models, discounting affects the solution through a multiplicative term in the denominator. With SDC problems, when the cost of crossing the threshold is a shock to instant utility, discounting affects the optimal solution in a quasi-linear way (see Nævdal and Vislie, 2008). Given that inertia risk is a hybrid of the two structures, it is interesting that the role of discounting in eq. (23) is also hybridised.

The steady state in eqs. (23)-(24) only has economic relevance provided some optimal solutions in fact converge to it. This occurs *e.g.* when Proposition 1 holds:

Proposition 1. *Assume that* $\delta > c^{-1}$. *Then the system defined by* (10)*,* (18)*,* (19)*,* (20)*,* (21) *and* (22) *has at least two eigenvalues with negative real parts.*

A proof based on Gershgorin's circle theorem is given in online appendix B. Proposition 1 implies that by appropriate choice of integration constants, we can ensure that $(x(t),\, s(t))$ converges to $(x_{ss},\, s_{ss})$. It is worth noting that $\delta > c^{-1}$ is a sufficient and not a necessary condition for the existence of two eigenvalues with negative real part, so convergence of x and s may also occur if $\delta \leq c^{-1}$. Online appendix B presents a heuristic discussion of stability and argues that stability also holds when $\delta > c^{-1}$ and that oscillations are unlikely to be optimal. If the remaining three eigenvalues for (18)-(22) are positive in steady state, then the steady state is a saddle point. Using Gershgorin's circle theorem, one can also prove that if r, $r + \delta$ and $r + \gamma$ are sufficiently large, such positive eigenvalues exist. However, the algebraic complexity and size of these bounds are such that little insight is gained from reproducing them. We thus assume that parameters are such that the equilibrium is a saddle point.

Dynamic Analysis

Steady state analysis does not tell the whole story, so we must analyse the dynamics. We now illustrate that the steady state to which the system converges depends to some extent on initial conditions. For initial levels of stock and stress $(x(t),\, s(t))$ that are sufficiently low when regulation starts, the steady state in eq. (23) with the associated values of u_{ss} and s_{ss}, is the relevant steady state as $\dot{s} > 0$ for all t. However, if the pollution process has gone on long enough, inertia means that if $s(0)$ is large enough, then $s(t) > s_{ss}$ for some t. Unless the disaster occurs, we learn that there are safe values of s higher than s_{ss}, so there is no gain in trying to steer the system back to lower levels. Proposition 2 thus establishes a continuum of steady states in a segment of the line $\dot{s}$ (for proof, see online appendix B):

Proposition 2. *Assume that an optimal solution exists. Define the set $W = \{(x, s): (x_{ss} < x \le x(\infty)_{\beta=0}, s_{ss} < s \le s(\infty)_{\beta=0}, \dot{s} = \alpha x - \gamma s = 0)\}$. Then any (x, s) belonging to W is a steady state.*

In order to illustrate path dependency of optimal solutions, we first analyse a case where regulation starts when the initial condition $(x(0), s(0))$ is sufficiently low to converge to the steady state in eqs. (23) and (24) and then a case where convergence is to some point in W.

Figure 2 illustrates optimal paths for $x(t)$ and $s(t)$ when regulation starts at low pollution levels. If unregulated, the system would converge to $(x, s) = \left(u^0/\delta, \alpha u^0/\gamma\delta\right)$. Along the straight line from the origin $\dot{s} = 0$, there are three different steady states: (x_{ss}, s_{ss}) is the steady state from eq. (23) when the system is regulated under the risk of a jump in β; $(x(\infty)_{\beta=0}, s(\infty)_{\beta=0})$ is the steady state when there is no risk of a jump in B, because $\beta = 0$ or/and $\lambda = 0$; and $(x(\infty)_{\beta>0}, s(\infty)_{\beta>0})$ is the steady state after the jump has occurred. Along the $\dot{s} = 0$ line, $s = \alpha\gamma^{-1}x$. S^* indicates the true value of the threshold, which is unknown to the regulator. Curve A indicates the optimal path. When the threshold is crossed, the dashed part of curve A is no longer optimal, and the optimal path is instead the solid curve towards $(x(\infty)_{\beta>0}, s(\infty)_{\beta>0})$. Curve B indicates the path that a regulator unaware of the threshold would choose. When the threshold is crossed, the regulator adjusts the system away from the dashed curve towards $(x(\infty)_{\beta>0}, s(\infty)_{\beta>0})$ instead.
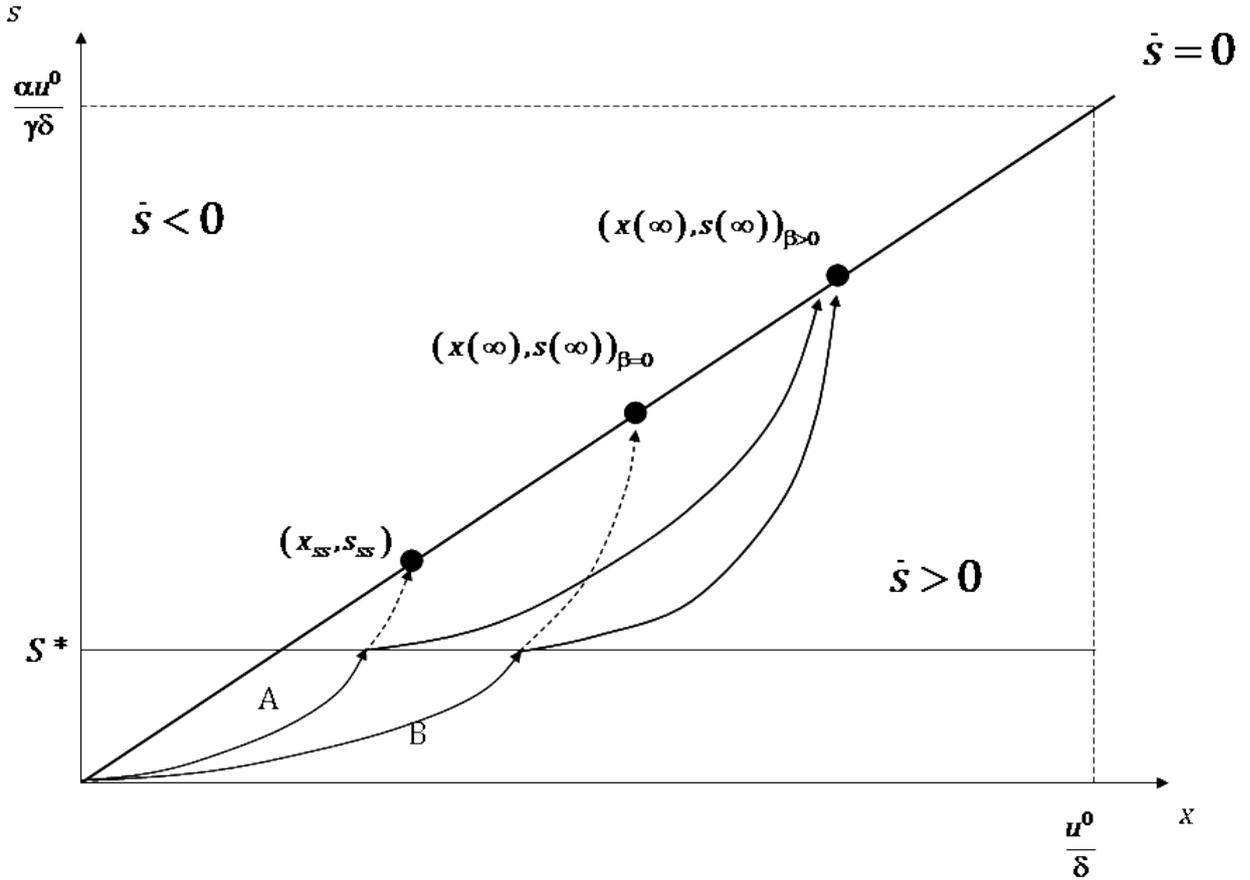
Figure 2: Comparison of the optimal policy (solid curves) taking risk into account (path A) and a suboptimal policy ignoring risk (B) when threshold S^* is crossed. The dashed curves indicate the optimal paths had the stress threshold not been reached.

Figure 3 illustrates a case where the true value of the threshold S^* lies above the steady state (x_{ss}, s_{ss}). In that case, following policy A, accounting for the risk, does not trigger the catastrophe, while following policy B, ignoring the risk, triggers the catastrophe. The regulator should then steer the system along the dashed curve towards $(x(\infty)_{\beta>0}, s(\infty)_{\beta>0})$.
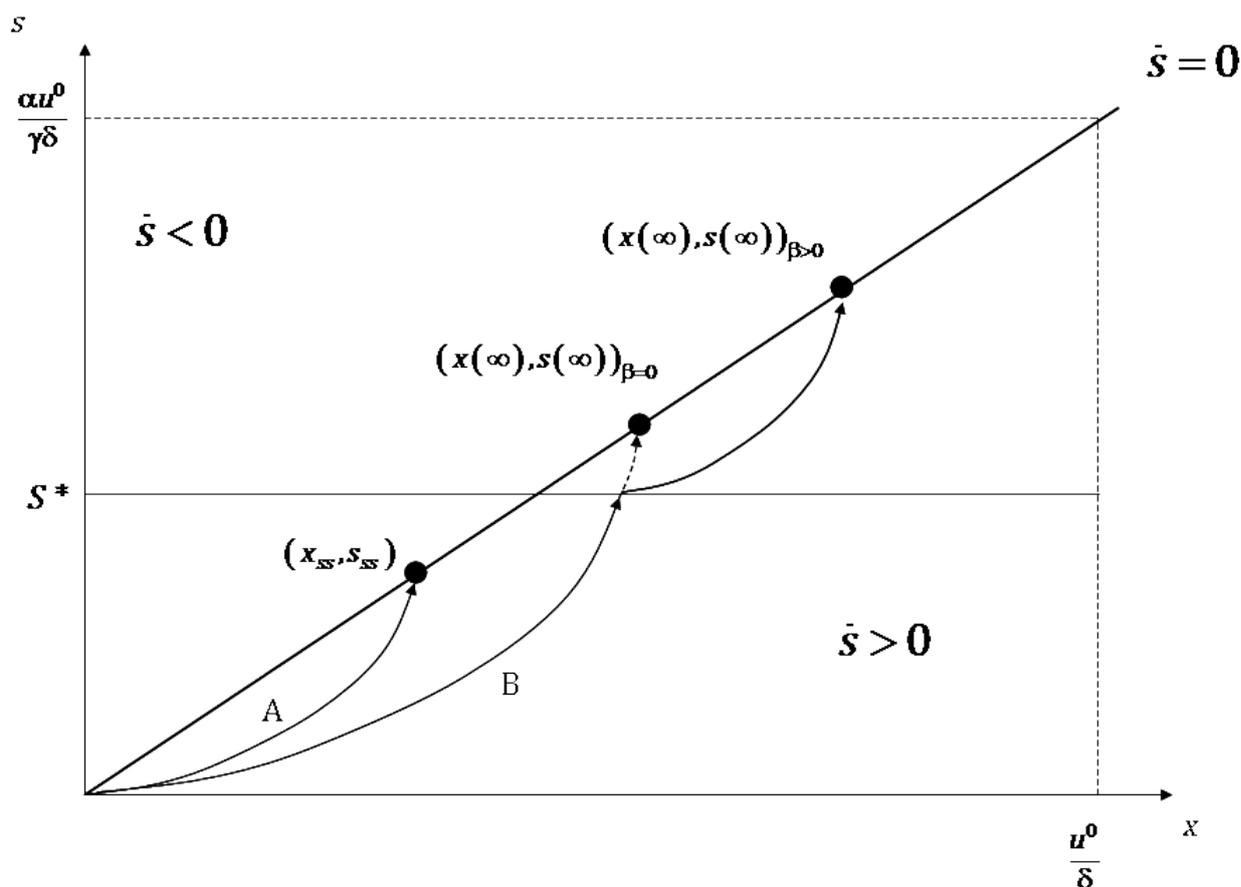
Figure 3: Comparison of the optimal policy, (solid curve *A*), which does not trigger the catastrophe, and the suboptimal policy (solid curve B), which does. The dashed curve indicates the optimal path had the stress threshold not been reached.

We give a heuristic description of the dynamics when regulation starts at relatively high pollution levels (see formal analysis and optimality conditions in online appendix C). On moving through (x, s)-space, as long as we do not cross the threshold we learn where it is *not* located. Any value of s that has been experienced previously is thus known to be safe. This means that, when the system has previously been imperfectly regulated, the optimal long-run equilibrium when regulation starts near the origin may no longer be relevant either, because some values of s are known to be safe, or inertia means that we are committed to experiencing values of $s > s_{ss}$. If $s(\infty)_{\beta=0} > s(0) > s_{ss}$, or if $s(0) < s_{ss}$ and inertia is large enough, then $s(t)$

increases above s_{ss} for some period, regardless of how $u(t)$ is chosen. To see the effect of inertia, we solve eqs. (5)-(6) with some arbitrary initial conditions $(x(t_1), s(t_1))$ and $u(t) = 0$:

$$x(t) = x(t_1)e^{-\delta(t-t_1)}$$
$$s(t) = \frac{e^{-\delta(t-t_1)}x(t_1)\alpha - e^{-\gamma(t-t_1)}\left(x(t_1)\alpha + s(t_1)(\delta-\gamma)\right)}{\gamma - \delta} . \tag{26}$$

The maximum value $\bar{s}$ of $s(t)$, conditional on $u(t) = 0$, over the interval $[t_1, \infty)$ is attained at $\bar{t}$ and given by:

$$\bar{t} = t_1 + \frac{1}{\gamma - \delta}\ln\left(\frac{\gamma\left(\alpha x(t_1)+(\delta-\gamma)s(t_1)\right)}{\delta\alpha x(t_1)}\right)$$
$$\bar{s} = x(t_1)\frac{\alpha}{\gamma}\left(\frac{\gamma\left(\alpha x(t_1)+(\delta-\gamma)s(t_1)\right)}{\delta\alpha x(t_1)}\right)^{-\frac{\delta}{\gamma-\delta}} .$$

The initial conditions $s(t_1)$ and $x(t_1)$ imply a commitment to accept a stress level $\bar{s} = s(\bar{t})$ larger than $s(t_1)$. A necessary condition for the steady state in eqs. (23) and (24) to be relevant is clearly that $s_{ss} < \bar{s}$. If this does not hold, then if the threshold is low enough, the catastrophe will be triggered regardless of action taken (*i.e.* even if $u(t) = 0$), because of inertia. We illustrate different possibilities in Figure 4. The unregulated path is illustrated by the arrow from the origin to the unregulated steady state $\left(u^0/\delta, \alpha u^0/\gamma\delta\right)$, which is the long-run equilibrium conditional on the catastrophe not being triggered and is given by:

$$\lim_{t\to\infty} x(t)_{u(t)=u^0} = \frac{u^0}{\delta}, \quad \lim_{t\to\infty} s(t)_{u(t)=u^0} = \frac{\alpha u^0}{\gamma\delta} . \tag{27}$$

We illustrate three qualitatively different types of optimal regulation paths I, II, and III, starting at different locations on the unregulated path. Path I occurs when regulation starts relatively early. In spite of inertia and the late start, $s(t)$ will remain below s_{ss} and (x_{ss}, s_{ss}) is the relevant steady state.

If regulation is delayed until point II, $s(t)$ will increase above s_{ss} even if u is set to zero for all t after regulation starts. The steady state (x_{ss}, s_{ss}) therefore loses its relevance, as there is no gain from forcing $s(t)$ down to s_{ss}. However, it is obviously not optimal to act as if there is no risk. The solution is to steer the system to some point $(\hat{x}, \hat{s})$ in the set W defined in Proposition 2, and let it remain there indefinitely. An important part of optimal regulation is to determine the exact location of $(\hat{x}, \hat{s})$, as this point is endogenous and will be the long-run equilibrium.
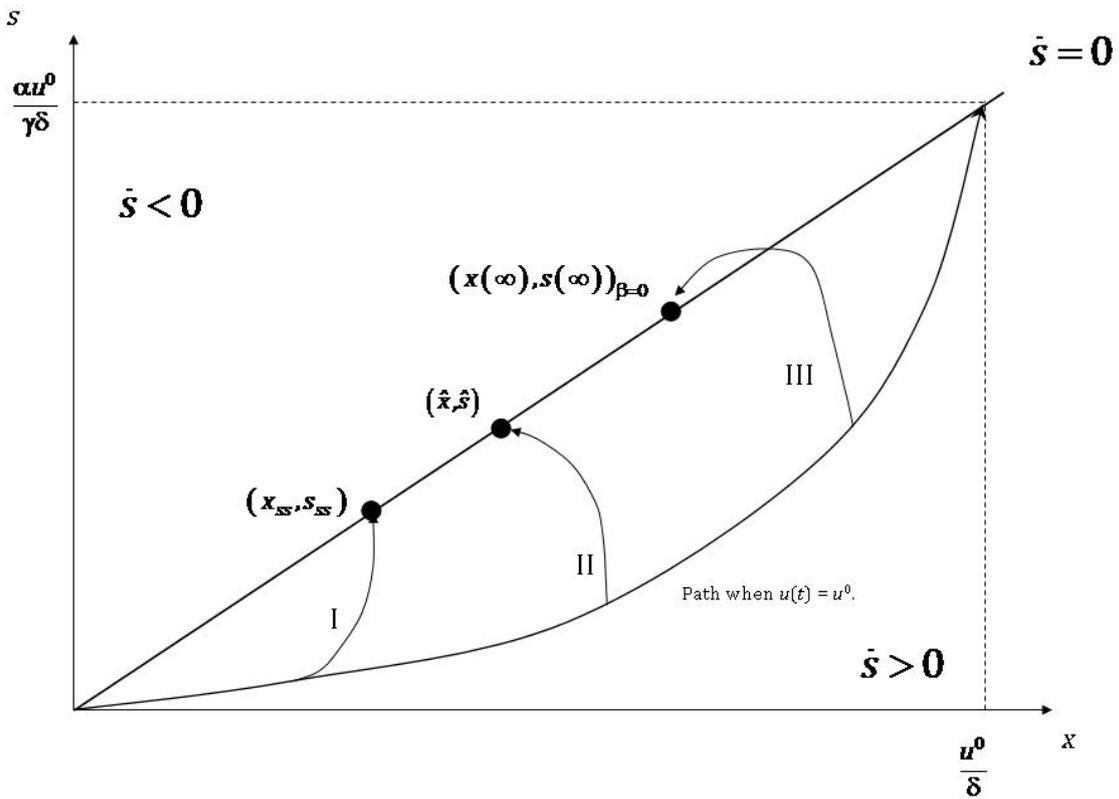
Figure 4: Three paths that are optimal given different initial conditions and conditional on not reaching the threshold.

Path III occurs when regulation starts so late that $s(t)$ becomes larger than the steady state value of s without risk, $s(\infty)_{\beta=0}$. Here again, tardy regulation means accepting the risk that

follows from inertia. However, in this case when $\dot{s} = 0$ is reached, $s(t)$ and $x(t)$ are both higher than the optimal steady state level without risk. If, by luck, the maximum value of $s(t)$ has been reached without catastrophe, then there is no more risk as it is optimal to let $s(t)$ converge to the lower level $s(\infty)_{\beta=0}$. As the hazard rate is proportional to $\dot{s}$, for paths I and II, the hazard rate converges towards zero as x and s approaches steady state. Path III overshoots the $\dot{s} = 0$ isocline and experiences a period where $\dot{s} < 0$, implying that the probability of disaster is zero before $x(t)$ and $s(t)$ converge.

Comparison between inertia risk, SDC and TDC

Comparisons of the steady state solution in the different approaches is challenging because different hazard rate processes mean different dynamic systems, *i.e.* 'comparing apples and oranges'. In fact, inertia risk involves an extra differential equation. Further, parameters such as λ are not directly comparable in the different models, as they have different units. However, we can compare the optimal steady state values of the state variable x for all three types.[8] This comparison should make sense regardless of model dimensions and other elements differentiating them. The TDC and SDC models share the following structure:

$$\max_{u(t)\geq 0} \mathbb{E}_\tau \left(\int_0^\infty \left(-Ax - \tfrac{c}{2}\left(u^0 - u \right)^2 \right) e^{-rt} dt \right) \tag{28}$$

subject to:

$$\dot{x} = u - \delta x + B, \ x(0) \text{ given } \dot{B} = 0, \ B(0) = 0, \ B(\tau^+) - B(\tau^-) = \beta. \tag{29}$$

They differ in the stochastic processes generating the catastrophe.

---

[8]For brevity, we do not show the calculations. They are straight-forward to reproduce with the optimality conditions given in online appendix A.

First, we compare models of inertia risk and TDC risk. TDC problems do not have an equation for stress and also differ from inertia risk problems by their stochastic processes. The stochastic process for the TDC problem is given by:

$$\lim_{dt \to 0} \frac{\Pr\left(\tau \in [t, t+dt] \mid \tau \geq t\right)}{dt} = \lambda_t(x). \tag{30}$$

Assuming for simplicity that $\lambda_t(x) = \lambda x$, we can compute the resulting steady state value, x_{ss}^{TDC}, to be compared with x_{ss}:

$$x_{ss}^{TDC} = \frac{u^0}{\delta} - F + \delta \frac{\sqrt{r^2 C + \lambda \delta (2r + \delta + 2G)}}{E}.$$

where

$$F = \frac{\left(r\delta + u^0 \lambda\right)}{\delta \lambda \left(1 + \dfrac{r}{r + \delta + 2G\lambda}\right)}. \tag{31}$$

$$E = \delta \lambda \left(\delta c (r + \delta)(2r + \delta) + 2A\lambda\right).$$

$$G = \frac{A}{c\delta(r + \delta)}.$$

In this set-up, the expression for x_{ss}^{TDC} is much more complex than x_{ss} in eq. (23) and there are structural differences. The abatement level in x_{ss}^{TDC} is given by $x_{ss}^{TDC} - u^0/\delta$ and responds non-linearly to changes in damage A because A enters both the numerator and denominator. With inertia risk, A only enters the numerator. The abatement level also depends in a non-trivial way on u^0, whereas abatement in x_{ss} in eq. (23) does not.

Note that in steady state, the hazard rate remains an important factor in the optimality conditions for TDC problems, but is zero with inertia risk. The hazard rate still influences the steady state solution with inertia risk, but only through its derivative with respect to state variables. See Nævdal (2016) for an analysis of how the hazard rate and its derivative affect optimal solutions.

While these examples represent highly stylised models, they illustrate that if inertia risk is the most realistic model, using TDC is neither a good approximation nor simplification. In particular, the comparative dynamics are different as parameters have different effects in the different models.

We now compare inertia risk and SDC risk. The SDC problem that is similar to our inertia risk problem has no equation for stress and a stochastic process given by:

$$\lim_{dt \to 0} \frac{\Pr\left(\tau \in [t, t+dt] \mid \tau \geq t\right)}{dt} = \lambda \max\left(\dot{x}, 0\right).$$
$$B\left(\tau^+\right) - B\left(\tau^-\right) = \beta.$$
(32)

The resulting steady state value of x is given by x_{SDC}^{ss}:

$$x_{ss}^{SDC} = \frac{u^0}{\delta} - \frac{A}{c\delta(r+\delta)} + \frac{r+\delta - \sqrt{(r+\delta)^2 + \frac{2A\beta\lambda^2}{c(r+\delta)}}}{\delta\lambda}.$$
(33)

The solution clearly has a similar structure to x_{ss} in eq. (23), although x_{ss} is slightly more complex than x_{ss}^{SDC}. This is perhaps not surprising, as SDC is a special case of inertia risk. If s is a very fast variable relative to x, then one can transform the inertia risk problem into a SDC problem. However, that would mean losing all the subtle dynamic effects that follow from the sluggishness of s. The difference between these two solutions depends on the size of the term $(r+\gamma)/\alpha$. If $(r+\gamma)/\alpha < 1$ then $x_{ss} < x_{ss}^{SDC}$, while $x_{ss} > x_{ss}^{SDC}$ if $(r+\gamma)/\alpha > 1$. Recall that α^{-1} is the expected time it takes for a unit of x to result in a unit of stress. If stress responds rapidly to changes in the state variable and decreases slowly after such disturbance (low natural cooling, γ), the optimal steady state will entail much lower temperature than without inertia and vice versa. In this problem, discounting reinforces this last effect, illustrating that impatience plays a similar role to the capacity to absorb stress for the steady state shock size. If $r + \gamma = \alpha$, the delays imposed by inertia are exactly compensated for by the degree of

cooling and the degree of impatience. In this case, the stock variable has the same steady state level with SDC as with inertia. However, it must be emphasised that inertia still means that the hazard rate is sluggish relative to the SDC case and therefore we approach the steady state on a different path even if $r + \gamma = \alpha$.

IV. Discussion

The results presented in section III are based on a particular model of climate change. Here, we discuss whether they can be generalised to other types of systems and dynamics.

While our application focuses on climate change, the model we use is so general that it could apply to many different pollution problems with a risk of regime shift. For example, the term u could represent impacts of CO_2 on ocean acidification, emission of gases affecting the ozone layer or inflow of nutrients to a lake or other type of aquiphere. Similarly, parameter δ could represent many different types of natural absorption of the pollutant. The stress equation is also so general that it is appropriate to many different cases where stress depends on the level of some particular stock variable. In fact, we chose this set-up for its generality and its capacity to describe traits particular to systems subject to potential regime shifts in the literature (see section II).

However, while our application can represent many types of pollution problems, there are problems for which it is less appropriate. Thus we also discuss to what extent this approach can be generalised. First, we consider problems with more interacting variables than the two in our model. We envisage a set-up where the variables x and s are vectors instead, in which case stability analysis may become quite complicated. The application studied is not well suited to represent resource management problems, in which case the resource, *e.g.* a fish stock or a forest, exhibits some kind of growth dynamics that we do not yet incorporate. The early regime shift literature focused mostly on shallow lake dynamics, but there has been

substantial expansion to model all kinds of resource use in particular fisheries, grasslands, forests, coral reefs, *etc*. These studies show that similar types of dynamics occur in all these systems and that many of the results obtained with shallow lakes can often be successfully applied to other systems. Generalisation would therefore be feasible and useful in our inertia risk approach. In most cases, it would probably be enough to replace the linear input *u* with some function of the stock dynamics itself that represents growth. Since we already feature regime shift dynamics through term *b* and the stress equation, this function would only need to incorporate some rather simple growth component, *e.g.* logistic growth.

Our approach depicts regime shift as a sharp change (sudden shift *B* in state on crossing a threshold) that is also irreversible. Note that representing a regime shift by a flip between 0 and *B* is the limit case of a more general model, which would incorporate a term $B(x^\theta/(1+x^\theta))$ when θ goes towards infinity. Then, for *x* < 1 this term goes towards 0 and for *x* > 1 it goes towards *B*. At the other extreme, high smoothness is best characterised by a linear change instead. This general formulation may allow for some reversibility.

We exclude the possibility of limit cycles (see online appendix B for motivation). However, recent research indicates that many systems, including several possible tipping elements of the climate system, are forced periodically at a time scale comparable to their internal dynamics (Bathiany *et al.*, 2018). This is a pertinent, but rather difficult, area for future research.

Finally, there may be situations for which a pure TDC approach would be more appropriate, but which still exhibit lags. Such problems with purely time-distributed catastrophes and lags would probably require a rather different approach that has yet to be defined.

We show that introducing the concept of system stress can be a useful way to model systems with inertia and potential risk of catastrophic shifts. Our model is not intended to produce reliable predictions of some real-world system, a near impossibility given their complex

adaptive system dynamics, but to highlight trade-offs and potential impacts of alternative choices in the study context. Specifically, our approach relies on the ability to measure stress and how it changes over time. This is closely related to the concept of resilience, which has proven quite difficult to measure despite many attempts. Moreover, many scholars argue that measuring resilience is not useful *per se*, but rather that it is a useful concept for applying to different problems in order to assess their propensity to undergo regime shifts of significance for their users (Walker and Salt, 2012). Stress is slightly different, *e.g.* in many real-world problems it may be feasible to measure some proxy of an aggregated level of ecosystem stress, *e.g.* the planetary boundaries literature proposes measuring these boundaries as particular levels of some "control variables"[9]. These include atmospheric CO_2 concentration for climate change, carbonate ion concentration for ocean acidification, stratospheric ozone concentration *etc.* (Rockström *et al.*, 2009). For many of these control variables the scientific community has identified continuously refined boundary values or risk thresholds, at which the risk of crossing a threshold becomes non-negligible (Steffen *et al.*, 2015). Following Margolis and Nævdal (2008) and Crépin and Folke (2015), and given the substantial risk that a global non-reversible catastrophe could occur, we argue that it is reasonable to take a precautionary approach and use these boundaries as a proxy for threshold values. In other systems, like coral reefs, a measure of stress could be the average temperature in the water column and a temperature threshold around 30°C for bleaching. However, identifying accurate indicators of stress and their threshold is a task that requires substantial scientific research and sometimes arbitrary choices about the most convenient parameters to measure at the moment, given available data. We are now in a much better position than we were 20

---

[9]These should not be confused with the control variables used in control theory. The confusion in terms originates from different scientific traditions in different disciplines.

years ago when it comes to understanding these types of dynamics in natural systems, but the path ahead remains long and tedious. One contribution of our paper is to highlight the kind of information needed to properly study the issues. In many cases, unfortunately, we are unaware of the threshold location and sometimes even of the risk of regime shifts. Resilience theory advocates nurturing general resilience to cope in these cases (Biggs *et al.*, 2012b). Stress may be difficult to measure precisely, but in some cases exact measurement may be unnecessary. Assume that, in the stress equation (6), we can observe x and that scientific theory gives us a good estimate of γ, the system's capacity to absorb stress. Assume also that the system had no stress at a specific point in history, $\hat{t} \le 0$. Then:

$$\frac{s(t)}{\alpha} = e^{\gamma t} \int_{\hat{t}}^{t} x(\eta) e^{-\gamma \eta} d\eta \,. \tag{34}$$

where the right-hand side is observable even if the left-hand side is not. Hence, if one is able to form a probability distribution over possible locations of the threshold divided by α, then neither $s(t)$ nor α need to be observed, as their ratio can be inferred from eq. (34).

V. Final remarks

We introduce *inertia risk*, a novel way to model dynamic catastrophic risk which we show is appropriate for many real-world situations. With inertia risk, we use a stochastic structure where the optimal probability of catastrophe may span the entire interval [0, 1]. Optimally managed inertia risk also exhibits path dependency and accounts for possible time lags. We illustrate our approach with a model of climate change, in which stress is substantially related to changes in temperature and decreases slowly. After this disturbance, the optimal steady state entails much lower risk than without inertia, so the optimal policy with inertia proves to be more precautionary than if inertia were not accounted for. Hence, managers who ignore these lagged effects would take more risk than necessary to end up in a situation of

31

undesirable climate change. This risk is not justified by the potential gains associated with that risky behaviour. On the other hand, managers who are aware of these processes may have more time to react before the catastrophe occurs and may actually be able to avert it. It is also worthwhile noting that, with inertia risk, the hazard rate will depend on different parameters than with SDC and TDC. Therefore, this paper provides valuable information about the kinds of variables that managers who want to monitor these types of system should measure.

In addition, the possibility of non-Markovian controls and learning in inertia risk gives rise to a continuum of long-run equilibria that to our knowledge does not appear in TDC models. This is particularly striking in comparison with TDCs where the long-run equilibrium prior to catastrophe may exhibit path dependency if there is a Skiba point in the model, but they are always Markovian. Overall, this means that using TDC or SDC models may lead to substantial errors in policy recommendations that cannot be corrected when later discovered because of the intrinsic path dependency inherent in the system. This in turn suggests that many models of resource exploitation with endogenous risk, including many climate models, resource management models and even models of knowledge spillover, should be revisited using an inertia risk approach.

References

Anthony, K. M. W., Anthony, P., Grosse, G., and Chanton, J. (2012), Geologic methane seeps along boundaries of Arctic permafrost thaw and melting glaciers, *Nature Geoscience*, *5*(6), 419-426.

Barbier, E. B. (2013), Wealth accounting, ecological capital and ecosystem services, *Environmental and Development Economics*, **18**, 2, 133–161.

Bathiany, S., Scheffer, M., Van Nes, E. H., Williamson, M. S., and Lenton, T. M. (2018), Abrupt Climate Change in an Oscillating World, *Scientific reports*, *8*(1), 5040.

Berkes, F., and Folke, C. (1998), Linking social and ecological systems for resilience and sustainability. *Linking social and ecological systems: management practices and social mechanisms for building resilience*, *1*(4).

Berry, K., Finnoff, D., Horan R.D., and Shogren J. F. (2015), Managing the endogenous risk of disease outbreaks with non-constant background risk, *Journal of Economic Dynamics and Control*, **51**, C, 166-179.

Biggs, R., Blenckner, T., Folke, C., Gordon, L. J., Norström, A., Nyström, M. and Peterson, G.D. (2012a), Regime shifts. *in* Hastings A. and Gross L., editors, *Encyclopedia of Theoretical Ecology*, University of California Press, Berkeley, United States.

Biggs, R., Carpenter, S.R. and Brock, W.A. (2009), Turning back from the brink: detecting an impending regime shift in time to avert it, *Proceedings of the National Academy of Sciences* 106, 3, 826-831.

Biggs, R., Schlüter, M., Biggs, D., Bohensky, E. L., BurnSilver, S., Cundill, G., ... and Leitch, A. M. (2012b), Toward principles for enhancing the resilience of ecosystem services, *Annual review of environment and resources*, *37*, 421-448.

Cai, Y., Judd, K. L. and Lontzek, T. S. (2013), The Social Cost of Stochastic and Irreversible Climate Change (January), NBER Working Paper No. w18704, Available at SSRN: https://ssrn.com/abstract=2202643

Carpenter, S.R., and Turner, M.G. (2000), Hares and Tortoises: Interactions of Fast and Slow Variables in Ecosystems. *Ecosystems* 3, 6, 495-497.

Clarke, H. R., and Reed, W.J. (1994), Consumption/pollution tradeoffs in an environment vulnerable to pollution-related catastrophic collapse, *Journal of Economic Dynamics and Control*, *18*(5), 991-1010.

Crépin, A.-S. (2007), Using Fast and Slow Processes to Manage Resources with Thresholds, *Environmental and Resource Economics*, 36, 191-213.

Crépin, A.-S., and Folke, C. (2015), The Economy, The Biosphere and Planetary Boundaries: Towards Biosphere Economics, *International Review of Environmental and Resource Economics* 8, 57-100.

Crutzen, P. J. (2002), Geology of mankind, *Nature*, *415*(6867), 23-23.

Doraszelski, U. (2003), An R&D race with knowledge accumulation, *RAND Journal of Economics*, 20-42.

Feichtinger, G., Novak, A. and Wirl, F. (1994), Limit cycles in intertemporal adjustment models – theory and applications, *Journal of Economic Dynamics and Control*, **18**, 353–380.

Gjerde, J., Grepperud, S., and Kverndokk, S. (1999), Optimal climate policy under the possibility of a catastrophe, *Resource and Energy Economics*, *21*(3), 289-317.

Goreau, T. J., and Hayes, R. L. (1994), Coral bleaching and ocean "hotspots." *Ambio* **23,** 176–180.

Graβ, D., Xepapadeas, A. and de Zeeuw, A. J. (2017), Optimal management of ecosystem services with pollution traps: The lake model revisited, *Journal of the Association of Environmental and Resource Economists*, 4(4),1121–1154.

Heal, G. (1984), Interactions between economy and climate: A framework for policy design under uncertainty, *Advances in applied microeconomics*, *3*, 151-168.

Hirota, M., Holmgren, M., Van Nes, E. H., and Scheffer, M. (2011), Global resilience of tropical forest and savanna to critical transitions, *Science*, 334(6053), 232-235.

Hughes, T. P. (1994), Catastrophes, phase shifts, and large-scale degradation of a Caribbean coral reef, *SCIENCE-NEW YORK THEN WASHINGTON-*, 1547-1547.

Joughin, I., Smith, B. E., and Medley, B. (2014). Marine ice sheet collapse potentially under way for the Thwaites Glacier Basin, West Antarctica, *Science*, *344*(6185), 735-738.

Le Kama, A. A., Pommeret, A., and Prieur, F. (2014), Optimal Emission Policy Under The Risk Of Irreversible Pollution, *Journal Of Public Economic Theory*, **16**, 6, 959-980.

Lemoine, D. and Traeger, C. (2014), Watch your step: Optimal policy in a tipping climate *American Economic Journal: Economic Policy* 6(1),137-166.

Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S., and Schellnhuber, H. J. (2008), Tipping elements in the Earth's climate system, *Proceedings of the national Academy of Sciences*, *105*(6), 1786-1793.

Levin, S. A., Xepapadeas, A., Crépin, A.-S., Norberg, J., de Zeeuw, A.J., Folke, C., Hughes, T.P., Arrow, K.J., Barrett, S, Daily, G. C., Ehrlich, P.R., Kautsky, N., Mäler, K.-G., Polasky, S., Troell, M., Vincent, J. and Walker, B. (2012), Social-ecological systems as complex adaptive systems: modeling and policy implications. *Environment and Development Economics,* 18, 111–132.

Liski, M. and F. Salanié (2016), Tipping points, delays, and the control of catastrophes. Paper presented at the SURED 2018 - Monte Verità Conference on Sustainable Resource Use and Economic Dynamics. Link

Lontzek, T. S., Y. Cai, K. L. Judd and T. M. Lenton, (2015), Stochastic integrated assessment of climate tipping points indicates the need for strict climate policy, *Nature Climate Change*, 5, 5, 441-444, Nature Research.

Margolis, M., and Nævdal, E. (2008), Safe minimum standards in dynamic resource problems: conditions for living on the edge of risk. *Environmental and Resource Economics*, *40*(3), 401-423.

Matthews, H. D., Gillett, N. P., Stott, P. A., and Zickfeld, K. (2009), The proportionality of global warming to cumulative carbon emissions. *Nature*, *459*(7248), 829.

Mäler, K.-G., Xepapadeas, A. and de Zeeuw, A. J. (2003), The economics of shallow lakes. *Environmental and Resource Economics*, 26(4), 603-624.

Nævdal, E. (2003), Optimal regulation of natural resources in the presence of irreversible threshold effects, *Natural Resource Modeling*, 16 (3), 305-333.

Nævdal, E. (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*, *30*(7), 1131-1158.

Nævdal, E., and Oppenheimer, M. (2007), The economics of the thermohaline circulation—A problem with multiple thresholds of unknown locations. *Resource and Energy Economics*, *29*(4), 262-283.

Nævdal, E. and Vislie, J. (2008), Climate Change, Catastrophic Risk and the Relative Unimportance of Discounting, Memo 28/2008, Department of Economics, University of Oslo. Link

Norström, A. V., Nyström, M., Lokrantz, J., and Folke, C. (2009), Alternative states on coral reefs: beyond coral-macroalgal phase shifts. *Mar Ecol Prog Ser*, *376*, 295-306.

van der Ploeg, F. and de Zeeuw, A. J. (2013), Climate Policy and Catastrophic Change: Be Prepared and Avert Risk, Oxcarre Working Papers 118. Oxford Centre for the Analysis of Resource Rich Economies, University of Oxford.

van der Ploeg, F. (2014), Abrupt positive feedback and the social cost of carbon, *European Economic Review*, **67**, 28-41.

Polasky, S., De Zeeuw, A., and Wagener, F.O.O. (2011), Optimal management with potential regime shifts, *Journal of Environmental Economics and Management*, 62(2), 229-240.

Reed, W. J., and Heras, H.E. (1992), The conservation and exploitation of vulnerable resources, *Bulletin of Mathematical Biology*, **54**, 2-3:185–207.

Rockström, J, W. Steffen, K. Noone, Å. Persson, F. S. Chapin, III, E. F. Lambin, T. M. Lenton, M. Scheffer, C. Folke, H. J. Schellnhuber, B. Nykvist, C. A. de Wit, T. Hughes,

S. van der Leeuw, H. Rodhe, S. Sörlin, P. K. Snyder, R. Costanza, U. Svedin, M. Falkenmark, L. Karlberg, R. W. Corell, V. J. Fabry, J. Hansen, B. Walker, D. Liverman, K. Richardson, P. Crutzen, and J. A. Foley. (2009), A safe operating space for humanity. *Nature*, **461**, 472-475.

Scheffer, M. and Carpenter, S.R. (2003), Catastrophic regime shifts in ecosystems: linking theory to observation. *Trends in ecology and evolution* 18(12), 648-656.

Seierstad, A. (2009), *Stochastic Control in Discrete and Continuous Time*, Springer.

Srinivasu, P.D.N. (2004) Regime shifts in eutrophied lakes: a mathematical study. *Ecological modelling*, 179(1),115–130.

Solomon, S., Ivy, D. J., Kinnison, D., Mills, M. J., Neely, R. R., and Schmidt, A. (2016), Emergence of healing in the Antarctic ozone layer, *Science*, aae0061.

Steffen, W. L., Sanderson, A., Tyson, P.D., Jäger, J., Matson, P.A., Moore III, B., Oldfield, F., Richardson, K., Schellnhuber, H.J., Turner II, B. L. and Wasson, R.J. (2004), *Global Change and the Earth System - A Planet Under Pressure*. Global Change - The IGBP Series. Springer-Verlag, Berlin (DEU).

Steffen, W, Richardson, K., Rockström, J., Cornell,s. E., Fetzer, I., Bennett, E.M., Biggs, R.O. Carpenter, S.R., de Vries, W., de Wit, C.A. Folke,C., Gerten, D., Heinke, J., Mace, G.M., Persson, L.M., Ramanathan,V., Reyers, B., and S. Sörlin. (2015), Planetary Boundaries: Guiding human development on a changing planet, *Science* 1259855.

Steffen, W., Rockström, J., Richardson, K., Lenton, T. M., Folke, C., Liverman, D. Summerhayes, C.P., Barnosky, A.D., Cornell, S.E., Crucifix, M., Donges, J.F., Fetzer, I., Lade, S.J. , Scheffer, M., Winkelmann, R., and Schellnhuber, H. J. (2018), Trajectories of the Earth Systen in the Anthropocene, *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1810141115

Tsur, Y., and Zemel, A. (1995), Uncertainty and irreversibility in groundwater resource management. *Journal of environmental economics and management*, *29*(2), 149-161.

Tsur, Y., and Zemel, A. (1998). Pollution control in an uncertain environment, *Journal of Economic Dynamics and Control*, *22*(6), 967-975.

Wagener, F. O. O. (2003), Skiba points and heteroclinic bifurcations, with applications to the shallow lake system. *Journal of economic dynamics and control* 27, 9, 1533-1561.

Walker, B., and Salt, D. (2012), *Resilience thinking: sustaining ecosystems and people in a changing world*. Island Press.

Walter, K. M., Zimov, S. A., Chanton, J. P., Verbyla, D., and Chapin, F. S. (2006), Methane bubbling from Siberian thaw lakes as a positive feedback to climate warming, *Nature*, *443*(7107), 71-75.